Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Research on compression technology of meteorological big data based on satellite communication

Ji Zhang

1035045713@qq.com

Shanghai Meteorological Observatory,Shanghai,China

Caijun Yue

Shanghai Climate Center

Liang Guan

Shanghai Meteorological Observatory,Shanghai,China

Zhiqiang Chen

Shanghai Marine Meteorological Center

Article

Keywords: dimensionality reduction plane octree coding, Compression algorithm, Symmetric cryptographic table

Posted Date: July 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10187955/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Research on compression technology of meteorological big data based on satellite communication

Ji Zhang[1], Caijun Yue[2], Liang Guan[1], Zhiqiang Chen[3]

1 Shanghai Meteorological Observatory,Shanghai,China

2 Shanghai Climate Center,Shanghai,China

3 Shanghai Marine Meteorological Center, Shanghai, China

Corresponding author: Ji Zhang, 1035045713@qq.com

ABSTRACT

Due to communication satellites are limited by word count and frequency, they cannot meet the needs of ocean-going ships for big data of numerical weather prediction. Therefore, a big data compression algorithm combining lossy compression and lossless compression is proposed in this paper. Lossy compression is based on the premise of preserving the core meteorological characteristics of different scale weather systems, and the dimensionality reduction plane octree coding is designed to eliminate redundancy. Lossless compression technology based on triple matrix pointer symmetric cryptographic table is designed on the premise of not changing the data volume. The calculation results of ECMWF numerical weather prediction data from July to December 2021 show that the average lossless compression rate is 0.21%, the average lossless compression rate is 5.28%, the average total compression rate is 0.112‰, and the single data in the compressed region (0°N-45°N, 90°E-135°E) is less than 1.7Kb. Using Tiantong satellite as the communication medium, further tests conducted on Yuezhanyuke No.2 research vessel during July 19-25, 2021 show that it takes 5.2 seconds and 166 seconds on average to complete the ocean-going data transmission of regional and global single compressed data sets, and 2.4 hours to complete the ocean-going data transmission of global compressed data sets for 10 days. In addition, after calculation, the above data can be converted into 1, 32 and 1664 pieces of Beidou-3 short message data respectively. It can be seen that the algorithm has the characteristics of small compressed data storage space and fast transmission speed.

Key words: dimensionality reduction plane octree coding; Compression algorithm; Symmetric cr yptographic table

1. Introduction

At present, China's Beidou-3 communication satellites and Tiantong-1 communication satellites have achieved global networking(Yan, X. et al. 2019), providing link support for global information transmission. Among them, the short message function of Beidou satellite has been widely used in meteorological data transmission, forecast and early warning information release, disaster weather monitoring and other occasions in Marine and remote areas with poor communication capability(Hu, J. et al. 2023; Li, Z. et al. 2023; Liu, B. et al. 2023), which can realize one-to-one or one-to-many communication in emergency situations(Gao,Y. et al. 2024), and the information transmission scope covers the whole world. However, the short message system of Beidou-3 has the problems of limited words and

frequency(Hu, J. et al. 2023; Hu, R. et al. 2019), Beidou-3 utilizes GEO satellites to provide the regional short message communication service to users in China and surrounding areas with Maximum length of a single message is 14000bit (approximately equivalent to 1000 Chinese characters), and Beidou-3 utilizes MEO satellites to provide global short message communication services to authorized users located on or 1000 km above the Earth's surfacewith Maximum length of a single message is 560bit (approximately equivalent to 40 Chinese characters)(Lu, J. et al. 2020; Jiang, K. et al. 2025). Tiantong-1 satellite communication system is widely used in various fields, including distance education, satellite broadcasting, agricultural monitoring, ocean monitoring and disaster warning, etc, With two-way communication function, it can realize the rapid transmission and sharing of information, and the communication covers the areas except the two latitudes above 76 degrees,The maximum transmission rate is 384Kbps, that is, 48Kb/s, the number of users has reached millions, and the user link works in the S-band, which is less affected by rain failure(Ji, M. 2018), and is suitable for safe and real-time data transmission under complex sea conditions. However, it is worth noting that Beidou satellite and Tiantong satellite have advantages and disadvantages in communication capabilities and use. Beidou satellite communication and navigation system has advantages of low cost and high security, but its data transmission rate and communication frequency are low, which cannot meet the data rate requirements of Kbit/s(Gao, Y. et al. 2024). Tiantong-1 satellite has a high transmission rate,The Network can be connected to the Public Switched Telephone Network (PSTN) and the Public Land Mobile Network (PLMN), such as implementation of 3G communication, however ,the device has high power consumption and large delay(Xiao, N. et al. 2018). Giving full play to the advantages of the two has important scientific significance and practical application value.

The requirements of large storage space and high transmission time are the significant characteristics of meteorological data. With the rapid development of global meteorological numerical forecasting technology, the spatio-temporal resolution of grid forecast data accurately describing atmospheric state is getting higher and higher, and the data storage is also showing exponential growth. For example,the European Centre for Medium-Range Weather Forecasts (ECMWF) has a land surface model spatial resolution of 0.125° and a10-day total forecast data volume of about 15G per start time. The Global Assimilation Prediction System developed by the CMA Earth System Modeling and Prediction Center (CMA-GFS) has a spatial resolution of 0.125° and a 10-day total forecast data volume of about 120G per start time. On the other hand, the navigation of ships at sea needs meteorological forecasting

products to provide scientific and technological support and service guarantee. From the comparison of data volume and satellite communication capability, it can be seen that although more and more ocean-going ships have begun to receive meteorological information around the world through email, voice, short message, data communication and other means, due to bandwidth and communication capability limitations, only simple data exchange can be completed, and ocean-going transmission of meteorological numerical forecast data with massive data cannot be realized(Xingyuan Y. et al. 2019). Data analysis has found that there are a large number of redundant data that are repeated or irrelevant to the performance of weather systems in numerical forecast data. It can be seen that how to compress and store massive numerical prediction data effectively has become one of the important research directions at present. Therefore, this study combines the data compression theory with the spatial scale characteristics of weather system, and designs a lossless compression algorithm of dimensionality reduction plane octree encoding and a lossless compression algorithm of triple matrix pointer symmetric cryptographic table, which realizes the large scale compression of numerical forecast data.

