https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Extended TCKF1D-Var framework for Mie–Raman Lidar Water Vapor Profiling in the Nocturnal Boundary Layer: Insights into Pre-precipitation Moisture Evolution

Qi Zhang[1,3], Tianmeng Chen[2*], and Jianping Guo[2*]

[1]Key Open Laboratory of Intelligent Meteorological Observation Technology, China Meteorological Administration, Beijing 100081, China

[2]State Key Laboratory of Severe Weather Meteorological Science and Technology & Specialized Meteorological Support Technology Research Center, Chinese Academy of Meteorological Sciences, Beijing 100081, China

[3]Engineering Technology Research and Development Center, China Huayun Meteorological Technology Group Co. Ltd., Beijing 100081, China

Correspondence to: Tianmeng Chen (chentm@cma.gov.cn) and Jianping Guo (jpguo@cma.gov.cn)

Abstract. Accurate characterization of boundary-layer water vapor prior to nocturnal heavy precipitation remains challenging due to limited observational capability. In this study, we build upon a previously developed and validated thermodynamic- and cloud-microphysics-constrained Kalman filter one-dimensional variational (TCKF1D-Var) framework by extending it to incorporate nitrogen and water vapor Raman channel observations from the China Meteorological Administration Mie–Raman lidar (MRL) network. A physics-informed lidar observation operator based on the classical Raman lidar formulation is developed, together with a data-driven calibration component to account for time-varying instrumental and aerosol-related uncertainties. In addition, process and observation error covariance matrices are dynamically estimated within the Kalman filter framework to enhance retrieval robustness. The method is evaluated against co-located radiosonde observations launched prior to nocturnal heavy precipitation events at 56 MRL–radiosonde co-located stations across China in 2025. The retrieved water vapor mass mixing ratio profiles, with a vertical resolution of 30 meters and a temporal resolution of 30 minutes, exhibit consistently reduced mean bias and root mean square error compared to ERA5 prior profiles, with the largest improvements found in the 1.2–3.0 km layer. Analysis of nocturnal heavy precipitation cases further demonstrates that the retrievals capture coherent pre-precipitation moisture evolution. These results highlight the potential of combining physically constrained retrieval frameworks with Raman lidar observations for improved monitoring of boundary-layer moisture.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

1 Introduction

Nocturnal precipitation over China is frequently characterized by abrupt initiation, large hourly rainfall totals, and high societal impact (Luo et al., 2016 and 2020; Chen et al., 2021; Zhang et al., 2025a; Gao et al., 2026). Previous studies have shown that rapid variations in boundary-layer moisture play a critical role in both the triggering and rapid intensification of nighttime precipitation (Li et al., 2026; Richardson et al., 2024; Sun et al., 2025). In particular, localized moisture accumulation and vertical redistribution within the boundary layer can precondition the atmosphere for deep convection (Ahmed et al., 2020; Kirshbaum et al., 2018; Behrendt et al., 2011; Wulfmeyer et al., 2011), even in the absence of strong large-scale forcing (Kirshbaum 2011; Lu et al., 2025). These features highlight the necessity of monitoring high-resolution temporal and vertical variations of boundary-layer water vapor during nighttime. However, capturing such rapid and fine-scale moisture evolution remains challenging for existing observing systems (Richter et al., 2026; Mayer et al., 2012; Oue et al., 2022; Wulfmeyer et al., 2008 and 2015).

Ground-based active remote sensing provides a promising pathway to address this limitation. Notably, Raman lidars operating at a laser wavelength of 354.7 nm can simultaneously detect nitrogen Raman signals at 386.7 nm and water vapor Raman signals at 407.5 nm, enabling the profiling capability of atmospheric water vapor (Vaughan et al., 1988; Di Girolamo et al., 2017; Lange et al., 2019 and 2025; Whiteman et al., 2006 and 2010; Wulfmeyer et al., 2010; Behrendt et al., 2002). Compared with passive instruments such as ground-based microwave radiometers and hyperspectral infrared sounders, Raman lidars offer substantially higher vertical resolution and are capable of resolving detailed moisture structures within the boundary layer (Wulfmeyer et al., 2015; Gambacorta et al., 2025). In addition, Raman lidar offers more direct water vapor measurements with less dependence on a priori assumptions compared to passive remote sensing techniques (Whiteman et al., 2012; Foth and Pospichal; 2017; Filioglou et al., 2017). These advantages make 354.7 nm Raman lidar systems an ideal observational tool for investigating boundary-layer moisture evolution associated with nocturnal precipitation.

To achieve sufficient signal-to-noise ratio, Raman lidar systems typically require pulse energies exceeding 100 mJ (Wandinger et al., 2015). However, recent technological advances in micro-pulse laser design have enabled an alternative measurement paradigm based on low single-pulse energy and high repetition rate operation. Specifically, improvements in ultraviolet average power (1–5 W) and reduced beam divergence (0.3–0.5 mrad), combined with the use of large-aperture telescopes (0.3–0.4 meters diameter), allow micro-pulse laser-based Raman lidars to compensate for reduced pulse energy through high pulse accumulation, thereby enabling the retrieval of vertical profiles of atmospheric thermodynamic parameters, including water vapor and temperature, under both daytime and nighttime conditions (Di Girolamo et al., 2023). Despite these advances, the Mie–Raman lidar (MRL) network deployed by the China Meteorological Administration (CMA) operates with a pulse energy of 0.6 mJ at 354.7 nm and repetation rate of 1000 Hz (Shao et al., 2025; Zhang et al., 2025b).

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Although such systems inherently benefit from high repetition rates, the resulting observations are still subject to significant noise, particularly at short temporal scales, posing challenges for retrieving reliable water vapor mass mixing ratio profiles using conventional inversion approaches. Therefore, there is a clear need for advanced retrieval methods that can fully exploit the nitrogen and water vapor Raman channel observations from these lidars and provide robust estimates of boundary-layer water vapor structure.

In the previous work, we developed a thermodynamic- and cloud-microphysics-constrained Kalman filter one-dimensional variational (TCKF1D-Var) retrieval framework (Zhang et al., 2026a) applied to ground-based microwave radiometer (GMWR) observations to retrieve temperature, humidity, and hydrometeor profiles with high accuracy. This method incorporates moist thermodynamic constraints by using virtual potential temperature as the control variable and improves retrieval stability under cloudy and precipitating conditions by coupling the cloud mircophysics scheme with the cost function. The retrieved profiles have demonstrated strong capability in capturing the evolution of boundary-layer thermodynamic conditions prior to the initiation of heavy precipitation.

This study extends the TCKF1D-Var framework to incorporate nitrogen and water vapor Raman channel observations from the MRL network deployed by the CMA. By integrating these measurements within a physically constrained retrieval system, we derive boundary-layer water vapor mass mixing ratio profiles with a vertical resolution of 30 meters and a temporal resolution of 30 minutes. The performance of the proposed method is evaluated using collocated radiosonde observations from 56 stations across China in 2025. Furthermore, we assess the capability of the retrieved profiles to capture pre-precipitation boundary-layer moisture signals by analyzing nocturnal heavy precipitation events observed at the same stations. The remainder of this paper is organized as follows. Section 2 describes the observational instruments, datasets, and the extensions of the TCKF1D-Var framework developed in this study. Section 3 presents the validation of the retrieved water vapor mass mixing ratio profiles and demonstrates their capability to capture boundary-layer moisture variations prior to heavy precipitation. Finally, Section 4 provides a summary and concluding remarks.

2 Instrument, Data, and Method

2.1 Mie–Raman lidar

Starting in 2021, CMA initiated the deployment of MRLs and associated instruments at key observational stations (Zhang et al., 2025c; Shao et al., 2025). By the end of 2025, a total of 56 radiosonde stations participating in international data exchange had been co-located with MRLs across China, each providing more than one year of operational observations (Figure 1). The MRL system operates at laser wavelengths of 354.7, 532.1, and 1064.1 nm. The receiver is composed of eight detection channels, among which three are Raman channels centered at 386.7, 407.5, and 607.6 nm, enabling the

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

retrieval of atmospheric water vapor and nitrogen Raman backscatter signals. Detailed system specifications are summarized in Table1.

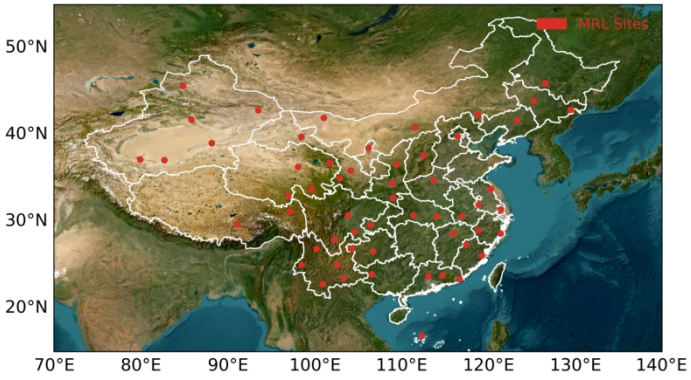

Figure 1: Spatial distribution of the MRL systems co-located with radiosonde stations across China as of the end of 2025. A total of 56 stations participating in international data exchange are equipped with MRL systems, each providing more than one year of continuous observations. The topographical basemap is retrieved from ArcGIS Online powered by Esri (Esri et al., 2025).

Table 1: Technical specifications of the MRL system

Laser	Emission Wavelength	354.7 nm	532.1 nm	1064.1 nm
	Single Pulse Enery	0.6 mJ	1.5 mJ	1.8 mJ
	Repetition Rate		1000 Hz	
	Beam Divergence	0.3 mrad	0.3 mrad	0.5 mrad
Telescope	Type		Cassegrain	
	Primary Mirror Diameter		30 cm	
	Field of View		1 mrad	
Raman Channel Interference Filter	Center Wavelength	386.7 nm	407.5 nm	607.6 nm
	Bandwidth	0.5 nm	1 nm	1 nm
	Peak transmission	80%	90%	90%
	Out-of-band rejection	OD7	OD6	OD 7
Raman Channel Detector	Type		PMT	
	Quantum Efficiency		40%	
	Spectral Range		300 nm – 720 nm	
	Dark Count Rate		100 counts/second	
	Maximum Linear Count Rate		1.5×10^6 counts/second	

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

	Pulse Pair Resolution	30 ns
Measurement Capability	Blind Range Height	150 m
	Effective Resolution	30 m

2.2 Data

2.2.1 In-situ observations

Routine radiosonde observations were conducted at the co-located stations, with soundings launched twice daily at approximately 00:00 and 12:00 UTC. The measurements provide temperature with a resolution of 0.1 K and an accuracy of 0.5 K, relative humidity with a resolution of 1% and an accuracy of 5%, and pressure with a resolution of 0.1 hPa and an accuracy of 0.5 hPa (Yao et al., 2025; Cao et al., 2025). These observations serve as independent reference data for evaluating the accuracy of thermodynamic profiles retrieved from the MRL. Hourly accumulated precipitation from tipping bucket rain gauges was used to identify extreme precipitation cases in Section 3.2, with an absolute measurement bias of ± 0.2 mm under weak rainfall conditions (≤ 4 mm/h) and a relative bias within ±4% during heavy precipitation.

