remote sensing

MDPI

Article

Comparative Analysis and Optimal Selection of Calibration Functions in Pure Rotational Raman Lidar Technique

Yinghong Yu [1], Siying Chen [1,2,*], Wangshu Tan [1], Rongzheng Cao [1], Yixuan Xie [1], He Chen [1,2], Pan Guo [1], Jie Yu [1], Rui Hu [1], Haokai Yang [1] and Xin Li [3]

[1] School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China; yh_yu@bit.edu.cn (Y.Y.); tanws@bit.edu.cn (W.T.); caorz@bit.edu.cn (R.C.); yixuan@bit.edu.cn (Y.X.); shinianshao@bit.edu.cn (H.C.); guopan@bit.edu.cn (P.G.); bityujie@bit.edu.cn (J.Y.); hu_rui@bit.edu.cn (R.H.); haok_yang@bit.edu.cn (H.Y.)
[2] Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing 314019, China
[3] Academy of Military Sciences, Academy of Military Medicine Sciences, Beijing 100085, China; lixin2006011649@163.com
* Correspondence: csy@bit.edu.cn

Abstract: The pure rotational Raman (PRR) lidar technique relies on calibration functions (CFs) to extract temperature information from raw detection data. The choice of CF significantly impacts the accuracy of the retrieved temperature. In this study, we propose a method that combines multiple Monte Carlo simulation experiments with a statistical analysis, and we first conduct simulated comparisons of the calibration effects of different CFs while considering the impact of noise. We categorized ten common CFs into four groups based on their functional form and the number of calibration coefficients. Based on functional form, specifically, we defined $1/T = \mathrm{f}(\ln Q)$ as a forward calibration function (FCF) and $\ln Q = \mathrm{g}(1/T)$ as a backward calibration function (BCF). Here, T denotes temperature, and Q denotes the signal intensity ratio. Their performance within and outside the calibration interval is compared across different integration times, smoothing methods, and reference temperature ranges. The results indicate that CFs of the same category exhibit similar calibration effects, while those of different categories exhibit notable differences. Within the calibration interval, the FCF performs better, especially with more coefficients. However, outside the calibration interval, the linear calibration function (which can be considered a two-coefficient FCF) has an obvious advantage. Conclusions based on the simulation results are validated with actual data, and the factors influencing calibration errors are discussed. Utilizing these findings to guide CF selection can enhance the accuracy and stability of PRR lidar detection.

Keywords: pure rotational Raman temperature measurement technology; calibration function comparison; Monte Carlo method; least squares fitting

check for updates

Citation: Yu, Y.; Chen, S.; Tan, W.; Cao, R.; Xie, Y.; Chen, H.; Guo, P.; Yu, J.; Hu, R.; Yang, H.; et al. Comparative Analysis and Optimal Selection of Calibration Functions in Pure Rotational Raman Lidar Technique. *Remote Sens.* **2024**, *16*, 3690. https://doi.org/10.3390/rs16193690

Academic Editors: Dmitry Efremenko and Alexander Kokhanovsky

Received: 27 August 2024
Revised: 29 September 2024
Accepted: 30 September 2024
Published: 3 October 2024

1. Introduction

Atmospheric temperature is a crucial factor in the atmospheric environment. The troposphere, which is the most closely related to human activities, has rapidly changing temperatures and a complex change mechanism. Lidar, with its high temporal and spatial resolution and precision, is extensively employed for the detection of atmospheric temperatures. Among various lidar technologies, pure rotational Raman (PRR) lidar stands out as the preferred method for detecting vertical temperature profiles in the troposphere and lower stratosphere [1,2].

The PRR temperature measurement technique was initially proposed by Cooney [3], and it is based on the different temperature dependences of PRR spectral lines [4]. By comparing two PRR signals with opposite temperature dependences, a quantity solely dependent on atmospheric temperature is obtained, which is then used for temperature retrieval. Prior to temperature retrieval, PRR lidar needs to be calibrated; hence, it is also called the relative temperature measurement technique. If only one high-order and one low-order PRR spectral line are received independently, then the signal intensity ratio Q of the high- and low-quantum-number channels can be expressed as $\exp(\alpha + \beta/T)$. However, in practical applications, whether using double-grating spectrometers (DGs) or interference filters (IFs), it is challenging to extract independent PRR spectral lines [5–11]. At this time, Q can no longer be directly expressed as a simple function of temperature T [12], necessitating the use of a calibration function (CF) to approximate the relationship between Q and T so as to invert the detection signal into an atmospheric temperature profile [1].

Several CFs have been proposed to approximate the relationship between Q and T. In 1983, Arshinov et al. [12] proposed a method to represent the intensity of the actual received multi-spectral line signal with a single spectral line intensity, resulting in a first-order linear CF. Behrendt et al. [13] extended the exponential term of the linear CF to a quadratic form in 2000, resulting in a second-order CF with a lower theoretical error. Di Girolamo et al. [14] further extended the exponential term to a cubic form, yielding a theoretically more accurate third-order CF. In addition, specific functions [15,16] and inversion methods [17–19] have been proposed for particular situations. For example, Chen et al. [15] proposed a CF that maintains good accuracy and robustness under a low signal-to-noise ratio (SNR), and a CF proposed by He et al. [16] could reduce the error by about 50% compared with commonly used functions under a weak SNR. These CFs and inversion methods each have their own advantages, but they are only applicable in specific cases. In general, first- and second-order CFs remain commonly used [1,16], although they are also special forms derived under certain assumptions.

To better align with actual conditions, Gerasimov et al. [20,21] conducted a large analysis and derivation of the relationship between the signal intensity ratio Q of multi-spectral line reception and temperature T, resulting in general calibration function 1 (GCF1) for PRR lidar, and GCF2 was obtained from the special case of its integer power approximation. GCF1 and GCF2 correspond to Equations (1) and (2), respectively. Most CFs are special cases of these two GCFs. Theoretically, the GCF accounts for the collision broadening of PRR spectral lines from all atmospheric molecules, thus achieving a greater calibration accuracy. However, as the GCF is an infinite series, it cannot be directly used for temperature inversion, and, thus, its special forms are utilized to calculate temperature profiles. In 2018, Gerasimov [22] introduced nine nonlinear CFs, as shown in Table 1. Gerasimov compared the calibration effects of these nine CFs through simulations and validated them with actual data; however, the conclusions drawn are only applicable to specific lidar systems. Subsequently, Gerasimov [23] further simulated and compared the calibration effects of various CFs under different spectral filter bandwidths and laser central wavelength settings. Other studies also employed similar methods to determine the most suitable CF for specific systems [24–26]. However, these analyses did not consider the impact of noise and, thus, do not accurately reflect practical conditions.

$$\ln Q(T) = \cdots + \frac{\alpha_{-3}}{T\sqrt{T}} + \frac{\alpha_{-2}}{T} + \frac{\alpha_{-1}}{\sqrt{T}} + \alpha_0 + \alpha_1\sqrt{T} + \alpha_2 T + \cdots = \sum_{n=-\infty}^{\infty} \alpha_n T^{\frac{n}{2}} \tag{1}$$

$$\ln Q(T) \approx \cdots + \frac{\beta_{-3}}{T^3} + \frac{\beta_{-2}}{T^2} + \frac{\beta_{-1}}{T} + \beta_0 + \beta_1 T + \beta_2 T^2 + \cdots = \sum_{n=-\infty}^{\infty} \beta_n T^n \tag{2}$$

When utilizing PRR technology for temperature detection in practice, it is inevitably affected by noise. Since the backscatter signal detected by PRR lidar is relatively weak, photon-counting detectors are generally employed. In lidar systems that use photon counting for detection, shot noise is the primary source of random noise [27,28]. During the PRR temperature measurement process, shot noise and the calibration procedure are the most influential factors affecting the detection results [29]. Therefore, analyzing the calibration effects of different CFs under the influence of shot noise can better reflect the actual situation. In practical detection scenarios, the SNR and corresponding atmospheric temperature range of the retrieved data typically differ from those of the calibration data. It is common practice to select an interval with a higher SNR for calibration. All of these will lead to calibration results used for temperature retrieval extending beyond the calibration interval. Therefore, it is essential to analyze the performance of each CF beyond the calibration interval. However, previous studies have neither conducted comparative analyses of CFs under the influence of noise nor examined their extrapolation performance.

Table 1. Nonlinear CFs in nine special cases of GCFs. In the table, y denotes ln$Q$, x denotes $1/T$, and u denotes $1/\sqrt{T}$. a, b, c, and d are the coefficients to be calibrated.

Three-Coefficient Calibration Functions		Four-Coefficient Calibration Functions
Direct special cases of GCF2	CF1: y = a + bx + cx^2 CF2: y = a + bx + c/x	Indirect special cases of GCF2 CF7: x = a + by + cy^2 + dy^3
Indirect special cases of GCF2	CF3: x = a + by + cy^2 CF4: x = a + by + c/y	CF8: x = a + by + cy^2 + d/y CF9: x = a + by + c/y + d/y^2
Direct special cases of GCF1	CF5: y = a + bu + cu^2 CF6: y = a + bu + c/u	

In this paper, the effects of different CFs within and outside the calibration interval are compared and analyzed under the influence of noise. The primary focus is on shot noise, incorporating Poisson random processes during signal simulation. Due to the randomness introduced by the Poisson process, multiple simulations are conducted using the Monte Carlo (MC) method [30,31], followed by a statistical analysis. Additionally, the calibration performance of each CF under various integration times, smoothing methods, and reference temperature ranges is analyzed using MC experiments. This paper is organized as follows: In Section 2, the principle of signal generation and temperature inversion is introduced, and the implementation of the MC and statistical methods is explained in detail. In Section 3, the calibration effects of different CFs are compared using simulated data following the methods described in Section 2. In Section 4, the simulation results from Section 3 are validated with actual observation data, and the factors affecting calibration results and their interdependencies are further analyzed. Finally, Section 5 concludes this paper.

2. Methods

2.1. Echo Signal Simulation

Firstly, a single-pulse echo signal is simulated according to the lidar equation. The number of echo photons detected via PRR scattering can be expressed as

$$S(z) = E_0 \Delta z A \xi \frac{\lambda_0}{hc} \frac{O(z)}{z^2} \beta(z) \tau(z) + S_\mathrm{d} + S_\mathrm{b} \tag{3}$$

where $S(z)$ denotes the number photons received by the lidar at range z. E_0 denotes the emitted energy of the laser pulse, λ_0 is the central wavelength of the laser, Δz is the range resolution of the system, A is the effective receiving area of the telescope, h is the Planck constant, and c is the speed of light. ξ is the overall efficiency of the lidar system, including the optical efficiency of all components in the transmission and receiving optical systems, as well as the detection efficiency of the electronic devices. $O(z)$ denotes the geometric overlap factor, primarily caused by the incomplete overlap between the laser beam cross-section and the field of view received by the telescope. $\beta(z)$ denotes the backscatter coefficient for PRR scattering by atmospheric molecules, $\tau(z)$ denotes the two-way transmittance of the laser through the atmosphere, S_d is the dark noise of the detector, and S_b is the noise generated by atmospheric background light.

The equation for calculating $\beta(z)$ is

$$\beta(z) = \sum_{i=\mathrm{O_2,N_2}} \sum_{J_i} N(z) \eta_i \tau_\mathrm{RR}(J_i) \left(\frac{d\sigma}{d\Omega}\right)_\pi^{\mathrm{RR},i} (J_i) \tag{4}$$

where $N(z)$ is the number density of atmospheric molecules, and η_i denotes the volumetric fraction of nitrogen or oxygen in the atmosphere, with nitrogen accounting for 78.08% and oxygen accounting for 20.95%. The term J_i indicates different rotational quantum numbers (of nitrogen or oxygen molecules), $\tau_{\mathrm{RR}}(J_i)$ is the reception efficiency of the lidar system for the J_i corresponding spectral lines in the Raman spectrum, and $(d\sigma/d\Omega)_{\pi}^{\mathrm{RR},i}(J_i)$ refers to the PRR differential backscattering cross-section corresponding to J_i. The method for calculating the differential scattering cross-section of single lines in the PRR spectrum is detailed in [32–34].

$\tau(z)$ can be obtained by integrating the extinction coefficient along the transmission path, with the calculation equation given by

$$\tau(\mathrm{z}) = \exp\left[-\int_0^z \alpha(z', \lambda_0)dz'\right] \exp\left[-\int_0^z \alpha(z', \lambda_{\mathrm{r}})dz'\right] \quad (5)$$

where the first term represents the loss of outgoing light with wavelength λ_0 transmitted to range z, and the second term represents the loss of backscattered light with wavelength λ_{r} transmitted to the receiving telescope. α is the atmospheric extinction coefficient, for which the calculation equation is as follows:

$$\alpha(z, \lambda) = \alpha_{\mathrm{a}}(z, \lambda) + \alpha_{\mathrm{m}}(z, \lambda) \quad (6)$$

where α_{a} denotes the extinction coefficient of aerosol in the atmosphere, and α_{m} denotes the extinction coefficient of atmospheric molecules. In this study, the extinction coefficients of atmospheric molecules and aerosols are simulated based on the standard atmospheric model. The detailed calculation methods can be found in [35].

S_{d} can be calculated by integrating time Δt and dark count rate S_{d0} according to the following equation:

$$S_{\mathrm{d}} = S_{\mathrm{d0}}\Delta t \quad (7)$$

The equation for calculating S_{b} is as follows:

$$S_{\mathrm{b}} = I_{\mathrm{b}} A \Delta\lambda \Delta t \xi_{\tau} \xi_{\mathrm{d}} \frac{\phi^2 \pi}{4} \frac{\lambda_{\mathrm{r}}}{hc} \quad (8)$$

where I_{b} denotes the luminance of atmospheric background radiation, the typical value of which can be set to 0.149 $\mathrm{w}/(\mathrm{m}^2\mathrm{sr}\cdot\mathrm{nm})$ during the day. The atmospheric background radiation during nighttime detection is approximately 0.1% of the daytime value [36], which can generally be neglected, but it should be considered when considering the impact of noise. $\Delta\lambda$ is the transmittance spectral bandwidth of the receiving channel, ξ_{τ} is the optical efficiency of the receiving optical system, ξ_{d} is the detector efficiency, and ϕ is the field of view of the receiving telescope.