2. Data and compression technology

Data compression algorithms are divided into two types: lossless compression and lossy compression(Tintó Prims, O. et al. 2023). Lossless compression is a compression algorithm that reduces the size of data files without reducing data quality. It can retain all information in the source data, but the compression ratio is not high. Common algorithms include Huffman coding, Lempel-Ziv-Welch(LZW) compression, Run-Length Coding (RLE) compression, L777 and L888 compression algorithms(Mohammadi, H. et al. 2022; Mahammad, F. S. et al. 2020; Franklin, M. et al. 2022). Lossy compression means that after using compressed data for reconstruction, the obtained data is different from the source data, but it will not cause people to misunderstand the information expressed by the source data. Lossy compression can obtain a higher compression ratio and is suitable for the original data. The data has too high accuracy and density, and there is redundant information(Smith, J. et al. 2012; Tintó Prims, O. et al. 2023; Yu, Z. et al. 2021).

This paper designs a big data compression algorithm combining lossy compression and lossless compression, a dimensionality reduction planar octree coding lossy compression algorithm and a lossless compression algorithm of a triple matrix pointer symmetric cryptographic table(Ryabko, B. 2023) (FIG. 1).

The selected data are ECMWF Global High Resolution Numerical Prediction Model data (0.125°X0.125°) (hereinafter referred to as ECMWF Numerical Prediction Model).

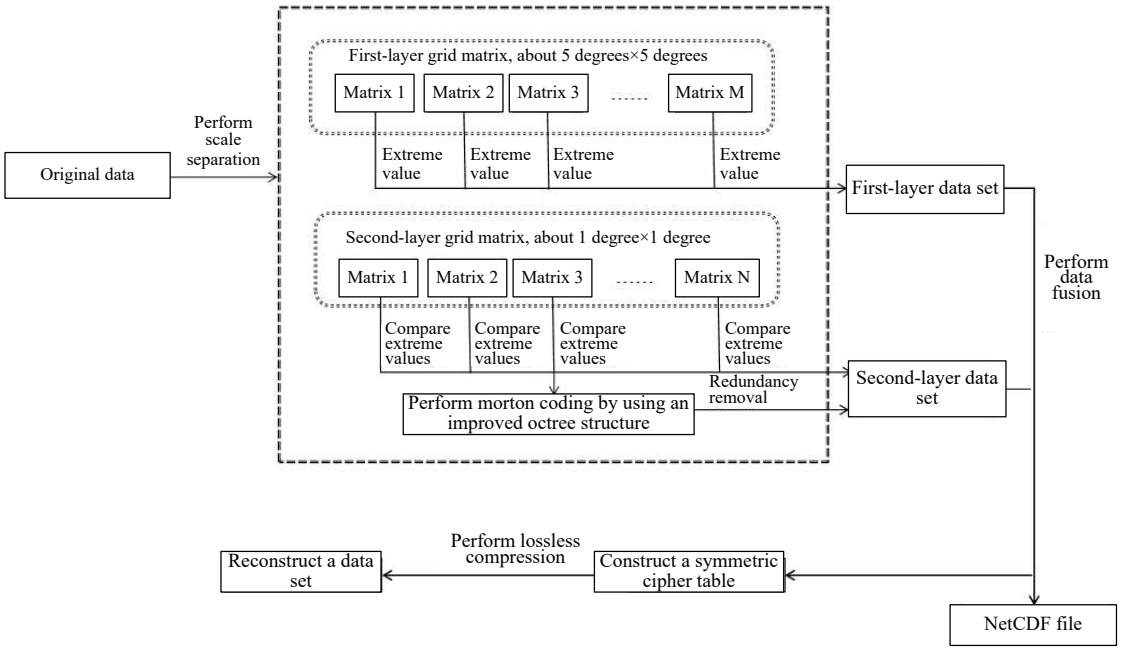


Fig. 1. Data flow of big data compression algorithm.

a. Primary lossy compression

1) DATA PROCESSING AND SCALE SEPARATION

Point cloud DATA is a discrete data set represented in a three-dimensional coordinate system, consisting of a series of three-dimensional coordinates {(X,Y,Z),DATA}. ECMWF Numerical Prediction Model data is a typical point cloud data with this feature, and the three-dimensional coordinates have a strict correspondence with the geographic coordinates in both horizontal and vertical directions. The distribution of its data reflects the spatial scale and intensity characteristics of the weather system. When designing the data compression scheme, this paper fully considers the spatial scale characteristics of the weather system on the basis of the data compression theory, and establishes the point cloud big data compression scheme for the purpose of capturing the weather system. The point cloud DATA used in this paper has only one vertical layer, namely {(X,Y,Z=1), DATA}. Firstly, the multi-layer grid nesting scheme is used to decompose spatial data and obtain the hash grid dataset of weather systems at different scales (FIG. 2). Then the redundant information is removed by using reduced dimension octree coding. Finally, the point cloud void generated by the compressed data is processed by the data fusion method, and the lossy compressed hash grid data set is obtained. The data set greatly

saves physical storage space while reducing data transfer time. The technical flow of lossy data compression scheme is shown in FIG. 3.Paste text here, if needed.