2.2.2 Fifth Generation ECMWF (ERA5) Reanalysis

The ERA5 reanalysis dataset (Hersbach et al., 2020) provides dynamically consistent and physically constrained atmospheric fields by assimilating a wide range of satellite and in situ observations within a state-of-the-art numerical weather prediction framework . Owing to its widespread use and well-documented accuracy, ERA5 is adopted as the a priori state for the retrieval of water vapor mass mixing ratio profiles in this study. Specifically, profiles of temperature, specific humidity, pressure, and geopotential height are extracted and interpolated to the station locations using nearest-neighbor spatial interpolation. The temporal resolution is further refined using inverse distance weighting method.

2.3 Extension of TCKF1D-Var to MRL observations

In this study, we retain the core design of the previously developed GMWR-oriented TCKF1D-Var framework (Figure 2 a and b , Zhang et al., 2026a). Specifically, virtual potential temperature (θ_v) is adopted as the control variable, defined as:

$$\theta_v = T(\frac{P_0}{P})^{R_d/C_p}(1 + 0.61q_v - q_c), \quad (1)$$

where P is air pressure, T is temperature, q_v and q_c are the water vapor and cloud water mixing ratios, respectively, and R_d and C_p are the gas constant and specific heat capacity of dry air at constant pressure. The ratio-based cost function $J_{(x)}$ is also preserved and expressed as:

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

$$J_{(x)} = (\frac{H_{(x)}}{y} - 1)^2 + (\frac{\theta_{\nu(x)}}{\theta_{\nu(x_o)}} - 1)^2, \quad (2)$$

where y represents the instrument observation at a given time; in this study, it is defined as the ratio of the water vapor Raman signal at 407.5 nm (PR_{H_2O}) to the nitrogen Raman signal at 386.7 nm (PR_{N_2}). The term x_o denotes the a priori profile, and $H_{(x)}$ is the lidar observation operator that maps the atmospheric state x to the corresponding simulated observation. The minimization of the cost function is consistently performed using the L-BFGS-B algorithm (Gerber and Furrer, 2019). To enable accurate retrieval of water vapor mass mixing ratio profiles from low single-pulse-energy Raman lidar observations, we introduce two key methodological extensions (Figure 2c) to the TCKF1D-Var framework. These include (i) the development of a A physics-informed lidar observation operator with random forest calibration (Section 2.3.1) for nitrogen and water vapor Raman signals; and (ii) a Kalman filter–based observation fusion strategy with dynamically updated process and observation error covariance matrices (Section 2.3.2). Together, these advances substantially enhance the capability of the framework to exploit MRL observations, leading to robust and physically consistent water vapor retrievals.

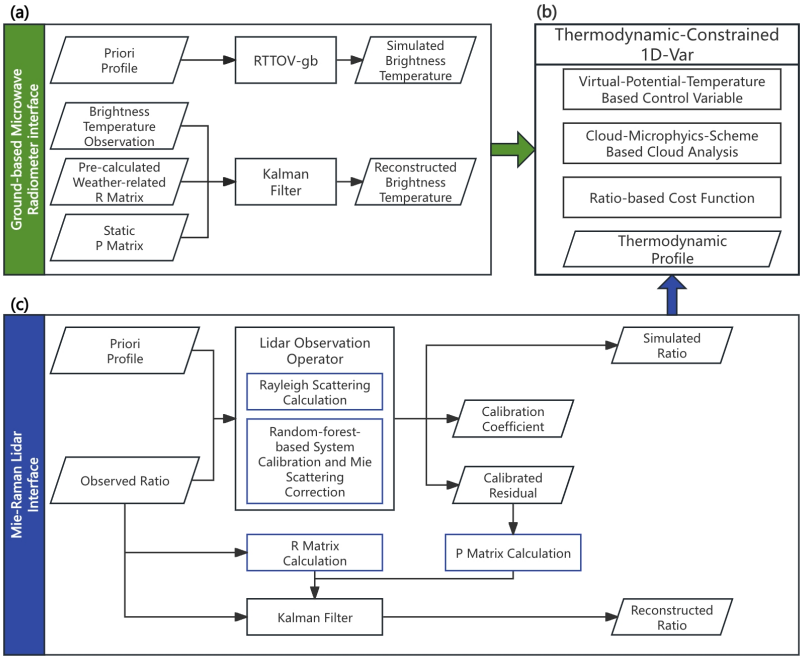

Figure 2: Workflow of TCKF1D-Var framework

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

2.3.1 Lidar observation operator

The lidar observation operator is constructed following the classical Raman lidar formulation described by Ansmann et al. (1992), Whiteman et al. (1992), and Wulfmeyer and Behrendt (2021). The water vapor Raman signal at 407.5 nm can be expressed as

$$PR_{H_2O} = K_{H_2O}\frac{O(z)}{z^2}N_{H_2O}(z)\frac{d\sigma_{H_2O}(\pi)}{d\Omega}\ exp\left\{-\int_0^z\left[\alpha^{mol}_{H_2O}(z)+\alpha^{aer}_{H_2O}(z)+\alpha^{mol}_{\lambda_0}(z)+\alpha^{aer}_{\lambda_0}(z)\right]dz\right\}, \quad (3)$$

where $O(z)$ denotes the overlap function, K_{H_2O} is the system efficiency factor provided by the instrument manufacturer, and $N_{H_2O}(z)$ is the water vapor number density at level z. The term $\frac{d\sigma_{H_2O}(\pi)}{d\Omega}$ represents the range-independent differential Raman backscatter cross section. The extinction coefficients $\alpha^{mol}(z)$ and $\alpha^{aer}(z)$ describe molecular and aerosol extinction at the Raman (407.5 nm) and emitted (354.7 nm) wavelengths, respectively. Similarly, the nitrogen Raman signal at 386.7 nm is given by

$$PR_{N_2} = K_{N_2}\frac{O(z)}{z^2}N_{N_2}(z)\frac{d\sigma_{N_2}(\pi)}{d\Omega}\ exp\left\{-\int_0^z\left[\alpha^{mol}_{N_2}(z)+\alpha^{aer}_{N_2}(z)+\alpha^{mol}_{\lambda_0}(z)+\alpha^{aer}_{\lambda_0}(z)\right]dz\right\}, \quad (4)$$

where K_{N_2} includes all range-independent system parameters for the nitrogen Raman channel, and $N_{N_2}(z)$ is the nitrogen number density.

Taking the ratio of Eqs. (3) and (4) eliminates common system- and geometry-dependent terms, yielding

$$H = \frac{PR_{H_2O}}{PR_{N_2}} = C_R\frac{N_{H_2O}(z)}{N_{N_2}(z)}\frac{d\sigma_{H_2O}(\pi)/d\Omega}{d\sigma_{N_2}(\pi)/d\Omega}\frac{exp\left[\int_0^z\alpha^{mol}_{N_2}(z)dz\right]}{exp\left[\int_0^z\alpha^{mol}_{H_2O}(z)dz\right]}\frac{exp\left[\int_0^z\alpha^{aer}_{N_2}(z)dz\right]}{exp\left[\int_0^z\alpha^{aer}_{H_2O}(z)dz\right]}, \quad (5)$$

where $C_R = \frac{K_{H_2O}}{K_{N_2}}$ is an instrument-dependent constant. The ratio $\frac{N_{H_2O}(z)}{N_{N_2}(z)}$ represents the number density of water vapor molecules relative to that of nitrogen molecules in air. Considering that nitrogen constitutes a nearly constant fraction (78.08%) of dry air, this ratio can be expressed as

$$\frac{N_{H_2O}(z)}{N_{N_2}(z)} = \frac{m_{H_2O}/M_{H_2O}}{(m_{air}/M_{air})\times 0.7808}, \quad (6)$$

where m_{H_2O} and (m_{air} denote the mass densities of water vapor and dry air, respectively, and $M_{H_2O} = 18.015\ g/mol$ and $M_{air} = 28.97\ g/mol$ are their corresponding molar masses. The ratio of mass densities can be conveniently expressed in terms of the water vapor mass mixing ratio ($qvapor$) as

$$\frac{m_{H_2O}}{m_{air}} = \frac{qvapor}{1-qvapor}, \quad (7)$$

Substituting Eq. (7) into Eq. (6), the molecular number density ratio becomes

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

$$\frac{N_{H_2O}(z)}{N_{N_2}(z)} = \frac{qvapor}{1-qvapor} \times \frac{28.97\ g/mol}{18.015\ g/mol} \times \frac{1}{0.7808} \approx 2.06 \frac{qvapor}{1-qvapor}. \quad (8)$$

The ratio of Raman cross sections $\frac{d\sigma_{H_2O}(\pi)/d\Omega}{d\sigma_{N_2}(\pi)/d\Omega}$ is assumed constant and set to 1.3 in this study, based on typical values of $6.5 \times 10^{-31}\ cm^2/sr$ and $4.8 \times 10^{-31}\ cm^2/sr$ for water vapor and nitrogen, respectively. The molecular extinction term is parameterized as

$$\alpha_{\lambda}^{mol} = \alpha_{\lambda,\ std}^{mol} \frac{P(z)}{P_0} \frac{T_0}{T(z)}, \quad (9)$$

where $\alpha_{\lambda,\ std}^{mol}$ is the standard-atmosphere extinction coefficient. $P_0 = 101325\ Pa$ and $T_0 = 288.15K$ represent the reference air pressure and temperature of the standard atmosphere, respectively. In this study, values of $7.0 \times 10^{-5} m^{-1}$ and $6.0 \times 10^{-5} m^{-1}$ are used for 386.7 nm and 407.5 nm, respectively. This leads to

$$\frac{exp\left[\int_0^z \alpha_{N_2}^{mol}(z)dz\right]}{exp\left[\int_0^z \alpha_{H_2O}^{mol}(z)dz\right]} = exp\left[10^{-5} \int_0^z \frac{P(z)}{P_0} \frac{T_0}{T(z)} dz\right]. \quad (10)$$

