



# Article A Microwave Radiometer Residual Inversion Neural Network Based on a Deadband Conditioning Model

Yuxin Zhao, Changzhe Wu \*, Peng Wu, Kexin Zhu ២ and Xiong Deng ២

College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China; zhaoyuxin@hrbeu.edu.cn (Y.Z.); wupengdemo2000@hrbeu.edu.cn (P.W.); zhukexin@hrbeu.edu.cn (K.Z.); xiongdeng407@hrbeu.edu.cn (X.D.)

\* Correspondence: 2015041018@hrbeu.edu.cn

Abstract: Microwave radiometers are passive remote sensing devices that are widely used in marine atmospheric observations. The accuracy of its inversion of temperature and humidity profiles is an important indicator of its performance. Back Propagation (BP) neural networks are widely used in the study of microwave radiometer inversion problems. However, the BP network which is carried by the radiometer inversion suffers from profile data collapse. To address this, this study introduced a residual network to improve the accuracy of water vapor vertical profiles. Aiming at the problem of large inversion temperature error due to the effect of turbulence on the light-travel phase induced by stationary fronts along the seashore in the subtropical monsoon climate region, we used historical data to establish the seasonal a priori mean profile and design a dead-zone residual adjustment model. The accuracy of the residual network and the deadband-adjusted residual network was verified using the meteorological records of the Taizhou region from 2013–2018, with the experimental data and BP hierarchical network as the comparison term. We found no data collapse in the temperature and humidity profile inversion results of the residual network. Relative to the initial BP hierarchical algorithm, where the error of water vapor in the range 6–10 km was reduced by 80%, the dead zone residual adjustment model in the inverse-temperature phenomenon reduced the sum of squares error by 21%, compared with the ordinary residual network inversion results. Our findings provide new insights into the accuracy improvement of radiometer remote sensing.

**Keywords:** microwave radiometer; seaside meteorological observations; residual neural network; inverse temperature; atmospheric profile inversion

# 1. Introduction

Microwave radiometers play an important role in land-based observatories on the seashore, as well as in ship-based observatories. Their passive observation and strong real-time characteristics enable them to monitor the marine atmospheric environment rapidly. However, as the applications deepen, various diversified remote sensing needs also have great demands on the accuracy of the radiometer.

The radiometer receives a brightness temperature signal from the atmosphere and inverts it to obtain water vapor and temperature profiles. The improvement of the accuracy of the inversion has greatly improved the accuracy of the radiometer measurements. There are two main methods for improving inversion accuracy, physical and statistical methods. Neural networks use historical data to form a mathematical model, which is a method derived from statistical methods. The physical inversion method obtains the corresponding atmospheric vertical profiles by solving radiative transfer equations. Westwater et al. (1968) proposed the use of the root-mean-square method to solve atmospheric parameter vertical profile [1]. Meanwhile, Smith (1970) proposed Smith's physical iterative method, which does not depend on historical sounding data and instead uses a stepwise approximation to derive atmospheric temperature vertical profile with good results [2]. Xiong (2016) used



Citation: Zhao, Y.; Wu, C.; Wu, P.; Zhu, K.; Deng, X. A Microwave Radiometer Residual Inversion Neural Network Based on a Deadband Conditioning Model. J. *Mar. Sci. Eng.* 2023, *11*, 1887. https://doi.org/10.3390/ imse11101887

Academic Editor: Fausto Pedro García Márquez

Received: 29 August 2023 Revised: 13 September 2023 Accepted: 26 September 2023 Published: 28 September 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). the 1D variational method to invert atmospheric parameters, and this algorithm yielded significantly better inversions of high-altitude temperature and humidity than the data from the RPG microwave radiometer infusion self-contained algorithm [3]. Yang et al. (2018) also used the 1D variational algorithm to invert the atmospheric temperature and water vapor parameters under clear-sky conditions. They found that the average error of temperature below 4 km was less than 0.2 K [4]. However, the physical inversion method is time-consuming and lacks real-time capability as a meteorological remote sensing device.

With the development of neural networks, neural network inversion has recently been increasingly applied to radiometer inversion algorithms. Shi et al. (2017) downscaled the retrieved  $0.25^{\circ} \times 0.25^{\circ}$  grid total precipitable water (TPW) to a  $0.05^{\circ} \times 0.05^{\circ}$  grid based on the original TPW algorithm to obtain a refined TPW, with a root mean square error (RMSE) of 3.45 mm and a correlation coefficient of 0.95 [5]. Che (2019) improved the retrieval of atmospheric temperature and relative humidity profiles by combining active and passive remote sensing. A maximum RMSE reduction of 2.2 K and 16% for the temperature and humidity profile, respectively, was obtained using a ground-based microwave radiometer and millimeter-wave cloud radar. This indicates that the accuracy of the radiometer can be greatly improved using composite remote sensing techniques [6]. Segal–Rozenhaimer et al. (2018) proposed a neural network-based algorithm for retrieving liquid low-ocean stratocumulus cloud microphysical property parameters from airborne multi-angle polarimetric radiometer measurements and achieved better results [7]. Rohit (2017) proposed a forecast model using the KLURT index (Unitless and dimensionless index) to predict convection and found that the forecast model performed well, with a forecast efficiency of 75%, false alarm rate of 35%, and lead time of 1 h; these values are comparable to those obtained using other forecasting techniques [8]. Yan et al. (2020) developed a deep learning method called batch normalized and robust neural network (BRNN). Compared with conventional backpropagation neural networks, BRNN reduced overfitting and more accurately described the nonlinear relationship between microwave radiometer measurements and atmospheric structure data [9]. H. Md. Azamathulla et al. utilized two different studies using Artificial Neural Networks (ANN) and Gene Expression Programming (GEP) in order to predict the atmospheric temperature in Tabuk. Atmospheric pressure, rainfall, relative humidity and wind speed were used as input variables for the developed model. The experiments yielded better results [10]. In this algorithm, both the land surface temperature (LST)and polarimetric land surface emissivity (LSE)were obtained with limited input data under almost all-weather conditions. Zhao et al. (2019) proposed a numerical correction algorithm based on different frequency-weighting functions to improve the accuracy of atmospheric temperature profiles for clear and cloudy days. In particular, the RMSE and mean absolute error (MAE) of the temperature profiles below 2 km were reduced by more than 50% [11] Zhao et al. (2018) proposed another stratification method to improve the efficiency and accuracy of the training network for obtaining tropospheric water vapor and temperature distributions. Above 6 km altitude, the RMSEs of the temperature and water vapor distributions of the layered method were reduced by 25.6% and 26.2%, respectively, and efficiency was improved by 20 times, compared with the traditional method [12].

Among the abovementioned methods, neural network inversion is mainly performed by replacing the network model, adding training parameters, and improving the network algorithm. For the network model, the BRNN and BP networks can be fitted effectively to reduce the inversion error to some extent. Regarding the training parameters, adding millimeter-wave cloud radar data and optical effect modules as network inputs can effectively improve accuracy. Although the improvement of the algorithm can improve the network inversion accuracy to some extent, it has little effect when other factors such as seasonal variations interfere. This can lead to large errors in the algorithm when dealing with severe contour oscillations. Moreover, accuracy is not the only criterion for measuring network performance. Sufficient timeliness and stability must be ensured for real-time inversion radiometer networks. It is difficult to obtain auxiliary parameters of other observation equipment in areas where radiometer accidental observation equipment is not available, which makes the method of auxiliary observation difficult. However, systematic studies and quantitative analytical models are lacking in this regard.

Based on the hierarchical BP and neural network (NN) inversion network, this study proposes a priori mean profile adjustment model for the dead zone based on the residual network to address the above problems. This is a type of network that uses historical data to generate a priori mean profile and temperature difference labels, with ground meteorological elements and brightness temperature data as input and without requiring additional upper air meteorological elements. Regarding the network stability problem, the causes of network collapse were analyzed, and an attempt was made to address this by introducing a residual module. For the large-scale profile oscillation problem, this study attributes part of the cause to the inverse temperature phenomenon, proposes a turbulence model based on fronts, and discusses the changes in the radiative transfer equation parameters using the atmospheric turbulence model to demonstrate the influence of inverse temperature on radiometer observations. The seasonal priori mean profile was also generated using historical data to improve accuracy under the inverse temperature phenomenon using a dead zone adjustment network based on temperature difference labels. Two experiments were designed to verify the resolution of the abovementioned issues using the sounding data from the Taizhou weather station from January 2012 to March 2018 as the true value and radiometer observation of brightness temperatures in the same period as the measurement data. The experimental results were then analyzed.

This paper is structured as follows. First, we introduce the phenomenon of BP network collapse using a radiative transfer equation. Then, we discuss the residual network (ResNet) and conduct experiments while analyzing the experiments. A model analysis and correction experiments for the inverse-temperature problem are then presented. Lastly, Section 5 presents the conclusion.

# 2. Problems of Inversion Vertical Profile in BP Layered Networks

This section is divided into two parts. The first part describes the working principle of inversion of the microwave radiometer of the MP3000A model and verifies the advantages of the BP hierarchical network at the level of training time. At the end of this section, the data crash phenomenon during the inversion of the BP hierarchical network is addressed. In the second part, a network-level theoretical analysis is conducted to derive the causes of this phenomenon based on the data collapse phenomenon that occurred in the first subsection.