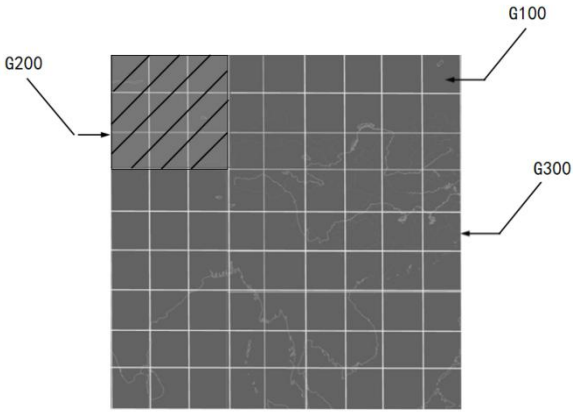


Fig. 2.schematic diagram of three-layer grid nesting

(First layer grid G100, second layer grid G200, third layer grid G300)

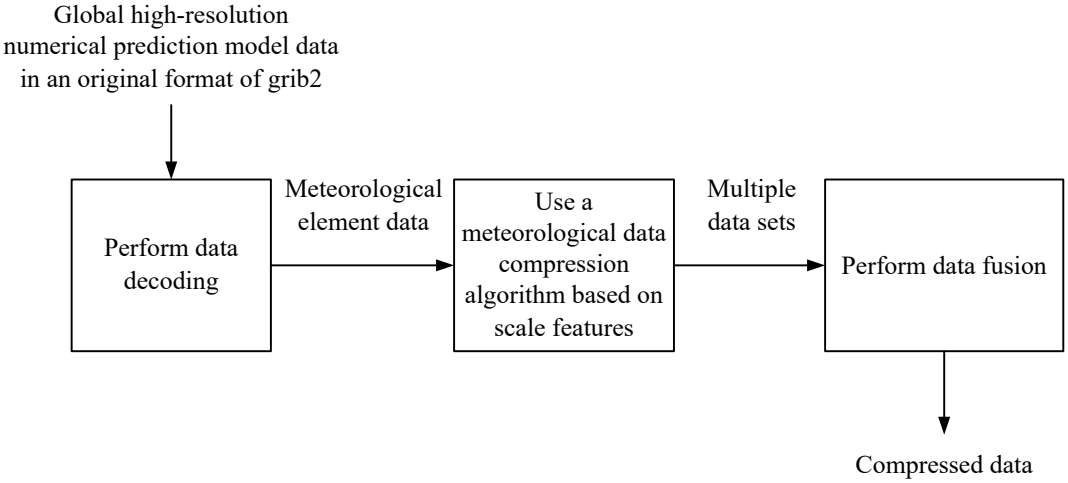


Fig. 3.Technical flow of lossy compression scheme

In this paper, data processing refers to the decoding of scalar ECMWF Numerical Prediction Model data satisfying the two conditions of single factor and continuous change, such as mean sea level pressure (PRMSL), 2m temperature, relative humidity, etc. Considering the urgent need of mean sea level pressure data for ocean-going ships, the point cloud big data compression scheme in this paper takes mean sea level pressure as the research object.

Scale separation is to construct the data matrix according to the difference of the horizontal range of the weather system, and design the corresponding big data compression scheme. In

this study, the scale separation of synoptic scale and mesoscale weather systems is mainly carried out, which is divided into the following steps:

(1)Based on the spatial resolution of the original ECMWF Numerical Prediction Model data fbl_in, the horizontal grid number num_lon and num_lat, and the positioning information of the central point, the initial latitude and longitude grid positioning information was calculated, recorded as one-layer grid G100, and the sea level pressure data of all grid points were read and assigned to PRMSL according to formula (1). Store in the im_data matrix.

$$\begin{aligned} &\mathrm{PRMSL}(m,n) = \mathrm{round}(\mathrm{PRMSL}(m,n)) \\ &m = 180/\mathrm{fbl_in}, \quad n = 360/\mathrm{fbl_in} \end{aligned} \tag{1}$$

round() means round.

(2) The grid boundary information equivalent to $SCALE = SS$ and $SCALE = LS$ in the horizontal direction is obtained respectively. The method is to select the smallest positive integer value $Interal(i)$ near $[\mathrm{SCALE}/(\mathrm{fbl_in} \times 100)]$, while satisfying the following conditions:

$$\begin{aligned} &\mathrm{Interal(i)} = [\mathrm{SCALE}/(\mathrm{fbl_in} \times 100)] \pm \mathrm{i}, \mathrm{i} = 1,2.....\mathrm{N} \\ &\mathrm{num_lon\%Interal(i)} = 0 \\ &\mathrm{num_lat\%Interal(i)} = 0 \end{aligned} \tag{2}$$

Set $\mathrm{ITL} = \mathrm{Interal(i)}$ to obtain grid boundary information equivalent to $\mathrm{SCALE} = \mathrm{SS}$ and $\mathrm{SCALE} = \mathrm{LS}$, denoted as two-layer grid G200 and three-layer grid G300, they form three layers grid nested with G100. It is found that G200 and G300 are both upscaled grids of G100, that is, each G200 and G300 grid is composed of several G100 grids. Where LS and SS represent the synoptic scale and the mesoscale weather system, respectively, and the unit is km. In the formula, % is the residual operator, and [] represents rounding by removing decimals.

(2) EXTRACTION OF CORE EIGENVALUES

After calculation, it is found that there are obvious differences in the amount of embedded PRMSL grid matrix data (G100 grid) in the above different scale grid (G200 grid and G300 grid), and there are also great differences in the range of numerical intervals. Therefore, this

paper establishes core meteorological characteristic value extraction schemes according to the differences in the amount of embedded matrix data in the grids, as follows:

For the G300 grid, we sort the PRMSL data values of the G100 grid matrix contained in each grid, obtain the MAX and MIN values, and record the corresponding latitude and longitude information in the G100 grid. Traverse the entire G300 grid and obtain the PRMSL point cloud dataset of the irregular grid matrix corresponding to the SCALE=LS scale, which is denoted as the hash grid dataset DDLarge.