The aerosol extinction term is parameterized using the Ångström power law (Ångström, 1929):

$$\alpha_{\lambda}^{aer} = \alpha_{355}^{aer} \left(\frac{\lambda}{355.0}\right)^{-A}, \quad (11)$$

where A is the Ångström exponent derived from 354.7 nm and 532 nm aerosol extinction measurements:

$$A = -\frac{ln(\alpha_{355}^{aer}/\alpha_{532}^{aer})}{ln(355.0/532.0)}, \quad (12)$$

Thus, the aerosol contribution becomes

$$\frac{exp\left[\int_0^z \alpha_{N_2}^{aer}(z)dz\right]}{exp\left[\int_0^z \alpha_{H_2O}^{aer}(z)dz\right]} = exp\left\{\alpha_{355}^{aer} \int_0^z \left[\left(\frac{386.7}{355.0}\right)^{-A} - \left(\frac{407.5}{355.0}\right)^{-A}\right] dz\right\}, \quad (13)$$

Substituting Eqs. (8), (10), and (13) into Eq. (5), the final lidar observation operator is obtained as

$$H = 2.68 \times C_R \cdot \frac{qvapor}{1-qvapor} \cdot exp\left\{10^{-5} \int_0^z \frac{P(z)}{P_0} \frac{T_0}{T(z)} dz + \alpha_{355}^{aer} \int_0^z \left[\left(\frac{386.7}{355.0}\right)^{-A} - \left(\frac{407.5}{355.0}\right)^{-A}\right] dz\right\}, \quad (14)$$

which depends only on instrument constants and directly observed quantities. The coefficient 2.68 results from the combination of the molecular number density ratio (Eq. 8) and the assumed Raman cross-section ratio.

Although the C_R in Eq. (14) can, in principle, be determined from factory-calibrated system constants, such an approach is not suitable for long-term observations. Continuous operation of the MRL system leads to gradual degradation in optical efficiency, resulting in inconsistencies between the initial calibration constants and the actual instrument state. Consequently, the direct use of factory-derived C_R introduces systematic biases in the retrieved thermodynamic profiles. In addition, the aerosol extinction coefficient at 355 nm provided by the MRL has a temporal resolution of 5 minutes, which is coarser than

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

the 1-minute resolution of the nitrogen and water vapor Raman signals. Incorporating this lower-resolution product into the Mie scattering term in Eq. (14) may degrade retrieval accuracy, particularly under rapidly evolving atmospheric conditions. To address these limitations, a random forest (Belgiu and Drăguţ, 2016) based calibration strategy is introduced to account for the combined effects of the system calibration coefficient and unresolved aerosol extinction-related uncertainties. The random forest model is trained using lidar-observed Raman signals within a moving temporal window preceding the retrieval time, and no external thermodynamic observations (e.g., radiosonde data) are used in the training process, thereby avoiding potential information leakage in the validation. The modified observation operator can thus be expressed as

$$H = 2.68\frac{qvapor}{1-qvapor} \cdot exp\left\{10^{-5}\int_0^z \frac{P(z)}{P_0}\frac{T_0}{T(z)}dz\right\} \cdot RandomForest(y), \quad (15)$$

where $RandomForest$ denotes the random forest model with 100 trees and a squared-error split criterion, and is applied independently at each vertical level using time-series features derived from the Raman signal ratio within the temporal window. It should be emphasized that the random forest component serves only as a multiplicative correction to the physically derived operator, rather than replacing its underlying physical structure. In this way, the physical interpretability of the observation operator is preserved while improving its adaptability to time-varying instrument performance and atmospheric conditions.

At this stage, the lidar observation operator is established. In addition to providing the simulated ratio (y_{sim}) derived from the atmospheric state profiles, the operator also outputs a calibration coefficient ($Coef_{Cal}$) profile and calibrated residual (Res_{Cal}) profiles, thereby enabling a comprehensive characterization of the forward-model uncertainty. Specifically, the $Coef_{Cal}$ profile represents the combined term $C_R \cdot exp\left\{\alpha_{355}^{aer}\int_0^z\left[(\frac{386.7}{355.0})^{-A}-(\frac{407.5}{355.0})^{-A}\right]dz\right\}$ in Eq. (14), which is estimated using the random forest algorithm constrained by the MRL Raman channel observations. Meanwhile, the Res_{Cal} profiles is defined as the difference between the MRL-observed ratio and the simulated ratio within a time window preceding the thermodynamic profile retrieval time. This residual quantifies the discrepancy between observations and the operator-simulated signal given the atmospheric state, and is therefore interpreted as the process error of the lidar observation operator.

2.3.2 Kalman filter–based observation fusion

Kalman filtering provides a recursive Bayesian framework for optimally combining model predictions and observations under uncertainty (Kalman, 1960). By explicitly accounting for process and observation errors through their respective covariance matrices, it enables sequential updating of the system state while minimizing estimation variance. One of its key advantages lies in its ability to handle noisy and incomplete measurements, while maintaining temporal consistency in the retrieved variables. In addition, the Kalman filter naturally incorporates prior information and propagates uncertainty forward in time, making it suitable for retrieving atmospheric thermodynamic profile low-single-pulse-energy MRL

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

observation where both measurement noise and model errors are non-negligible. Consequently, dynamically updating the process error covariance (P) matrix and the observation error covariance (R) matrix becomes essential. For low-single-pulse-energy MRL Raman channel observations, both error sources exhibit strong temporal variability: the R is influenced by rapidly changing measurement uncertainty associated with environmental contamination and signal attenuation, while the P reflects model inadequacies that vary with atmospheric conditions and the performance of the lidar observation operator. Therefore, by dynamically estimating and updating both covariance matrices, the Kalman filter can more accurately represent the time-dependent error structure, thereby improving the reliability and physical consistency of the fused atmospheric state estimates.

In this study, the R matrix is estimated using the variability of the MRL-observed ratio relative to its short-term mean within a time window preceding the thermodynamic profile retrieval time. Specifically, the deviations between the instantaneous ratio (y) and its temporal mean over this period ($\overline{y}$) are used to quantify the observation uncertainty as

$$R = \frac{(y-\overline{y})^T \bullet (y-y)}{N_{obs}}, \tag{16}$$

where the superscript T denotes the transpose operator, and N_{obs} represents the number of MRL Raman channel observations within a time window prior to the retrieval. The P matrix is estimated using an analogous formulation but based on the Res_{Cal} derived from the lidar observation operator, computed as

$$P = \frac{Res_{Cal}^T \bullet Res_{Cal}}{N_{obs}}. \tag{17}$$

In this study, both the diagonal and off-diagonal elements of the R and P matrices are retained to account for vertical error correlations associated with turbulent mixing within the boundary layer, which are expected to influence the structure of the retrieved water vapor mass mixing ratio profiles (Foken et al., 2012).

3 Results and discussion

3.1 Retrieval temporal resolution selection

Previous studies (Laly et al., 2024 and 2025) have suggested that the temporal resolution for water vapor mass mixing ratio profile retrieval using the 407.5 and 386.7 nm Raman channels typically ranges from 15 to 30 minutes. However, owing to differences in instrumental configurations and retrieval strategies, the optimal temporal resolution adopted in this study cannot be directly prescribed and must instead be determined through sensitivity experiments based on representative cases. Given that this study focuses on the evolution of boundary-layer water vapor prior to the initiation of short-duration intense nocturnal precipitation, radiosonde observations at 12:00 UTC preceding the triggering time are selected as the reference.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Corresponding water vapor mass mixing ratio profiles with temporal resolutions of 10, 20, and 30 minutes are constructed from MRL observations, and their performance is evaluated using the mean bias (MB) and root mean squared error ($RMSE$), defined as $MB = \overline{(radiosonde - retrieval)}$ and $RMSE = \sqrt{\overline{(radiosnonde - retrieval)^2}}$, respectively, as the primary metrics to identify the temporal resolution most suitable for the MRL observations and the TCKF1D-Var framework. Furthermore, based on hourly accumulated precipitation observations co-located with 56 radiosonde stations in 2025, heavy precipitation events at each station are identified using the 99th percentile of the station-specific annual hourly precipitation distribution as the threshold. Nighttime extreme precipitation events are then selected from these events by restricting to periods with a solar elevation angle lower than −6°. The selected 39 cases, including station locations, occurrence times, and hourly rainfall intensity, are summarized in Table 2.

Table 2: Summary of the selected case studies used for the sensitivity experiments

Station ID	Latitude (degree North)	Longitude (degree East)	Occurance Time (UTC)	Hourly Accumulated Precipitation (mm)	Radiosonde Launching Time (Hours Pre-Onset)
50953	45.93	126.56	2025-08-02 13:00	52.0	1
			2025-08-05 19:00	35.1	7
53772	37.62	112.57	2025-07-04 13:00	61.1	1
			2025-08-19 19:00	15.7	7
54161	43.90	125.21	2025-07-24 19:00	15.2	7
54218	42.30	118.83	2025-07-26 19:00	10.7	7
56187	30.74	103.86	2025-06-17 18:00	30.2	6
			2025-07-25 15:00	38.7	3
56492	28.80	104.60	2025-07-04 16:00	23.9	4
			2025-08-09 16:00	25.5	4
			2025-08-10 16:00	26.9	4
56571	27.90	102.26	2025-06-18 20:00	19.4	8
			2025-09-18 20:00	20.9	8
57083	34.71	113.65	2025-08-05 16:00	23.8	4
57245	32.71	109.03	2025-08-09 13:00	16.3	1
			2025-08-21 19:00	18.2	7

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

57461	30.73	111.36	2025-08-07 15:00	31.9	3
			2025-08-09 12:00	51.4	0
			2025-08-30 13:00	35.0	1
			2025-09-05 15:00	47.0	3
57516	29.57	106.46	2025-07-08 17:00	51.4	5
			2025-08-06 16:00	35.4	4
			2025-09-05 17:00	33.5	5
58150	33.74	120.29	2025-08-25 17:00	38.7	5
			2025-08-30 18:00	21.4	6
			2025-09-23 19:00	25.6	7
58362	31.40	121.45	2025-10-17 21:00	38.9	9
58424	30.61	116.96	2025-07-16 19:00	28.1	7
			2025-08-10 15:00	19.3	3
58633	29.00	118.90	2025-04-11 19:00	24.6	7
			2025-05-22 19:00	19.3	7
58725	27.33	117.46	2025-06-15 14:00	36.7	2
58847	26.08	119.28	2025-05-18 13:00	23.4	1
			2025-07-20 20:00	35.9	8
59211	23.90	106.60	2025-05-22 14:00	25.7	2
59280	23.71	113.08	2025-08-05 21:00	35.6	9
			2025-09-20 20:00	32.4	8
59293	23.79	114.72	2025-07-31 16:00	73.0	4
59316	23.38	116.67	2025-07-21 17:00	33.3	5

The mean error profiles derived from the sensitivity experiments (Figure 3a) indicate a clear dependence of retrieval performance on the chosen temporal resolution within the TCKF1D-Var framework. Among the tested configurations, the 30 minutes temporal resolution (purple) yields the smallest MB, followed by the 20 minutes configuration (green), while the 10 minutes resolution (red) exhibits the largest MB. Despite these differences, all three experiments show consistently lower mean errors than the ERA5 prior profile (black), demonstrating the capability of the proposed framework to effectively improve water vapor mass mixing ratio retrievals. Consistent behavior is also observed in the $RMSE$ profiles

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

(Figure 3b), where the measurement uncertainty associated with the 30 minutes temporal resolution (purple) is significantly reduced compared to those obtained with 20 minutes (green), 10 minutes (red), and the ERA5 prior (black). This suggests that extending the temporal averaging window enhances the water vapor information content and leads to more stable retrievals. Considering both the mean error and measurement uncertainty, a 30-minute time window preceding the thermodynamic profile retrieval time is therefore selected as the optimal configuration for the TCKF1D-Var framework in this study.