After simulating with Equation (3), a set of single-pulse signals containing background noise and dark noise is obtained. This set is considered the expectation of single-pulse echo photons. To introduce shot noise, multiple Poisson randomizations are applied based on the required integration time, resulting in multiple sets of single-pulse signals that include shot noise.

2.2. Temperature Inversion

The whole process of temperature inversion from lidar signals can be divided into three modules: data preprocessing, CF fitting, and temperature retrieval.

Data preprocessing primarily involves integrating and denoising the raw signal. First, the integration time is set based on the system configuration and the detection purpose, after which the raw signal is then accumulated over the specified integration period. The position where the signal intensity ceases to decrease with an increasing altitude is referred to as the detection limit of the system. Signals beyond this detection limit are considered background noise, including contributions from atmospheric background radiance and the noise generated by the receiving electronic system. Even after subtracting the background noise, the integrated signal still contains random noise. A common denoising technique is the moving window average method, which is categorized into fixed-window smoothing [13,14,24,37] and variable-window smoothing [21,38]. While integration and smoothing can enhance the SNR, they also reduce the temporal and spatial resolution of the system and may introduce some unquantifiable errors [29]. In practice, it is essential to balance the trade-offs between the SNR and temporal–spatial resolution by choosing an appropriate integration time and smoothing methods.

For the preprocessed signal, a calibration interval (i.e., the range corresponding to the calibration sample points) is selected, followed by the fitting of the CF. The calibration interval should be selected in a region that is unaffected by the geometric overlap factor at low altitudes and the low SNR at high altitudes. For the lidar system used in this study, the influence range of the geometric overlap factor is below approximately 1 km [39]. Long-term detection results also show that the temperature detection results in the range of 0–1 km have large errors. Therefore, in this study, the calibration interval is chosen to be above 1 km, extending to the range where the pre-smoothed SNR reaches 10 (approximately 5 km). Here, the 1 km equal altitude position is relative to the working position of the lidar system. When processing simulation data, the calibration interval is consistently set to 1–5 km. As previously mentioned, the GCFs cannot be directly applied for calibration. Gerasimov et al. [20–22] proposed that a specific form of CFs, suitable for calibration, should meet two conditions: the inclusion of a linear term and the ability to derive a physically meaningful solution. Under these conditions, nine specific CFs are retained, as shown in Table 1. Although many other specialized CFs exist, they either fail to meet the aforementioned conditions, are computationally complex, or provide only marginal improvements in calibration accuracy. Therefore, they are not included in this study.

In addition to the nine nonlinear CFs in Table 1, a linear CF is also included for a comparative analysis. Based on the number of calibration coefficients and structural characteristics, the 10 CFs are classified, and the classification results are shown in Table 2. To illustrate, the three-coefficient forward calibration function (3c-FCF) is taken as an example, which includes three coefficients to be calibrated and uses x (i.e., $1/T$) as the dependent variable. When performing temperature retrieval, the temperature value (i.e., T) can be directly obtained from the calibration results. Conversely, the three-coefficient backward calibration function (3c-BCF) uses y (i.e., $\ln Q$) as the dependent variable and requires an inverse solution for temperature retrieval. The logic for the four-coefficient forward calibration function (4c-FCF) is consistent with the above, while the linear calibration function (LCF) does not require the same distinction as the nonlinear CFs. For the preprocessed signal, 10 different CFs (i.e., CF0–CF9) are used for calibration, resulting in calibration coefficients for each one.

Table 2. Classification of a linear CF and nine nonlinear CFs. In the table, y denotes $\ln Q$, x denotes $1/T$, and u denotes $1/\sqrt{T}$. a, b, c, and d are the coefficients to be calibrated.

Type 1: LCF	**Type 2: 3c-BCF**	**Type 3: 3c-FCF**	**Type 4: 4c-FCF**
CF0: $x = a + by$	CF1: $y = a + bx + cx^2$ CF2: $y = a + bx + c/x$ CF3: $y = a + bu + cu^2$ CF4: $y = a + bu + c/u$	CF5: $x = a + by + cy^2$ CF6: $x = a + by + c/y$	CF7: $x = a + by + cy^2 + dy^3$ CF8: $x = a + by + cy^2 + d/y$ CF9: $x = a + by + c/y + d/y^2$

Each CF in Table 2 corresponds to a specific temperature retrieval function (TRF). The temperature retrieval functions TRF0–TRF9, corresponding to CF0–CF9, are all expressed in the form of temperature T as a function of $\ln Q$, which can be directly derived by solving the calibration equations in Table 2. For any CF, once the calibration coefficients are obtained, the corresponding TRF is determined. By substituting the signal intensity ratio at a given position into the TRF, the temperature at that location can be obtained, enabling the retrieval of the temperature profile. For the N sets of single-pulse signals simulated in Section 2.1, temperature retrieval is performed using different CFs following the aforementioned method. The calibration error (i.e., the difference between the retrieved temperature and the reference temperature) is then calculated to compare the calibration accuracy of the various CFs.

2.3. MC Experiments and Statistical Methods

The entire process, from the simulation of single-pulse signals to the calculation of calibration errors for various CFs, constitutes a single MC simulation experiment. The flow of one MC experiment is illustrated in the yellow box in Figure 1. A concise summary of the complete MC simulation experiment is as follows: Using the specified system parameters and atmospheric conditions, simulate the single-pulse echo signals based on Equation (3), treating them as the expected number of single-pulse echo photons. Firstly, perform N Poisson randomizations to generate N sets of single-pulse signals that include shot noise. Then, apply the three modules outlined in Section 2.2—signal preprocessing, CF fitting, and temperature retrieval—to process the signals and compute the retrieved temperatures, and, finally, calculate the calibration/retrieval errors for various CFs.

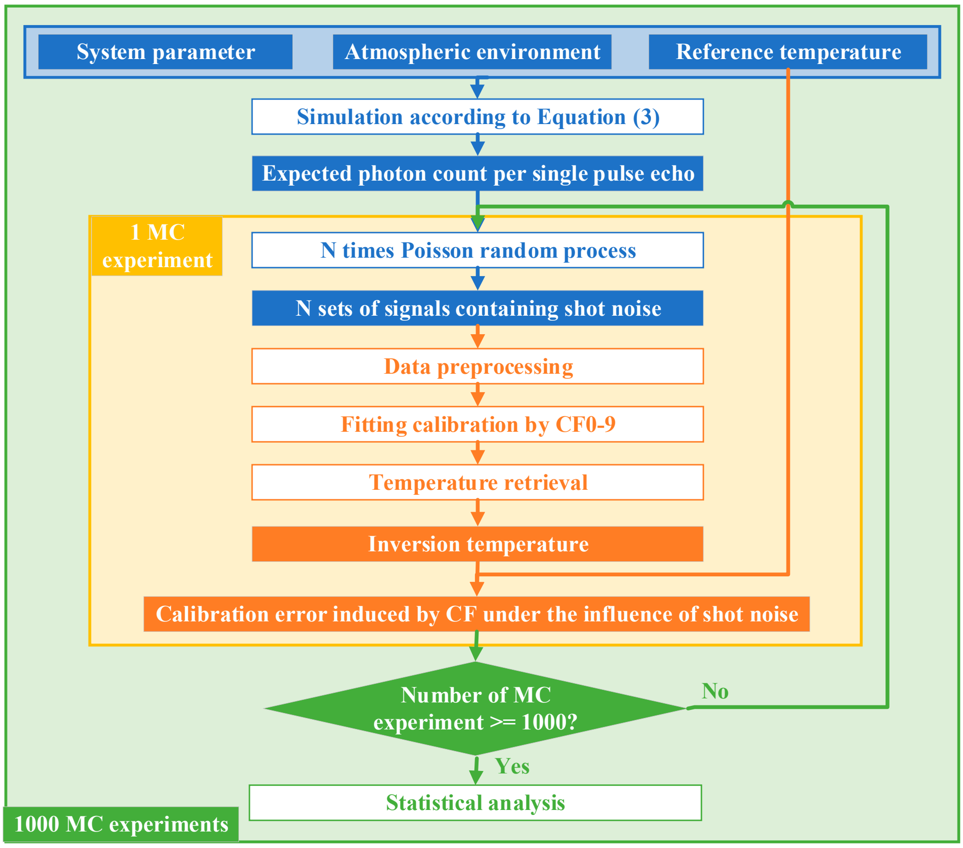


Figure 1. A flowchart illustrating the process of the statistical analysis of the calibration performance of different CFs based on MC simulation experiments.

Due to the introduction of a Poisson random process, the results obtained from a single MC simulation are stochastic. Therefore, multiple MC simulations are conducted under the same experimental conditions to obtain average results and distribution probabilities. In addition, before each MC simulation experiment, a random seed is set to make the experimental results repeatable.

The number of repeated experiments is a critical parameter in MC experiments. The JCGM 2008 [40] document, jointly authored by the International Bureau of Weights and Measures (BIPM) and others, provides guidance on the number of MC simulations. It mentions that, typically, 10^6 MC trials are necessary to achieve a 95% confidence level in the output results. However, conducting 10^6 simulations requires substantial time and computational resources. Therefore, an experiment was conducted to investigate the relationship between the accuracy of the output results and the number of trials. For different numbers of MC simulations, the relative deviation between the output results at various distance gates and the results obtained from 10^6 trials was calculated. The mean absolute relative deviation (MARD) and standard deviation of the relative deviation (SDRD) were then statistically analyzed, as shown in Table 3. When the number of trials was 1000, the MARD relative to the results of the 10^6 trials was 0.959%. To reduce the MARD by an order of magnitude, the number of trials must be increased by two orders of magnitude. Thus, considering both output accuracy and computational cost, 1000 MC simulations were selected in this study, which can still provide results with a high degree of credibility.

Table 3. The differences between the output results obtained from various numbers of MC trials and those obtained from 10^6 MC trials.

MC Trials Number	**5×10^5**	**10^5**	**5×10^4**	**10^4**	**5000**	**1000**	**500**
MARD to 10^6	0.000524	0.00103	0.00146	0.00310	0.00414	0.00959	0.0134
SDRD to 10^6	0.000559	0.00109	0.00160	0.00332	0.00439	0.0103	0.0146

For each preset condition, 1000 MC simulation experiments are conducted. As a result, for each CF, there are 1000 corresponding calibration errors on each range gate. The mean absolute error (MAE) [41] and the standard deviation of the error (SDE) [42] of these 1000 sets of calibration results are then calculated. The formula for MAE is given in Equation (9), which is a precise measure of the average magnitude of errors and can accurately reflect the size of the actual prediction error. The formula for SDE is provided in Equation (10), which is the arithmetic square root of the variance, reflecting the degree of dispersion among individual data points in a dataset:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - x^{\text{ref.}}\right| \tag{9}$$

$$SDE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \tag{10}$$

where $x^{\text{ref.}}$ denotes the reference true value, and $\overline{x}$ is the mean of the dataset.

3. Simulation Results

3.1. Simulation Parameters and Calibration Errors without Noise

The major parameters of the system used in the simulations are listed in Table 4. This model was set based on the PRR lidar system from the Beijing Institute of Technology, Beijing, China [43]. When calculating the backscatter cross-section, the contributions of the first 24 spectral lines of both the Stokes and anti-Stokes branches in the PRR spectrum to the high- and low-order PRR channels were considered. The PRR spectra of the N_2 and O_2 molecules at 200 K and 280 K are shown in Figure 2. Figure 2 also presents the spectral transmission function (STF) curve used in the simulations, which was set based on the spectrometer-measured curves of the double-grating polychromator arranged by the Beijing Institute of Technology. In our simulation, we utilized data from the 1976 U.S. Standard Atmosphere model, using its tropospheric temperature profile as the reference temperature profile and its atmospheric pressure data to calculate the molecular number density (i.e., N in Equation (4)).

Table 4. Major parameters for PRR lidar system simulations.

System Configuration	Parameter Value
Laser	Nd: YAG
Wavelength	532 nm
Laser pulse energy	60 mJ
Pulse repetition rate	20 Hz
Telescope effective diameter	0.2 m
Total optics efficiency	0.5
Detector quantum efficiency	0.1
Detector dark count	100 s^{-1}

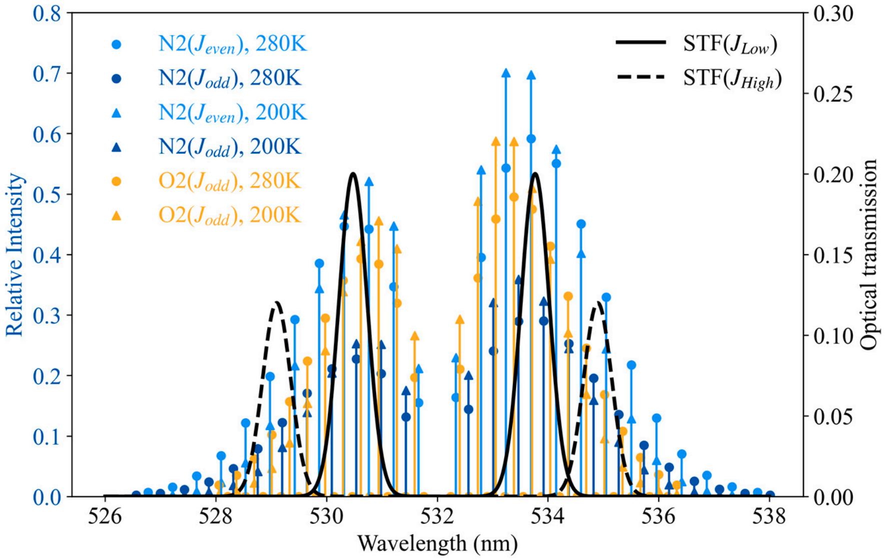


Figure 2. The PRR spectra of N2 and O2 at T = 200 K (blue and yellow triangles, respectively) and T = 280 K (blue and yellow circles, respectively). The even- and odd-order PRR spectra of N2 are distinguished by light blue and dark blue, respectively. J_{evev} and J_{odd} represent the even and odd order of the rotation quantum numbers, respectively. The wavelength of incident light is 532 nm. The black dashed and solid lines represent the transmission curves of the receiver, with the passband centers at 529.10, 530.48, 533.77, and 534.90 nm, and the full width at half maximum (FWHM) of the curves is 0.6 nm. The maximum transmittance of the low- and high-order channels is 20% and 12%, respectively.

