

# The application of data fusion method in the analysis of ocean and meteorology observation data

## Abstract

Data fusion is a new technology of data processing for multi - source ocean meteorology data. In order to clearly understand how to apply this technology in the data processing of ocean meteorology observation, the author first divides the fusion methods into three categories, i.e., pixel-level fusion, feature-level fusion and decision-making level fusion. Then, the corresponding methods of each data fusion are discussed in detail. Afterward, the SST fusion data from China Meteorological administrative is analyzed in order to explain that the fusion data is applied in numerical forecast model on the China's Bohai sea region. Finally, the authors believe that the data fusion technology of multi-source observation data has a good application value when the observation data is processed into grid data of ocean meteorology, which can be used as the grid truth-value of numerical model inspection. In addition, the fusion technology at decision-making level can also develop with the intelligent technology, and deeply mine the application of data fusion technology to the ocean meteorology observation data.

**Keyword:** data fusion, ocean meteorology, observation data, data processing

Volume 3 Issue 3 - 2019

Wang Wei,<sup>1</sup> Liu Zhi-jie,<sup>2</sup> Zhao Yue,<sup>3</sup> Gong Xiao-qing<sup>1</sup>

<sup>1</sup>Tianjin Institute of Meteorological Science, China

<sup>2</sup>Meteorology Administrative of Xiqing District, China

<sup>3</sup>Meteorology Administrative of Jinnan District, China

**Correspondence:** Wang Wei, Tianjin Institute of Meteorological Science, Tianjin, China, Tel 086-022-23358783, Email wwei356@163.com

**Received:** May 13, 2019 | **Published:** June 03, 2019

## Introduction

Recently, data fusion technology emerges in China as a new data processing method. Because of the maturity and fast development of new observation technology, researchers can obtain more and more high-resolution real-time observation data to study the interaction between ocean and atmosphere. The technology of data fusion can help researchers quickly extract rich information of meso-scale weather from multi-source observation data. Moreover, the analysis field of integrating various observation data has significant advantages of analyzing ocean weather situation, precipitation analysis, physical quantity diagnosis and so on, and is much more objective than the data of the single atmospheric sounding because the fusion data contains comprehensive information of various atmospheric observation.<sup>1,2</sup> The fusion information of abundant meso-scale observation has a good application prospect in improving the spatial- temporal resolution of observation data, cloud information and precipitation prediction.<sup>3</sup> The fusion data can also represent the meso-scale characteristics of local weather system, and describe meso-scale 3d structure and evolution in detail, which presents fine variation characteristic of meso-scale weather factor so that researchers could discover some new meso-scale weather information.

This paper mainly discusses the feasibility of data fusion and the application of data fusion methods in ocean meteorology. At present, more and more researcher pay attention on the data fusion technology of ocean meteorology observation, but few researchers systematically discuss the data fusion theory of ocean meteorology. The author's comprehensive discussion can facilitate the application of data fusion technology in ocean meteorological observation.

## Data fusion method of two dimension ocean meteorology data

Data fusion technology is a kind of information processing technology that can automatically analyze and synthesize several observation information acquired by computer in time sequence under certain criteria, so as to complete the required decision-making and

evaluation tasks. Although the concept of data fusion started in the early 1970s, the actual progress and development of the technology happened in the 1980s, especially in recent years, which has attracted worldwide attention.<sup>4,5</sup> Data fusion technology is often applied in image processing to merge the crude and fine resolution images into one image which is used to improve the crude image texture and remain the large image coverage.<sup>6,7</sup>

At present, data fusion technology has also been applied to the grid data processing of ocean meteorological observation. The basic principle of data fusion method in ocean meteorological observation is to estimate variable X on the basis of two independent and unbiased estimates of variables X1 and X2, so as to obtain a new weighted estimate of X. The accuracy of X1 and X2 may not be equal, and the variables are allowed to have different mean square deviations so that the estimate of merged X should have a minimum mean square deviation. Data fusion methods of ocean meteorological observation can be divided into three categories: pixel-level fusion, feature-level fusion and decision-making level fusion.