# 2.1. Radiometer Inversion Principles and Data Crash Problems

The radiation transmission equation to be inverted by a microwave radiometer is [13]:

$$T_B(0) = T_B(\infty)e^{-\tau(0,\infty)} + \int_0^\infty k_a(r) \cdot T(r) \cdot e^{-\tau(0,r)} dr$$
(1)

 $k_a$  is the sum of oxygen and water vapor absorption coefficients, in the formula  $T_B(0)$  denotes the brightness temperature of all microwave radiation over the surface, r denotes the altitude, T(r) denotes the atmospheric temperature at altitude  $\tau(0,r)$  denotes the optical thickness from altitude r to the ground. The first term on the right indicates the data after the attenuation of the cosmic background radiation brightness temperature, and the second term indicates the cumulative sum of all the thin atmospheric layers of microwave radiation brightness temperature above the Earth's surface reaching the Earth's surface after the attenuation [14].

The microwave radiative transfer equation is an integral equation, and the integral has an embedded exponential integral, which belongs to the first class of the Fredholm equation. The solution of the equation is a pathological equation problem for which an analytical solution cannot be given directly. Since the exact analytical solution of Equation (1) cannot be derived, BP networks are often used to solve the inversion problem of radiometers as they have some inverse fitting capability for such equations. As a traditional neural network, the BP neural network has a nonlinear fitting ability.

The radiometer model used in this paper is the MP3000A radiometer, which contains surface meteorological sensors (Met Sensors) to measure the air temperature, relative humidity and air pressure around the equipment. Therefore, the four elements of surface meteorology (temperature, relative humidity, pressure and altitude) can be added to the network input to ensure the stability of the inversion network. The built-in neural network used in the radiometer is the Stuttgart neural network (NN network) used for the inversion of the brightness temperature information. This type of NN network is a patchwork of multiple single-layer network groups of perceptrons. A standard back-propagation algorithm is used for training and a standard feed-forward network is used to derive the profiles. The profile output is divided into 58 layers, with one data output every 50 m from 0 to 500 m altitude, one data output every 100 m from 500 m to 2 km altitude, and one data output every 250 m from 2 km to 10 km. The number of layers of these independent measurements (eigenvalues) is 58 layers. The MP3000A radiometer used in this paper has 22 microwave detection channels with the following center frequencies: 22.235, 22.500, 23.035, 23.835, 25.000, 26.235, 28.000, 30.000, 51.250, 51.760, 52.280, 52.800, 53.340, 53.850 GHz. The radiometer has three observation directions, N, S and Z. Since the zenith Z observation angle resolution is too low, the observation values of N and S directions are close, so the observation values of S direction are used as the experimental data in this paper. The experiments are also conducted in an environment with weak electromagnetic interference. The temperature data received by these channels and the output obtained by radiometer inversion are the temperature degree profile and the water vapor profile.

Conventional radiometer built-in network uses ground four elements and level 1 file (bright temperature data transformed by a photoelectric signal obtained from a radiometer receiver) as input, temperature at 58 altitudes, and water vapor density as output. This network has a good accuracy at low altitudes but lacks sufficient resolution at high altitudes due to low resolution. In addition, BP neural network which is an added error backpropagation NN network structure, as a radiometer internal training network can also be used for inversion calculation. Using the sounding data, it is feasible to target the improvement of the output resolution of the BP network. The outputs used in this paper are 26 inputs as well as 100 outputs. 100 outputs have a height interval of 100 m, but the training time of the network is greatly increased as a result.

To solve the problem of excessive training time, the output of 0–10 km was divided into three layers as follows:0–2 km, 2–6 km, and 6–10 km [12]. The layered method can better extract data features and more accurately invert atmospheric temperature and water vapor data, as well as accelerate the convergence speed of the network. The setting of the hidden layer nodes of the neural network is related to its performance, with too few leading to low accuracy, and too much prolonging training time. In order to verify the training time advantage of this network, the experiment is enjoyed by using Taizhou data according to the following design.

The data format used in Table 1 is referenced by Zhao. et al. (2019) [12]. A total of 2000 random brightness temperature vertical profiles were selected for inversion. In order to ensure the accuracy of the training time, the average value of 10 training sessions was selected as the result of the training time comparison experiment. The inversion results are shown in Figure 1.

Table 1. Node configuration diagram of the BP hierarchical network.

	Number of Input Nodes	Number of Hiding Nodes	Number of Output Nodes
Bottom network	26	20	20
Middle layer network	26	40	40
High-level network	26	40	40



**Figure 1.** The inversion experiments of BP hierarchical network and traditional hierarchical network. (**a**,**b**) show the inversion comparison experiments of temperature and water vapor respectively. (**c**) shows the training time comparison of BP hierarchical network layer 1–3 and NN network.

From the above figure (Figure 1), The inverse accuracy of the BP layered network and NN network are similar, and the quantitative analysis of the accuracy of the two networks will be discussed in detail in Section 4. The training time of each layer in the BP layered network is much smaller than that of the NN network. The conclusion that can be drawn from this is that the BP hierarchical network can effectively improve the network training efficiency within a certain error range. However, with an increase in the number of inversions, the BP network will have data collapses (Figure 2).



**Figure 2.** Inverse anomaly crash data plot. (a) The set of anomalous data in the data range: the horizontal coordinate represents the data date number, and the vertical coordinate represents the height of the crash data. (b) The crash vertical profile diagram: parts enclosed by the red dashed box represent the data crash area.

According to Figure 2a, the BP hierarchical network in the red-boxed area presents a large error and the error arises in a step rather than an asymptotic manner. Figure 2b shows the set of collapsed data derived from a separate 2000-group inversion experiment. The reason for the high degree of non-overlap in Figure 2a,b is that this experiment was repeated twice under the same conditions (same 2000 sets of brightness temperature profiles, same BP hierarchical network). From this, we can conclude that inversion data crashes are random and uncertain for the same parameter configuration of the neural network trained and inverted for the same set of brightness temperature data. Based on this phenomenon of data collapse, this paper speculates that this may be caused by the internal structure of the BP network. The following paper will elaborate on the feasibility of this argument from the perspective of BP network inversion.

# 2.2. Theoretical Analysis of Data Collapse Phenomenon Based on Gradient Disappearance and Gradient Explosion

The microwave radiation transmission equation is an integral equation, and the integral has an embedded exponential integral, which belongs to the first class of the Fredholm equation. The solution of the equation is a pathological problem and cannot be directly given as an analytical solution, but only the approximate analytical solution can be obtained. Therefore, Equation (1) can be approximated as follows:

$$T_B \xrightarrow{A} T_r$$
 (2)

where  $T_B$  denotes the brightness temperature from height 0 to maximum observation height.  $T_B$  is approximated infinitely by the *A* process to  $T_r$ ,  $T_r$  means the temperature at a certain altitude *r*. Therefore, Process A can be expressed as

$$T_r = A[T_B(0)] \cdot o_1(x) + o_2(x)$$
(3)

Because the brightness temperature transport Equation (1) cannot be resolved, two perturbations  $o_1(x)$ ,  $o_2(x)$  are added to process A, where  $o_1(x)$  is a random nonlinear perturbation and  $o_2(x)$  is a linear perturbation. The equation then becomes

$$T_r = A(T_B) \cdot o_1(x) + o_2(x)$$
(4)

Similarly, in the process of parsing and fitting the brightness temperature data by the BP network, the expected result  $T_r$  is

$$T_r \approx T_r$$
 (5)

Therefore, process *A* can be considered an expectation function. The inversion process for the individual neuron expansion of the BP network is described as follows:

$$T_r = f(T_B) \cdot \theta + b \tag{6}$$

where *f* is the activation function of the single layer  $\theta$ , and *b* is the error parameter. Corresponding to the two perturbations of Equation (4),

$$\lim_{x \to k_1} o_1(x) = \theta, \lim_{x \to k_2} o_2(x) = b$$
(7)

where k is denoted as  $o_1(x)$ ,  $o_2(x)$ , in which x converges to the network error parameter  $\theta$ , b at some corresponding  $k_1$ ,  $k_2$ .

For the network structure in brightness temperature inversion, all have the activation function  $f_n(T_B)$ , where n denotes the number of network layers, and the activation function for the n + 1 layer network update is as follows:

$$f_{n+1} = f(f_n \cdot \omega_n + 1 + b_n + 1)$$
(8)

where  $\omega_n$  is the *n*th layer of the built-in adjustable network. The gradient information is updated according to the chain-derivation rule to update the weight information of the *n*th hidden layer [15].

$$\Delta \omega = \frac{\partial Loss}{\partial \omega_2} = \frac{\partial Loss}{\partial f_n} \cdot \frac{\partial f_n}{\partial f_{n-1}} \cdot \dots \cdot \frac{\partial f_3}{\partial f_2} \cdot \frac{\partial f_2}{\partial \omega_2}$$
(9)

It is easy to derive:

$$\frac{\partial f_2}{\partial \omega_2} = f_1, \ \frac{\partial f_n}{\partial f_{n-1}} = g \tag{10}$$

Thus, *g* is the derivative of the activation function; if g > 1, the final derived gradient update increases exponentially as the number of layers increases, that is, a gradient explosion occurs, and if g < 1, then the derived gradient update information decays exponentially as the number of layers increases, that is, a gradient disappearance occurs [16].