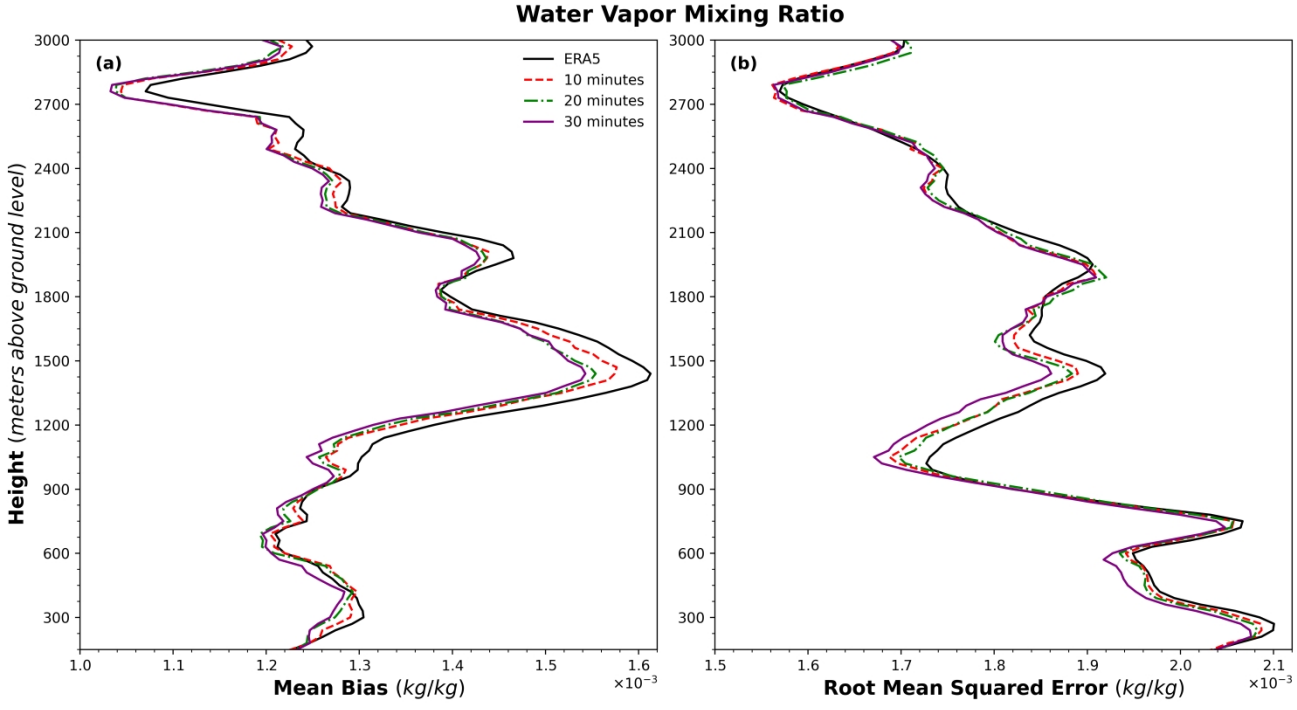

Figure 3: Vertical profiles of (a) mean bias (MB) and (b) root mean square error (RMSE) of the retrieved water vapor mass mixing ratio profiles obtained from the sensitivity experiments with temporal resolutions of 10 minutes (red), 20 minutes (green), and 30 minutes (purple), together with the ERA5 prior profile (black). The 30 minutes configuration yields the smallest MB and RMSE, followed by the 20 minutes and 10 minutes configurations, while all experiments show improved performance compared to the ERA5 prior.

To assess why the 30-minute temporal resolution yields the highest retrieval accuracy, we examine the relative magnitude of the effective observational information provided by the observations and the forward model error by introducing the innovation-to-forward model error ratio (IFR), defined as the ratio between the innovation and the observation operator error:

$$IFR = \frac{Innovation}{Error_{Operator}}, \tag{18}$$

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

where the $Innovation$ is defined as the difference between the reconstructed ratio obtained from MRL observations using the Kalman filter ($Ratio_{obs}^{reconstruct}$) and the simulated ratio derived from the lidar observation operator driven by the ERA5 prior profile ($Ratio_{sim}^{ERA5}$):

$$Innovation = \left| Ratio_{obs}^{reconstruct} - Ratio_{sim}^{ERA5} \right|, \quad (19)$$

This term represents the mismatch between the prior-driven forward simulation and the observation-constrained reconstruction, and thus reflects the information gain introduced by the observations. The observation operator error ($Error_{Operator}$) is defined as the difference between the reconstructed ratio ($Ratio_{obs}^{reconstruct}$) and the simulated ratio obtained using radiosonde profiles ($Ratio_{sim}^{Radiosonde}$):

$$Error_{Operator} = \left| Ratio_{obs}^{reconstruct} - Ratio_{sim}^{Radiosonde} \right|, \quad (20)$$

This quantity characterizes the discrepancy between the forward model simulation based on the best-available atmospheric state (radiosonde observations) and the reconstructed observation, thereby providing a measure of the forward model (observation operator) uncertainty. Based on these definitions, an IFR value greater than 1.0 indicates that the effective observational information exceeds the forward model error, suggesting that the observational signal is relatively strong compared to forward model uncertainties. Conversely, IFR values smaller than 1.0 indicate that the observational signal may be more strongly affected by forward model uncertainties, limiting the ability of the assimilation system to effectively adjust the prior profile. It should be noted that the IFR defined here is a diagnostic metric rather than a rigorous information-theoretic quantity. Its purpose is to provide a qualitative assessment of the relative magnitude of observational information compared to forward model uncertainty within the TCKF1D-Var framework. Therefore, the IFR is used primarily for interpretative analysis among different experimental configurations rather than as a strict quantitative measure of information content.

The vertical distributions of IFR averaged over the cases listed in Table 2 are presented in Figure 4 for the three sensitivity experiments with temporal resolutions of 10, 20, and 30 minutes. It is evident that the IFR profile associated with the 30 minutes configuration (purple) is generally higher than those of the 20 minutes (green) and 10 minutes (red) configurations. This indicates that the superior accuracy of the 30 min retrievals can be attributed to a more favorable balance between observational information and forward model error. Specifically, the longer Kalman filter time window enhances the reconstructed ratio, leading to larger innovations relative to the ERA5-based prior simulation. As a result, the observational information becomes less susceptible to being obscured by the noise of the lidar observation operator, thereby improving the overall retrieval performance. It should be noted, however, that the above analysis is limited to explaining why the 30-minute temporal resolution outperforms the 10- and 20-minute configurations, and does not imply a monotonic improvement

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

of retrieval accuracy with increasing Kalman filter window length. Although further increases in the temporal window may potentially enhance retrieval accuracy, they may simultaneously reduce the capability to capture rapid atmospheric variations within boundary layer preceding the onset of intense precipitation. This trade-off could ultimately degrade the performance of the TCKF1D-Var method in short-term severe weather early warning applications.

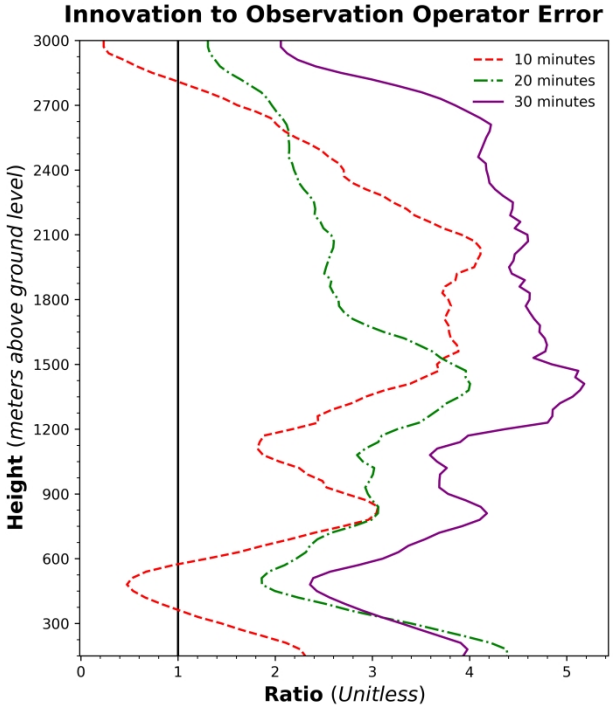

Figure 4: Vertical distributions of the mean innovation-to-observation operator error ratio (IFR) for the three sensitivity experiments with temporal resolutions of 10 minutes (red), 20 minutes (green), and 30 minutes (purple), averaged over the cases listed in Table 2. The IFR quantifies the relative magnitude of effective observational information with respect to forward model error. Values greater than 1.0 (black) indicate that the observational information dominates over the forward model uncertainty, whereas values smaller than 1.0 suggest that the information content is largely obscured by forward model errors.