The echo signals are simulated using the parameters mentioned above, and the absolute calibration (or temperature) error is calculated using similar methods in [22–26]. The absolute calibration error represents the difference between the temperature profile retrieved from the simulated signal intensity ratio Q by any CF in Table 2 and the reference temperature profile, without considering the impact of noise. The absolute calibration errors of each CF are shown in Figure 3. The results in Figure 3 show that the magnitude of the absolute calibration errors for the nine nonlinear CFs (i.e., CF1–9) obtained in this study is consistent with that in [22,23]. Furthermore, the four-coefficient CFs similarly yield smaller absolute calibration errors. Additionally, due to the specific STF used in this study, the values and ranking of the absolute calibration errors for each CF differ from the results in the references. This also validates the conclusion in the literature that "CFs' comparison results of lidar systems with different configurations (i.e., laser wavelength and STF) are different." The absolute calibration errors produced by each CF in this study, ranked from smallest to largest, are as follows: CF7 < CF9 < CF5 < CF8 < CF1 < CF3 < CF4 < CF2 < CF6 < CF0. CF7 has the smallest error (less than 2.5×10^{-5} K), while CF5 has the smallest error among the three-coefficient CFs (less than 2×10^{-3} K).

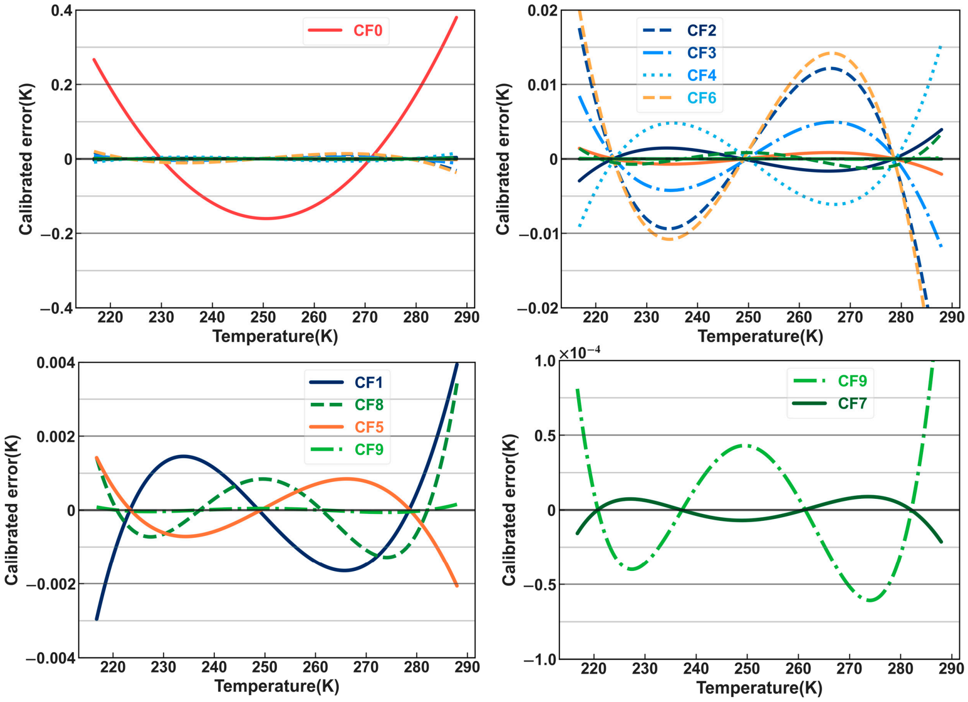


Figure 3. Simulation results: an intercomparison of absolute calibration (or temperature) errors produced using CF0–9 without considering noise.

However, as previously discussed, comparing absolute calibration errors without considering noise effects is not sufficiently reflective of real-world conditions. The following presents a statistical comparison of calibration errors for different CFs while considering the influence of noise (including shot noise, background light noise, and dark noise), that is, the MC simulation and statistical comparison results obtained according to the methods described in Section 2 of this paper.

3.2. MC Simulation and Statistical Results

According to the simulation method described in Section 2, in addition to the previously discussed system parameters and atmospheric model data, it is also necessary to configure the integration time and smoothing method in the data preprocessing module. We set the integration time to 60 min, corresponding to the Poisson random number N of 72,000 in Figure 1. We applied an initial window size of 5 and used a smoothing method where the sliding window expands by 1 point on each side of the central point for every 20 data points, which is referred to as variable sliding window method 1 (VSW-M1). Specifically, within the height range of 0 to 600 m, the sliding window is 150 m, and, for every 600 m increase in detection height, the sliding window expands by 30 m on both the upper and lower sides. VSW-M1 was proposed after testing various smoothing methods from [13,14,21,24,37,38] on actual data and considering the impact on the SNR and the temporal–spatial resolution.

After setting the integration time and smoothing method, 1000 MC experiments were conducted following the workflow shown in Figure 1. The MAE and SDE of the calibration (or temperature) results in both the calibration interval and the extrapolation interval (i.e., outside the calibration range) were statistically obtained, as shown in Figure 4. The significance of these two statistics was mentioned in Section 2.3. The MAE can reflect the magnitude of the actual prediction error, with smaller values indicating a higher calibration accuracy. The SDE can reflect the degree of dispersion among individuals in a set of data, with smaller values indicating a higher stability of the calibration results. In the figure, curves of the same color system represent the statistical results for the same category of

CFs: Red represents LCF (CF0), blue represents 3c-BCF (CF1–4), orange represents 3c-FCF (CF5–6), and green represents 4c-FCF (CF7–9).

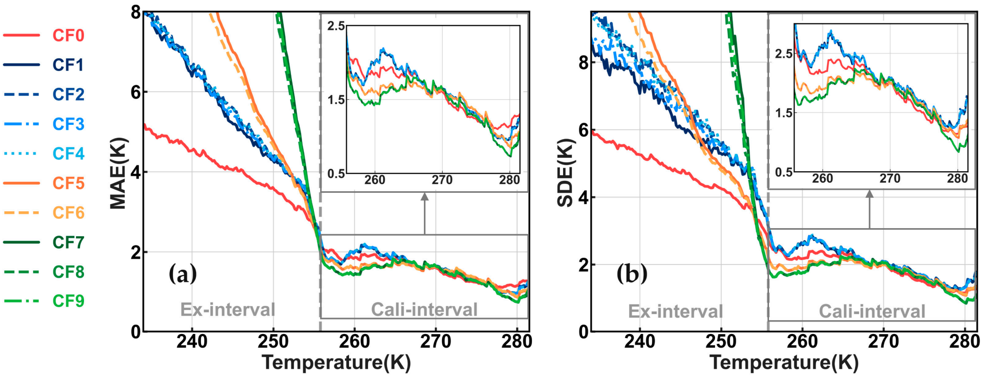
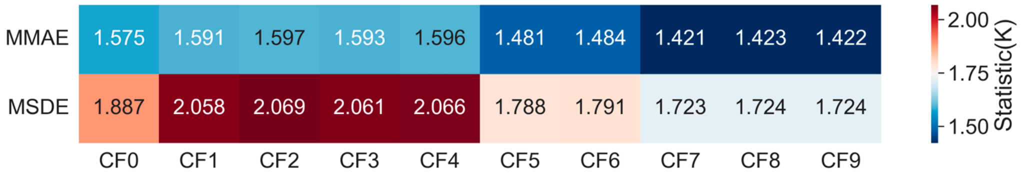


Figure 4. A single-pulse signal was simulated under the specified system parameters and the 1976 standard atmospheric model. MC simulation experiments were conducted using 72,000 random Poisson counts and the VSW-M1 smoothing method. An intercomparison was conducted between the calibration (or temperature) error statistics generated by CF0–9 after 1000 MC simulation experiments. (**a**) The MAE of the calibration results generated using CF0–9; (**b**) the SDE of the calibration results generated using CF0–9. "Cali-interval" represents the calibration interval, and "Ex-interval" represents the extrapolation interval. The small gray box in the upper-right corner of each subgraph shows a vertical magnification of the results within the calibration interval.

As shown in Figure 4, whether in the calibration interval or extrapolation interval, the MAE and SDE curves for the calibration results generated by the same category of CFs are quite similar, while the differences between the different categories of CFs are obvious. In the extrapolation interval, the MAE generated by different CFs, ranked from smallest to largest, is LCF < 3c-BCF < 3c-FCF < 4c-FCF. In the range near the calibration interval (corresponding to a temperature range of 246–256 K), the SDE generated by 3c-BCF is greater than that generated by 3c-FCF. Apart from the above scenario, the ranking of the SDE generated by the different CFs is consistent with that of the MAE. This indicates that, in the extrapolation interval, the LCF has a distinct advantage in both calibration accuracy and stability.

Within the calibration interval, each CF generates MAE and SDE curves that exhibit a local minimum at both ends, and the differences in the middle region (corresponding to the 265–275 K temperature range) are relatively indistinguishable. In the higher-altitude regions within the calibration interval (corresponding to temperatures below 265 K), the order from smallest to largest is 4c-FCF < 3c-FCF < LCF < 3c-BCF. Overall, throughout the entire calibration interval, the MAE and SDE generated by 4c-FCF appear to be the smallest across most range gates.

To provide a clearer comparison, the MAE and SDE at different range gates within the calibration interval were averaged to obtain the mean of the mean absolute error (MMAE) and the mean standard deviation of the error (MSDE). The results are presented in Figure 5. Whether the MMAE or MSDE, the differences among the CFs of the same category are in the order of 10^{-3} K, while the differences between the CFs of different categories are in the order of 10^{-1} K (up to 0.335 K). There is a significant disparity in the MMAE and MSDE between the different categories of CFs, ranked from smallest to largest as follows: 4c-FCF < 3c-FCF < LCF < 3c-BCF. In contrast, the differences among the CFs in the same category are nearly negligible. Combined with the results in Figure 4, it can be concluded that, within the calibration interval, 4c-FCF exhibits the best calibration performance among the four types of CFs.



	CF0	CF1	CF2	CF3	CF4	CF5	CF6	CF7	CF8	CF9
MMAE	1.575	1.591	1.597	1.593	1.596	1.481	1.484	1.421	1.423	1.422
MSDE	1.887	2.058	2.069	2.061	2.066	1.788	1.791	1.723	1.724	1.724

Figure 5. Statistical results: heatmap of the MMAE and MSDE of calibration (or temperature) results generated by CF0–9 within the calibration interval.

3.3. Comparison under Different Conditions

When considering the impact of noise, the SNR significantly influences the calibration results. The integration and smoothing operations in the preprocessing module have a substantial effect on the SNR. Additionally, we are interested in understanding how different atmospheric reference temperature ranges and their corresponding atmospheric pressures affect calibration performance. Therefore, we conducted simulation comparisons of the calibration effects of different CFs under various integration times, smoothing methods, and reference temperature ranges.

3.3.1. Different Integration Times

Different integration times were selected, specifically 90, 75, 60, 45, 30, and 15 min, corresponding to the Poisson random number N of 108,000, 90,000, 72,000, 54,000, 36,000, and 18,000, respectively, in Figure 1. All other settings remained consistent with those described in Section 3.2. Under each integration time setting, 1000 MC experiments were conducted following the process outlined in Figure 1. The MAE and SDE of the calibration (or temperature) results were statistically obtained, as shown in Figures 6 and 7, respectively. The comparison results between CFs of the same category and different categories are similar to those discussed in Section 3.2, and they are not reiterated here.

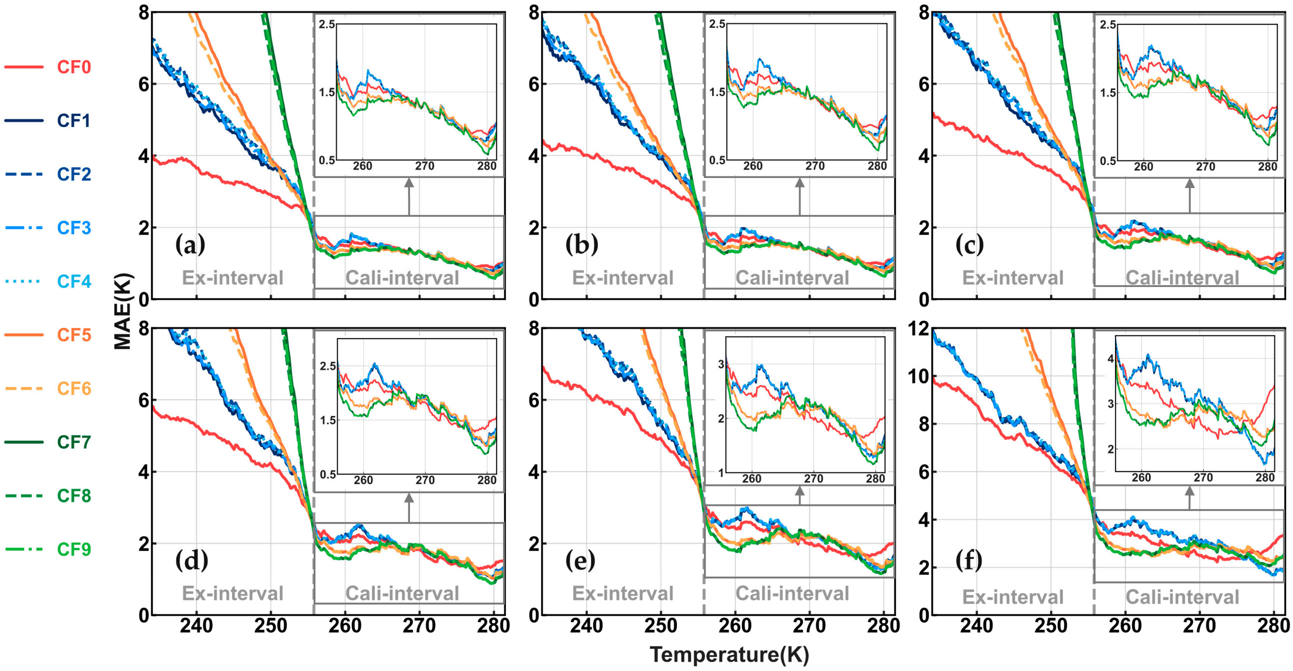


Figure 6. The system parameters, atmospheric model, and smoothing method are consistent with those used in Figure 4. Under different integration times, the MAE statistics for the calibration results were generated through 1000 MC simulation experiments: (**a**) 90 min; (**b**) 75 min; (**c**) 60 min; (**d**) 45 min; (**e**) 30 min; and (**f**) 15 min.