From Equations (4) and (7), it can be seen that  $o_1(x)$ ,  $o_2(x)$  parameter variation is random, which depends on the meteorology of the radiometer observation point as well as the observation error caused by the observation error. For the random parameter k, there exists K satisfying.

$$K = |k_1 - k_2| \tag{11}$$

The best fit is achieved when set *X* of *x* satisfies the following conditions:

$$x \in \{X | K \to 0, \{x \to k_1 \cup x \to k_2\}\}$$

$$(12)$$

In other words, the best fit is achieved when the parameter difference K converges to 0, while x converges to either  $k_1$  or  $k_2$ . Conversely, when

$$x \in \{ \hat{X} | K \gg 0, \{ x \neq k_1 \cup x \neq k_2 \} \}$$
(13)

That is, in terms of the observed parameters, the worst fit is achieved when the parameter difference *K* is much larger than 0, while *x* is not equal to any one of  $k_1$ ,  $k_2$ .

From Equations (8) and (9), it can be seen that in terms of network parameters, if g > 1, then the final derived gradient update increases exponentially when the number of layers increases, that is, a gradient explosion occurs. The large error increase caused by the gradient explosion, then it can be concluded that when the network training and inversion satisfy the above situation simultaneously, which means that:

$$\left\{x \in \hat{X}\right\} \cap \left\{g \in (1, +\infty)\right\} \tag{14}$$

A significant increase in error occurs when the network parameters and observed parameters satisfy both the gradient explosion and worst fit effect, respectively. This phenomenon leads to data collapse in hierarchical BP networks.

From the point of view of the network, the speed of learning varies greatly from layer to layer, as shown by the fact that the layers close to the output in the network learn very well and the layers close to the input learn very slowly. In some cases, to increase the training time, the first few layers' weights are similar to the difference between the values initialized at random. Therefore, the root cause of the gradient disappearance and explosion lies in the training law of the back-propagation error, and the essence lies in the structure of the BP network.

# 3. Inversion of Residual Network Based on Microwave Remote Sensing Data

In response to the collapse of the BP hierarchical network proposed in the previous section, this section proposes a residual network for the correction. First, the suppression of the gradient explosion by the residual module is analyzed theoretically, and then, experiments are designed to screen the optimal network parameters. Pairwise comparison tests were conducted by using this parameter to verify the fitting results of the residual network. A series of problems arising from these experiments were analyzed. Further validation was carried out to address the issues that arose in the experiments.

#### 3.1. Role of Residual Networks for Gradient Explosion

From the analysis of the BP network, it can be concluded that the network collapse depends on both the observed parameters and the network parameters that simultaneously satisfy the two conditions at the same time. The network collapse problem can be alleviated when network parameters are selected to effectively circumvent problems such as gradient explosions. To solve this problem, we used a residual module added to the BP hierarchical network.

The residual network has a good ability to solve gradient explosions and gradient disappearances. For the input x of the stacked layer structure, the learning feature is denoted as H(x). Compared to the original feature F(x) + x, the residual feature is F(x) = H(x) - x. This is because residual learning is easier than learning the original feature directly. When the residuals are zero, the stacking layer only performs constant mapping at this point and the network performance does not degrade, thus achieving better performance.

Firstly, the residual unit can be expressed as below [17]:

$$y_l = h(x_l) + F(x_l, W_l)$$

$$x_{l+1} = g(y_l) \tag{15}$$

where  $x_l$  and  $x_{l+1}$  denote the input and output of the lth residual unit, respectively.

Each residual unit generally contains a multilayer structure. *F* is the residual function, which denotes the learned residual; while  $x_l = h(x_l)$  denotes constant mapping; and *g* is the ReLU activation function. Based on the above equation, the learning features from shallow *l* to deep *L* were obtained as follows:

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i)$$
(16)

By applying the chain derivative rule in Equation (9) into Equation (16) [18]:

$$\frac{\partial Loss}{\partial x_l} = \frac{\partial Loss}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_l} \cdot \left(1 + \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} F(x_i, W_i)\right)$$
(17)

The first factor,  $\frac{\partial Loss}{\partial x_l}$  indicates the gradient of the loss function arriving at *L*. The 1 in parentheses indicates that the short-circuit mechanism can propagate the gradient without loss, and since the residual gradient is not always -1, the presence of a 1 does not cause the gradient to vanish or explode, even if the value is small.

As shown in Figure 3, the residual network (ResNet or res) incorporates residual units through a short-circuiting mechanism. The changes are mainly reflected in the fact that ResNet directly uses the convolution of stride = 2 for downsampling and replaces the fully connected layer with a global average pooling layer. To maintain the complexity of the network, the number of ResNet feature maps is doubled for every half reduction in size. Where residual learning refers to the addition of a short-circuiting mechanism between every two layers of the normal network [19].



Figure 3. Schematic diagram of the residual neural network used in the comparison test.

ResNet uses two types of residual units, corresponding to shallow and deep networks. For short-circuit connections, when the input and output dimensions are the same, the input can be added directly to the output. However, when the dimensions are not identical (corresponding to dimension doubling), they cannot be added directly. There are two main strategies to achieve this.

- 1. Use zero padding to increase the dimension; when you generally have to do a down-sample first, you can use strde = 2 pooling, which will not increase the parameters.
- Using a new mapping (projection shortcut), generally using a 1 × 1 convolution, which increases the number of parameters but also increases the amount of computation. A projection shortcut can be used in addition to the direct use of constant mapping.

In this study, due to the network used in this experiment, the input and output dimensions are the same but different in size. We used the shallow network projection shortcut convolution mode [20].

# 3.2. Network Parameters Selection

This experiment bridges the above analysis and conducts a comparison experiment at the same time and space. The parameters of the BP network were adjusted based on the Taizhou brightness temperature data, and the residual network module was added to this experiment. The information used in this study was the sounding data from the Taizhou weather station from January 2012 to March 2018. In addition, to increase the stability of the network, the four measurable elements on the ground (temperature, relative humidity, pressure, and height) are added to the inputs of the neural network; the above 26 inputs are used as the input values of the neural network. The output samples include temperature vertical profiles and water vapor content vertical profiles, and the generated vertical profiles range from the ground to a height of 10 km height with 100 m accuracy, the temperature and water vapor content vertical profiles from the ground to the 10 km range are 101 layers each, where the ground data do not need to be obtained by inversion and can be obtained by field measurements, so the output is 100 nodes.

The experimental design was divided into two parts: the first part was the network parameter judgment. Two criteria, loss and network consumption time, were mainly used to determine the best network parameters for the parameters listed in the text. The second part is an experiment for comparing the performance of the residual network and the BP hierarchical network. The optimal network parameters obtained in the previous section were mainly used to add the residual module to invert the brightness temperature data.

This section serves as a selection experiment for the network parameters and is a prelude to the comparison experiment. This experiment compares the loss and network consumption time by enumerating possible network parameters and finding the least computationally expensive parameter between them.

The number of network layers was configured according to hidden layers 1 to 4, and for each layer, the network had the same number of nodes set with an adaptive learning rate. The pre-experimental parameters were set as follows:

The experiments were performed according to the network parameters listed in Table 2. The first 2000 brightness temperature distributions in the dataset in date order are selected as the training set input and the temperature distribution is selected as the output [21]. For the representation aspect, networks with one, two, three, and four hidden layers are referred to as Networks I, II, III, and IV, respectively. The number of nodes under different hidden layers is indicated using the corner scale, such as the number of nodes in two hidden layers 10–30–30. In node number 1, the network is indicated as I -1. The experimental results are as follows [12].

Number of Nodes							
Node Number	1	2	3	4	5	6	7
Bottom	10	15	18	20	22	25	30
Middle Level	30	35	38	40	42	45	50
High Level	30	35	38	40	42	45	50

**Table 2.** Assignment of nodes within the network in the optimal parameter screening experiment of the BP hierarchical network.

Regarding the experimental results in Figure 4, in terms of the time cost, as the number of layers and nodes of the network increased, the training time also increased. The training time cost of increasing the number of layers was greater than that of increasing the number of nodes, which is obvious in multi-layer and multi-node networks; thus, both time and loss factors should be considered. From a real-time perspective, it is necessary that the network training fitting speed is sufficiently fast and the time cost is sufficiently small. In terms of accuracy, radiometer inversion of brightness temperature should be sufficiently accurate to describe the atmospheric profile. In terms of a sufficiently small loss, network number IV-6 has the smallest loss function; however, its training time was as high as 85.129. Compared with network II-4, which had a loss of 0.9863, the network had a loss reduced by 0.1641; however, its training time was increased by 338.7%.

These findings represent the results of all network runs in the descending order of loss. After network II-4, the training loss converged, but training time continued to increase. In summary, network number II-4 was used as the main BP hierarchical network structure in the residual network comparison experiment. In other words, there were 40 nodes in the middle and top layers, 20 nodes in the bottom layer, two hidden layers, the same number of nodes in each layer, and a fully connected BP neural network between adjacent layers.





**Figure 4.** Experimental results of the neural network parameter screening experiment. (**a**) part shows the training time for different network parameters and the (**b**) part shows the training loss (Loss) for different network parameters.