## Data fusion method of pixel level

Pixel-level fusion is performed directly on the collected raw data level, and data synthesis and analysis are carried out before the original measurements of various sensors are deeply preprocessed. Pixel-level fusion usually uses the centralized fusion system to complete the fusion process. This data fusion is a low-level fusion, for example, an imaging of sensor generating validates the target attributes by processing fuzzy image of several pixels. The pixel-level fusion methods include Poisson method<sup>8</sup> and Kalman method.<sup>9,10</sup>

- i. The data fusion of Poisson equation. Data fusion method uses the Poisson equation in which S is the anomaly of the observation variable and the symbol  $\Phi$  is the anomaly of fusion data. The  $\Phi$  is calculated as follow:

$$\nabla^2 \Phi = \rho \quad (1)$$

The  $\rho$  in the above equation is defined as follow:

$$\rho = \nabla^2 S \tag{2}$$

In the application, the area of grid data is usually greater than the observation data, so the boundary values are set to neighbor value, and cloud coverage is set to missing value.

ii. The data fusion of Kalman method. Kalman method is an effective data fusion method. If the fusion data are sampled at discrete points with time sequence, data fusion of kalman method is assumed that the n-dimensional linear dynamic system and the m-dimensional linear observation system could be described by the following different equation :

$$X_k = \Phi_{k,k-1} X_{k-1} + W_{k-1} \tag{3}$$

$$Y_k = H_k X_k + V \tag{4}$$

Where,  $X_k$  is n-dimensional vector which represents the system state at moment k.  $\Phi_{k,k-1}$  is n-by-n matrix called the state transition matrix of the system, which reflects the transformation of the system from the state k-1 to the state k at sampling time.  $W_k$  is n-dimensional vector that represents the random disturbance impacting on the system at moment k, which is called model noise.  $\{W_k\}$  is generally assumed to be a sequence of gaussian white noise with a known zero mean and covariance matrix  $Q_k$ .  $Y_k$  is the m-dimensional observation vector.  $H_k$  is m-by-n observation matrix which represents the transition from the state variable  $X_k$  to the observed variable  $Y_k$ .  $V_k$  is m-dimensional observation noise vector which is assumed that  $\{V_k\}$  is a sequence of gaussian white noise with known zero mean and covariance matrix  $R_k$ .

After derivation, the recursive formula of filtering can be obtained as follows:

$$\tilde{X}_k = \Phi_{k,k-1} \tilde{X}_{k-1} + G_k [Y_k - H_k \Phi_{k,k-1} \tilde{X}_{k-1}] \tag{5}$$

$$G_k = P_k H_k^T [H_k P_k H_k^T + R_k]^{-1} \tag{6}$$

$$P_{k+1} = \Phi_{k+1,k} C_k \Phi_{k+1,k}^T + Q_k \tag{7}$$

$$C_k = (I - G_k H_k) P_k \tag{8}$$

Where,  $Q_k$  is the covariance matrix of n-by-n model noise  $W_k$ ;  $R_k$  is the covariance matrix of m-by-m observation noise  $V_k$ ;  $G_k$  is n-by-m gain matrix;  $\tilde{X}_k$  is the filtered estimation of the n-dimensional vector at the moment k;  $C_k$  is the covariance matrix of n-by-n estimated error.

As the above mentioned formula, the fusion data estimation  $\tilde{X}_k$  can be recursively calculated at each time interval according to equation  $\tilde{X}_0 = E\{X_0\}$  and given  $P_0$  by means of the known matrix  $Q_k, R_k, H_k, \Phi_{k,k-1}$  and the observed value  $Y_k$  at time k.