3.2 Thermodynamic profile general performance

According to the statistics summarized in Table 2, the lower bound of nighttime extreme precipitation occurring at the 56 MRL – radiosonde co-located stations in 2025 is 10.7 mm. Therefore, a threshold of hourly accumulated precipitation exceeding 10 mm during nighttime is adopted to define heavy precipitation events, allowing for an expanded sample set and a more general evaluation. The statistical distribution of the enlarged samples is presented in Table 3. For events with hourly precipitation of 10–20 mm and those exceeding 30 mm, the lead-time between the 12:00 UTC radiosonde launches and the heavy precipitation onset time ranges from 0 to 9 hours. In contrast, for events with precipitation between 20 and 30 mm, this time difference ranges from 0 to 8 hours. The number of events within the 10 – 20 mm category (170 cases) is substantially larger than those in the 20–30 mm (28 cases) and ≥30 mm (26 cases) categories. Applying the same evaluation

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

approach described in Section 3.1 to the selected cases in Table 3, the vertical distributions of the MB with 95% confidence intervals (Figure 5a) and the $RMSE$ (Figure 5b) are obtained for the TCKF1D-Var water vapor mass mixing ratio profiles with a temporal resolution of 30 minutes, as well as for the ERA5 prior profiles. Consistent with the results shown in Figure 3, the evaluation based on the expanded sample set further confirms that the TCKF1D-Var retrievals exhibit lower MB and $RMSE$ than the ERA5 prior profiles. This demonstrates that, at a temporal resolution of 30 minutes, the TCKF1D-Var framework, constrained by MRL Raman channel observations, has a clear capability to improve the ERA5 prior water vapor mass mixing ratio profiles. In addition, the results based on the expanded samples indicate that the improvement in water vapor mass mixing ratio retrieval is more pronounced within the 1200–3000 meters altitude above ground level compared to below 1200 m. This feature is broadly consistent with the findings in Figure 3, although the altitude separating the two regimes (previously around 1050 m in Figure 3) shows a slight upward shift in the present analysis.

Table 3: Statistical distribution of the expanded sample set of nighttime heavy precipitation events

Radiosonde Launching Lead-time (hours)	Precipitation Amount		
	10 – 20 mm	20 – 30 mm	above 30 mm
0	15	3	3
1	23	1	3
2	16	2	3
3	16	5	4
4	14	6	1
5	14	1	4
6	12	2	1
7	22	6	2
8	17	2	1
9	21	0	4

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

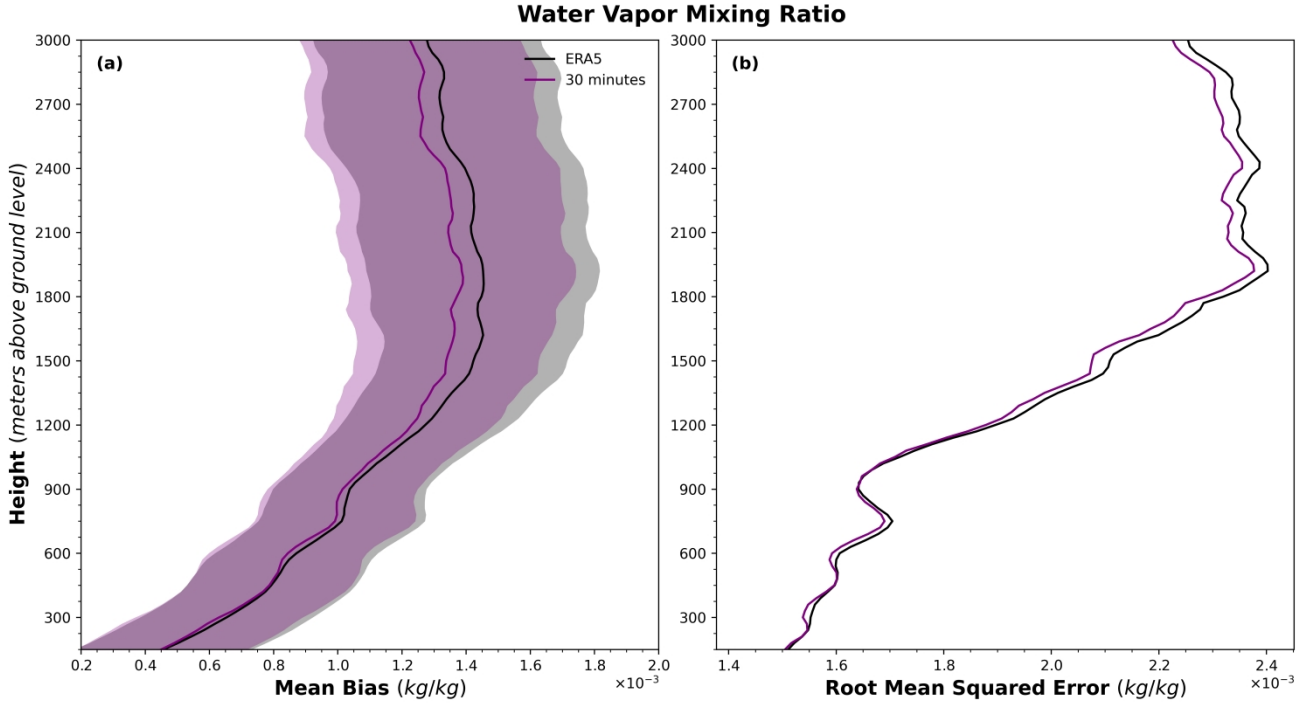

Figure 5: Vertical distributions of (a) mean bias with 95% confidence intervals and (b) root-mean-square error (RMSE) for water vapor mass mixing ratio profiles derived from the TCKF1D-Var method with a temporal resolution of 30 minutes (purple) and from ERA5 prior profiles (black), based on the expanded sample set of nighttime heavy precipitation events defined in Table 3. The shaded areas represent the 95% confidence intervals.

To investigate the dependence of retrieval accuracy on precipitation intensity, the MB with 95% confidence intervals and $RMSE$ profiles are calculated separately for three categories of nighttime heavy precipitation events defined in Table 3. For the 10–20 mm category, the resulting MB (Figure 6a) and $RMSE$ (Figure 6b) are highly consistent with those obtained from the full sample (Figure 5). This similarity is primarily attributed to the dominance of this category, which accounts for approximately 75.8% of the total sample size. In contrast, the results for the 20–30 mm category show more pronounced differences compared to the full-sample statistics. Specifically, while the full-sample MB exhibits a vertical structure characterized by lower values near the surface (approximately 0.55×10^{-3} kg/kg) and higher values aloft (approximately 1.4×10^{-3} kg/kg), the 20–30 mm category displays a more vertically uniform distribution (Figure 6c), with mean bias values generally ranging between 1.5×10^{-3} kg/kg and 2.3×10^{-3} kg/kg. Similarly, the $RMSE$ for the full sample increases with height (from 1.5×10^{-3} kg/kg to 2.35×10^{-3} kg/kg), whereas the 20–30 mm category (Figure 6d) shows values between 2.0×10^{-3} kg/kg and 2.75×10^{-3} kg/kg, with a notable reduction in $RMSE$ within the 900–1200 meters layer. The most distinct differences are observed for the ≥30 mm category. The characteristic vertical pattern of MB observed in the full sample (lower near the surface and higher aloft) disappears entirely (Figure 6e). Moreover, the 95% confidence intervals are broader than those of the full sample and exhibit an opposite vertical tendency, with larger uncertainties near the surface

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

and smaller ones at higher altitudes. In addition, a reversal is found in the $RMSE$ vertical structure (Figure 6f), which changes from increasing with height (in the full sample) to decreasing with height for the ≥30 mm category. Despite these substantial differences across precipitation intensity classes, the overall conclusion remains robust: the TCKF1D-Var water vapor mass mixing ratio profiles with a temporal resolution of 30 minutes consistently outperform the ERA5 prior profiles in terms of both MB and $RMSE$.