# 3.3. Experiments and Conclusions

This subsection focuses on the experimental design of the comparison experiments between the residual networks and the BP hierarchical neural networks. The experiment aimed to analyze the performance of the two networks in terms of fitting by comparing their inversion accuracy and stability for the same dataset. The experiments were designed using five years of historical-sounding data in the Taizhou area. Atmospheric temperature and water vapor density profiles were obtained via hierarchical inversion using a hierarchical BP neural network and residual neural network. The two networks were also compared and analyzed using simultaneous air data samples as real values.

# 3.3.1. Experimental Design

The optimal parameters of the BP layered network given in the previous subsection were used to obtain the residual network in the comparison experiment by adding the residual module between the non-adjacent layers of the same layer inversion network based on the BP network.

For the 4382 sets of brightness temperature data measured at the Taizhou radiometer observation site, invalid data were screened out, and 3339 available data were obtained as data sources, of which, 2000 data points were used as the training set for the two networks, and 1339 as the test set for the experiment. The input was a  $26 \times 1$  column vector, consisting of the four elements of surface meteorology (temperature, relative humidity, pressure, and altitude) and brightness temperature data obtained on the day of the data. The output was the water vapor and temperature vertical profiles of the atmosphere on that day. The output profiles were all  $100 \times 1$  column vectors. Each adjacent data point in the column vector represents an atmospheric parameter at 100-m intervals.

# 3.3.2. Water Vapor Profile Analysis

A comparison of the output water vapor profile values of the two networks with the actual sounding data is shown in Figure 5a–d.



**Figure 5.** Comparison of single-day profiles of inverse water vapor profiles generated by the NN network, BP network and residual network.

The water vapor density gradually decreased with increasing altitude (Figure 5). Based on Figure 5d, in the case of smoother water vapor variation, the three networks have similar fitting abilities. However, in the remaining three plots, the fitting ability of the network shows a different trend. As shown at the low level (0-2 km), the error of the residual neural network is larger, and compared with the BP neural network, the mean square error and the average absolute error of the residual neural network are large, with the maximum value up to  $0.7 \text{ g/m}^3$ . However, the error is still smaller compared to the conventional NN network, which has a maximum error of  $1.1 \text{ g/m}^3$  at this height. At the middle level (2–6 km), the BP network and the traditional NN network errors are more similar, with error bands floating around  $0.8 \text{ g/m}^3$ , the error of the residual neural network is significantly reduced, with the error not exceeding  $0.5 \text{ g/m}^3$ , and the mean square error and the average absolute error are smaller than those of the BP. In the upper layer (6–10 km), the residual network has an obvious advantage, with the error not exceeding 0.1 g/m<sup>3</sup>. NN networks and BP networks are similar in accuracy, the error of the BP neural network is always between 0.3-0.5 g/m<sup>3</sup>, and the residual neural network has a great advantage in terms of mean square error and average absolute error.

To more accurately represent the water vapor error values, logarithmic coordinates are used in the above figure. The residual network outperformed the traditional BP hierarchical network for all the errors at full height (Figure 6). For a more accurate comparison, the crash data of the BP network were removed from all experimental data analysis. Regarding the inversion of the annual atmospheric water vapor density profile, in the low layer (0–2 km), The NN network is similar to the residual neural network, the error of the BP neural network did not exceed 0.6 g/m<sup>3</sup>, whereas the maximum error of the residual neural network reached 0.87 g/m<sup>3</sup>, slightly higher than that of the BP neural network; the mean square error of the BP network was  $1.38 \times 10^{-5}$  better than that of the residual neural network ( $2.26 \times 10^{-5}$ ). In the range of experimental data, the BP network has higher accuracy due to the BP network and NN network at this altitude. This indicates that the high-resolution network at low altitude (conventional NN network), compared to the BP network, plays no role in improving the accuracy of high-resolution observations at this

altitude. In the middle layer (2–6 km), the NN network and BP network are closer in error data, the maximum error of the BP neural network reached 0.63 g/m<sup>3</sup>, whereas that of the residual neural network was 3.06 g/m<sup>3</sup>, slightly better than that of the BP neural network. The maximum error of the BP neural network reached 0.63 g/m<sup>3</sup>, whereas the error of the residual neural network did not exceed 0.5 g/m<sup>3</sup>, which was slightly better than the BP neural network; the mean square error of the residual network was  $3.06 \times 10^{-6}$ , which was better than the mean square error of the BP network ( $1.86 \times 10^{-5}$ ). Lastly, in the upper layer (6–10 km) the residual network has obvious advantages, with the maximum error not exceeding 0.06 g/m<sup>3</sup>. Meanwhile, the error of the BP network and NN network ranged between 0.3 and 0.5 g/m<sup>3</sup>. The mean square error of the residual network was only  $7.01 \times 10^{-8}$ , which was better than that of the other neural network at  $2.12 \times 10^{-5}$ . Based on the above experimental data, it can be concluded that the residual network outperforms the other two networks at observation altitudes above 2 km, and this advantage is more obvious at altitudes above 6 km. In terms of stability, there were 21 collapses in the BP layered network, and no collapse occurred after the addition of the residual module.



**Figure 6.** Full-year mean squared error (MSE) and mean absolute error (MAE) data of the residual network and BP network inversion of the water vapor profile for 2018. Among them, (**a**–**d**) is the error data of 0–10 km, 0–2 km, 2–6 km, and 6–10 km, respectively. The left coordinates of the left side of the four statistical plots are the MSE error scale, and the right side is the MAE error scale.

# 3.3.3. Temperature Profile Analysis

By analyzing and comparing the temperature profile output values of the two networks with the actual sounding data, the mean absolute error (MAE) and mean square error (MSE) of the inversion data of the residual network, BP neural network and NN network for the entire year 2018 were determined.

The residual neural network outperformed the BP neural network and NN network at all levels, but this accuracy advantage is not obvious. (Figure 7). At a height range of 6–10 km, the advantage of the residual network over the BP network was not obvious. The error of the three networks decreases in the order of NN network, BP network, and residual network, but the reduced error value is insignificant compared to the size of the data itself. For atmospheric temperature, at the lower level (0–2 km), the maximum error of the three networks did not exceed 3 K, and the mean square error of the residual network was 1.3, which was better than the mean square error of the BP network of 1.7. At the middle level (2–6 km), the maximum error of the three networks did not exceed 4.5 K, and the mean square error of the residual network was 4.0, which was slightly better than the

mean square error of the BP network of 4.2 and the NN network of 4.25. In the upper layer (6–10 km), the maximum error of the three networks did not exceed 3 K, and the mean square error of the residual network was 2.8, which was slightly better than the mean square error of the BP network of 2.9. In terms of stability, the BP hierarchical network crashed 17 times, and no crashes occurred after adding the residual module.



**Figure 7.** Full-year mean squared error (MSE) and mean absolute error (MAE) data of the NN network, BP network and Res network inversion of the temperature profile for 2018. Among them, (a-d) is the error data of 0–10 km, 0–2 km, 2–6 km, and 6–10 km, respectively.

In general, the addition of a residual module increased stability. No data collapse occurred within the experimental data range. In the water vapor density profile inversion, the effect was not obvious in the lower layers but significant in the upper layers. It is generally better than the other neural networks and closer to the sounding data values. Although the residual neural network has a slight advantage in terms of temperature inversion, the improvement is small. As in Figure 8, the atmospheric temperature in the region in the red dashed box increases with height, resulting in poor inversion of the NN network, the BP network and the residual neural network in this region, the maximum error exceeds 5 k. However, in the inversion of temperature contours, for the inverse temperature phenomenon, the inversion of all three networks showed some degree of inaccuracy. The analysis suggests that the main reason may be due to the climate in Taizhou. The error itself is not a network error or a fitting problem, but a poor fit in the case of an abnormal change in the slope of the vertical profiles, as evidenced by the lack of a significant difference between the three networks. This may be caused by the relatively small percentage of inverse temperature phenomenon data in the training set. An experiment will be designed to prove this conclusion.

3.3.4. Results and Analysis of Network Overfitting Experiments under Inverse Temperature Conditions

As shown in Figure 9b, in the temperature thermogram, the temperature tends to be flat in the summer under the influence of non-typhoon, for example. As a subtropical monsoon climate, Taizhou is in this climate most of the time. As shown in Figure 9a, inversions are generated more frequently in winter. The statistical results show that the inversion temperature data only accounts for 12% of the total data volume, which is under-represented in the network training. However, the network generated by extracting the inverse temperature data for separate training has two problems [22]. First, the generation of inverse

temperature cannot be predicted in real-time meteorological observations. Second, the amount of data is so small that the neural network for inverse temperature can be overfitted even at very shallow network layers. To verify the overfitting problem. In this paper, all inverse temperature profiles in the range of true values are selected for the experiments. To determine that the overfitting is caused by the amount of data, this experiment is divided into three groups according to the different proportions of the training and test sets. The ratio of the training set and the test set is divided into three groups a, b, and c according to 9:1; 8:2; and 7:3. The absolute errors of the training and test sets were compared and the experimental results are shown below:



**Figure 8.** (**a**–**d**) Comparison of the single-day profiles of the inverse temperature profiles of the NN network, BP network and residual network. The areas with larger errors are marked by the red dashed boxes in (**a**,**c**,**d**).

