• NEWS & VIEWS •

BUPTCMCC-6G-DataAI-Maritime: a configurable channel dataset of offshore scenario for AI research

Lei TIAN[1], Jikun DU[1], Zihang DING[1], Jianhua ZHANG[1*], Yuanzhi HE[2] & Lantu GUO[3]

[1]*State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*Academy of Systems Engineering, Academy of Military Sciences, Beijing 100141, China*
[3]*China Research Institute of Radiowave Propagation, Qingdao 266107, China*

Introduction of BUPTCMCC-6G-DataAI-Maritime. A space-air-ground-sea integrated network is one of the key visions of the sixth generation (6G) mobile communication system to achieve ubiquitous connectivity. 6G also encompasses integrated sensing and communication [1], as well as integrated AI and communication [2], which aims to enhance the network intelligence and provide mobile communication services with higher reliability. AI enables communication systems to achieve real-time environmental perception and prediction by continuously monitoring dynamic changes. However, implementing AI effectively to optimize communication performance and address the complexity of channels requires substantial data input [3].

The generation of channel multipath parameters mostly relies on channel measurement or ray tracing (RT) simulations for the existing dataset. These channel datasets primarily focus on terrestrial urban scenarios, while they lack adequate support for maritime scenarios [4]. Compared with terrestrial mobile communications, maritime communications face complex and variable sea conditions and meteorological factors, which significantly impact channel characteristics and communication performance. BUPTCMCC-6G-DataAI-Maritime, based on the BUPTCMCC-6G-DataAI+ dataset [5], has been released as the first publicly available and generative channel dataset for maritime scenarios. This dataset bridges the gap between the existing channel datasets and the need for maritime specific channel data.

The construction method of dataset. During the simulation of signal propagation based on RT, the accurate reconstruction of the environment and accurate electrical parameters of materials are both of great importance for the calculation of channel multipath parameters.

• Sea surface undulation model. The ocean wave spectrum, which is usually obtained by scholars through long-term observations of sea surface characteristics, can simulate ups and downs of ocean waves in a more realistic manner. JONSWAP (Joint North Sea Wave Project) is a typical deep ocean wave spectrum and unsuitable for shallow waters. As waves travel from deep to shallow areas, their shape and characteristics change. Thus, for offshore scenarios, we use the TMA (Texel-Marsen-Arsloe) spec-

trum, which is derived by multiplying JONSWAP with a function that accounts for water depth [6]. The double Debye equation is commonly used to calculate the dielectric constant of seawater because it has a wide applicable frequency range and relatively high accuracy [7].

• Simulation of propagation mechanism. As the large size of the offshore scenario and in order to depict the undulations of the sea surface in more detail, the sea surface is divided into grids with a size of 1 m × 1 m. When the wind speed over the sea surface is low, the undulation of the sea surface is small, and the sea surface can be regarded as a plane at this time. The reflection occurring on the sea surface can be regarded as a specular reflection. However, when the wind speed increases, the undulation of the sea surface also becomes violent. At this time, the sea surface is a rough plane, and the roughness coefficient needs to be added in the reflection calculation [8]. The Physical Optics (PO) algorithm can calculate the scattered electric field based on the induced current on the target surface [9]. The sea surface can be considered a rough surface, and thus scattering frequently accompanies reflections on it. Once the scattering zone is defined, the current for each small facet within this zone can be calculated [8]. We assume that the path of scattering is from TX → center of scattering surface element → RX in order to obtain the angular parameters. The dataset is generated using the methods mentioned above, and is compared with the measured data to verify the effectiveness of the proposed dataset [10].

The scenario of dataset. The offshore scenario is 2200 m long and 600 m wide, and it comprises two distinct sets of base stations and user grids. Specifically, the height of base stations is set at 30 and 8 m, respectively, while the height of user grids is fixed at 5 m. Each base station and user grid is equipped with different types of antennas. The simulation frequency of the scenario is 3.5 GHz. Under these circumstances, the undulation of the sea surface is the main factor affecting signal propagation. At the same time, the dataset encompasses three distinct sea states: smooth, moderate, and rough [8]. As shown in Figure 1, the direct path is represented by the red dotted line, whereas the reflection paths are depicted by red solid lines and the scattering paths by