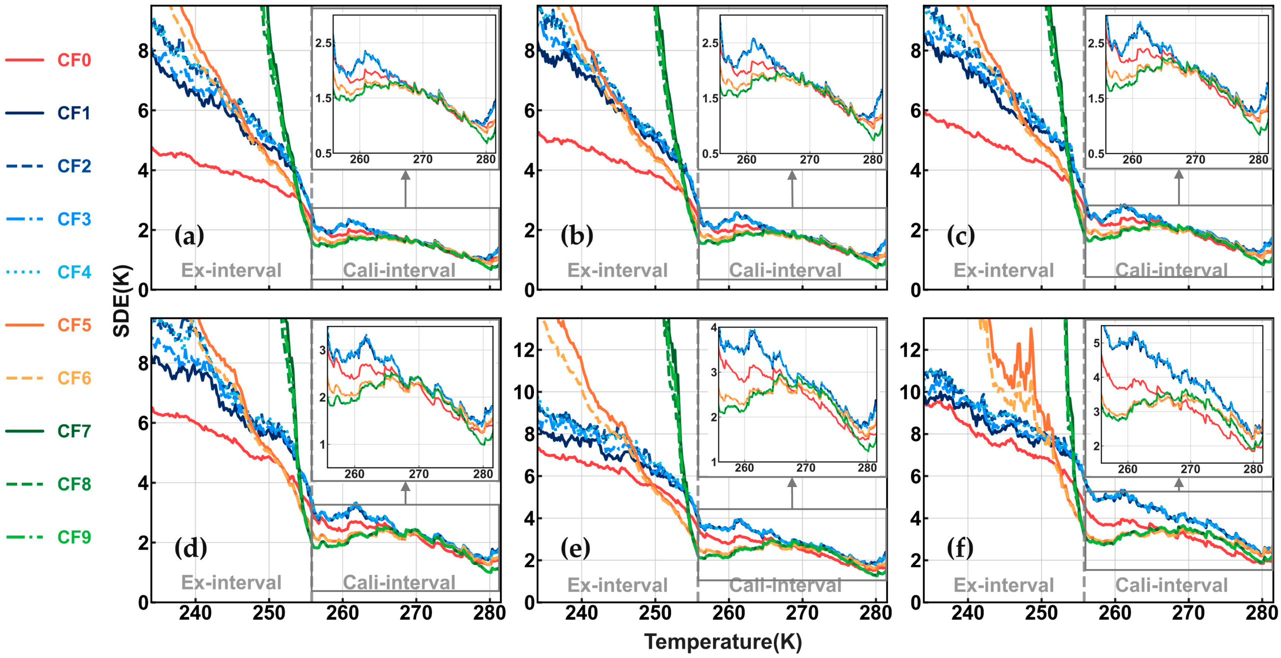


Figure 7. All settings are identical to those in Figure 6, presenting the SDE statistics for the calibration results under different integration times: (**a**) 90 min; (**b**) 75 min; (**c**) 60 min; (**d**) 45 min; (**e**) 30 min; and (**f**) 15 min.

As the integration time decreases (i.e., as the SNR drops), the MAE and SDE values generated by any CF increase, and the differences between the different categories of CFs become more obvious. Regardless of the integration time, the differences in the MAE and SDE between the CFs in the extrapolation interval remain consistent with those discussed in Section 3.2, with the LCF consistently providing the best calibration performance among the four categories of CFs. If the acceptable error is set to 5 K, the LCF can be extrapolated below 245 K when the integration time is not less than 45 min. However, when the integration time is 30 or 15 min, the errors in the extrapolation interval are almost always greater than 5 K. For comparison, the absolute calibration errors calculated using CF0–9 without considering noise are presented in Figure 8. The results show that the absolute errors generated by each CF within the 20 K extrapolation range (i.e., 235–255 K) are small, with LCF < 0.4 K and other nonlinear CFs (CF1–9) < 0.05 K. This indicates that noise significantly amplifies the calibration errors of all CFs in the extrapolation interval. Under the influence of noise, the LCF performs the best in the extrapolation interval, with the extrapolation error decreasing as the SNR increases.

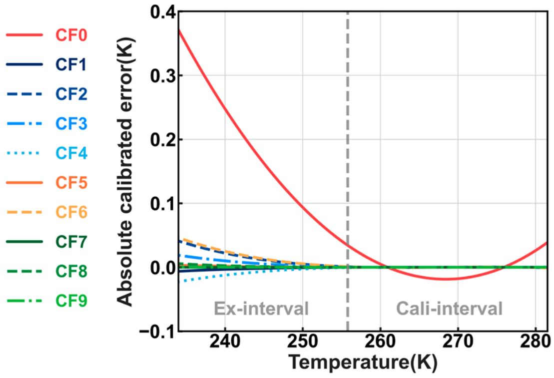
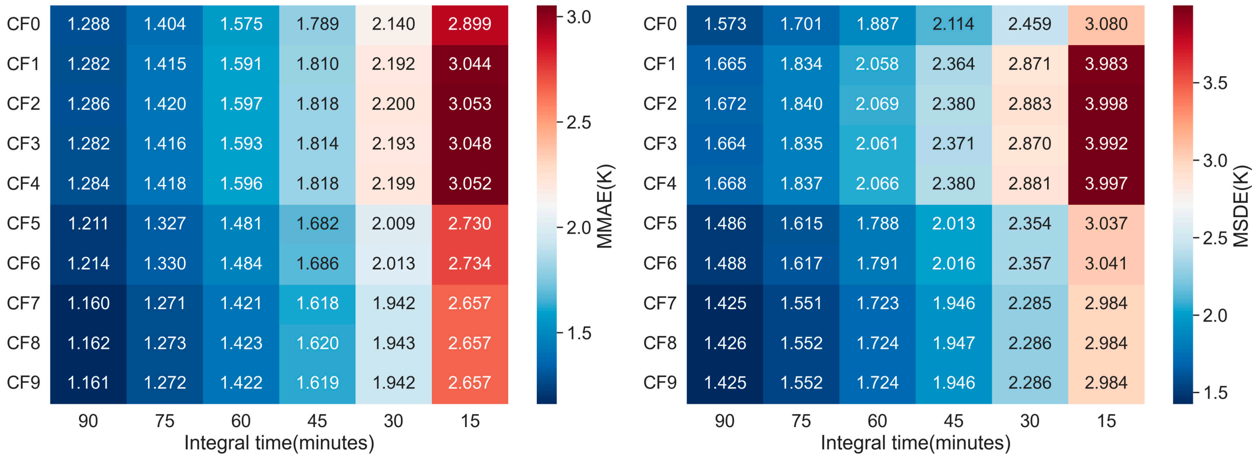


Figure 8. The absolute calibration errors for both the calibration and extrapolation intervals calculated using CF0–9 without considering noise and by using the 1–5 km range as the calibration interval.

Within the calibration interval, the shorter the integration time, the more obvious the difference of the MAE and SDE generated by varying CFs in the middle region (corresponding to the 265–275 K temperature range), and LCF begins to show certain advantages. However, when observing the MAE and SDE across the entire calibration interval, the average level of 4c-FCF should still be the lowest. Furthermore, as shown in Figure 7, regardless of the integration time, the SDE generated by 3c-BCF is almost always the largest within the calibration interval, which indicates poor stability when using this CF category for temperature retrieval.

The MAE and SDE at different range gates within the calibration interval are then averaged to obtain the MMAE and MSDE, as shown in Figure 9. When the integration time is 90 min (during which, the SNR in the calibration interval is relatively high), the LCF has the largest MMAE; however, the difference from 3c-BCF does not exceed 0.006 K. For the other five integration times, the overall magnitude and ranking of the MMAE and MSDE for each CF remain consistent with those described in Section 3.2. By combining the results in Figures 6 and 7, we observe that, regardless of the integration time, 4c-FCF consistently performs the best within the calibration interval. Additionally, as the integration time decreases, the difference in the MMAE and MSDE between 3c-BCF and the other CFs increases, further indicating that noise has a significant impact on 3c-BCF.



	90	75	60	45	30	15
CF0	1.288	1.404	1.575	1.789	2.140	2.899
CF1	1.282	1.415	1.591	1.810	2.192	3.044
CF2	1.286	1.420	1.597	1.818	2.200	3.053
CF3	1.282	1.416	1.593	1.814	2.193	3.048
CF4	1.284	1.418	1.596	1.818	2.199	3.052
CF5	1.211	1.327	1.481	1.682	2.009	2.730
CF6	1.214	1.330	1.484	1.686	2.013	2.734
CF7	1.160	1.271	1.421	1.618	1.942	2.657
CF8	1.162	1.273	1.423	1.620	1.943	2.657
CF9	1.161	1.272	1.422	1.619	1.942	2.657

	90	75	60	45	30	15
CF0	1.573	1.701	1.887	2.114	2.459	3.080
CF1	1.665	1.834	2.058	2.364	2.871	3.983
CF2	1.672	1.840	2.069	2.380	2.883	3.998
CF3	1.664	1.835	2.061	2.371	2.870	3.992
CF4	1.668	1.837	2.066	2.380	2.881	3.997
CF5	1.486	1.615	1.788	2.013	2.354	3.037
CF6	1.488	1.617	1.791	2.016	2.357	3.041
CF7	1.425	1.551	1.723	1.946	2.285	2.984
CF8	1.426	1.552	1.724	1.947	2.286	2.984
CF9	1.425	1.552	1.724	1.946	2.286	2.984

Figure 9. Heatmap of the MMAE and MSDE for calibration results generated by CF0–9 within the calibration interval and under different integration times.

3.3.2. Different Smoothing Methods

The six smoothing methods selected were as follows: no smoothing; three fixed sliding window methods with window sizes of 11, 15, and 21 (11 FSW, 15 FSW, and 21 FSW); VSW-M1 mentioned in Section 3.2; and variable sliding window method 2 (VSW-M2) referenced in [21], which uses an initial window size of 3 and adds 1 point to each side of the center point for every 10 data points. All other settings remained consistent with those described in Section 3.2. Under the various smoothing methods, 1000 MC simulations were conducted following the process outlined in Figure 1. The MAE and SDE of the calibration results were statistically obtained, as shown in Figures 10 and 11, respectively.

The results in the figures indicate that, as the level of smoothing increases (as shown in Figures 10 and 11, where the smoothing level gradually increases from [a] to [f], except for in subfigure [e]), the MAE and SDE values generated by each CF decrease. Compared with the unsmoothed results, smoothing not only significantly reduces the MAE and SDE, but also alters the shape of the result curves. After smoothing, the MAE and SDE curves for each CF within the calibration interval exhibit a characteristic with a local minimum at both ends and a nearly uniform shape in the middle region.

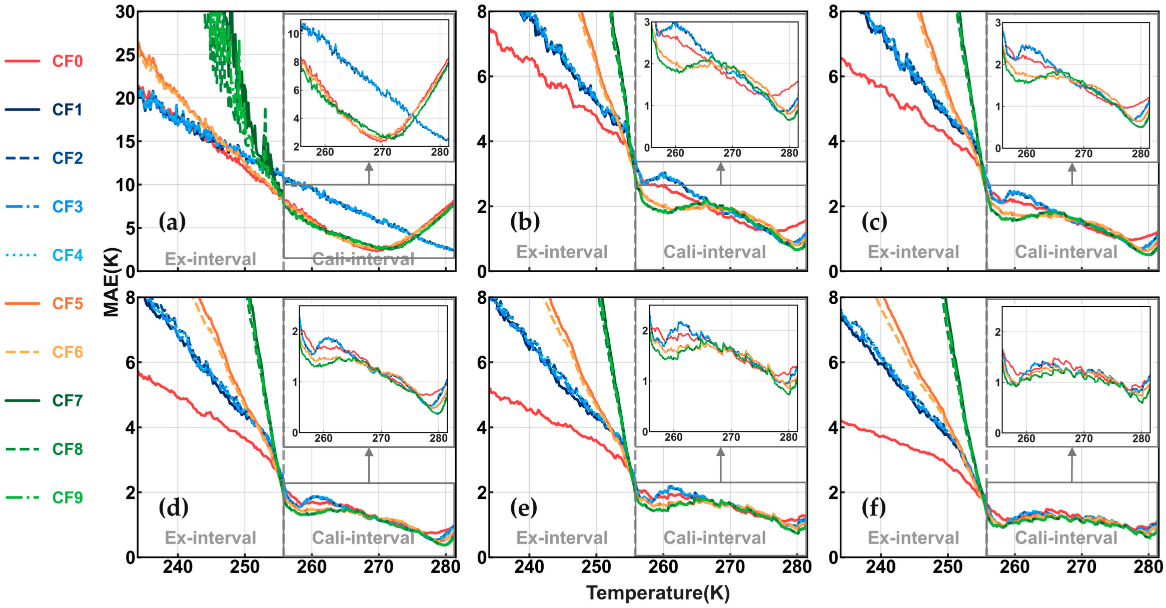


Figure 10. The system parameters, atmospheric model, and integration time are consistent with those used in Figure 4. Under different smoothing methods, the MAE statistics for the calibration results were generated through 1000 MC simulation experiments: (**a**) no smoothing; (**b**) 11 FSW; (**c**) 15 FSW; (**d**) 21 FSW; (**e**) VSW-M1; and (**f**) VSW-M2.