From the above figure (Figure 10), the decrease in the number of practice sets leads to a decrease in the learning ability of the network, but this learning ability is based on overfitting. This is demonstrated by the fact that the relative errors of the training sets of the three networks gradually increase as the proportion of training sets gradually decreases, in which the training set errors of the NN network, BP network and residual network increase by 30%, 17% and 20%, respectively, and the test set errors are relatively stable. Meanwhile, the residual network has some overfitting resistance, and the error of the residual network has the smallest error compared with the other two networks in all groups of laboratories. However, the error of the test set is still at a high level. This indicates that the adjustment ability of the residual network is still unable to effectively correct the overfitting caused by too small samples. This overfitting of shallow networks is caused by the amount of data, and adding residual modules, for example, is not effective in reducing this effect.



**Figure 9.** Winter and summer 2013 profile heat map. (**a**) The temperature heat map within 0–10 km for some winter months from January to February 2013. (**b**) The temperature heat map within 0–10 km for the summer months from July to August 2013.



**Figure 10.** Experimental results of the overfitting network experiment. Where NN, BP and Res are the conventional NN network, BP hierarchical network and residual network, respectively. (**a**–**c**) represent the three groups assigned according to the proportion of the training set test set as illustrated above, respectively.

#### 3.4. Results and Discussion

The above experiments show that the residual neural network has a great improvement in stability. In addition, the error has been reduced to some extent. From the point of view of single-day contours, the residual network reduces the oscillations of the inversion contours, which makes them closer to the real values. the accuracy of water vapor density is greatly improved compared to the NN network and BP hierarchical network. From the dataset point of view, the effect of errors in the crash data dataset is tremendous. Often the effect of a few data collapse contours on the seasonal error mean can be much larger than the accumulation of errors during normal inversion. Therefore, the residual network eliminates the crashed data in this respect by controlling the network stability and suppressing the gradient vanishing and gradient explosion, resulting in a reduction in the error. The radiometer used in this paper is the MP3000A, and the NN network is loaded on the MP-3000A. In contrast to other methods of improving accuracy, the residual network is improved from the BP neural network and also the BP network is a variant of the initial network NN network loaded by the radiometer. So the NN-BP layering-ResNet process is fully compatible in the engineering sense. Authors believe that directly applying other networks would lead to larger errors in the radiometer when faced with changing areas due to a lack of multi-location training data, as well as potential conflicting command issues due to compatibility resulting from loading the network on the host computer. However, there is still some error in the temperature inversion due to the presence of the inverse temperature phenomenon. Regarding this problem, the chapter proposes a certain method to solve this error, the inverse temperature samples are extracted and trained separately, and the experimental results show that due to the small samples, the overfitting produced by the three networks is extremely obvious. This proves that the effect of inverse temperature on temperature inversion cannot be well corrected simply by improving the networks.

## 4. Deadband Regulation Residual Network Based on Inverse Temperature

This section addresses the issue of poor fitting in the case of anomalous changes in the profile slope. A turbulent light range model was designed, and the effect of the inversion temperature on radiometer observations was discussed. Because the inversion temperature process appears regionally and temporally, seasonal stratification based on height stratification is required. Based on this idea, the dead zone was a frequently regulated model proposed in this study. Comparative experiments were conducted to verify the accuracy of the model, and the experimental results were analyzed.

# 4.1. Principle of Inverse Temperature

The inverse temperature phenomenon occurs when the atmospheric profile does not decrease with increasing altitude in accordance with actual local meteorological conditions, producing an extreme phenomenon of increasing temperature with increasing altitude in certain altitude ranges. This is mainly due to the urban heat island effect, extreme weather, and cold fronts generated by the convergence of warm and cold air currents, cloud top, Energy exchange between sea and land, energy exchange between land and air, etc. The inverse temperature model discussed in this subsection is based on frontal turbulence.

In this paper, only the gas turbulence phenomenon in the atmosphere is considered, the cloud layer and the temperature variation factor at the top of the cloud are not considered in the model. A schematic diagram of the front formed by the Siberian high pressure and the oceanic warm current in Taizhou is shown in Figure 11a. The other irrelevant terms are removed and abstracted to the model diagram of the observatory front in Figure 11b. The radiometer observation range is exactly in the middle of the cold and warm air masses. Owing to the different rotation directions of the cold and warm cyclones, a more complex turbulence phenomenon is generated in the frontal region in the middle of the air masses. At the temperature sampling point at height r (for considering only the irradiance reception in the zenith direction), the radiometer receives microwave information from the conical air column in Figure 11c. The conical air column was downscaled to obtain the side view in Figure 11d, which is then used as the main model in this subsection.

For a certain observation range to receive the sum of microwaves is superimposed from a hemisphere, in which the radiometer is located under the coverage of cold air masses. When the observed altitude *r* covers the cold air masses located below and warm air masses located above, an inverse temperature condition occurs. In the model on the right,  $l_1$ ,  $l_2$ , and  $l_3$  indicate three different turbulences;  $O_\rho$  indicates the equivalent brightness temperature observation point for the entire surface at the center of the profile circle;  $O_0$  indicates the



vertical observation position at the equivalent height r, where no turbulence occurs; and  $\rho$  denotes the distance between the two points [23].

**Figure 11.** A schematic diagram of the inverse temperature turbulence model. (**a**,**b**) The measurement range of the observatory and locations of the cold and warm air masses. (**c**) The radiometer measurement ranges at a certain altitude. (**d**) The turbulent light range model.

- 1. When  $l \ll \rho$ , such small-scale inhomogeneities have little effect on the phase difference between the two points. This is mainly attributed to the fact that their own undulations are not large, and the number of such small-scale inhomogeneities experienced by the two rays on a longer propagation path should be statistically equivalent.
- 2. When  $l \gg \rho$ , such large-scale inhomogeneities also have little effect on the phase difference between the two points, because generally cover the propagation paths of both rays. The two rays can be assumed to experience the same phase change.
- 3. When  $l \approx \rho$ , the inhomogeneous scale, which is similar to the distance between two points, has the greatest effect on the phase. The difference in the position of the light relative to the inhomogeneous region and the difference in the number of inhomogeneous regions on the two optical paths have a significant effect on the phase difference.

Therefore, when analyzing the phase difference of the optical range of two points, it is necessary to focus on the influence of turbulent vortices at a similar distance  $\rho$  between the two points. Now, there are *i* turbulent vortex distances  $\rho$  on the two points, and its

two-point refractive index distribution for *n*, *n*<sup> $\prime$ </sup>. Then, the turbulence caused by the  $\rho$  distance on the wave number  $k = \frac{2\pi}{\lambda}$ , the phase difference is [24]

$$dS_i = k\rho(n - n') \tag{18}$$

Their mean values are [25]

$$\langle dS_i \rangle = 0 \tag{19}$$

Variance is [26]

$$\left\langle dS_{i}^{2}\right\rangle = k^{2}\rho^{2}\left\langle \left(n-n^{\prime}\right)^{2}\right\rangle$$
 (20)

Now defined [27]

$$D_n(\rho) = \left\langle (n - n')^2 \right\rangle \tag{21}$$

Combining Equations (18) and (19), we obtain

$$\left\langle dS_i^2 \right\rangle = k^2 \rho^2 D_n(\rho) \tag{22}$$

Over the entire propagation path, the number of vortex turbulences of  $l \approx \rho$ , where *L* is the total length of the optical range and L = r. Then, for the total phase difference, we have [28]

$$\Delta S = \sum_{i=1}^{N} dS_i \tag{23}$$

The total phase variance is

$$\Delta S^2 = N \left\langle dS_i^2 \right\rangle = N k^2 \rho^2 D_n(\rho) \tag{24}$$

Therefore, the phase structure function is proportional to the refraction structure function as follows [24]:

$$D_s(\rho) = const \cdot k^2 \rho^2 D_n(\rho) \tag{25}$$

When the observation position is located in the turbulent inertia region, refractive index constructors can be obtained as phase constructors [24]:

$$D_s(\rho) = const \cdot C_n^2 k^2 L \rho^{\frac{5}{3}}, l_0 \ll \rho \ll L_0$$
(26)

where  $C_n^2$  is the refractive index construction factor [25]:

$$C_n^2 = \left(79 \times 10^{-6} \frac{p(r)}{T(r)^2}\right)^2 C_T^2(c)$$
(27)

where p(r) and T(r) represent the pressure and temperature, respectively, at height *r*.

Under uniform isotropy, there exists a temperature structure function  $C_T^2(c)$  as follows [26]:

$$C_T^2(r) = \frac{\langle [T(x) - T(x+c)] \rangle}{c^{2/3}}$$
(28)

where *c* represents the molecular spacing.

When the minimum vortex turbulence scale  $l_0$  was larger than the observed distance, only the effect of this vortex turbulence was considered because the minimum vortex turbulence had an internal scale. Similar to Equation (26), the phase difference is

$$dS_i = kl_0(n - n') \tag{29}$$

The minimum number of vortex turbulences along the entire propagation path is [27]

$$N = \frac{L}{l_0} \tag{30}$$

Thus, the phase constructor [29] becomes

$$D_s(\rho) = const \cdot C_n^2 l_0^{-\frac{1}{3}} k^2 L \rho^2, \rho \ll l_0$$
(31)

The angle of arrival, which is closely related to the transverse phase difference, is shown in the following diagram Figure 12.