For G200 grid, the feature value extraction scheme of G300 grid is repeated, and a set of hash grid data set corresponding to $SCALE = SS$ is also obtained. After calculation, it is found that there are a large number of $MAX = MIN$ cases in the G200 grid corresponding to this data set. Therefore, it is necessary to optimize the core meteorological feature value extraction scheme under G200 grid. Specifically, when $MAX(PRMSL) = MIN(PRMSL)$ is encountered in a G200 grid, the PRMSL data value of the center of this grid is extracted, and the latitude and longitude information and PRMSL grid point value of the data in a layer of grid G100 are recorded. Therefore, a set of irregular grid matrix PRMSL point cloud data set corresponding to G200 grid is generated, which is denoted as hash grid intermediate data set DDSmallTMP, and a compression scheme is further designed.

3) DIMENSION REDUCTION PLANE OCTREE CODING

After extracting the eigenvalues of the above scheme, it is found that a large number of the same data are still piled up in the dataset DDSmallTMP, so it is necessary to further design a redundancy compression scheme to highlight the expression of the eigenvalues of the weather system and compress the data volume.

Octree is a tree-like data structure used to describe three-dimensional space. Each node represents the volume element of a cube, and each node has eight child nodes. Octree coding uses Morton coding to represent leaf node with data and the path from root node to leaf node in three-dimensional space, and regards each three-dimensional point as a volume element(Liu, D. et al. 2020).

In this paper, arbitrary grid points in DDSmallTMP data set are taken as nodes, eight sub-nodes are constructed to form a dimensionally reduced plane octree structure of horizontal structure, and Morton coding is performed on this dimensionally reduced plane octree structure to obtain the octree coding of each node. After removing the point information encoded by octree as

‘00000000’, a set of irregular grid PRMSL dataset DDSmall is obtained, which realizes the de-duplication and de-redundancy of DDSmallTMP dataset.

Considering that the amount of data contained in each G200 grid corresponding to DDSmallTMP dataset is 1 or 2, the point closest to the node is selected to participate in Morton coding when constructing child nodes. As shown in FIG. 4, when Morton coding is carried out using the dimensionally reduced plane octree structure, for any *node*(i, j) in DDSmallTMP data set, the eight positions of upper left, upper, upper right, left, right, lower left, lower and lower right are defined as $childnode(i, j,1)$、$childnode(i, j,2)$、$childnode(i, j,3)$、$childnode(i, j,4)$、$childnode(i, j,5)$、$childnode(i, j,6)$、$childnode(i, j,7)$ and $childnode(i, j,8)$, respectively. The corresponding codes of each child node are $cd(i, j,1)$、$cd(i, j,2)$、$cd(i, j,3)$、$cd(i, j,4)$、$cd(i, j,5)$、$cd(i, j,6)$、$cd(i, j,7)$ and $cd(i, j,8)$. Define the code of this node as $CODE(i, j)$, then:

$$value(i, j, n) = PRMSL(i, j, n) - PRMSL(childnode(i, j, n)), n \in [1,8]$$

$$If\ value(i, j, n) = 0,\ then\ cd(i, j, n) = 0,\ else\ cd(i, j, n) = 1$$

$$CODE(i, j) = \prod_{n=1}^{8} cd(i, j, n) \tag{3}$$

That is, the code of a node is equal to the sequence of the code of the child nodes in the eight directions of the upper left, upper, upper right, left, right, lower left, lower, lower right. Data points with child node <8 are not processed.

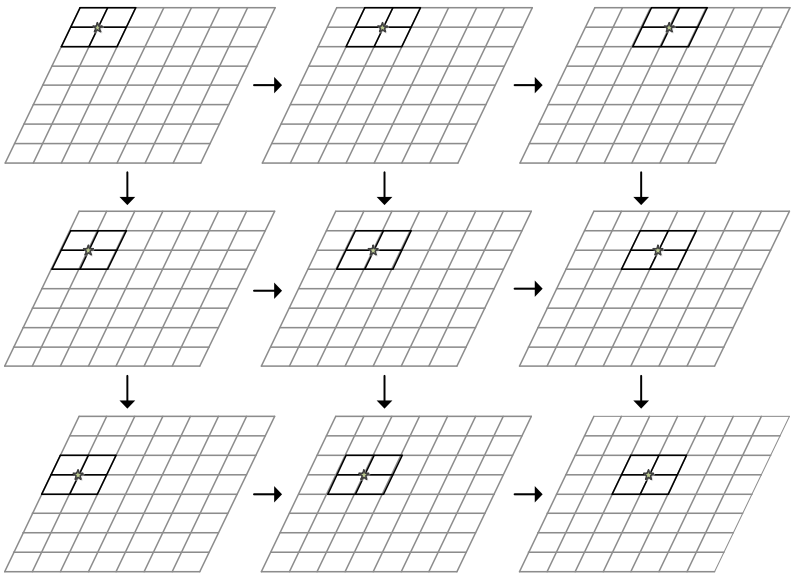


Fig. 4. Schematic diagram of Reduced dimensional plane octree structure

FIG. 5 illustrates the result of the layer-by-layer encoding of nodes 1023hPa, 1022hPa, and 1023hPa. The corresponding child nodes codes are (1,0,1,1,0,0,0, 0), (1,1,0,1,1,1,1,1), (1, 0, 0, 0, 0, 0, 1), and the Morton codes of each node obtained are CODE='10110000', '11011111', '10000001', respectively. By the same token, the Morton CODE of the entire DDSmallTMP dataset can be obtained. After removing the point information of CODE='00000000', the irregular hash grid dataset DDSmall after redundancy is removed can be obtained.