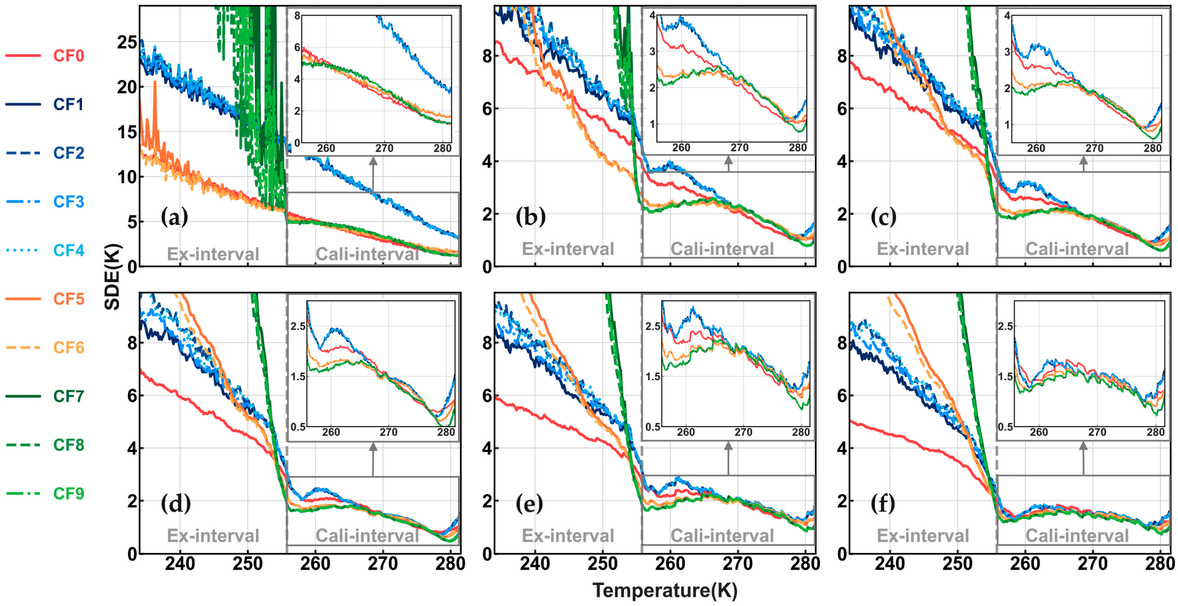
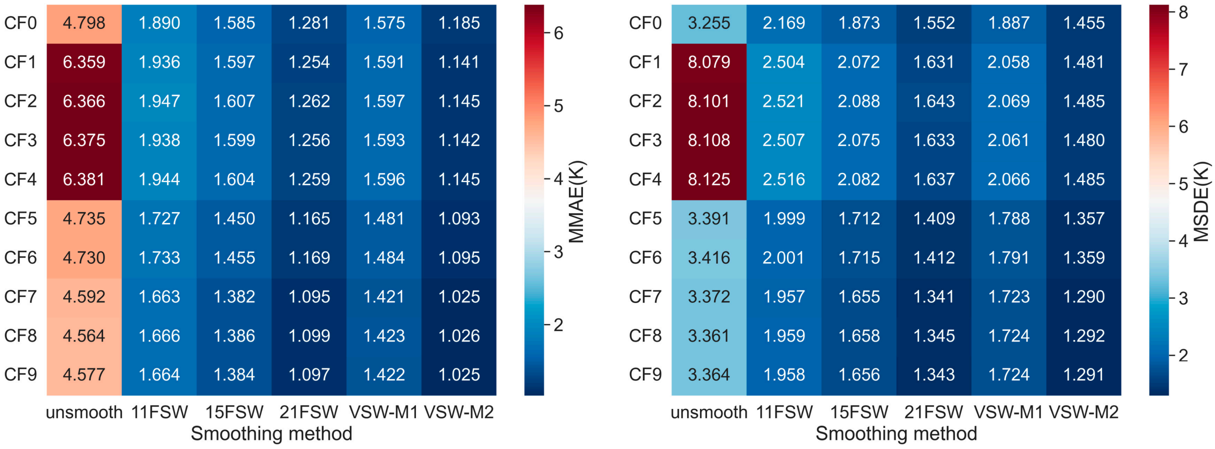


Figure 11. All settings are identical to those in Figure 10, presenting the SDE statistics for the calibration results under different smoothing methods: (**a**) no smoothing; (**b**) 11 FSW; (**c**) 15 FSW; (**d**) 21 FSW; (**e**) VSW-M1; and (**f**) VSW-M2.

In the extrapolation interval, the comparison of the CFs remains consistent with the results discussed in Section 3.2, with the LCF providing the best calibration performance among the four categories of CFs. Within the calibration interval, when comparing the MAE and SDE produced by each CF, it appears that the average performance of 4c-FCF is still the best. When using VSW-M2, areas with a lower SNR experience greater smoothing, resulting in the calibration results of each CF across the entire calibration interval exhibiting high accuracy and low dispersion. However, this smoothing method significantly reduces the vertical resolution of the inversion results, and, hence, VSW-M2 was not selected for processing the actual data in this study.

Similarly, the MMAE and MSDE within the calibration interval were calculated, and the results are presented in Figure 12. The overall magnitude and ranking of these values remained consistent with those in Section 3.2. When the VSW-M2 smoothing method was applied (where the SNR within the calibration interval is relatively high), the MMAE of the LCF was the largest, but the difference from 3c-BCF did not exceed 0.044 K. Under the other five smoothing methods, the overall magnitude and ranking of the MMAE and MSDE for each CF were consistent with the findings in Section 3.2. By combining the results in Figures 10 and 11, we found that, regardless of the smoothing method used, 4c-FCF consistently performed the best within the calibration interval.



	unsmooth	11FSW	15FSW	21FSW	VSW-M1	VSW-M2
CF0	4.798	1.890	1.585	1.281	1.575	1.185
CF1	6.359	1.936	1.597	1.254	1.591	1.141
CF2	6.366	1.947	1.607	1.262	1.597	1.145
CF3	6.375	1.938	1.599	1.256	1.593	1.142
CF4	6.381	1.944	1.604	1.259	1.596	1.145
CF5	4.735	1.727	1.450	1.165	1.481	1.093
CF6	4.730	1.733	1.455	1.169	1.484	1.095
CF7	4.592	1.663	1.382	1.095	1.421	1.025
CF8	4.564	1.666	1.386	1.099	1.423	1.026
CF9	4.577	1.664	1.384	1.097	1.422	1.025

	unsmooth	11FSW	15FSW	21FSW	VSW-M1	VSW-M2
CF0	3.255	2.169	1.873	1.552	1.887	1.455
CF1	8.079	2.504	2.072	1.631	2.058	1.481
CF2	8.101	2.521	2.088	1.643	2.069	1.485
CF3	8.108	2.507	2.075	1.633	2.061	1.480
CF4	8.125	2.516	2.082	1.637	2.066	1.485
CF5	3.391	1.999	1.712	1.409	1.788	1.357
CF6	3.416	2.001	1.715	1.412	1.791	1.359
CF7	3.372	1.957	1.655	1.341	1.723	1.290
CF8	3.361	1.959	1.658	1.345	1.724	1.292
CF9	3.364	1.958	1.656	1.343	1.724	1.291

Figure 12. Heatmap of the MMAE and MSDE for calibration results generated by CF0–9 within the calibration interval and under different smoothing methods.

Without smoothing, the SNR within the calibration interval remained above 10 (an SNR greater than 10 is generally considered reliable data [44]), but the MMAE and MSDE obtained were excessively large, especially for 3c-BCF. Moreover, as the degree of smoothing decreased (indicating a lower SNR), the difference in the MMAE between 3c-BCF and the other categories of CFs became more obvious. This indicates that, for the system model established in this study, smoothing can effectively improve calibration accuracy. Additionally, among the different CFs, the calibration results of 3c-BCF are the most susceptible to noise.

3.3.3. Different Reference Temperature Ranges

The calibration performance of the CFs was compared under different atmospheric reference temperature ranges and corresponding atmospheric pressures. Radiosonde temperature and pressure data were used instead of the standard atmospheric model for the simulations. Radiosonde data from 15 July 2023, 15 October 2023, and 15 January 2024 were selected to simulate the atmospheric temperature and pressure conditions in summer, spring/autumn, and winter, respectively. All other settings remained consistent with those described in Section 3.2. A total of 1000 MC simulations were conducted following the process outlined in Figure 1, and the MAE and SDE of the calibration results were statistically obtained, as shown in Figures 13 and 14, respectively. The comparison results between the CFs of the same category and different categories are similar to those discussed in Section 3.2, and they are not reiterated here. Notably, for any given CF, the MAE and SDE within the calibration range appear to follow the trend of 2024.1.15 < 2023.10.15 < 2023.7.15; that is, the lower the atmospheric reference temperature range, the smaller the calibration error statistics.

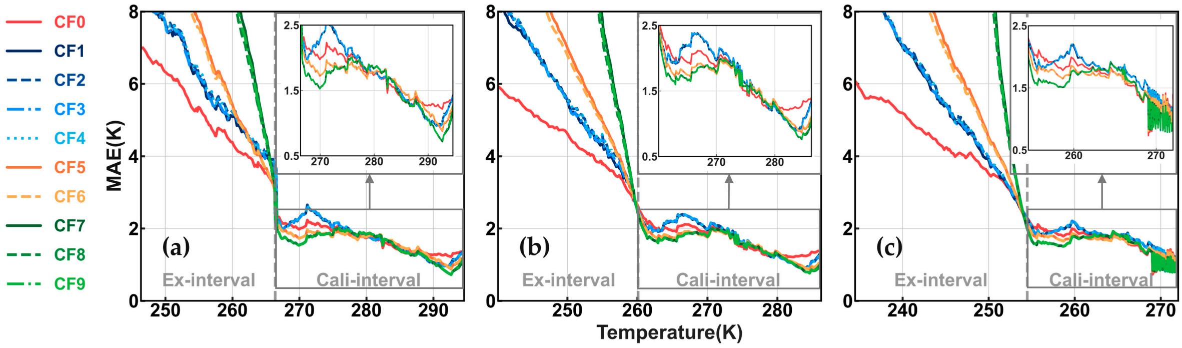


Figure 13. The system parameters, integration time, and smoothing method are consistent with those used in Figure 4. Under different atmospheric reference temperature ranges, the MAE statistics for the calibration results were generated through 1000 MC simulation experiments. (**a**) 15 July 2023; (**b**) 15 October 2023; and (**c**) 15 January 2024.

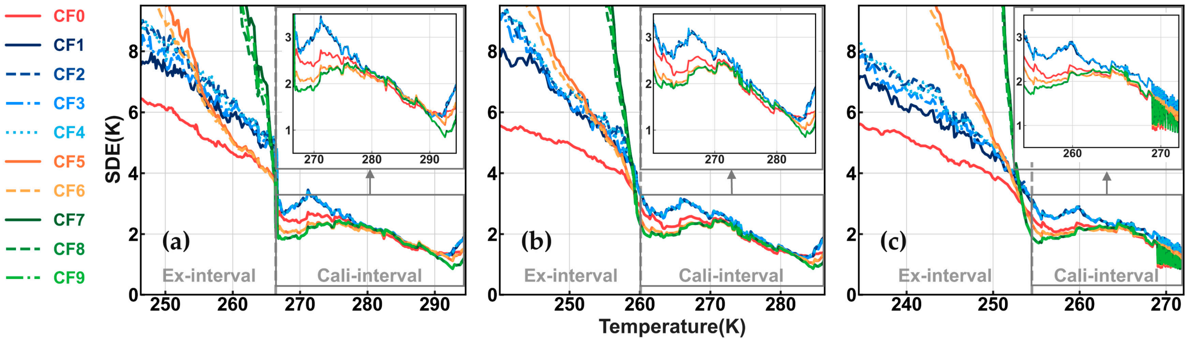


Figure 14. All settings are identical to those in Figure 13, presenting the SDE statistics for the calibration results under different atmospheric reference temperature ranges. (**a**) 15 July 2023; (**b**) 15 October 2023; and (**c**) 15 January 2024.

Similarly, the MMAE and MSDE within the calibration interval are calculated, and the results are presented in Figure 15. The overall magnitude and ranking of these values remain consistent with those in Section 3.2. By combining the results in Figures 13 and 14, the conclusion that 4c-FCF performs the best within the calibration interval is reaffirmed. Additionally, we focus on the calibration performance of the same CF under different reference temperature ranges. The changes in the MMAE and MSDE values clearly indicate that, for any CF, better calibration results are achieved when the overall temperature within the corresponding reference temperature range is lower.

To analyze the reasons, we simulated the scattering cross-sections of the PRR spectral lines at different temperatures. Then, by using the STF curves shown in Figure 2, we calculated the spectral line intensities received by the high- and low-order channels. This enabled us to determine the PRR spectral line intensity ratio (LIR) between the high- and low-order channels at various temperatures. If the variation in the LIR with temperature is greater within a certain range, it indicates a higher temperature sensitivity of the system in that range, leading to more accurate temperature retrieval. To compare the LIR variations across different temperatures, the first-order difference sequence of the LIR was calculated, and the results are shown in Figure 16. It can be observed that, within the temperature range of 200–350 K (including the temperature range studied), the lower the temperature, the greater the first-order difference of the LIR, indicating a higher temperature sensitivity of the PRR spectrum and, thus, resulting in fewer detection errors.