### Data fusion method of feature level

Feature-level fusion belongs to the middle layer fusion. It first extracts the original information from the sensor, and then comprehensively analyzes and processes the feature information. The advantage of feature-level fusion is that it could achieve considerable information compression which is conducive to real-time processing, and the extracted features are directly related to the decision analysis so the fusion results can give the best feature information required by

the decision analysis. Feature-level fusion usually adopts distributed or centralized fusion system. For example, data fusion of multi-scale grid is often used in ocean meteorological data fusion. This algorithm is based on the sequential variation method of the multi-scale grid algorithm. All the temporal-spatial information is utilized by the multi-scale grid method. The target functional on each grid is as follows:

$$J^{(n)} = \frac{1}{2} X^{(n)T} X^{(n)} + \frac{1}{2} (H^{(n)} X^{(n)} - Y^{(n)})^T O^{(n-1)} (H^{(n)} X^{(n)} - Y^{(n)}) \quad (n=1, 2, 3, \dots, N) \tag{9}$$

Where,  $Y=Y_{obs}-HX_b$ ,  $X=X_a-X_b$ ,  $O$  is the error covariance matrix of the observation field;  $X_b$  is background field vector;  $X_a$  is the analysis field vector;  $Y_{obs}$  is the observed constant vector;  $H$  is the bilinear interpolation operator from the model grid to the observation point;  $X$  represents the correction vector of model vector which is calculated from the variable assimilation system.  $Y$  is the difference between the observation field and the model field.  $n$  represents the  $n^{th}$  mesh, and  $N$  is the scale number of mesh.

$$Y^{(1)} = Y^{obs} - H^{(1)} X^b \quad (n=1) \tag{10}$$

$$Y^{(n)} = Y^{(n-1)} - H^{(n-1)} X^{(n-1)} \quad (n=2, 3, \dots, N) \tag{11}$$

The multi-scale analysis starts at the crude level. When  $n=1$ ,  $Y(1)$  is the deviation of mapping the observation and model background field to the observation position, and  $X(n-1)$  is the solution or approximate solution of  $J(n-1)$ . After solving  $J(n-1)$ ,  $X(n-1)$  is interpolated into a finer mesh at the  $n^{th}$  layer.  $Y(n)$  can be calculated by formula 11<sup>th</sup>, and  $X(n)$  can be solved by  $J(n)$  minimization. The final analysis result is the superposition of the analysis results of multi-scale grid.

$$X^a = X^b + X_L = X^b + \sum_{n=1}^N X^{(n)} \tag{12}$$

### Data fusion method of decision-making level

The data fusion of decision-making level integrates different types of sensors observing the same target, and each sensor completes basic data processing locally. The method includes preprocessing, feature extraction, recognition or judgment. Its goal is to establish a preliminary conclusion on the observed target. Then, the decision-making level fusion is made through association processing, and finally the joint inference result is obtained. The main methods contain the decision tree method and BAYES inference method.

### Available data fusion type of ocean meteorology observation data

**The fusion of radar data:** The data fusion of weather radar is a complicated process because different types of radar data need different processing method in fusion process. Any coordinate format of radar data must generate the intermediate files which are finally unified into the local rectangular coordinate format on the grid, and then the information of various radar data is fused into the grid observation by fusion method.

**The fusion of satellite data:** The data fusion of satellite data is usually to combine satellite brightness temperature data with sounding data and surface data to obtain the cloud analysis data. Another data fusion of satellite data is cloud-derived wind fusion, which is first processed into the intermediate file required by data fusion system, and then the

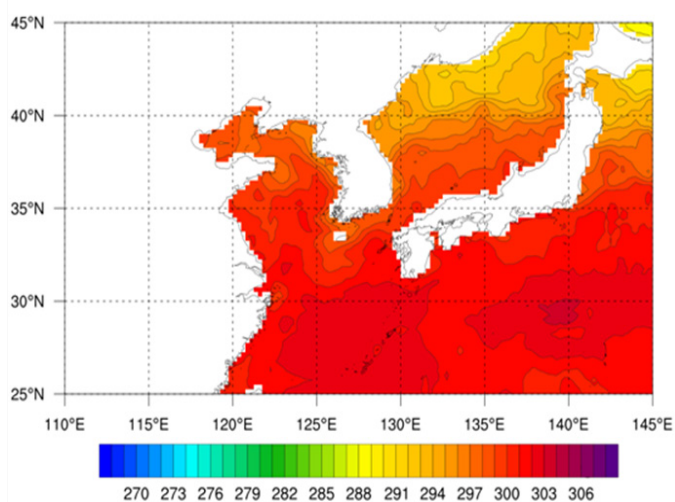
fusion wind field together with other observation data is generated by the fusion system.