Figure 12. Phase, angle of arrival, and optical range model.

As depicted in Figure 12, the quantitative relationship between the phase difference  $\Delta S$  and the optical range difference  $\Delta L$  at two observation points at a distance  $\rho$  from each other is [28]

$$k\Delta L = \Delta S \tag{32}$$

The resulting quantitative relationship between the angle of arrival at baseline  $\rho$  and the phase difference  $\Delta S$  is [30]

$$\alpha = \frac{\Delta L}{\rho} = \frac{\Delta S}{k\rho} \tag{33}$$

According to the undulating variance of the angle of arrival  $\alpha$  for the two cases where the observation distance lies within the turbulent inertia zone and the observation area distance is much smaller than that under turbulent de-scaling is [30]

$$\left\langle \alpha^{2} \right\rangle = \begin{cases} const \cdot C_{n}^{2} L l_{0}^{-\frac{1}{3}}, \ \rho \ll l_{0} \\ const \cdot C_{n}^{2} L \rho^{-\frac{1}{3}}, \ l_{0} \ll \rho \ll L_{0} \end{cases}$$
(34)

Bringing Equations (27) and (28) into the above equation yields

$$\left\langle \alpha^{2} \right\rangle = \begin{cases} const \cdot \left( 79 \times 10^{-6} \frac{p(r)}{T(r)^{2}} \right)^{2} \frac{\langle [T(x) - T(x+c)] \rangle}{c^{2/3}} L l_{0}^{-\frac{1}{3}}, \ \rho \ll l_{0} \\ const \cdot \left( 79 \times 10^{-6} \frac{p(r)}{T(r)^{2}} \right)^{2} \frac{\langle [T(x) - T(x+c)] \rangle}{c^{2/3}} L \rho^{-\frac{1}{3}}, \ l_{0} \ll \rho \ll L_{0} \end{cases}$$
(35)

For the equivalent brightness temperature point  $O_{\rho}$  in turbulence, its actual optical range  $\hat{L}(O_{\rho})$  is [31]:

$$\hat{L}(O_{\rho}) = r - \Delta L = r - \alpha \rho \tag{36}$$

In Section 2, Equation (1), the optical thickness  $\tau(0, r)$  is expressed as [13]

$$\tau(0,r) = \int_0^r K_a(r)r \tag{37}$$

For the observation point  $O_{\rho}$  at the propagation to the radiometer sensor, the brightness temperature size of  $T_B(0)$  when the theoretical distance is r [31], the actual distance  $\hat{L}(O_{\rho})$ .

Therefore,  $O_{\rho}$  at the propagation to the radiometer brightness temperature size of  $T_B(0)$  distance should have the following transmission equation:

$$T_B(0) = T_B(\infty)e^{-\tau_\rho(\Delta L,\infty)} + \int_{\Delta L}^{\infty} k_{a\rho}(r) \cdot T(r) \cdot e^{-\tau_\rho(\Delta L,r)} dr$$
(38)

where  $k_{a\rho}$  represents the refractive index and  $\tau_{\rho}(\Delta L, \infty)$  represents the optical thickness at turbulence. Then, the actual radiometer receives the brightness temperature data at turbulence  $O_{\rho}$  as follows:

$$T_B(0) = T_B(\infty)e^{-\tau_{\rho}(0,\infty)} + \int_0^\infty k_{a\rho}(r) \cdot T(r) \cdot e^{-\tau_{\rho}(0,r)} dr$$
(39)

It follows that

$$T_B(0) \neq T_B(0) \tag{40}$$

Combining Equation (39) into Equation (4), we have

$$T_r = A(T_B(0)) \cdot o_1(x) + o_2(x)$$
(41)

$$\hat{T}_r = A\left(T_B(0)\right) \cdot o_1(x) + o_2(x) \tag{42}$$

Available in:

$$T_r \neq \hat{T}_r \tag{43}$$

Under this model we can derive the following hypothetical inference: the turbulence affects the light range by affecting the refractive index and other factors to produce a phase difference, which causes the brightness temperature measured by the radiometer to change, and thus affects the determination of the temperature profile. Because the angle of arrival *a* is related to the size of the turbulence, air pressure, and temperature, the temperature and pressure cannot be used to compensate for the light range. However, the essence of turbulence formation is frontal airflow, and the formation of fronts is due to the convergence of cold and warm air masses. The convergence of cold and warm air masses was caused by the monsoons. Taizhou has a subtropical monsoon climate, and its winter is mainly caused by the combination of two air masses: the Pacific Ocean (warm air mass) and Siberia (cold air mass) [32–35]. The high-pressure air mass from Siberia flows south to meet the warm and humid ocean airflow and forms fronts along the coast. Therefore, the a priori mean profile can be formed using different seasons, and further corrections can be made after using the a priori mean profile as a baseline, which is described in detail in the next chapter.

#### 4.2. Deadband Regulation Model

Ideally, in the troposphere, the temperature decreases uniformly with increasing height [36]. Historical data showed that the coastal atmospheric condition in the Taizhou area drops by approximately 0.6 K for every 100 m increase in height. Thus, a standard-temperature a priori mean profile can be constructed [37–39].

The method of calculating the average temperature can effectively eliminate temperature fluctuations caused by extreme weather, thus obtaining a more representative standard temperature a priori mean profile. Calculating the average monthly temperature is a more accurate practice; however, considering that the amount of data is limited, the calculated results may not be generalizable, and the magnitude of sea temperature variation is small. Therefore, we calculated the average temperature by season as the standard temperature a priori mean profile.

For the definition of the four quarters, this study follows the standard seasonal definition [40]. The average temperature profiles were also calculated.

The background temperature profile can be considered to represent the trend of normal weather temperature with height (Figure 13). The difference between the single-day data

and background temperature profile is made layer by layer, which is used as the criterion for judging the inverse temperature, and the magnitude of the difference (herein, inverse temperature evaluation label) can be used to judge the degree of inverse temperature. Most altitudes where the inverse temperature is located are distributed above 2000 m; thus, it is only necessary to apply the labeling method to that section of altitude to reduce network training time and shorten the number of training sets for a single network, without affecting the correction accuracy. This substantially improves operational efficiency. As shown in Figure 13b, for the temperature profile in the inverse of the residual network, there exists

 $T_r(r_1)$  at a height of  $r_1$ . Then, there must be  $T_r^b(r_1)$  on the corresponding a priori mean profile, such that the inverse temperature evaluation label  $T_{tab}(r_1)$  on  $r_1$  height in this inversion can be expressed as



**Figure 13.** (**a**): The mean a priori mean profile vertical profiles for the four seasons. (**b**): A schematic illustration of the temperature labels using the inverse vertical profiles of the residual network for data from winter No. 23 as an example.

Considering the fluctuation in temperature and special extreme weather, the inverse temperature evaluation label needs to be further processed based on the difference to reduce the influence of special weather. The method used was to set the dead zone and saturation area, and there were dead, regulation, and breakdown zones in the regulation. In this study, three schemes were designed as follows:

Regarding the selection of data, 0.2 K and 0.8 K are empirical data derived from actual observations, and 8 K is the maximum error of the BP network (non-collapse state) and the network of residuals for all temperature inversion experiments.

From Figure 14, the following can be observed: Model 1 has a dead zone of  $\pm 0.2$  K, a breakdown zone of  $\pm 8$  K, and the rest is part of the regulation zone. Model 2 has a dead zone within  $\pm 0.5$  K, a breakdown zone outside of the 8 K range, and a regulation zone in the rest of the region. Model 3 has a dead zone within  $\pm 0.5$  K, no breakdown zone and the rest is the regulation zone.

$$T_{tab}(r_1) = T_r(r_1) - T_r^b(r_1)$$
(44)



**Figure 14.** Three methods of the deadband regulation model. The horizontal coordinate is the inverse temperature label  $T_{tab}(r_1)$  obtained in the inversion, and the vertical coordinate indicates the magnitude of the adjustment temperature. The dashes in the figure then indicate the regulation for different inversion temperature labels.

- 1. When  $T_{tab}(r_1)$  falls within the dead zone, the temperature change is within the error band, does not belong to the inverse-temperature category, and no treatment is performed.
- 2. When  $T_{tab}(r_1)$  falls within the regulation zone, the temperature-change region is in the inverse-temperature region and is regulated. In this study, regulation was carried out according to the curve with a slope of 1.
- 3. When  $T_{tab}(r_1)$  falls in the breakdown region, the temperature change is extremely drastic. However, to prevent a dramatic change in climate on a single day from exceeding the mean a priori mean profile processing range and ensure data authenticity, regulation was carried out in the training set, according to the maximum value of the regulation area.

# 4.3. Experimental Analysis

An experiment was designed to compare and select the best of these three scenarios. Based on the residual network in Section 3, four ground-based meteorological elements and brightness temperature data are input and atmospheric temperature profiles are output. The obtained atmospheric profiles are brought into the quilt to determine the temperature labels. The temperature labels are processed by each adjustment model to obtain three sets of adjusted temperature profiles. The obtained temperature profiles are compared with the output data of the re-integrated network and the experimental results are analyzed.