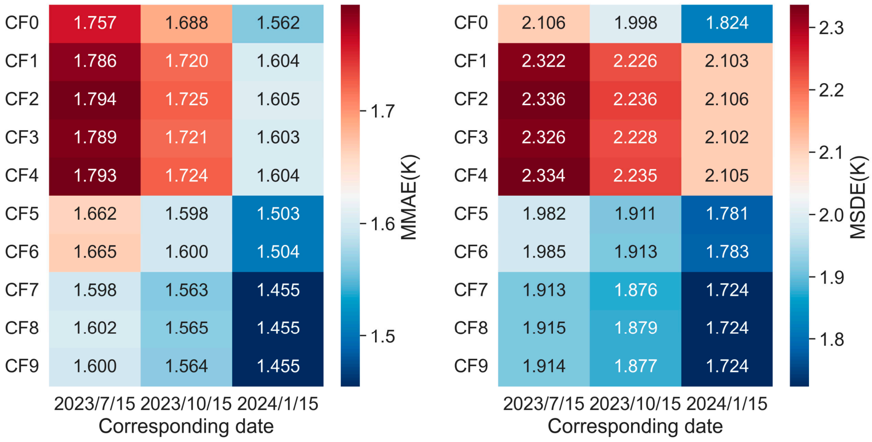


Figure 15. Heatmap of the MMAE and MSDE for calibration results generated by CF0–9 within the calibration interval and under different atmospheric reference temperature ranges.

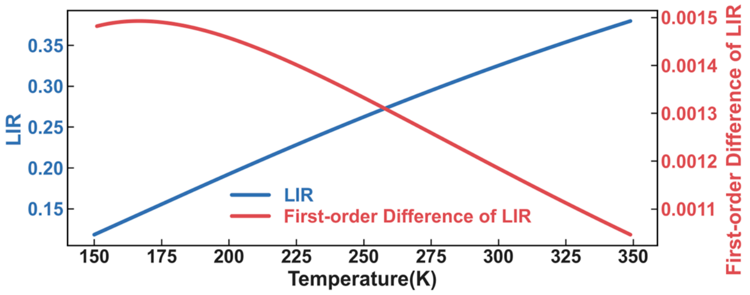


Figure 16. The PRR spectral line intensity ratio (LIR) received by high- and low-order channels at different temperatures, as well as the first-order difference sequence of the LIR.

4. Verification and Discussion

This paper presents a statistical comparison of the calibration effects of different CFs in both calibration and extrapolation intervals through MC simulation experiments conducted with various settings. Section 3 presents the statistical results of these simulations. This section validates the simulation results using actual observational data and attempts to discuss the factors that influence the calibration results.

4.1. Actual Data Verification

The raw data used were obtained by the PRR lidar system at the Beijing Institute of Technology in Beijing, China. This system is used for nighttime tropospheric temperature detection, with a range resolution of 30 m, and its main system parameters are consistent with those in the simulations. Calibration was performed using radiosonde data from station 54,511 in the southern suburbs of Beijing (39.93°N, 116.28°E), which launches radiosondes twice daily (at 00:00 and 12:00 UTC). As the PRR lidar system operates only at night, radiosonde data at 12:00 UTC (i.e., 20:00 Beijing time) were used for validation. Echo signals of the closest 60 min interval were integrated (i.e., accumulating 72,000 pulses) and smoothed using the VSW-M1 method, as described in this study. For systems with a low laser power, a smaller fixed-size sliding window can be applied to micro-average the signal intensity ratio Q in order to reduce statistical fluctuations and improve retrieval accuracy [21]. However, as this study focused on comparing the calibration effects of different CFs, a uniform processing method was sufficient. Hence, Q was not micro-averaged here. The height range from above 1 km to an SNR of 10 before smoothing was selected as the calibration interval. After calibration using CF0–9, the corresponding TRF was used to retrieve temperature profiles. The results are shown in Figure 17.

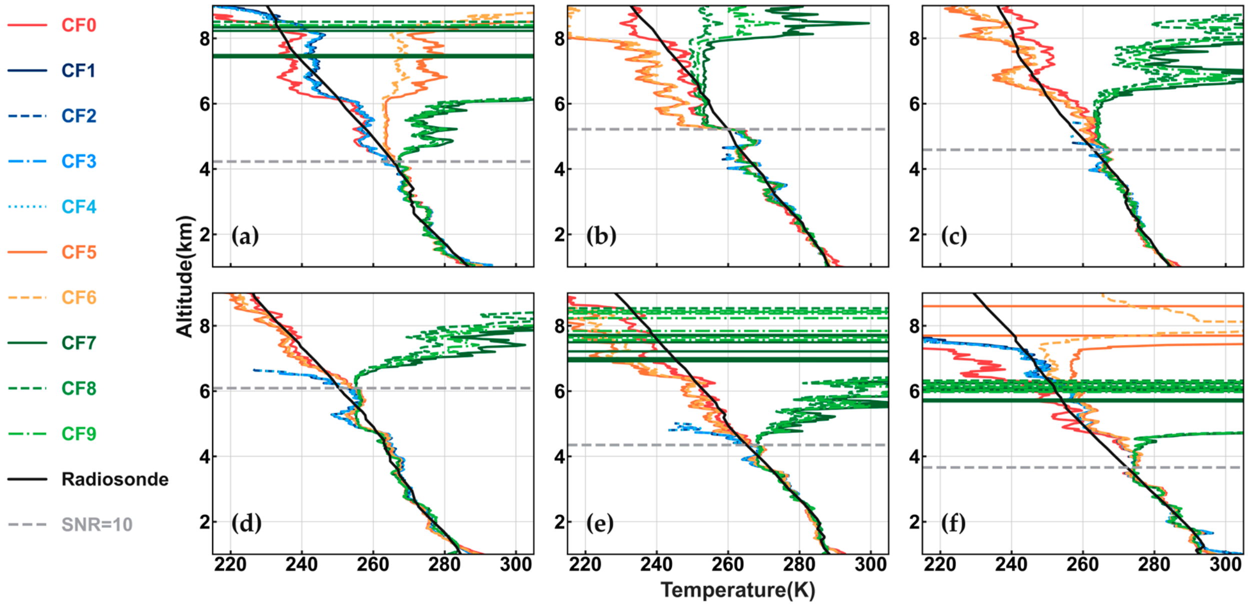
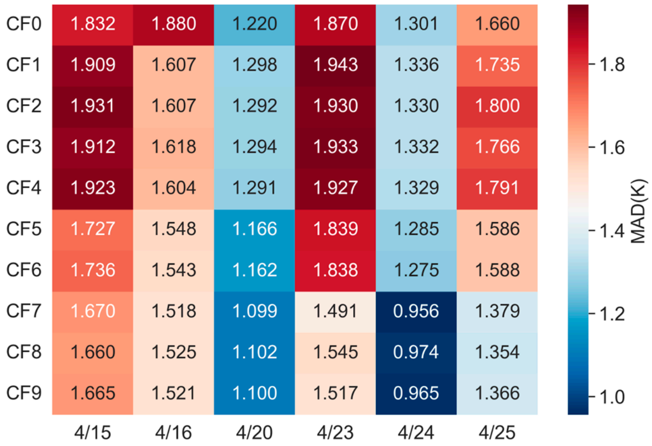


Figure 17. Actual data inversion results: a comparison between temperature profiles retrieved using various TRFs (calibrated by the corresponding CFs) and radiosonde temperature profiles. (**a**) 15 April 2024; (**b**) 16 April 2024; (**c**) 20 April 2024; (**d**) 23 April 2024; (**e**) 24 April 2024; and (**f**) 25 April 2024.

The temperature profiles for six days between 15 and 25 April 2024, retrieved using CF0–9, are shown alongside the corresponding radiosonde data in Figure 17. As the displayed temperature profiles are derived from the calibration data, the deviation from the radiosonde temperatures represents the actual calibration error. Additionally, the smaller the error, the better the calibration performance. The results indicate that the calibration effects are similar within the same category of CFs. In the extrapolation interval, however, the calibration (or temperature) results generated by different categories of CFs are obviously different. When using 3c-BCF for temperature retrieval, solutions that lack physical significance are often encountered. This issue is observed in four out of the six days shown and is a common problem encountered in actual data inversion. Excluding this case, the calibration errors in the extrapolation interval, ranked from smallest to largest, are LCF < 3c-BCF < 3c-FCF < 4c-FCF. This is consistent with the results obtained from the simulation experiments.

Figure 17 shows that, within the calibration interval, there is no significant difference in the calibration (or temperature) results generated by the different CFs. To compare the differences, the mean absolute deviation (MAD) between the inversion temperatures and radiosonde temperatures at various range gates within the calibration interval is calculated for each CF, and the results are shown in Figure 18. It can be seen that the average level of actual calibration errors is similar to the statistical results obtained from the MC simulation experiments in this study, and it is significantly greater than the absolute calibration errors calculated without considering noise (as shown in Figure 3). The MAD generated by same category of CFs is relatively close, while there are significant differences in the MAD between different categories of CFs. Except for April 16, where the LCF exceeds 3c-BCF, the ranking of the MAD from smallest to largest is as follows: 4c-FCF < 3c-FCF < LCF < 3c-BCF. These results further verify the reliability of the simulation results.

Each calibration process of the actual data can correspond to a single MC simulation experiment, and the results have some randomness. Therefore, it is reasonable for some actual data calibration results to not perfectly align with the statistical results of the simulation experiment. Moreover, the actual detection process is affected by random noise in addition to shot noise. Therefore, the calibration results of six different sets of actual data can only verify some simulation results. However, they do offer support for the simulation results that we aim to verify. A significantly larger amount of actual data may be required to verify more detailed patterns or comparative results.



	4/15	4/16	4/20	4/23	4/24	4/25
CF0	1.832	1.880	1.220	1.870	1.301	1.660
CF1	1.909	1.607	1.298	1.943	1.336	1.735
CF2	1.931	1.607	1.292	1.930	1.330	1.800
CF3	1.912	1.618	1.294	1.933	1.332	1.766
CF4	1.923	1.604	1.291	1.927	1.329	1.791
CF5	1.727	1.548	1.166	1.839	1.285	1.586
CF6	1.736	1.543	1.162	1.838	1.275	1.588
CF7	1.670	1.518	1.099	1.491	0.956	1.379
CF8	1.660	1.525	1.102	1.545	0.974	1.354
CF9	1.665	1.521	1.100	1.517	0.965	1.366

Figure 18. Statistical results: heatmap of the MAD between the retrieved temperatures generated using CF0–9 and radiosonde temperatures within the calibration interval.

4.2. Discussion on the Causes of Calibration Errors

By analyzing the results, we posit that the calibration error generated by a CF is the coupling result of several factors. These factors include shot noise, least square fitting, the specific CF itself, and the temperature sensitivity of PRR spectral lines. The following specifically analyzes the impact of these factors and discusses their coupling effects through some examples.

4.2.1. The Impact of Shot Noise

In all MC simulation experiments in Section 3.3, the Q-T relationship curves for CFs of the same category are very similar. Thus, CF1, CF5, and CF7 are used to represent 3c-BCF, 3c-FCF, and 4c-FCF, respectively, for an analysis. Without considering noise, the fitting sample points (i.e., the corresponding distribution points between Q and T) and the resulting Q-T relationship curves are shown in Figure 19a–d. When considering noise, an arbitrary MC simulation experiment, without smoothing, is selected from Section 3.3.2, and the results are shown in Figure 19e–h. The results indicate that, in the absence of noise, the fitting sample points exhibit a near-linear distribution. However, under the influence of noise, the sample points exhibit random dispersion, weakening the constraints on the fitting results and causing the fitting curve to deviate from the ideal results obtained without noise.

To assess whether the impact of shot noise on the fitting sample points and fitting results is widespread, simulations were conducted with various levels of noise influence. Different noise levels were simulated by adjusting the signal strength and smoothing methods. The signal strength was altered through system parameter adjustments while keeping the other settings consistent with those in the MC experiments shown in Figure 19, and the results of simulations with signal magnifications of 20× and 40× are presented in Figure 20. The results show that, as the signal strength increases, the divergence of the fitting sample points gradually decreases, and the fitting curves gradually approach the ideal result. However, even with a 40× signal amplification, where the SNR of the fitting sample points is sufficiently high (well above actual data levels), the sample points still exhibit random divergence, and the fitting curves remain deviated from the ideal result to some extent. Figure 21 shows the results obtained using different smoothing methods, specifically from a random simulation in Section 3.3.2 utilizing VSW-M1 and 21 FSW. Compared with the unsmoothed results (Figure 19e–h), both smoothing operations reduce the degree of divergence among the sample points to some extent. However, after smoothing, the distribution of the sample points resembles an irregular sawtooth pattern, which significantly deviates from the near-linear distribution observed under ideal conditions. It can be seen that, regardless of whether the signal is enhanced or smoothing is applied, the impact of shot noise persists and cannot be overlooked. Under the influence of shot noise, the fitting sample points no longer exhibit near-linear distribution characteristics. At this point, it may be necessary to use a least squares analysis to re-evaluate the calibration performance of the different CFs.