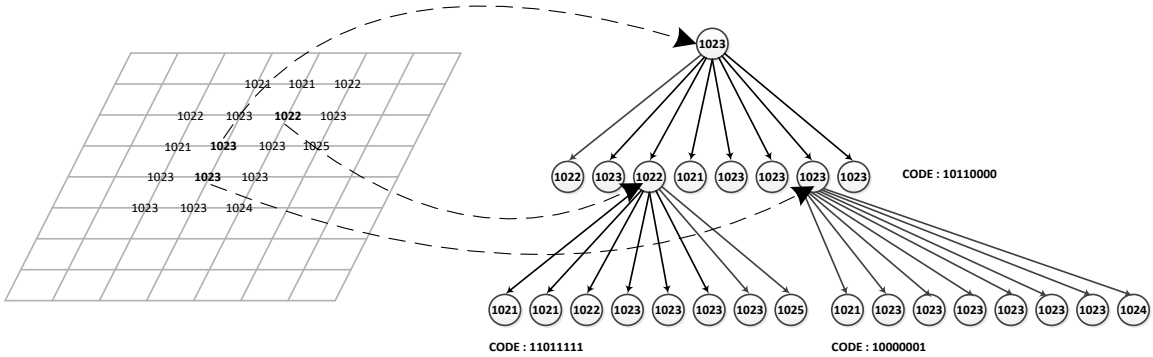


FIG. 5 The coding case of dimensionality reduction plane octree represented by Morton code

4) DATA FUSION AND POINT CLOUD DATA HOLES PROCESSING

Large point cloud data holes are likely to appear in the DDSmall data set obtained by using dimensionality reduction plane octree coding to remove redundancy. When using this data set for isobar analysis, weather system deformation or weather system boundary error will occur. Therefore, it is necessary to carry out special processing on the point cloud data holes of DDSmall data set. This paper mainly deals with the situation that the amount of data in the grid is less than 2 after DDSmall data set covers G300 grid. Specifically, First, the DDSmall data set is spatially positioned and embedded on the G300 grid. Then, the whole G300 grid is traversed. For the grid with less than 2 data, fill the corresponding DDLarge data set with data, and the duplicate data is filtered out. This data set is the final data set after lossy compression, converted into the required format and stored for shore use.

b.Secondary lossless compression

In order to shorten the satellite communication time, ensure the data transmission efficiency and reduce the packet loss rate, a triple matrix pointer symmetric cipher table is designed to lossless compress the lossy compression DDate data set mentioned above.

1) PRESSURE VALUE TRANSFORMATION

Due to $PRMSL \in (600hPa, 1400hPa)$

Let $PRMSL = PRMSL - 500$ (4)

Get PRMSL（100hPa，900hPa）,The value of PRMSL changes from a 4-bit integer to a 3-bit integer, and the data storage capacity is reduced.

2) MATRIX SLICE

Define the global grid matrix QQ with latitude and longitude as coordinates, slice the matrix QQ according to the interval range of 45°*45° (FIG. 6), divide it into $4*8=32$ regions, and represent by QB (i,j) code, where $i \in [1,4]$, $j \in [1,8]$.

Let $QB(i,j) = i \times 10 + j$

For example, $QB(2,7) = 27$, representing the range (0-45°N, 90-135°E) (FIG. 6).

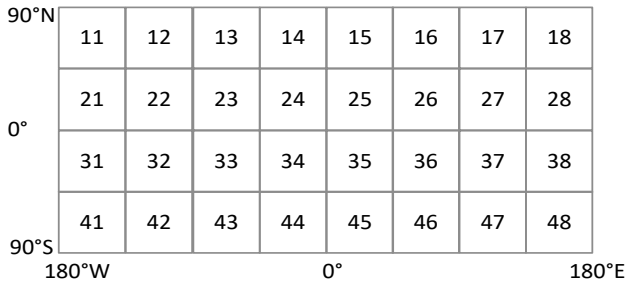


FIG. 6 Global grid matrix slicing and region coding

3)TRIPLE MATRIX DEFINITION AND SYMMETRIC CRYPTOGRAPHIC TABLE DESIGN

Symmetric cryptography is a cryptosystem in which the same key is used for encryption and decryption. The decryption algorithm is the inverse process of the encryption algorithm, using substitution and permutation techniques. In this paper, a lossless compression algorithm based on symmetric cryptographic table is designed. After the compressed data is transmitted to the ship by satellite, it is extracted and restored to the required data format. The specific algorithm is as follows:

Take LS=500km, SS=100km, and encode QB(i,j) in any region to design symmetric cryptographic table and establish data compression algorithm.

Let MTX_G300 represents the G300 grid matrix contained in the QB(i,j) region (5°* 5°), MTX_G200 represents the G200 grid matrix contained in a G300 grid matrix (1°* 1°),

MTX_G100 represents the G100 grid matrix contained in a G200 grid matrix(0.125°*0.125°), The grid matrix sizes of each layer are shown in Table 1.