**The fusion of earth-surface and sea-surface data:** The available fusion information of surface observation data is very abundant, which includes METAR, buoy/ship data, local high temporal-resolution observation data and ground-observation of GPS station. At the same time, the surface fusion data can also utilize quality control program of the Kalman filter to fill the inaccurate observed values with an estimated value which is helpful to make up data density in time sequence.

**The fusion of sounding data:** The data fusion of sounding data could merge information of radio sonde, radiometer and wind profiler. This fusion data can provide fine grid information of weather elements such as temperature, humidity and wind.

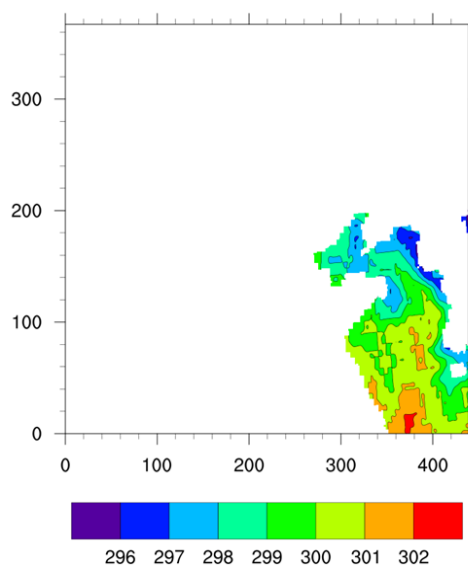
### The example of SST fusion data applied in operation

**SST fusion data of China Meteorology Administration:** The fusion system of ocean data in China Meteorological Administration (CMA) adopts the multi-scale grid fusion. In the procedure of SST fusion, deviation correction and data fusion of SST observation data are realized by means of MetOp-B/AVHRR retrieval data, GCOM-W1/AMSR2 retrieval data, FY-3C/VIRR retrieval data, buoy and ship of SST observation. The coverage of the global SST fusion products is within (90°S–90°N, 180°W and 180°E), the spatial resolution is 0.25°x0.25°, and time resolution is 1 day. Figure 1 shows part of SST the fusion data covering Yellow and Bohai sea region.



**Figure 1** The region of SST in Yellow and Bohai sea from the fusion data of CMA. Color map is degree of SST, unit is K.

**The application of SST fusion data in weather forecast model:** At present, the Tianjin Meteorological Administration could obtain daily SST fusion data from CMA by Internet network. After the SST fusion data received, it is processed into the format of WRF model by WPS in Tianjin Meteorological Administration. Coverage of WRF model in Tianjin is (30.86N–47.56N, 99.76E–130.24E), and its grid distance is 5km. Figure 2 shows the distribution of SST fusion data within the coverage of WRF model. Due to lack of fine SST on Yellow and Bohai Sea in China, SST fusion data could make the regional weather numerical models greatly improve the SST initial field in Yellow and Bohai sea. The improved initial field of SST is conducive to promoting the prediction accuracy of meso-scale weather in coastal area.



**Figure 2** The coverage of WRF in Tianjin Meteorology Administration, the color show the distribution of SST processed by WPS in Yellow and Bohai sea, unit is K. The number of abscissa and ordinate are grid point number of WRF coverage.

### Discussion

The data fusion of multiple observation data could get some information which is close to the grid data truth-value of atmospheric and oceanic elements. Data fusion is different from data assimilation. Data assimilation needs to consider the adaptability of the weather forecast model to the observed data. But data fusion does not need to consider the adaptability of the weather forecast model to the fusion data. Data fusion technology focuses mainly on the effective absorption of data information from various observation data, so that the true-value of atmospheric and oceanic elements could be acquired at the place where is lack of observation in-situ data. The data fusion center can merge not only the information of multiple sensors, and also the information from multiple sensors and observation fact of human-machine interface. This data fusion method is usually belonging to decision-making level fusion.