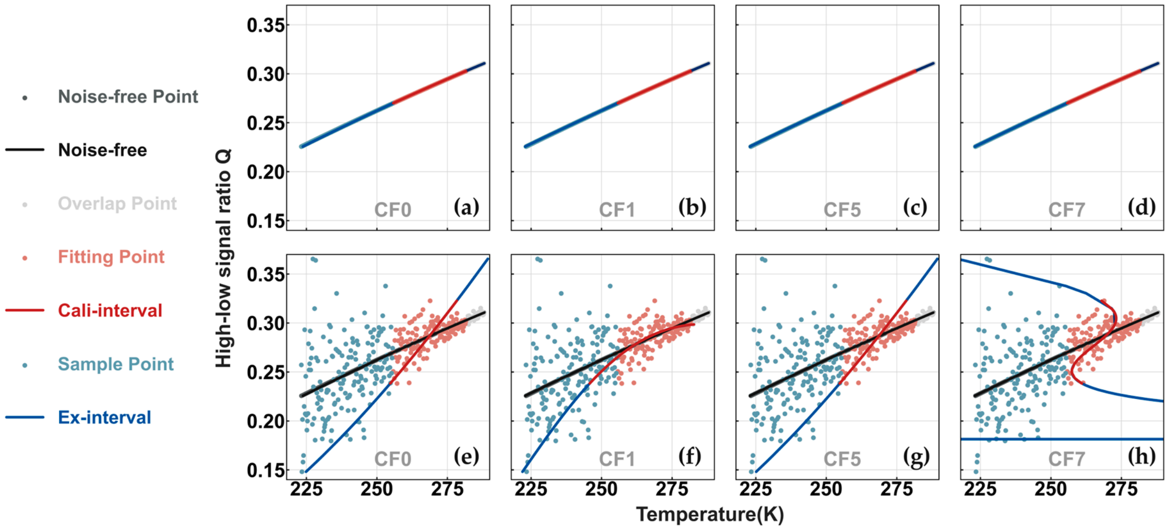


Figure 19. A comparison of the fitting sample points and the fitting results of different CFs, both with and without considering noise. Panels (**a**–**d**) show the fitting results for CF0, CF1, CF5, and CF7 without considering noise effects. Panels (**e**–**h**) present the fitting results for CF0, CF1, CF5, and CF7 under the influence of noise, using an arbitrary MC simulation result from Section 3.3.2 that did not apply smoothing. The dark gray, light gray, red, and blue circles represent the samples without noise, affected by overlap factors, used for calibration fitting, and outside the calibration interval, respectively. The black, red, and blue curves denote the fitting results without noise, within the calibration interval, and within the extrapolation interval, respectively.

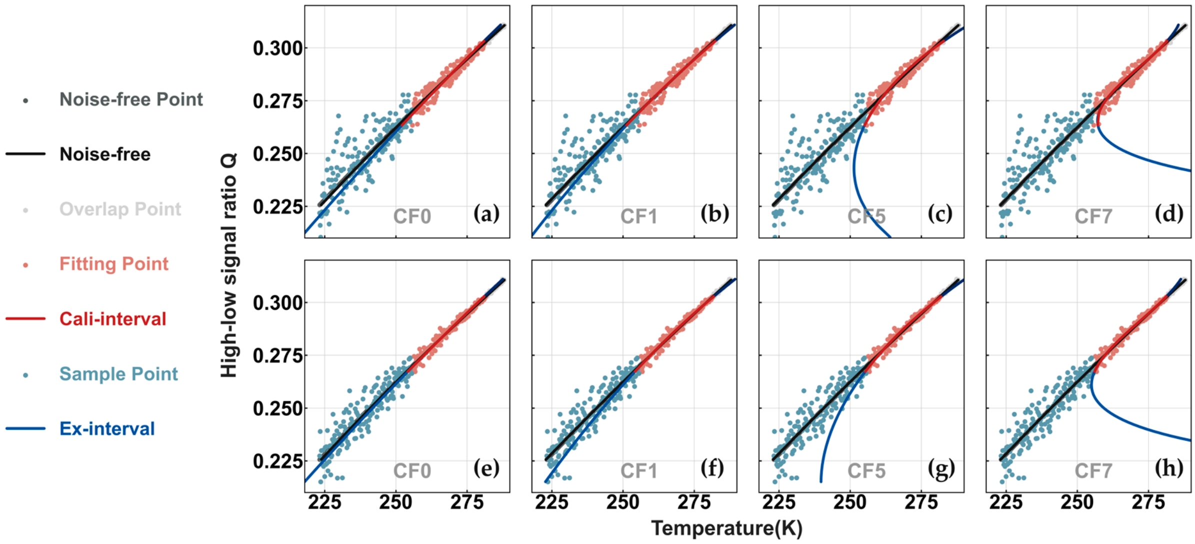


Figure 20. Same as Figure 19, but for different signal strength. Panels (**a**–**d**) show the fitting results for a signal amplified by a factor of 20. Panels (**e**–**h**) present the fitting results for a signal amplified by a factor of 40. In both cases, the results are presented for CF0, CF1, CF5, and CF7.

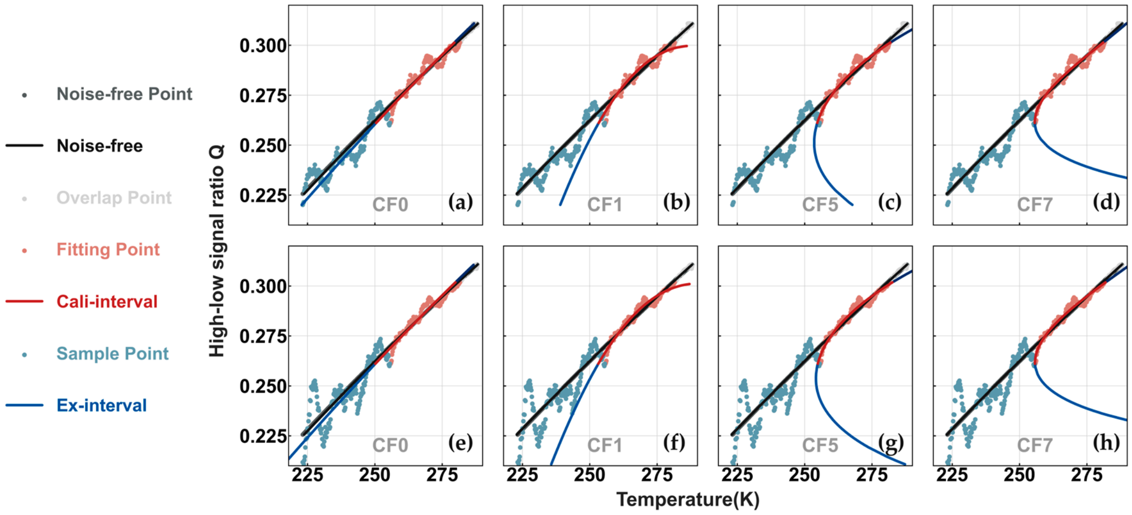


Figure 21. Same as Figure 19, but for different smoothing methods. Panels (**a**–**d**) show one instance of an MC simulation using VSW-M1 from Section 3.3.2. Panels (**e**–**h**) show one instance of using 21 FSW from the same section. In both cases, the results are presented for CF0, CF1, CF5, and CF7.

4.2.2. The Impact of Least Squares Fitting and Shot Noise

The least squares method [42] seeks the optimal function fit for data by minimizing the sum of squared errors. This best function fit is for the data being fitted. The data in the calibration interval are the data to be fitted. They are near linear without considering noise, whereas shot noise introduces divergence and randomness into the data distribution. In this context, fitting $\ln Q$ to $1/T$ using the FCF (including the LCF) can directly seek the minimum residual of the target value T, and this is expected to generate smaller errors than fitting $1/T$ to $\ln Q$ and then calculating T using the BCF. Additionally, due to the divergence and randomness of the fitting objects, a higher-order FCF should achieve smaller average errors. Both simulated and actual data show that, within the calibration interval, the calibration errors of various CFs are ranked as 4c-FCF < 3c-FCF < LCF < 3c-BCF. This is consistent with the above analysis.

Additionally, due to the characteristics of the least squares method, linear fitting results in a straight line (no inflection points), while a three-coefficient fit has one inflection point, and a four-coefficient fit has two. The greater the number of inflection points, the higher the sensitivity of the fitted curve to noise. Figures 19 and 20 show that, as the SNR decreases (indicating a greater shot noise impact), the curvature of the fitted curves of all the CFs, except for the LCF, becomes greater, with more noticeable changes as the number of inflection points increases. In our simulations, the curvature of the 3c-BCF fitting curve is consistently smaller, making it the closest to linear among the three nonlinear CFs. Furthermore, generally, the curvature of the fitting curves for various CFs follows the order of LCF < 3c-BCF < 3c-FCF < 4c-FCF. In an ideal (no-noise) situation, the fitting sample points are nearly linear. Although the distribution is not normal due to the Poisson process, it still centers on this near-linear relationship. Therefore, the greater the curvature of the fitted curve, the more it deviates from the sample points in the extrapolation interval, which helps to explain the different calibration results of the various CFs in the extrapolation interval.

The following discussion uses the statistical results obtained from the MC simulations. Figure 22 shows the mean and standard deviation of the calibration errors generated using CF0–9 from the MC simulations in Section 3.2. The mean error, obtained from multiple MC experiments with the Poisson process added, differs from the error distribution in the ideal (no-noise) case. Figure 3 shows the absolute calibration errors of the CFs in a noise-free scenario: Linear CF errors are under 0.4 K, nonlinear CF errors are under 0.03 K, and all CF error curves are close to 0. However, the statistical mean error curves lose these characteristics and have one fewer inflection point than the absolute calibration error curves. This is the manifestation of the impact of shot noise, which changes the fitting sample points and makes the fitting target relation deviate from the ideal scenario. Among the four CF categories, the mean error curve of 3c-BCF is the closest to zero, possibly because it is the closest to the GCF in form. However, it consistently generates the largest error standard deviation, likely because it does not directly seek the minimum residual of T, leading to greater variability and uncertainty in the calibration results for this category of CFs.

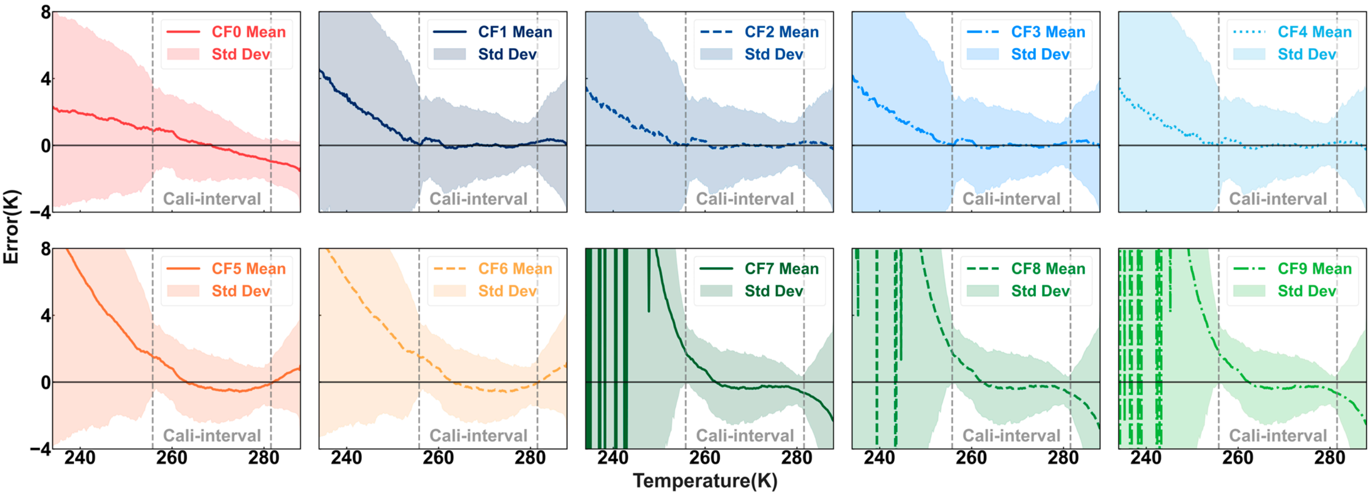


Figure 22. Statistical results obtained from Section 3.2: the mean and standard deviation of calibration errors generated using CF0–9. "Cali-interval" refers to the calibration interval.

4.2.3. Other Factors and Coupling Effects

For the system model in this paper, considering the impact of noise, the calibration results of the CFs in the same category are considered to be similar. However, as shown in Figures 5, 9, 12 and 15, there are still variations in the MMAE and MSDE among the CFs in the same category. Specifically, for 3c-BCF, the sequence is generally CF1 < CF3 < CF4 < CF2; for 3c-FCF, the sequence is generally CF5 < CF6; and, for 4c-FCF, the sequence is generally CF7 < CF9 < CF8. Additionally, the internal differences among the same category of CFs in 4c-FCF are the smallest. Without noise, the ranking of and differences in the absolute calibration errors (in Figure 3) among the CFs in the same category are consistent with these findings. In the preceding analysis, it was suggested that the shot noise and least squares method significantly impact the actual calibration results, potentially obscuring the effects of the specific CFs themselves. Nevertheless, the differences observed among the CFs in the same category reflect the impact of the specific CF itself, which, along with other factors, contributes to the actual calibration errors.

In addition to the effects of the shot noise, least squares fitting, and specific CF itself on the actual calibration errors discussed above, Section 3.3.3 discusses the impact of PRR spectral temperature sensitivity. This suggests that the calibration error in the PPR temperature measurement is the result of multiple interacting factors. As discussed above, shot noise and least squares fitting are likely major influential factors. However, the interaction and mutual constraints between these factors make it difficult to isolate and analyze their individual effects on calibration results. Furthermore, while some examples can demonstrate the effects of the specific CF itself and PRR spectral temperature sensitivity, it remains difficult to analyze the extent of their impact separately. Shot noise and least squares fitting appear to establish a significant baseline for calibration errors, upon which other factors exert their influence. However, whether these factors are simply additive or interact in more complex ways remains uncertain. It can be observed that the coupling effects of various factors on calibration results are highly complex, which makes the decoupling process difficult.