	MTX_G300	MTX_G200	MTX_G100
Matrix size	9*9	5*5	8*8
Resolution	5°* 5°	1°* 1°	0.125°* 0.125°
Numeric digit	2	2	2
Number of bits	16	16	16

Table 1. QB(i,j) region matrix size and number of bits (45°*45°)

The MTX_G100, MTX_G200 and MTX_G300 matrix values are defined as follows:

$$MTX_G300 = \begin{vmatrix} 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 \\ 21 & 22 & 23 & 24 & 25 & 26 & 27 & 28 & 29 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 91 & 92 & 93 & 94 & 95 & 96 & 97 & 98 & 99 \end{vmatrix}$$

$$MTX_G200 = \begin{vmatrix} 11 & 12 & 13 & 14 & 15 \\ 21 & 22 & 23 & 24 & 25 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 51 & 52 & 53 & 54 & 55 \end{vmatrix}$$

$$MTX_G300 = \begin{vmatrix} 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \\ 21 & 22 & 23 & 24 & 25 & 26 & 27 & 28 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 81 & 82 & 83 & 84 & 85 & 86 & 87 & 88 \end{vmatrix} \quad (5)$$

The pointer symmetric cryptographic table is designed as follows:

When PRMSL(m,n) is non empty, let:

$$a1(m,n) = L_G100(PRMSL(m,n))$$

279 $$a2(m,n) = L_G200(PRMSL(m,n))$$

280 $$a3(m,n) = L_G300(PRMSL(m,n))$$

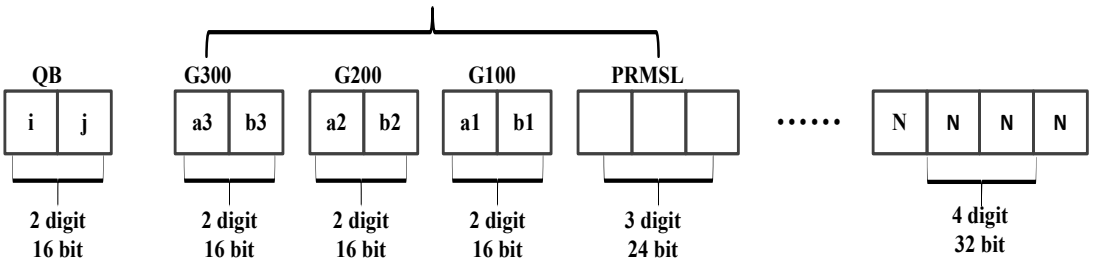
281 $$b1(m,n) = C_G100(PRMSL(m,n))$$

282 $$b2(m,n) = C_G200(PRMSL(m,n))$$

283 $$b3(m,n) = C_G300(PRMSL(m,n))$$

284 a1(m,n), a2(m,n) and a3(m,n) are the rows of PRMSL(m,n) in the MTX_G100, MTX_G200
285 and MTX_G300 matrices, respectively. b1(m,n), b2(m,n) and b3(m,n) are the columns of
286 PRMSL(m,n) in the MTX_G100, MTX_G200 and MTX_G300 matrices, respectively.

287 The compressed data stream is arranged in the following order, with data endings identified
288 by NNNN(FIG.7):



289

290 FIG. 7 Format diagram of compressed data information flow

291 Specifically:

292 QB(i, j)

293 $$MTX_G300(a3(1,1), b3(1,1))$$

294 $$MTX_G200(a2(1,1), b2(1,1))$$

295 $$MTX_G100(a1(1,1), b1(1,1))$$ PRMSL(1,1)

296 $$MTX_G300(a3(1,2), b3(1,2))$$

297 $$MTX_G200(a2(1,2), b2(1,2))$$

298 $$MTX_G100(a1(1,2), b1(1,2))$$ *PRMSL*(1,2)

299 .....

300 $$MTX_G300(a3(2,1), b3(2,1))$$

MTX _ G200(a2(2,1), b2(2,1))

MTX _ G100(a1(2,1), b1(2,1)) PRMSL(2,1)

MTX _ G300(a3(2,2), b3(2,2))

MTX _ G200(a2(2,2), b2(2,2))

MTX _ G100(a1(2,2), b1(2,2)) PRMSL(2,2)

.....

MTX _ G300(a3(m,1), b3(m,1))

MTX _ G200(a2(m,1), b2(m,1))

MTX _ G100(a1(m,1), b1(m,1)) PRMSL(m,1)

MTX _ G300(a3(m,2), b3(m,2))

MTX _ G200(a2(m,2), b2(m,2))

MTX _ G100(a1(m,2), b1(m,2)) PRMSL(m,2)

.....

MTX _ G300(a3(m, n), b3(m, n))

MTX _ G200(a2(m, n), b2(m, n))

MTX _ G100(a1(m, n), b1(m, n)) PRMSL(m, n)

NNNN

Among them, PRMSL (m, n) are all non empty, otherwise they will not participate in encoding. The above data is transmitted to the ship by satellite, decrypted by the symmetric cryptographic table and converted into the required data format.

3. Compression effect

a. Compression ratio

Compression ratio is the ratio of the storage space occupied by data after compression to that occupied before compression. According to the above compression algorithm, the ECMWF Numerical Prediction Model data from July to December 2021 are calculated. It is found that the size of the single file of ECMWF Numerical Prediction Model data in the original GRIB2

format is about 280Mb, which is about 13.9Mb after single element extraction by PRMSL(QB(2,7)). It is 29-31Kb after primary lossy compression, with an average lossy compression rate of 0.21%. After secondary lossless compression,it is about 1.6Kb, and the average lossless compression ratio is 5.28%. The total compression rate of data remained stable between 0.1‰ and 0.13‰, and the average total compression rate was 0.112‰, that is, the total data volume was reduced by nearly 10,000 times after two compressions (Table 2).