---

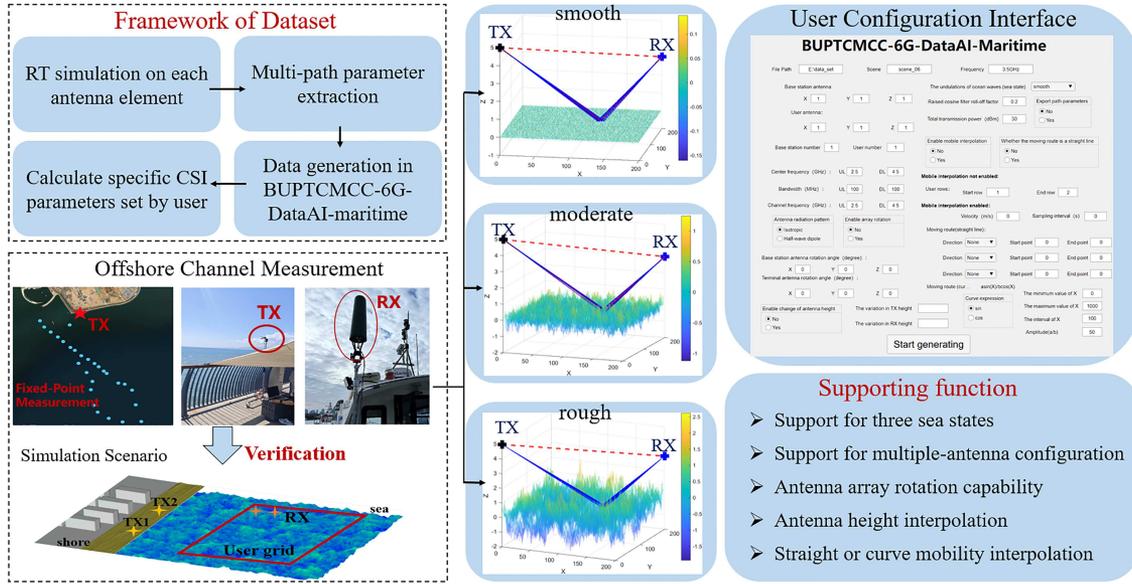* Corresponding author (email: jhzhang@bupt.edu.cn)

Figure 1 (Color online) Overall implementation process of BUPTCMCC-6G-DataAI-Maritime.

blue solid lines. However, the reflection paths are often obscured due to the large number of scattering paths that tend to cluster around them. Moreover, given the stochastic nature of sea waves and the principles of RT, the reflection paths may not always be present in certain scenarios. Therefore, under different sea conditions, the number and power of reflection and scattering paths are bound to be affected to some extent.

The function of dataset. The dataset contains a total of more than 53000 sets of channel characteristic parameters, and each set includes parameters such as time delay, angle, and power. Traditional datasets offer static data combinations that are not customizable to meet individual user requirements. Conversely, BUPTCMCC-6G-DataAI-Maritime enables users to dynamically adjust antenna configurations, moving paths, and sea conditions. For dataset generation, it can hold up to 64 antennas, arranged in a 4 × 4 × 4 3D configuration. Yet in this part, users can customize the antenna count and whether to apply height interpolation, with the custom count not exceeding per-direction or overall maxima. We support both linear and curved moving paths, and the curved paths are typically defined using sin or cos functions. Users can freely configure the moving speed and sampling time interval. Based on the user input parameters, we support the generation of specific CSI (channel state information) to facilitate subsequent research, such as beam prediction and resource allocation [3].

Conclusion. BUPTCMCC-6G-DataAI-Maritime overcomes the challenges in obtaining accurate channel information and meets the data requirements for AI-enabled communication in maritime scenarios. The inclusion of a versatile data generator enhances the datasets utility, allowing researchers to generate customized data by freely configuring parameters such as antenna configurations, sea states, and user mobility patterns. The dataset provides the fundamental data basis for maritime communication research, thereby facilitating the optimization and improvement of 6G maritime communication development.

Access methods. BUPTCMCC-6G-DataAI-Maritime can be downloaded from the website: https://scc.bupt.edu.cn/ dataset-public/datasets/34.

References
1 Zhao C S, Zhang J H, Zhang Y X, et al. BUPTCMCC-6G-CMG+: a GBSM-based ISAC standard channel model generator. Sci China Inf Sci, 2025, 68: 150304
2 Wang H, Zhang J, Nie G, et al. Digital twin channel for 6G: concepts, architectures and potential applications. IEEE Commun Mag, 2025, 63: 24–30
3 Shen Z B, Yu L, Zhang Y X, et al. DataAI-6G: a system parameters configurable channel dataset for AI-6G research. In: Proceedings of IEEE GlobeCom Workshops, 2023. 1910–1915
4 Zhang J H, Yu L, Liu S Y, et al. Wireless environmental information theory: a new paradigm towards 6G online and proactive environment intelligence communication. 2024. ArXiv:2412.11479
5 Yu L, Zhang J H, Han S F, et al. BUPTCMCC-6G-DataAI+: a generative channel dataset for 6G AI air-interface research. Sci China Inf Sci, 2025, 68: 197301
6 Bouws E, Günther H, Rosenthal W, et al. Similarity of the wind wave spectrum in finite depth water: 1. Spectral form. J Geophys Res Oceans, 1985, 90: 975–986
7 Meissner T, Wentz F J. The complex dielectric constant of pure and sea water from microwave satellite observations. IEEE Trans Geosci Remote Sens, 2004, 42: 1836–1849
8 Perez-Fontan F, Pastoriza-Santos V, Machado F, et al. A wideband satellite maritime channel model simulator. IEEE Trans Antennas Propagat, 2021, 70: 2114–2127
9 Miao H, Zhang J, Tang P, et al. Sub-6 GHz to mmWave for 5G-advanced and beyond: channel measurements, characteristics and impact on system performance. IEEE J Sel Areas Commun, 2023, 41: 1945–1960
10 Zhang J Q, Tian L, Tang P, et al. Measurement and analysis of wideband wireless channel in shore-to-ship scenarios. In: Proceedings of IEEE 102nd Vehicular Technology Conference (VTC2025-Fall), Chengdu, 2025. 1–6