5. Conclusions

The selection of CF significantly impacts the accuracy and stability of PRR lidar temperature measurements. In this study, for the first time, a comparative analysis of the calibration effects of 10 common CFs was conducted while considering the influence of noise. The CFs were classified based on their functional forms and the number of calibration coefficients. To assess the impact of noise (including background light noise, dark noise, and shot noise), Poisson processes were incorporated into the simulations, introducing randomness into the results of individual MC simulations. Therefore, multiple MC simulations were conducted to statistically compare the calibration effects of the different CFs. These simulation experiments were performed under various preset conditions to obtain comparative results for different scenarios. Additionally, the simulated results were verified using actual lidar detection data.

Through simulations and validation with actual data, we found the following: (1) The calibration errors obtained statistically, considering the influence of noise, were closely aligned with the actual calibration errors, both of which were significantly larger than the absolute calibration errors calculated without considering noise. (2) Under the CF classification in this study, the calibration effects were similar within the same CF category, but differed significantly among categories, allowing for a comparison and the selection of CFs by category. (3) In the extrapolation interval, the calibration effects of the CFs were ranked as follows: LCF > 3c-BCF > 3c-FCF > 4c-FCF. (4) Within the calibration interval, the calibration effects of the CFs were ranked as follows: 4c-FCF > 3c-FCF > LCF > 3c-BCF. Consequently, from both simulation and practical perspectives, the LCF exhibited the best extrapolation performance (i.e., it performed the best in the extrapolation interval), while 4c-FCF provided the best performance within the calibration interval.

Based on the results of this study, the following suggestions are proposed: If the calibration results need to involve an extrapolation interval during inversion, the LCF is recommended; otherwise, 4c-FCF is preferred. It is important to note that what is typically considered as within the calibration interval may, in fact, involve some degree of extrapolation due to shifts in the atmospheric temperature range to be detected and changes in the echo signal conditions. However, if the extrapolation interval is relatively small compared with the calibration interval, this situation can be approximately treated as one that does not involve extrapolation. In the concrete implementation process, it is still necessary for the calibration implementer to make a manual judgment. As previously discussed, shot noise and least squares fitting are likely significant factors affecting calibration results, and the comparison results between different CFs show low sensitivity to system parameters. Therefore, the conclusions of this study should be applicable beyond our system. In the process of using PRR lidar for temperature detection, the conclusions of this study can be referenced to select CFs, so as to achieve more accurate and stable atmospheric vertical temperature profile detections.

The actual data used for verification in this study are limited. More actual data will be accumulated in the future to examine the relatively minor simulation results, such as differences within the same CF category. Furthermore, this study included an in-depth analysis of the coupling relationships among various influencing factors, revealing the complexity of these interactions. In the future, quantitatively describing the impact of different factors on calibration errors will be a highly challenging, but valuable task. This will contribute to further improving calibration accuracy and, consequently, enhancing the data quality of Raman temperature measurement lidar systems.

Author Contributions: Conceptualization, Y.Y. and S.C.; methodology, Y.Y., W.T. and R.C.; software, Y.Y.; validation, Y.Y., H.C. and P.G.; investigation, Y.Y.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.Y., W.T., R.C., Y.X., J.Y., R.H., H.Y. and X.L.; project administration, S.C. and H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (2023YFB3905404).

Data Availability Statement: The Raman lidar data and radiosonde data can be provided for non-commercial research purposes upon request (csy@bit.edu.cn).

Acknowledgments: The authors thank the China Meteorological Administration (CMA) for the radiosonde data.

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

References

1. Behrendt, A. Temperature measurements with lidar. In *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*; Springer: New York, NY, USA, 2005; pp. 273–305.
2. Wulfmeyer, V.; Hardesty, R.M.; Turner, D.D.; Behrendt, A.; Cadeddu, M.P.; Di Girolamo, P.; Schlüssel, P.; Van Baelen, J.; Zus, F. A review of the remote sensing of lower tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles. *Rev. Geophys.* **2015**, *53*, 819–895. [CrossRef]
3. Cooney, J. Measurement of atmospheric temperature profiles by Raman backscatter. *J. Appl. Meteorol. Climatol.* **1972**, *11*, 108–112. [CrossRef]
4. Zeyn, J.; Lahmann, W.; Weitkamp, C. Remote daytime measurements of tropospheric temperature profiles with a rotational Raman lidar. *Opt. Lett.* **1996**, *21*, 1301–1303. [CrossRef]
5. Dinoev, T.; Simeonov, V.; Arshinov, Y.; Bobrovnikov, S.; Ristori, P.; Calpini, B.; Parlange, M.; Bergh, H.v.D. Raman lidar for meteorological observations, RALMO–Part 1: Instrument description. *Atmos. Meas. Tech.* **2013**, *6*, 1329–1346. [CrossRef]
6. Li, Y.; Lin, X.; Song, S.; Yang, Y.; Cheng, X.; Chen, Z.; Liu, L.; Xia, Y.; Xiong, J.; Gong, S.; et al. A Combined Rotational Raman–Rayleigh Lidar for Atmospheric Temperature Measurements Over 5–80 km With Self-Calibration. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7055–7065. [CrossRef]
7. Liu, F.; Yi, F. Lidar-measured atmospheric N 2 vibrational-rotational Raman spectra and consequent temperature retrieval. *Opt. Express* **2014**, *22*, 27833–27844. [CrossRef]
8. Nedeljkovic, D.; Hauchecorne, A.; Chanin, M.L. Rotational Raman lidar to measure the atmospheric temperature from the ground to 30 km. *IEEE Trans. Geosci. Remote Sens.* **1993**, *31*, 90–101. [CrossRef]
9. Alpers, M.; Eixmann, R.; Fricke-Begemann, C.; Gerding, M.; Höffner, J. Temperature lidar measurements from 1 to 105 km altitude using resonance, Rayleigh, and Rotational Raman scattering. *Atmos. Chem. Phys.* **2004**, *4*, 793–800. [CrossRef]
10. Achtert, P.; Khaplanov, M.; Khosrawi, F.; Gumbel, J. Pure rotational-Raman channels of the Esrange lidar for temperature and particle extinction measurements in the troposphere and lower stratosphere. *Atmos. Meas. Tech.* **2013**, *6*, 91–98. [CrossRef]
11. Hammann, E.; Behrendt, A.; Le Mounier, F.; Wulfmeyer, V. Temperature profiling of the atmospheric boundary layer with rotational Raman lidar during the HD (CP) 2 Observational Prototype Experiment. *Atmos. Chem. Phys.* **2015**, *15*, 2867–2881. [CrossRef]
12. Arshinov, Y.F.; Bobrovnikov, S.M.; Zuev, V.E.; Mitev, V.M. Atmospheric temperature measurements using a pure rotational Raman lidar. *Appl. Opt.* **1983**, *22*, 2984–2990. [CrossRef] [PubMed]
13. Behrendt, A.; Reichardt, J. Atmospheric temperature profiling in the presence of clouds with a pure rotational Raman lidar by use of an interference-filter-based polychromator. *Appl. Opt.* **2000**, *39*, 1372–1378. [CrossRef] [PubMed]
14. Di Girolamo, P.; Marchese, R.; Whiteman, D.N.; Demoz, B.B. Rotational Raman Lidar measurements of atmospheric temperature in the UV. *Geophys. Res. Lett.* **2004**, *31*, 1–5. [CrossRef]
15. Chen, H.; Chen, S.; Zhang, Y.; Guo, P.; Chen, H.; Chen, B. Robust calibration method for pure rotational Raman lidar temperature measurement. *Opt. Express* **2015**, *23*, 21232–21242. [CrossRef] [PubMed]
16. He, J.; Chen, S.; Zhang, Y.; Guo, P.; Chen, H. A novel calibration method for pure rotational Raman lidar temperature profiling. *J. Geophys. Res.-Atmos.* **2018**, *123*, 10925–10934. [CrossRef]
17. Mahagammulla Gamage, S.; Sica, R.J.; Martucci, G.; Haefele, A. Retrieval of temperature from a multiple channel pure rotational Raman backscatter lidar using an optimal estimation method. *Atmos. Meas. Tech.* **2019**, *12*, 5801–5816. [CrossRef]
18. Su, J.; McCormick, M.P.; Lei, L. New technique to retrieve tropospheric temperature using vibrational and rotational Raman backscattering. *Earth Space Sci.* **2020**, *7*, e2019EA000817. [CrossRef]
19. Farhani, G.; Martucci, G.; Roberts, T.; Haefele, A.; Sica, R.J. A Bayesian neural network approach for tropospheric temperature retrievals from a lidar instrument. *Int. J. Remote Sens.* **2023**, *44*, 1611–1627. [CrossRef]
20. Gerasimov, V.V.; Zuev, V.V. Analytical calibration functions for the pure rotational Raman lidar technique. *Opt. Express* **2016**, *24*, 5136–5151. [CrossRef]
21. Zuev, V.V.; Gerasimov, V.V.; Pravdin, V.L.; Pavlinskiy, A.V.; Nakhtigalova, D.P. Tropospheric temperature measurements with the pure rotational Raman lidar technique using nonlinear calibration functions. *Atmos. Meas. Tech.* **2017**, *10*, 315–332. [CrossRef]
22. Gerasimov, V.V. Comparative analysis of calibration functions in the pure rotational Raman lidar technique. *Appl. Phys. B* **2018**, *124*, 134. [CrossRef]
23. Gerasimov, V.V. A simulation comparison of calibration functions for different sets of spectral filter passbands in the traditional pure rotational Raman lidar technique. *Appl. Phys. B* **2020**, *126*, 184. [CrossRef]
24. Jia, J.; Yi, F. Atmospheric temperature measurements at altitudes of 5–30 km with a double-grating-based pure rotational Raman lidar. *Appl. Opt.* **2014**, *53*, 5330–5343. [CrossRef]

25. Gerasimov, V.V. Errors of Pure Rotational Raman Lidar Absolute Calibration Due to Collisional Line Broadening. *Atmos. Ocean. Opt.* **2022**, *35*, 576–583. [CrossRef]
26. Yuan, L. Research on Retrieval and Calibration Technology of Rotating Raman Temperature Measurement LiDAR. Master's Thesis, Xi'an University of Technology, Xi'an, China, 2017.
27. Liu, Z.; Hunt, W.; Vaughan, M.; Hostetler, C.; McGill, M.; Powell, K.; Hu, Y. Estimating random errors due to shot noise in backscatter lidar observations. *Appl. Opt.* **2006**, *45*, 4437–4447. [CrossRef]
28. Hayman, M.; Stillwell, R.A.; Spuler, S.M. Optimization of linear signal processing in photon counting lidar using Poisson thinning. *Opt. Lett.* **2020**, *45*, 5213–5216. [CrossRef]
29. Chen, S.; Cao, R.; Tan, W.; Xie, Y.; Chen, H.; Guo, P.; Yu, Y.; Yu, J.; Yao, S. Uncertainty of Pure Rotational Raman–Rayleigh Lidar for Temperature Measurement in Middle-to-Upper Atmosphere: Simulation Method. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–14. [CrossRef]
30. Harrison, R.L. Introduction to monte carlo simulation. In *AIP Conference Proceeding*; NIH Public Access: Washington, DC, USA, 2010; pp. 17–21.
31. Menčík, J.; Mencik, J. Monte Carlo simulation method. In *Concise Reliability for Engineers*; InTech: Rijeka, Croatia, 2016; pp. 107–116.
32. Wandinger, U. Raman lidar. In *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*; Springer: New York, NY, USA, 2005; pp. 241–271.
33. Penney, C.M.; Peters, R.S.; Lapp, M. Absolute rotational Raman cross sections for N_2, O_2, and CO_2. *JOSA* **1974**, *64*, 712–716. [CrossRef]
34. Vaughan, G.; Wareing, D.P.; Pepler, S.J.; Thomas, L.; Mitev, V. Atmospheric temperature measurements made by rotational Raman scattering. *Appl. Opt.* **1993**, *32*, 2758–2764. [CrossRef]
35. He, J. Inersion of Atmospheric Temperature and Aerosol Optical Properties under Low Signal-to-Noise Ratio and Air Polluted Environment Based on Raman-Mie Lidar. Ph.D. Thesis, Beijing Institute of Technology, Beijing, China, 2020.
36. Li, W. Research on Observation Methods and Spatial Distribution Characteristics of Urban Night Light Pollution. Master's Thesis, Dalian University of Technology, Dalian, China, 2017.
37. Behrendt, A.; Nakamura, T.; Onishi, M.; Baumgart, R.; Tsuda, T. Combined Raman lidar for the measurement of atmospheric temperature, water vapor, particle extinction coefficient, and particle backscatter coefficient. *Appl. Opt.* **2002**, *41*, 7657–7666. [CrossRef]
38. Behrendt, A.; Nakamura, T.; Tsuda, T. Combined temperature lidar for measurements in the troposphere, stratosphere, and mesosphere. *Appl. Opt.* **2004**, *43*, 2930–2939. [CrossRef] [PubMed]
39. Qi, B.; Chen, S.; Zhang, Y.; Chen, H.; Guo, P. Geometric Form Factor Retrieval Method for Ground-Based Lidar Based on Ground-Based and Space-Borne Synchronous Observation Data. *Chin. J. Lasers* **2017**, *44*, 910003.
40. Available online: https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c?version=1.5&download=true (accessed on 18 July 2024).
41. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
42. Lee, D.K.; In, J.; Lee, S. Standard deviation and standard error of the mean. *Korean J. Anesthesiol.* **2015**, *68*, 220. [CrossRef]
43. Chen, S.; Qiu, Z.; Zhang, Y.; Chen, H.; Wang, Y. A pure rotational Raman lidar using double-grating monochromator for temperature profile detection. *J. Quant. Spectrosc. Radiat. Transf.* **2011**, *112*, 304–309. [CrossRef]
44. Yan, Q.; Wang, Y.; Gao, T.; Gao, F.; Di, H.; Song, Y.; Hua, D. Optimized retrieval method for atmospheric temperature profiling based on rotational Raman lidar. *Appl. Opt.* **2019**, *58*, 5170–5178. [CrossRef]