Data series	PRMSL data bytes（Kb）	lossy compression bytes(Kb)	Lossless compression bytes(Kb)	Compression ratio Lose%	Compression ratio Loseless%	Total compression ratio ‰
1	13900	29.0	1.63	0.209	5.6	0.12
2	13900	28.6	1.46	0.206	5.1	0.11
3	13900	29.4	1.67	0.211	5.68	0.12
4	13900	30.5	1.50	0.219	4.92	0.11
5	13900	28.8	1.48	0.207	5.14	0.11
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Average lossy compression ratio: 0.21% Average lossless compression ratio: 5.28% Average total compression ratio：0.112‰						

Table 2. Data compression rate

4. Scientific research ship test

Relying on the Yue Zhan Yu Ke No.2 research ship, carrying the Tiantong-1 satellite mobile communication terminal, a four-day ocean-going data transmission test was carried out on the coast of Zhanjiang and the South China Sea. Using the S-band link resources of China's Tiantong-1 communication satellite, the Marine mobile communication terminal is carried on the Yue Zhan Yu Ke No.2 research ship to realize two-way real-time data communication between the research ship and the shore. Among them, the forward link is the satellite to the research ship, mainly through the command information to transmit satellite data; The reverse link is from the research ship to the satellite, and the satellite data is uploaded through the antenna of the communication terminal. The selected data are ECMWF Numerical Forecast

Model data from July 19-25, 2021 and corresponding primary lossy and secondary lossless compression data (QB(2,7)). It is found that the reception of satellite signals by ground mobile communication terminals is greatly affected by the surrounding environment, and there are phenomena such as signal interruption and too long transmission time in the near shore, while the success rate of short data transmission is significantly higher than that of long data, and the signal sending and receiving in open water is significantly more stable than that in the near shore, and the data transmission capacity in the ocean is significantly better than that in the near shore, meeting the requirements of stable signal transmission in the ocean.

As show in Table 3, the uncompressed ECMWF Numerical Prediction Model data has a large difference in the data transmission time between the two ends of the ship and the shore by Tiantong-1 communication satellite, 578.2 seconds on average, with multiple data outages and transmission failures. The average transmission time of a lossy compressed data is 12.4 seconds, and the transmission time is still different. After lossless compression, the average data transmission time is only 5.2 seconds, without transmission failure and data transmission is stable.

The test results show that it takes 5.2 seconds and 166 seconds on average to complete the ocean data transmission of regional and global single compressed data sets, and 2.4 hours to complete the ocean data transmission of 10-day global compressed data sets. In addition, after calculation, the above data can be converted into 1, 32 and 1664 Beidou-3 short message data respectively (each short message within the region is 14000bit/min). The corresponding data transmission time is 1 minute, 52 minutes and 28 hours respectively, and the timeliness fully meets the needs of users.

data series	category	position	time	number of bytes	duration（s）	average duration（s）
01		ZJ	20211018	13.9Mb	393	
01		ZJ	20211018	13.9Mb	3600+, interrupt	
	Uncompression					578.2
02		ZJ	20211018	13.9Mb	401	
03		ZJ	20211018	13.9Mb	interrupt	

04		SCS	20211019	13.9Mb	785	
04		SCS	20211019	13.9Mb	858	
05		SCS	20211020	13.9Mb	454	
		⋮	⋮	⋮	⋮	
11		ZJ	20211018	29Kb	38	
12		ZJ	20211019	28.6Kb	7	
13	Lossy	SCS	20211019	29.4Kb	8	12.4
14	compression	SCS	20211020	30.5Kb	3	
15		SCS	20211021	28.8Kb	6	
		⋮	⋮	⋮	⋮	
21		ZJ	20211018	1.63Kb	6	
22		SCS	20211019	1.46Kb	3	
23	Lossless	SCS	20211019	1.67Kb	8	5.2
24	compression	SCS	20211020	1.5Kb	7	
25		SCS	20211021	1.48Kb	2	
		⋮	⋮	⋮	⋮	

Table 3. Scientific test of data transmission.

Data range: Northwest Pacific Ocean, 0-45°N, 90-135°E. elements:PRMSL(mean sea level pressure). ZJ:Zhanjiang ,SCS:The South China Sea

5. Conclusions

In this paper, based on synoptic scale characteristics, we design dimensionality reduction plane octree coding, triple matrix nesting and pointer symmetric cryptographic table, and propose a big data compression algorithm combining lossless compression and lossless compression, which realizes large scale compression of ECMWF Numerical Prediction Model data. According to the calculation, the file storage of compressed data is less than 1.7Kb (QB(2,7) area), and the total compression rate of data is stable at 0.1‰-0.13‰. The stability, reliability and timeliness of data transmission between shore and ship before and after data compression are compared by using Tiantong-1 communication satellite through field science tests. It is

found that the average transmission time of ECMWF Numerical Forecast Model data in QB(2,7) region is only 5.2 seconds after compression, and it is stable and reliable.

According to a total of 52 forecast files of ECMWF numerical forecast data in 10 days, it takes 270 seconds and 2.4 hours to transmit compressed regional and global data using Tiantong-1 communication satellite as the medium, and 32 and 1664 pieces of data using Beidou-3 communication satellite short message system as the medium to transmit the above data, respectively. And the timeliness fully meets the needs of ocean communication. The data transmitted from the shore end to the ship end can be decrypted by the symmetric cryptographic table for data analysis, graphic drawing, system integration, etc., so as to realize the practical application of the ship end numerical forecast product. The invention patent applied for according to this algorithm has been authorized by the State Intellectual Property Office of China.

Data Availability Statement

Data sets generated during the current study are available from the corresponding author on reasonable request. The compressed data can be obtained from J. Zhang but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Competing Interests

The author declares no competing interests.

Author Contributions

Ji Zhang:Writing – Original draft, Conceptualization, Methodology, Investigation, Formal Analysis.

Caijun yue:Writing – Review & Editing, Supervision.

Liang Guan:Writing – Original draft, Software.

Zhiqiang Chen:Writing – Review & Editing, Supervision.

Intellectual Property Statement

The meteorological big data compression technology proposed in this manuscript aims to address the transmission bottleneck caused by limited and unstable bandwidth in offshore satellite communications for marine weather forecasting. It has been granted two invention patents in China and the Netherlands, respectively.