The fusion technology of meteorological data can be developed along the intelligent direction. For example, data fusion can extract symptom information and match between the symptom and the knowledge in the knowledge library under the action of reasoning machine so that data fusion system makes the diagnosis decision of ocean disaster, and also provide valuable service information to users. The diagnosis system of ocean disaster weather based on information fusion could add a self-learning module which could feed back to the knowledge library so as to modify the corresponding confidence factor and update the knowledge library. At the same time, the self-learning module can also obtain new knowledge and draw an inference from the knowledge of the knowledge library and the interaction between users' questions and dynamic responses of the system. The data fusion system with self-learning function can develop into an intelligent data fusion technology in ocean meteorology.

## Conclusion

The authors think that it is a promising research area of ocean meteorological data analysis based on data fusion technology. In the region of few observation stations, data fusion is an effective method to get the actual grid data. The fusion data of sea surface temperature from China Meteorology Administration has been applied to the ocean meteorological operation of Yellow and Bohai Sea in Tianjin Meteorology Administration, which is conducive to improve the forecast accuracy of the numerical weather prediction model of ocean meteorology in the Huang and Bohai Sea. Although the data fusion technology of ocean meteorology has been developing and the methods of data fusion still need to be improved, the data fusion method can quickly provide more ocean meteorological information with data fusion technology to the artificial intelligent development, which is helpful to ocean meteorology research and ocean weather forecast.

## Acknowledgments

None.

## Conflicts of interest

The work is supported by the funds of Collaborative Innovation of Meteorological Science and Technique in Huang-Bohai Region (QYXM01801) and the key operational special plan of China Meteorological Administrative (YBGJXM) (2019) (01-09).

## References

1. Zhao TB, Fu CB. Preliminary comparison and analysis between ERA-40, NCEP-2 reanalysis and observations over China. *Climatic and Environmental Research (in Chinese)*. 2006;11(1):14–32.
2. Zhou HF, Guo PW, Zhai J, et al. Application of LAPS analysis field in rainstorm mesoscale analysis. *Plateau Meteorology (in Chinese)*. 2010;29(2):461–470.
3. Tory KJ, Montgomery MT, Davidson NE. Prediction and diagnosis of tropical cyclone formation in an NWP system. Part I: The critical role of vortex enhancement in deep convection. *J Atmos Sci*. 2006;63:3077–3090.
4. Bahador K, Khamis A, Fakhreddine OK, et al. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*. 2013;14(1):28–44.
5. Bancroft JB, Lachapelle G. Data fusion algorithms for multiple inertial measurement units. *Sensors*. 2011;11(7):6771–6798.
6. Hilker T, Wulder MA, Coops NC, et al. A new data fusion model for high spatial-and temporal-resolution mapping of forest disturbance based on landsat and modis. *Remote Sens. Environ*. 2009;113(8):1613–1627.
7. ZHANG W, Li AN, Jin HA, et al. An enhanced spatial and temporal data fusion model for fusing landsat and modis surface reflectance to generate high temporal landsat-like data. *Remote Sens*. 2013;5:5346–5368.
8. Wang W, Wu DZ, Qu P, et al. The retrieve and data fusion of the sea surface temperature with Feng-yun geostationary satellite data. *Journal of Water Resources and Ocean Science*. 2016;5(4):53–63.
9. Shi SX, Xia DW, Yu G. Method of fusing the data reversed from satellite remote-sensing with the assimilation data for sea surface temperature. *ACTA Oceanologica Sinica*. 2005;27(4):31–37.
10. Wang YZ, Gu AL, Qu LQ. Merging sea surface temperature observed by satellite infrared and microwave radiometers using kalman filter. *Periodical of Ocean University of China*. 2010;40(12):126–130.