To demonstrate the effect of the dead zone model on the regulation of the inversetemperature state. For a single-day vertical profile comparison, the vertical profiles from standard clearing and inversion-temperature phenomena were selected.

The experimental analysis was set to two categories: the presence of inverse temperature and the absence of inverse temperature phenomenon (Figure 15). The comparative experimental error data in an environment without inverse temperature are shown in Table 3.



**Figure 15.** Results of the comparison experiments for the dead zone regulation model. (**a**,**b**): The temperature profile experiments under a standard clear sky in summer without inversion. (**c**,**d**) Experimental maps of the atmospheric profile under an inverse temperature condition.

Selected Program	SSE	MSE	MAE
Model 1	17.6261	1.0657	0.8705
Model 2	16.5719	0.6393	0.6052
Model 3	16.1983	0.5800	0.5132
Resnet	17.5381	0.9885	0.8972

**Table 3.** Comparison of the errors obtained by the three conditioning models and residual network in the comparison test without an inverse temperature phenomenon for the full range of 2018 data.

SSE: sum of squares error; MSE: mean square error; MAE: mean absolute error.

For the experiments without the inverse temperature, the errors of the various adjustment methods were not significant. The data from the inversion of the residual network were sufficient to fit the true-value vertical profiles. The real temperature vertical profile data were more similar to the priori mean profile vertical profile data formed by historical data. This resulted in the residual network inversion data not differing significantly from the priori mean profile data, and the data labels were mostly located in the dead zone and rarely in the conditioning or saturation zones. The SSE difference between the three models and the residual network was less than 1.43, and the error band fluctuation did not exceed 7.98%. This indicates that the fitting abilities of the four methods were extremely close. Because model 1 has the shortest dead zone interval, it is most affected by the priori mean profile; therefore, the adjustment zone has some influence on the vertical profile accuracy. It is presented in the data as having the largest error among all four terms. Because Models 2 and 3 have the same dead zone interval but different saturation areas, the error values are closer, and Model 3 is better. Overall, Model 3 was better for the inversion of the atmospheric temperature profile in the absence of inversion.

For the atmospheric profile of the inverse temperature phenomenon, the comparative test error data are listed in the following table.

As shown in Table 4, the oscillation of the true value vertical profiles in the inverse temperature case is extremely dramatic, which is manifested by a significant decrease in the fitting ability of the three models and the residual neural network. Compared with the

case without inverse temperature, the SSE of the residual network increased by 668.63%, and the three models also had a similar magnitude of error growth. However, Model 3 still maintains a better fitting ability, and its SSE error improves by 21.15% relative to the base residual network. The ability to describe the vertical profile oscillation caused by the a priori mean profile is reflected at this time, which is demonstrated by the fact that the MAE of all three models is smaller than that of the residual network, with the best effect of Model 3, which has a 13.03% error reduction. The error data of Models 2 and 3 are similar, but the saturation area is different, resulting in different maximum adjustment abilities of the vertical profiles.

Selected Program	SSE	MSE	MAE
Model 1	152.7484	3.8187	1.5407
Model 2	183.0371	3.1759	1.5388
Model 3	106.2985	2.6575	1.3542
Resnet	134.8040	3.3701	1.5571

**Table 4.** Comparison of the errors obtained for the three conditioning models and the residual network in the 2018 full volume data in the comparison test in the presence of temperature phenomena.

# 4.4. Results and Discussion

First, we analyze the effect of inversions on the radiometer, which are generated by turbulence in stationary fronts where warm and cold air currents converge, by changing the light path and thus affecting the radiometer observations. The analysis reveals that the inversions caused by stationary fronts are seasonal, so in this paper, we use the seasonal background field to correct the inversion results. Based on the seasonality this paper proposes three inversion models and comparative experiments are executed based on the models.

In summary, Model 3 works best in the range of experimental data and has a higher fitting ability than the other two models and the residual network. It can maintain real-value temperature profile tracking in the case of no inverse temperature and has the ability to describe the real-value profile in the case of inverse temperature. Because this model regulates the temperature, especially the inverse temperature phenomenon, the water vapor regulation still uses the residual network in Section 3, therefore, the water vapor accuracy remains the same.

# 5. Conclusions

Observations of tropospheric temperature and humidity profiles are of great importance. In a range of accuracy, a radiometer in 0–10 km can be a more accurate measurement of atmospheric temperature and humidity data. However, microwave remote sensing observation is non-contact and uncertain, and its essence is the inversion of the radiative transfer equation. This paper takes the inversion of traditional hierarchical BP networks as the basis. A residual network was proposed based on the data collapse problem of a BP hierarchical network. A dead zone adjustment model was added to the residual network to determine more accurate inverse temperature data. Experimental validation was carried out, and the following conclusions were drawn.

1. For the data collapse phenomenon of the BP hierarchical network, the stability of the residual network, comprising a traditional BP hierarchical network and a residual module, was greatly improved. No collapse occurred within the specified data range. Meanwhile, the water vapor profile accuracy was greatly improved at the upper level, and the error was within  $\pm 0.06$  g/m<sup>3</sup> at 6–10 km. However, using the residual network alone does not further improve the accuracy of the temperature profile based on the improved stability.

- 2. Based on the problem that the residual network has limited improvement in the accuracy of temperature inversion, the experimental results with large errors are analyzed in this paper. The analysis results point out that these errors are caused by the inverse temperature phenomenon to some extent. An inverse temperature phenomenon is an anomalous meteorological condition in which the temperature increases with height. Climate-based inversions are caused by the convergence of cold and warm air masses. In this paper, a possible cause of the inverse temperature observation error is proposed by modeling and analysis: turbulent disturbance generated by the cold and warm air masses leads to changes in the microwave phase difference, optical range, and angle of arrival, eventually affecting the optical thickness in the propagation of the brightness temperature signal. Based on the problem that the residual network has limited improvement in the accuracy of temperature inversion, the experimental results with large errors are analyzed in this paper. The analysis results point out that these errors are caused by the inverse temperature phenomenon to some extent. An inverse temperature phenomenon is an anomalous meteorological condition in which the temperature increases with height. Climate-based inversions are caused by the convergence of cold and warm air masses. This affects the transmission of the microwave signal, which in turn affects the accuracy of the radiometer observations.
- 3. Based on the climatic nature of the inverse temperature phenomenon, this study proposed a regulation model based on dead zone regulation. The model is a deadband regulation model based on seasonal a priori mean profiles. In the presence of the inverse temperature phenomenon, the MSE and MAE of Model 3 were 2.657 and 1.3542, respectively, which was better than those of the residual network.

Therefore, the residual neural network based on the layered BP network showed a better performance than the original network in terms of temperature and humidity profile fitting ability and stability. In addition, the temperature-profile fitting ability for the inverse temperature phenomenon was significantly improved after including the dead-zone adjustment model.

Nevertheless, this study had some limitations. First, the establishment of the priori mean profile lacks the variability of a large-scale climate. Second, the ability of the seasonal priori mean profile vertical profiles to respond to atmospheric changes remains insufficient.

**Author Contributions:** Conceptualization, methodology, Y.Z.; writing—original draft preparation, C.W.; visualization, K.Z.; investigation, P.W.; resources, funding acquisition, software, X.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Fundamental Research Funds for the Central Universitie. Funding number: 3072022YY0401.

Institutional Review Board Statement: Not applicable.

**Data Availability Statement:** Data availability. Data can be obtained by contacting the author Changzhe Wu (2015041018@hrbeu.edu.cn).

Acknowledgments: The authors would like to thank Yuan Xie, Songlin Zhang, Lei Yu and Honghui Li for their help on the algorithm. The reviewers' comments on improving the quality of the paper are also greatly appreciated. Special thanks also to the staff of Harbin Engineering University for their assistance in the field experiments.

Conflicts of Interest: The authors declare no conflict of interest.