1. Chinese Granted Invention Patent

Granted Patent No.: CN116683915B

Title:METEOROLOGICAL DATA COMPRESSION METHOD AND SYSTEM, ANDMEDIUM

Grant Date: 13/02/2024

Assignee: Shanghai Marine Meteorological Center

2. Dutch Granted National Patent

Patent No.: NL2037367

Title:METEOROLOGICAL DATA COMPRESSION METHOD AND SYSTEM, ANDMEDIUM

Grant Date: 25-07-2025

Issuing Office: Octrooicentrum Nederland

Assignee: Shanghai Marine Meteorological Center te Shanghai, China, CN

All authors of this manuscript, with the exception of Liang Guan, are listed as co-inventors of the aforementioned two patents. The core compression algorithm and transmission optimization scheme described in the paper fall entirely within the scope of the technical claims of these two granted patents. There are no other pending patent applications, licensing agreements, intellectual property disputes, or undisclosed proprietary rights related to this work. All authors confirm that all relevant intellectual property assets have been fully and transparently disclosed.

Funding

This work was supported by National Natural Science Foundation of China (Grant Nos. 41875059, 41875071).

REFERENCES

Franklin M., Ribeiro Junior, Reinaldo A.C., et al., 2022: Data reduction based on machine learning algorithms for fog computing in IoT smart agriculture. Biosystems Engineering, 223(B), 142-158. https://doi.org/10.1016/j.biosystemseng.2021.12.021.

Gao, Y., Zhang, Z., Wang, J., et al., 2024: Research on the meteorological data transmission system of civil aircraft based on BeiDou satellite. Journal of Physics: Conference Series, 2879. https://doi.org/10.1088/1742-6596/2879/1/012059.

Hu, R., Ju, L., & Chen, P., 2019: A security transmission system for Beidou short message based on SM9. J. Phys.: Conf. Ser., 1345. https://doi.org/10.1088/1742-6596/1345/2/022014.

Hu, J., Hong, Y., Jin, Q. et al., 2023: An Efficient Dual-Stage Compression Model for Maritime Safety Information Based on BeiDou Short-Message Communication. J. Mar. Sci. Eng., 11, 1521. https://doi.org/ 10.3390/jmse11081521.

Ji, M., 2018: Analysis of time and application of Tiantong-1 satellite mobile communication system. Satellite and Network(Chinese), 4, 42-43.

Jiang, K., Zhang, J., Xie, X., et al., 2025: Real-time precise orbit determination of LEO satellite based on BDS-3 global short message communication: simulation and verification. Meas. Sci. Technol., 36(6). https://doi.org/10.1088/1361-6501/add9b8.

Li, Z., Huang, J., Zhang, J. et al., 2023: Design of Power Amplifiers for BDS-3 Terminal Based on InGaP/GaAs HBT MMIC and LGA Technology. Micromachines, 14, 1995. https://doi.org/10.3390/ mi14111995.

Liu, B., Liu, A., Chen, L., et al., 2023: Smartphone-Based Positioning Augmented by BDSBAS Ionospheric Corrections. IEEE Geosci. Remote Sens. Lett., 20, 1-5. https://doi.org/10.1109/lgrs.2023.3257317.

Liu, D., Wang, T., Li, X. et al., 2020: A Multiresolution Vector Data Compression Algorithm Based on Space Division. ISPRS Int. J. Geo-Inf., 9(12), 721. https://doi.org/10.3390/ijgi9120721.

Lu, J., Guo, X., & Su, C., 2020: Global capabilities of BeiDou Navigation Satellite System. Lu et al. Satell Navig., 1, 27. https://doi.org/10.1186/s43020-020-00025-9.

Mahammad, F. S., & Viswanatham, V. M., 2020: Performance analysis of data compression algorithms for heterogeneous architecture through parallel approach. J Supercomput, 76, 2275–2288. https://doi.org/10.1007/s11227-018-2478-3.

Mohammadi, H., Ghaderzadeh, A., & Sheikh Ahmadi, A., 2022: A Novel Hybrid Medical Data Compression Using Huffman Coding and LZW in IoT. IETE Journal of Research, 69(11), 7831–7845. https://doi.org/10.1080/03772063.2022.2052985.

Painsky, A., Rosset, S., 2019: Lossless Compression of Random Forests. J. Comput. Sci. Technol., 34, 494-506. https://doi.org/10.1007/s11390-019-1921-0.

Ryabko, B., 2023: Unconditionally secure short key ciphers based on data compression and randomization. Des. Codes Cryptogr., 91, 2201–2212. https://doi.org/10.1007/s10623-023-01195-8.

Smith, J., Petrova, G., & Schaefer, S., 2012: Progressive encoding and compression of surfaces generated from point cloud data. Computers & Graphics, 36(5), 341-348. https://doi.org/10.1016/j.cag.2012.03.032.

Tintó Prims, O., Redl, R., Rautenhaus, M. et al., 2023: The effect of lossy compression of numerical weather prediction data on data analysis: a case study using enstools-compression 2023.11. Geosci. Model Dev., 17, 8909–8925. https://doi.org/10.5194/gmd-17-8909-2024.

Xiao, N., Wang, L., Zhang, Q., et al., 2018: GEOMulti-signal station synchronization technology in satellite mobile communication system. Radio communication technology(Chinese), 44(05), 453-457.

Yan, X., Huang, G., Zhang, Q., et al., 2019: Early analysis of precise orbit and clock offset determination for the satellites of the global BeiDou-3 system. Advances in Space Research, 3, 1270-1279, https://doi.org/10.1016/j.asr.2018.10.038.

Yu, Z., Li, D., Zhang, Z. et al., 2021: Lossy compression of Earth system model data based on a hierarchical tensor with Adaptive-HGFDR (v1.0). Geosci. Model Dev., 14, 875–887. https://doi.org/10.5194/gmd-14-875-2021.