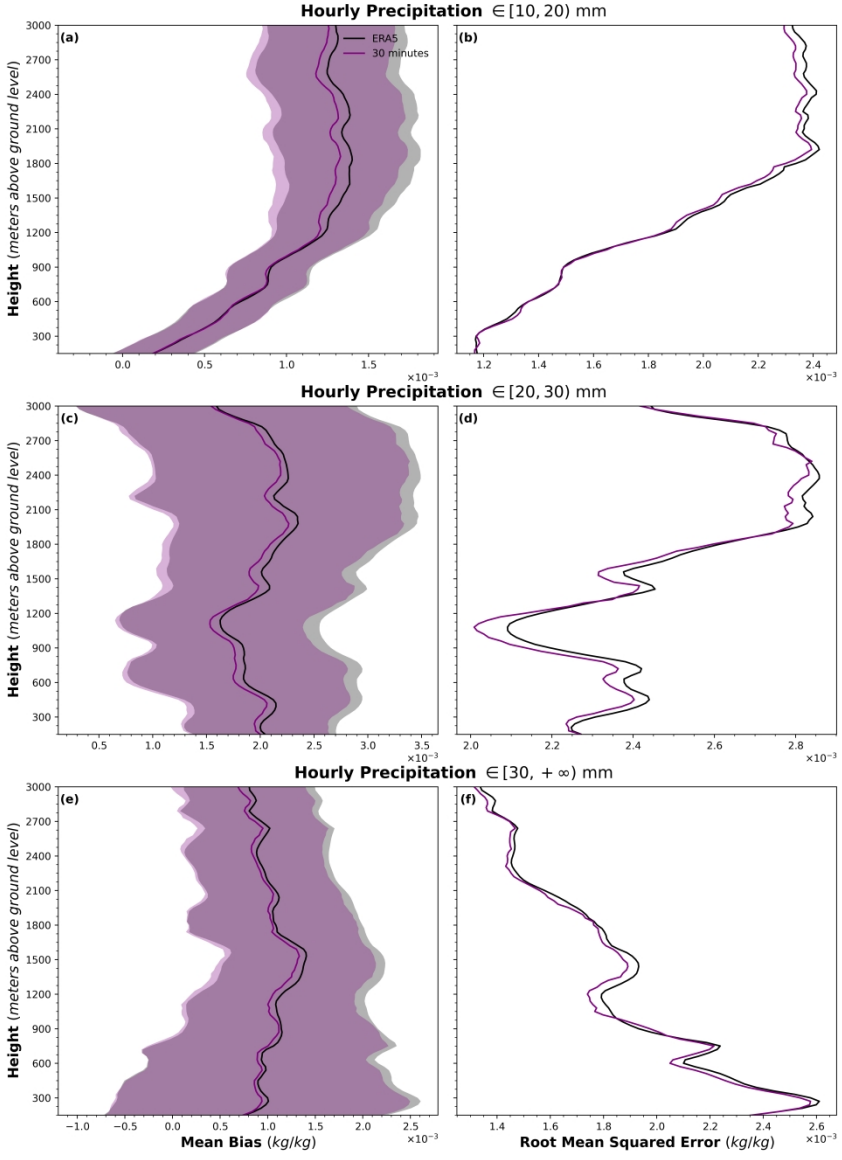

Figure 6: Vertical distributions of mean bias with 95% confidence intervals (a, c, e) and root mean squared error (b, d, f) for water vapor mass mixing ratio profiles derived from the TCKF1D-Var method with a temporal resolution of 30 minutes and from ERA5 prior profiles, stratified by precipitation intensity: (a, b) 10–20 mm, (c, d) 20–30 mm, and (e, f) ≥30 mm. The shaded areas represent the 95% confidence intervals.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Based on the results presented in Figures 4a, 5a, and 6a, c, and e, it is evident that the TCKF1D-Var method, constrained by MRL Raman channel observations, can correct the overall dry bias of the ERA5 prior water vapor mass mixing ratio profiles within the 0 – 3000 meters layer preceding nighttime heavy precipitation events. However, the relationship between the retrieved water vapor mass mixing ratio profile's analysis increment and the lead time relative to precipitation onset remains to be clarified. To investigate this aspect, the time–height evolution of the mean analysis increment (Figure 7), defined as the difference between the TCKF1D-Var retrievals and the ERA5 prior profiles, is calculated using the cases summarized in Table 3. As shown in Figure 7, within the 1200 – 3000 meters above ground level, the mean analysis increment remains consistently positive throughout the 0–9 h period prior to precipitation onset, indicating a persistent moistening correction to the ERA5 prior. Moreover, as the observation time approaches the onset of heavy precipitation, the magnitude of the analysis increment increases progressively, reaching a pronounced maximum within 2 hours before the onset, with values exceeding 1.0×10^{-4} kg/kg. In the 750–1200 meters layer, the temporal evolution of the analysis increment exhibits a more complex structure. The increment is positive at lead times earlier than 7 hours, becomes negative between 7 and 3 hours, and turns positive again during the final 3 hours before precipitation onset. The magnitude in this layer ranges between -0.25×10-4 kg/kg and 0.25×10-4 kg/kg, which is notably smaller than that in the 1200–3000 m layer. below 750 meters, the magnitude of the analysis increment is comparable to that within the 750–1200 m layer, but with a distinct temporal pattern: analysis increment remains positive from 9 to 2 hours before precipitation onset and becomes negative during the last 2 hours. These results indicate that the analysis increments introduced by the TCKF1D-Var method, which reflect the contribution of MRL Raman channel observations, exhibit clear temporal and vertical variability prior to heavy precipitation onset. Both the magnitude and vertical structure of the analysis increments reach their most pronounced state within approximately 2 hours before the onset of heavy precipitation.

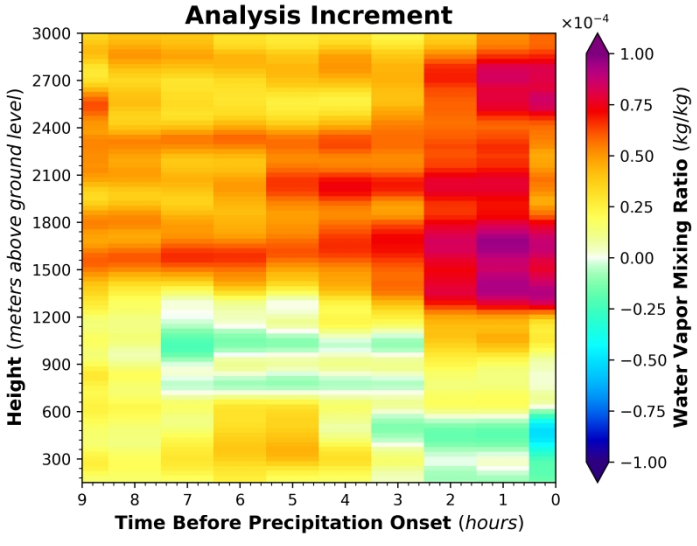

Figure 7: Time – height evolution of the mean analysis increment of water vapor mass mixing ratio profiles, defined as the difference between the TCKF1D-Var retrievals (30 minutes temporal resolution) and the ERA5 prior profiles.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

3.3 Observed nocturnal boundary-layer moisture evolution prior to precipitation onset

Although the statistical results presented in Section 3.2 reveal clear temporal and vertical variability in the analysis increments prior to nighttime heavy precipitation, case-based analyses are still required to verify whether these increments can be identified in individual events. To this end, two representative nocturnal heavy precipitation cases are selected for detailed investigation, corresponding to the maximum (Figure 8) and minimum (Figure 9) hourly rainfall intensities, respectively. The TCKF1D-Var retrieved water vapor mass mixing ratio profiles are compared with the ERA5 prior to further evaluate the performance of the proposed method under contrasting precipitation conditions.

The case with the maximum hourly rainfall intensity occurred at station 59293 at 16:00 UTC on 31 July 2025, with an hourly accumulated precipitation of 73.0 mm. A comparison between the TCKF1D-Var water vapor mass mixing ratio profiles at 30 minutes temporal resolution (Figure 8a) and the ERA5 prior profiles (Figure 8b) reveals clearly identifiable analysis increments, particularly within the $0-1500$ meters layer during the 2 hours preceding the onset of heavy precipitation. Relative to ERA5, the TCKF1D-Var retrievals provide a more detailed representation of the upward transport of water vapor in the 0–750 m layer, while also capturing enhanced temporal variability in the 750–1500 meters layer prior to precipitation onset. Further insight can be obtained from the evolution of the $Innovation$ (Figure 8c) and the reconstructed observations (Figure 8d). As the observation time approaches the onset of heavy precipitation, both the magnitude and vertical variability of the synthetic observations increase markedly, exhibiting a temporal–vertical structure consistent with that shown in Figures 8a and 8b. This consistency indicates that the reconstructed observations, derived from the raw Raman channel measurements of the MRL within the Kalman filtering approach, are sensitive to the evolution of the atmospheric state within the $0-3000$ meters layer prior to nocturnal heavy precipitation. In addition, the agreement between the $Innovation$ patterns and the corresponding structures in Figures 8a and 8b further supports the validity of the observation operator employed in this study.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

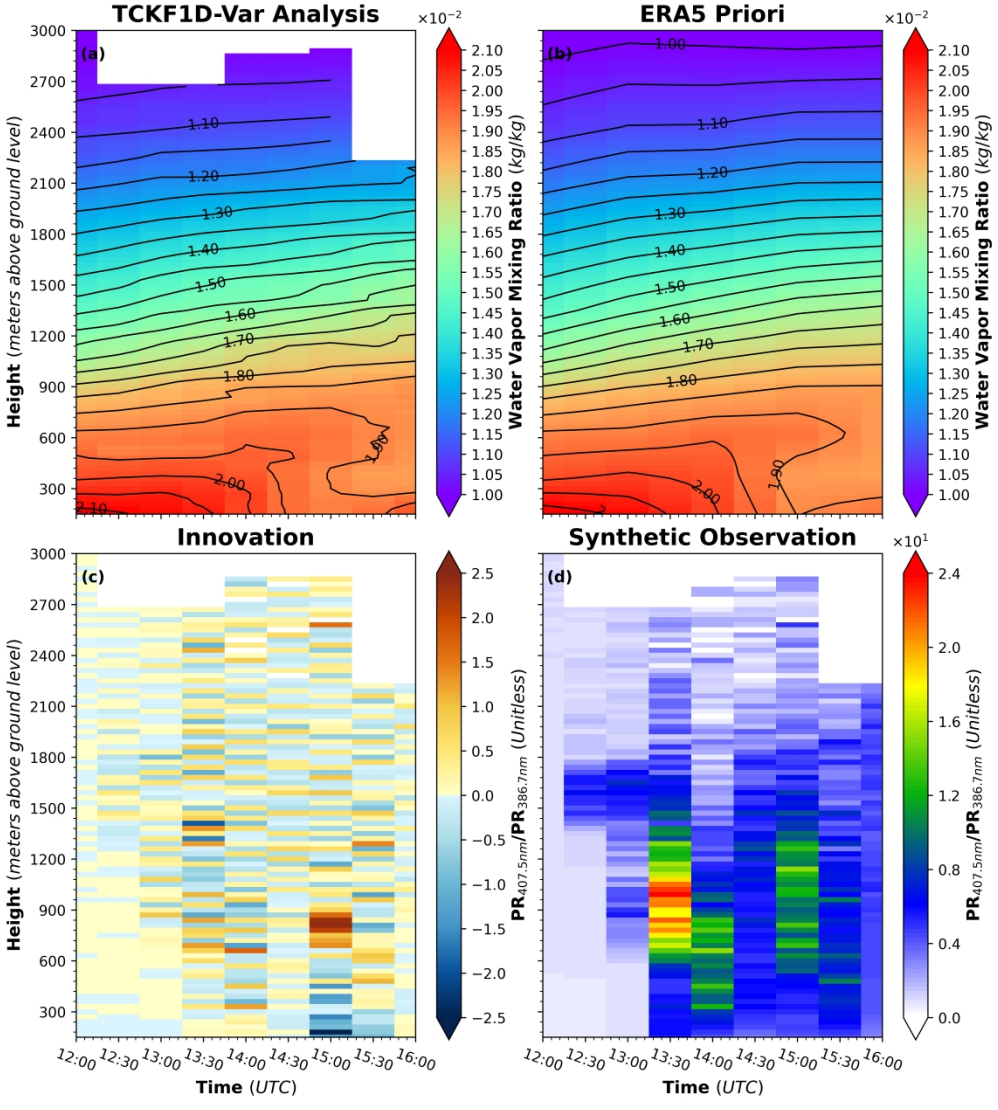

Figure 8: Time–height cross sections of (a) TCKF1D-Var retrieved water vapor mixing ratio at 30 minutes temporal resolution, (b) ERA5 prior water vapor mixing ratio, (c) *Innovation*, and (d) synthetic observations derived from the MRL Raman channel measurements, for the nighttime heavy precipitation case with maximum hourly rainfall intensity at station 59293 (16:00 UTC on 31 July 2025).