# References

- Westwater, E.R.; Strand, O.N. Statistical information content of radiation measurements used in indirect sensing. *J. Atmos. Sci.* 1968, 25, 750–758. [CrossRef]
- 2. Smith, W.L. Iterative solution of the radiative transfer equation for the temperature and absorbing gas profile of an atmosphere. *Appl. Opt.* **1970**, *9*, 1993–1999. [CrossRef] [PubMed]

- Xiong, S.W. Research on One-Dimensional Variational Inversion Algorithm for Ground-Based Microwave Radiometer. Master's Thesis, Huazhong University of Science and Technology, Wuhan, China, 2016. Available online: <a href="https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201801&filename=1016920464.nh">https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201801&filename=1016920464.nh</a> (accessed on 15 January 2023).
- 4. Yang, J.; Min, Q. Retrieval of atmospheric profiles in the New York State Mesonet using one-dimensional variational algorithm. *JGR Atmos.* **2018**, *123*, 7563–7575. [CrossRef]
- 5. Ji, D.; Shi, J.; Xiong, C.; Wang, T.; Zhang, Y. A total precipitable water retrieval method over land using the combination of passive microwave and optical remote sensing. *Remote Sens. Environ.* **2017**, *191*, 313–327. [CrossRef]
- 6. Che, Y.; Ma, S.; Xing, F.; Li, S.; Dai, Y. An improvement of the retrieval of temperature and relative humidity profiles from a combination of active and passive remote sensing. *Meteorol. Atmos. Phys.* **2019**, *131*, 681–695. [CrossRef]
- Segal-Rozenhaimer, M.; Miller, D.J.; Knobelspiesse, K.; Redemann, J.; Cairns, B.; Alexandrov, M.D. Development of neural network retrievals of liquid cloud properties from multi-angle polarimetric observations. *J. Quant. Spectrosc. Radiat. Transf.* 2018, 220, 39–51. [CrossRef]
- 8. Chakraborty, R.; Saha, U.; Singh, A.K.; Maitra, A. Association of atmospheric pollution and instability indices: A detailed investigation over an Indian urban metropolis. *Atmos. Res.* 2017, *196*, 83–96. [CrossRef]
- Yan, X.; Liang, C.; Jiang, Y.; Luo, N.; Zang, Z.; Li, Z. A Deep Learning Approach to Improve the Retrieval of Temperature and Humidity Profiles From a Ground-Based Microwave Radiometer. *IEEE Trans. Geosci. Remote Sens.* 2020, 58, 8427–8437. [CrossRef]
- 10. Azamathulla, H.M.; Rathnayake, U.; Shatnawi, A. Gene expression programming and artificial neural network to estimate atmospheric temperature in Tabuk, Saudi Arabia. *Appl. Water Sci.* **2018**, *8*, 184. [CrossRef]
- Zhao, Y.; Zhou, D.; Liu, C.; Wu, P.; Li, L.; Zhang, L.; Cheng, W. Numerical correction of atmospheric temperature profiles in clear and cloudy days. *Atmos. Res.* 2019, 217, 49–56. [CrossRef]
- 12. Zhao, Y.; Zhou, D.; Yan, H. An improved retrieval method of atmospheric parameter profiles based on the BP neural network. *Atmos. Res.* **2018**, *213*, 389–397. [CrossRef]
- 13. Aron Jazcilevich, D.A.; Fuentes-Gea, V. The random choice method in the numerical solution of the radiative transfer equation. *Environ. Softw.* **1994**, *9*, 23–31. [CrossRef]
- Belikovich, M.V.; Makarov, D.S.; Serov, E.A.; Kulikov, M.Y.; Feigin, A.M. Validation of atmospheric absorption models within the 20–60 GHz band by simultaneous radiosonde and microwave observations: The advantage of using ECS formalism. *Remote Sens.* 2022, 14, 6042. [CrossRef]
- 15. Wang, J.; Wen, Y.; Gou, Y.; Ye, Z.; Chen, H. Fractional-order gradient descent learning of BP neural networks with Caputo derivative. *Neural Netw.* **2017**, *89*, 19–30. [CrossRef]
- 16. Cai, J.; Yi, C. An adaptive gradient-descent-based neural networks for the on-line solution of linear time variant equations and its applications. *Inf. Sci.* 2023, 622, 34–45. [CrossRef]
- 17. Wei, M.; Jiang, W.; Hu, X. Residual storey drift estimation of the MDOF system with the weak storey under seismic excitations using the BP network. *Structures* **2023**, *48*, 465–477. [CrossRef]
- Han, Y.; Cao, L.; Geng, Z.; Ping, W.; Zuo, X.; Fan, J.; Wan, J.; Lu, G. Novel economy and carbon emissions prediction model of different countries or regions in the world for energy optimization using improved residual neural network. *Sci. Total Environ.* 2022, *860*, 160410. [CrossRef]
- 19. Shi, Y.; Xiong, L.; Qin, H.; Han, J.; Sun, Z. Seismic fragility analysis of LRB-isolated bridges considering the uncertainty of regional temperatures using BP neural networks. *Structures* **2022**, *44*, 566–578. [CrossRef]
- Sang, L.; Xu, M.; Qian, S.; Wu, X. Knowledge Graph enhanced Neural Collaborative Filtering with Residual Recurrent Network. Neurocomputing 2021, 454, 417–429. [CrossRef]
- 21. Yu, Z.; Ke, Z.; Ting, Z.; Liang, Z.; Xiong, D. Seabed sediments classification based on side-scan sonar images using dimensioninvariant residual network. *Appl. Ocean Res.* 2023, 130, 103429. [CrossRef]
- Gou, J.; Qu, S.; Guan, H.; Shi, P.; Zhang, Z.; Yang, H.; Liu, J.; Su, Z.; Han, X. Seasonal variation of transit time distribution and associated hydrological processes in a Moso bamboo watershed under the East Asian monsoon climate. *J. Hydrol.* 2023, 617 Pt B, 128912. [CrossRef]
- 23. Zhang, K.; Wang, F.; Weng, N.; Wu, X.; Li, X.; Luo, T. Optical turbulence characteristics in the upper troposphere–lower stratosphere over the Lhasa within the Asian summer monsoon anticyclone. *Remote Sens.* **2022**, 130, 4104. [CrossRef]
- Qiong, G.; Shi, Y.; Zong, J.; Lin, H.; Xiao, W. Structure of the refractive index distribution of the supersonic turbulent boundary layer. Opt. Lasers Eng. 2013, 51, 1113–1119. [CrossRef]
- 25. Sun, H.; Shi, H.; Chen, H.; Tang, G.; Sheng, C.; Che, K.; Chen, H. Evaluation of a method for calculating the height of the stable boundary layer based on wind profile lidar and turbulent fluxes. *Remote Sens.* **2021**, *13*, 3596. [CrossRef]
- 26. Ishimaru, A. The beam ease. In *Laser Beam Propagation in the Turbulent Atmosphere;* Stroheben, J.W., Ed.; Springer: Berlin/Heidelberg, Germany, 1978; pp. 120–170.
- 27. Banakh, V.A.; Smalikho, I.N.; Falits, A.V. Wind-temperature regime and wind turbulence in a stable boundary layer of the atmosphere: Case study. *Remote Sens.* 2020, *12*, 955. [CrossRef]
- 28. Banakh, V.A.; Smalikho, I.N. Lidar estimates of the anisotropy of wind turbulence in a stable atmospheric boundary layer. *Remote Sens.* **2019**, *11*, 2115. [CrossRef]
- 29. Pan, Y.; Zhao, M.; Zhang, M.; Dou, J.; Zhao, J.; Li, B.; Hu, Y. Propagation properties of rotationally symmetric power-exponentphase vortex beam through oceanic turbulence. *Opt. Laser Technol.* **2023**, *159*, 109024. [CrossRef]

- 30. Pomeau, Y.; Le Berre, M. Transition to turbulence or to periodic patterns in parallel flows. *Chaos Solitons Fract.* **2023**, *166*, 113019. [CrossRef]
- Tang, Y.; Duan, Z.; Yang, J.; Chen, Y. The probabilistic turbulence profiles of tropical cyclones in open and flat terrain. *J. Wind Eng. Ind. Aerodyn.* 2022, 228, 105107. [CrossRef]
- 32. Wu, J.; Shi, Z.; Yang, Y. Response of East Asian summer monsoon climate to North Atlantic meltwater during the Younger Dryas. *Quat. Sci. Rev.* 2022, 295, 107766. [CrossRef]
- 33. Rajeev, A.; Mahto, S.S.; Mishra, V. Climate warming and summer monsoon breaks drive compound dry and hot extremes in India. *iScience* 2022, 25, 105377. [CrossRef] [PubMed]
- Bakota, M.; Kos, S.; Mrak, Z.; Brčić, D. A new approach for improving GNSS geodetic position by reducing residual tropospheric error (RTE) based on surface meteorological data. *Remote Sens.* 2023, 15, 162. [CrossRef]
- 35. Xiao, X.; Weng, F. A comparison of information content at microwave to millimeter wave bands for atmospheric sounding. *Remote Sens.* 2022, 14, 6124. [CrossRef]
- 36. Zhang, L.; Tie, S.; He, Q.; Wang, W. Performance analysis of the temperature and humidity profiles retrieval for FY-3D/MWTHS in arctic regions. *Remote Sens.* **2022**, *14*, 5858. [CrossRef]
- Huang, P.; Guo, Q.; Han, C.; Zhang, C.; Yang, T.; Huang, S. An improved method combining ANN and 1D-Var for the retrieval of atmospheric temperature profiles from FY-4A/GIIRS hyperspectral data. *Remote Sens.* 2021, 13, 481. [CrossRef]
- Fang, X.; Guo, Z.; Jiang, D.; Zhang, W.; Zhang, R.; Li, M.; Wang, Y.; Zhang, T.; Miao, Y. No monsoon-dominated climate in northern subtropical Asia before 35 Ma. *Glob. Planet. Chang.* 2022, 218, 103970. [CrossRef]
- 39. Chase, B.M.; Boom, A.; Carr, A.S.; Reimer, P.J. Climate variability along the margin of the southern African monsoon region at the end of the African Humid Period. *Quat. Sci. Rev.* **2022**, *291*, 107663. [CrossRef]
- 40. Wu, C.-H.; Shiu, C.-J.; Tsai, I.-C.; Lee, S.-Y. Climatological changes in East Asian winter monsoon circulation in a warmer future. *Atmos. Res.* **2023**, *284*, 106593. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.