In contrast to the case with the maximum hourly rainfall intensity, the case with the minimum hourly rainfall intensity (Figure 9) occurred at station 54218 at 19:00 UTC on 26 July 2025, with an accumulated precipitation of 10.7 mm within one hour. A comparison between the TCKF1D-Var retrieved water vapor mass mixing ratio profiles and the ERA5 prior (Figures 9a and 9b) also reveals clearly identifiable analysis increments. However, these increments are primarily concentrated above 1200 meters and become most pronounced within the 2100–3000 meters layer. This vertical distribution is consistent with the statistical characteristics identified in Section 3.2 (Figure 6a). Similar to the maximum rainfall case, the

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Innovation exhibits a temporal–vertical structure that is consistent with those shown in Figures 9a and 9b. In addition, as the observation time approaches the onset of precipitation, both the magnitude and vertical variability of the synthetic observations (Figure 9d) increase progressively. These features are consistent with those identified in the maximum rainfall case, further demonstrating that the TCKF1D-Var framework, together with the MRL Raman channel observations, is capable of capturing coherent moisture structure variations prior to precipitation under different rainfall intensity conditions.

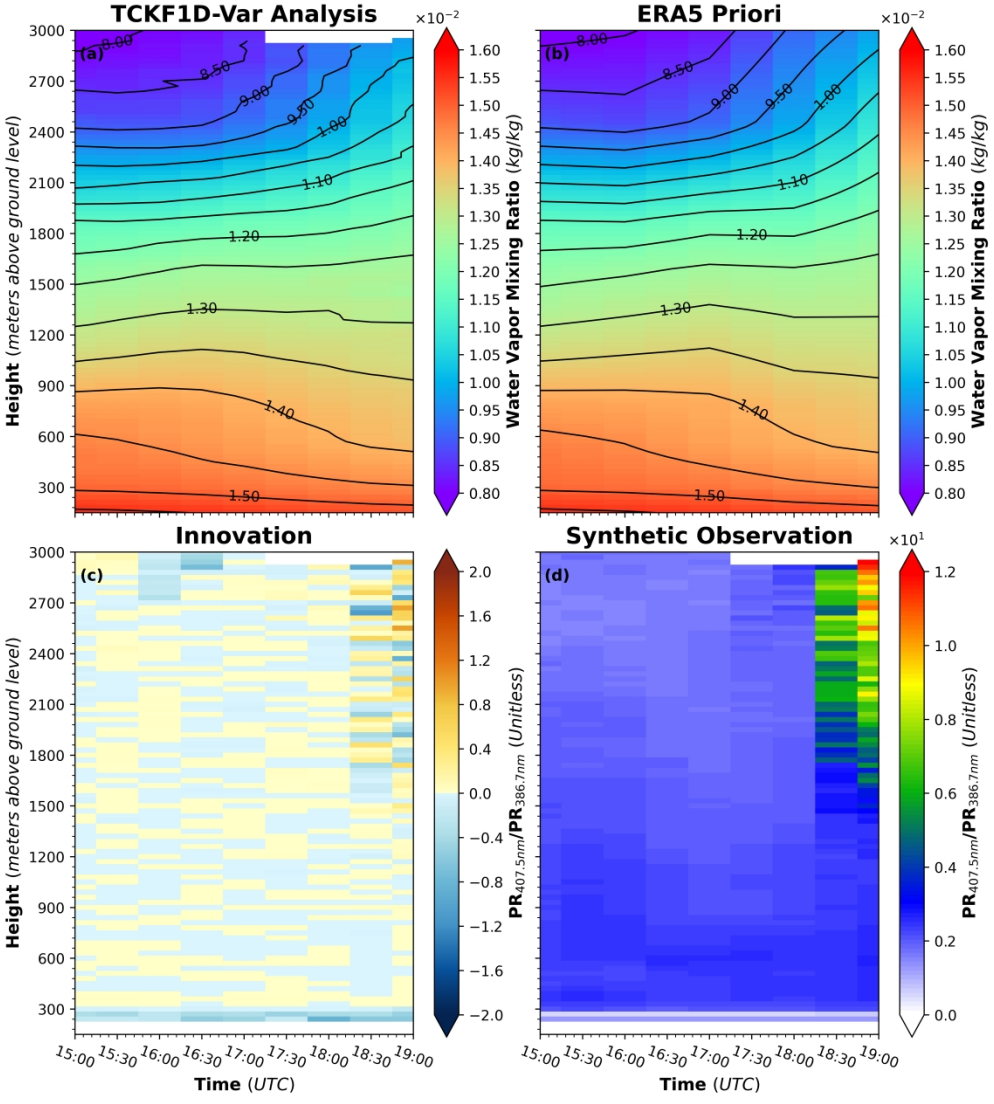

Figure 9: Time–height cross sections of (a) TCKF1D-Var retrieved water vapor mixing ratio at 30 min temporal resolution, (b) ERA5 prior water vapor mixing ratio, (c) *Innovation*, and (d) synthetic observations derived from the MRL Raman channel measurements, for the nighttime heavy precipitation case with minimum hourly rainfall intensity at station 54218 (19:00 UTC on 26 July 2025).

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

4 Summary and concluding remarks

In this study, we extend a previously developed and validated thermodynamic- and cloud-microphysics-constrained Kalman filter one-dimensional variational (TCKF1D-Var) framework to incorporate nitrogen and water vapor Raman channel observations from the Mie–Raman lidar (MRL) network of the China Meteorological Administration. A physics-informed lidar observation operator was developed based on the classical Raman lidar formulation. To account for time-varying instrumental performance and aerosol-related uncertainties, a data-driven calibration component was introduced as a multiplicative correction while preserving the underlying physical structure of the forward model. In addition, a Kalman filter–based observation fusion approach with dynamically estimated process and observation error covariance matrices was implemented to improve retrieval robustness under low signal-to-noise conditions.

The performance of the proposed method was evaluated against co-located radiosonde observations launched prior to nocturnal heavy precipitation events at 56 stations across China in 2025. The results demonstrate that the TCKF1D-Var retrievals, with a vertical resolution of 30 meters and a temporal resolution of 30 minutes, consistently reduce both mean bias and root mean square error relative to the ERA5 prior profiles. The improvement is relatively pronounced within the 1.2–3.0 km layer above ground level. Sensitivity experiments further indicate that the selected 30 minutes temporal window provides a favorable balance between observational information and forward-model uncertainty for the MRL observations.

Application to nocturnal heavy precipitation events shows that the retrieved water vapor mass mixing ratio profiles capture coherent temporal–vertical variations in boundary-layer moisture prior to precipitation onset. In particular, a persistent moistening signal relative to the ERA5 prior is identified within the 1.2–3.0 km layer, with increasing magnitude as the precipitation onset approaches. These results suggest that the integration of MRL Raman channel observations within a physically constrained retrieval framework can provide valuable information on pre-convective moisture evolution.

Despite the encouraging results, several limitations should be noted. First, the data-driven calibration component is designed to represent effective corrections to the observation operator, but it does not explicitly separate instrumental, aerosol, and model-related error sources. Second, the estimation of process and observation error covariance matrices relies on statistical approximations within a finite temporal window and may not fully capture all sources of uncertainty. Third, the retrieval performance is primarily evaluated against radiosonde observations; further assessment using additional independent datasets is needed to better characterize retrieval uncertainties under a wider range of atmospheric conditions. In addition, the choice of temporal resolution involves a trade-off between noise reduction and the ability to resolve rapid atmospheric variability, which may vary under different meteorological conditions.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Future work will focus on further improving the physical representation of the lidar observation operator, particularly through the development of a more explainable data-driven calibration model to better characterize time-varying instrumental and aerosol-related effects. In addition, efforts will be devoted to refining the estimation of process and observation error covariance matrices, with the aim of achieving a more realistic representation of observation and simulation uncertainties associated with boundary-layer turbulence. The extension of the framework to additional observational constraints will also be explored, including the integration of temperature-sensitive measurements such as rotational Raman channels from Raman lidars with atmospheric temperature profiling capability, as well as complementary observations from ground-based hyperspectral infrared sounders. Furthermore, the potential application of the retrieved water vapor mass mixing ratio profiles in data assimilation–numerical weather prediction systems will be investigated, particularly for improving the representation of pre-convective moisture conditions in short-term forecasts and severe weather early warning.

Code and Data availability

The exact version of the code and dataset used to produce the results presented in this study is archived on Zenodo (Zhang et al., 2026b; https://doi.org/10.5281/zenodo.19605193).

Author contributions

Tianmeng Chen prepared the observational data. Qi Zhang developed the TCKF1D-Var framework, performed the coding and data analysis, and drafted the manuscript. Tianmeng Chen revised the manuscript. Jianping Guo supervised the study as the principal investigator. Jianping Guo and Tianmeng Chen jointly secured major funding for this study, while Tianmeng Chen and Qi Zhang each secured additional independent funding from separate sources.

Competing interests

All authors delcare no competing interest.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Acknowledgements

To be added upon acceptance.

Financial support

This manuscript was jointly supported by the Ministry of Science and Technology of China under grant 2024YFC3013001, the National Natural Science Foundation of China (NSFC) under grant 42325501, the Chinese Academy of Meteorological Sciences under grant 2024Z003, and the Heavy Rainfall Research Foundation of China under grant BYKJ2025M24.

References

Ahmed, F., Adames, Á. F., and Neelin, J. D.: Deep convective adjustment of temperature and moisture, J. Atmos. Sci., 77, 2163–2186, https://doi.org/10.1175/JAS-D-19-0227.1, 2020.

Ansmann, A., Riebesell, M., Wandinger, U., et al.: Combined Raman elastic-backscatter lidar for vertical profiling of moisture, aerosol extinction, backscatter, and lidar ratio, Appl. Phys. B, 55, 18–28, https://doi.org/10.1007/BF00348608, 1992.

Behrendt, A., Nakamura, T., Onishi, M., Baumgart, R., and Tsuda, T.: Combined Raman lidar for atmospheric profiling, Appl. Opt., 41, 7657–7666, https://doi.org/10.1364/AO.41.007657, 2002.

Behrendt, A., Pal, S., Aoshima, F., et al.: Observation of convection initiation processes with a suite of state-of-the-art research instruments during COPS IOP 8b, Q. J. R. Meteorol. Soc., 137, 81–100, https://doi.org/10.1002/qj.758, 2011.

Belgiu, M. and Drăguţ, L.: Random forest in remote sensing, ISPRS J. Photogramm. Remote Sens., 114, 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011, 2016.

Cao, X., Guo, Q., Luo, H., Yang, R., Zhang, P., Guo, J., Wang, J., Xiao, D., Du, J., Sun, Z., Liu, S., Chen, S., and Huang, A.: Development and application of the round-trip drifting sounding system (RDSS), EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-2012, 2025.

Chen, Y., Wang, X., Huang, L., and Luo, Y.: Spatial and temporal characteristics of abrupt heavy rainfall events over Southwest China during 1981–2017, Int. J. Climatol., 41, 3286–3299, https://doi.org/10.1002/joc.7019, 2021.

Di Girolamo, P., Cacciani, M., Summa, D., et al.: Characterisation of boundary layer turbulent processes by Raman lidar BASIL, Atmos. Chem. Phys., 17, 745–767, https://doi.org/10.5194/acp-17-745-2017, 2017.

Di Girolamo, P., Franco, N., Di Paolantonio, M., Summa, D., and Dionisi, D.: Micro-pulse Raman lidar MARCO, Sensors, 23, 8262, https://doi.org/10.3390/s23198262, 2023.

Esri, DeLorme, HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, and MapmyIndia: World Imagery,

available at: https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9 (last access: 14 March 2026), 2025.

Filioglou, M., Nikandrova, A., Niemelä, S., et al.: Profiling water vapor mixing ratios in Finland by Raman lidar, Atmos. Meas. Tech., 10, 4303–4316, https://doi.org/10.5194/amt-10-4303-2017, 2017.

Foken, T., Aubinet, M., and Leuning, R.: The eddy covariance method, Springer, Dordrecht, https://doi.org/10.1007/978-94-007-2351-1_1, 2012.

Foth, A. and Pospichal, B.: Optimal estimation of water vapour profiles using Raman lidar and microwave radiometer, Atmos. Meas. Tech., 10, 3325–3344, https://doi.org/10.5194/amt-10-3325-2017, 2017.

Gambacorta, A., et al.: Improved planetary boundary layer sounding using hyperspectral microwave and lidar data fusion, IEEE Trans. Geosci. Remote Sens., 63, 1–23, https://doi.org/10.1109/TGRS.2025.3630972, 2025.

Gao, Z., Li, L., Zhao, C., Gao, G., Jiang, R., Yang, H., Liu, S., and Lu, P.: Impacts of distinct synoptic patterns on fine-scale precipitation characteristics in complex terrain of Southwestern China, Int. J. Climatol., e70315, https://doi.org/10.1002/joc.70315, 2026.

Gerber, F. and Furrer, R.: optimParallel: An R package providing a parallel version of the L-BFGS-B optimization method, R J., https://doi.org/10.32614/RJ-2019-030, 2019.

Hersbach, H., Bell, B., Berrisford, P., et al.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Kalman, R. E.: A new approach to linear filtering and prediction problems, Transactions of the ASME–Journal of Basic Engineering, 82, 35–45, https://doi.org/10.1115/1.3662552, 1960.

Kirshbaum, D. J., Adler, B., Kalthoff, N., Barthlott, C., and Serafin, S.: Moist orographic convection: Physical mechanisms and links to surface-exchange processes, Atmosphere, 9, 80, https://doi.org/10.3390/atmos9030080, 2018.

Kirshbaum, D. J.: Cloud-resolving simulations of deep convection over a heated mountain, J. Atmos. Sci., 68, 361–378, https://doi.org/10.1175/2010JAS3642.1, 2011.

Laly, F. and Chazette, P.: Comparative analysis of ERA5 and Raman lidar-derived moisture profiles in the framework of the WaLiNeAs field campaigns, Q. J. Roy. Meteor. Soc., 151, e5044, https://doi.org/10.1002/qj.5044, 2025.

Laly, F., Chazette, P., Totems, J., Lagarrigue, J., Forges, L., and Flamant, C.: Water vapor Raman lidar observations from multiple sites in the framework of WaLiNeAs, Earth Syst. Sci. Data, 16, 5579–5602, https://doi.org/10.5194/essd-16-5579-2024, 2024.

Lange, D., Behrendt, A., and Wulfmeyer, V.: Compact operational tropospheric Raman lidar, Geophys. Res. Lett., 46, 14844–14853, https://doi.org/10.1029/2019GL085774, 2019.

Lange, D., Behrendt, A., Senff, C., et al.: Water–vapor budget investigation using ground-based lidar, Bull. Atmos. Sci. Technol., 6, 24, https://doi.org/10.1007/s42865-025-00110-4, 2025.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Li, N., Guo, J., Guo, X., Chen, T., Zhang, Z., Tang, N., Wang, Y., Yang, H., Zheng, Y., and Zhou, Y.: On the nationwide variability of low-level jets prior to warm-season nocturnal rainfall in China revealed by radar wind profilers, Atmos. Chem. Phys., 26, 3339–3356, https://doi.org/10.5194/acp-26-3339-2026, 2026.

Lu, Y., Marsham, J. H., Parker, D. J., Klein, C. M., Taylor, C. M., Fang, J., and Tang, J.: Role of soil moisture gradients in favoring mesoscale convective systems in East China, Geophys. Res. Lett., 52, e2025GL117137, https://doi.org/10.1029/2025GL117137, 2025.

Luo, Y. L., Sun, J. S., Li, Y., et al.: Science and prediction of heavy rainfall over China: Research progress since the reform and opening-up of new China, J. Meteorol. Res., 34, 427–459, https://doi.org/10.1007/s13351-020-0006-x, 2020.

Luo, Y., Wu, M., Ren, F., Li, J., and Wong, W.: Synoptic situations of extreme hourly precipitation over China, J. Climate, 29, 8703–8719, https://doi.org/10.1175/JCLI-D-16-0057.1, 2016.

Mayer, S., Sandvik, A., Jonassen, M. O., et al.: Atmospheric profiling with the UAS SUMO, Meteorol. Atmos. Phys., 116, 15–26, https://doi.org/10.1007/s00703-010-0063-2, 2012.

Oue, M., Saleeby, S. M., Marinescu, P. J., Kollias, P., and van den Heever, S. C.: Optimizing radar scan strategies for tracking isolated deep convection, Atmos. Meas. Tech., 15, 4931–4950, https://doi.org/10.5194/amt-15-4931-2022, 2022.

Richardson, M. T., Kahn, B. H., and Kalmus, P. M.: Mesoscale air motion and thermodynamics predict heavy hourly U.S. precipitation, Commun. Earth Environ., 5, 472, https://doi.org/10.1038/s43247-024-01614-1, 2024.

Richter, J. H., Joseph, E., Arcodia, M. C., et al.: Earth system predictability across time scales for a resilient society, Bull. Amer. Meteorol. Soc., 107, E326–E351, https://doi.org/10.1175/BAMS-D-24-0155.1, 2026.

Shao, N., Wang, Q., Bu, Z., Yin, Z., Dai, Y., Chen, Y., and Wang, X.: China aerosol Raman lidar network (CARLNET), Remote Sens., 17, 414, https://doi.org/10.3390/rs17030414, 2025.

Sun, X., Yang, Z., and Niyogi, D.: Diurnal urban rainfall anomalies across different landscapes, Sci. Adv., 11, eads5046, https://doi.org/10.1126/sciadv.ads5046, 2025.

Vaughan, G., Wareing, D. P., Thomas, L., and Mitev, V.: Humidity measurements in the free troposphere using Raman backscatter, Q. J. R. Meteorol. Soc., 114, 1471–1484, https://doi.org/10.1002/qj.49711448406, 1988.

Wandinger, U.: Raman lidar, in: Lidar, Springer, New York, https://doi.org/10.1007/0-387-25101-4_9, 2005.

Whiteman, D. N., Cadirola, M., Venable, D., Calhoun, M., Miloshevich, L., Vermeesch, K., Twigg, L., Dirisu, A., Hurst, D., Hall, E., Jordan, A., and Vömel, H.: Correction technique for Raman water vapor lidar signal-dependent bias and suitability for water vapor trend monitoring in the upper troposphere, Atmos. Meas. Tech., 5, 2893–2916, https://doi.org/10.5194/amt-5-2893-2012, 2012.

Whiteman, D. N., Demoz, B., Rush, K., Schwemmer, G., Gentry, B., Di Girolamo, P., Comer, J., Veselovskii, I., Evans, K., Melfi, S. H., Wang, Z., Cadirola, M., Mielke, B., Venable, D., and Van Hove, T.: Raman lidar measurements during the International H2O Project. Part I: Instrumentation and analysis techniques, J. Atmos. Ocean. Technol., 23, 157–169, https://doi.org/10.1175/JTECH1838.1, 2006.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Whiteman, D. N., Melfi, S. H., and Ferrare, R. A.: Raman lidar system for the measurement of water vapor and aerosols in the Earth's atmosphere, Appl. Opt., 31, 3068–3082, https://doi.org/10.1364/AO.31.003068, 1992.

Whiteman, D. N., Rush, K., Rabenhorst, S., et al.: Airborne and ground-based measurements using a high-performance Raman lidar, J. Atmos. Ocean. Technol., 27, 1781–1801, https://doi.org/10.1175/2010JTECHA1391.1, 2010.

Wulfmeyer, V. and Behrendt, A.: Raman lidar for water vapor and temperature profiling, in: Springer Handbook of Atmospheric Measurements, edited by: Foken, T., Springer, Cham, https://doi.org/10.1007/978-3-030-52171-4_25, 2021.

Wulfmeyer, V., Behrendt, A., Bauer, H.-S., Kottmeier, C., Corsmeier, U., et al.: The convective and orographically induced precipitation study: A research and development project of the World Weather Research Program for improving quantitative precipitation forecasting in low-mountain regions, Bull. Am. Meteorol. Soc., 89, 1477–1486, https://doi.org/10.1175/2008BAMS2367.1, 2008.

Wulfmeyer, V., Behrendt, A., Kottmeier, C., et al.: The convective and orographically induced precipitation study (COPS): Scientific strategy and highlights, Q. J. R. Meteorol. Soc., 137, 3–30, https://doi.org/10.1002/qj.752, 2011.

Wulfmeyer, V., Hardesty, R. M., Turner, D. D., Behrendt, A., Cadeddu, M. P., Di Girolamo, P., Schlüssel, P., Van Baelen, J., and 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., 53, 819–895, https://doi.org/10.1002/2014RG000476, 2015.

Wulfmeyer, V., Hardesty, R. M., Turner, D. D., et al.: Review of remote sensing of lower tropospheric thermodynamic profiles, Rev. Geophys., 53, 819–895, https://doi.org/10.1002/2014RG000476, 2015.

Wulfmeyer, V., Pal, S., Turner, D. D., et al.: Can water vapour Raman lidar resolve turbulent variables?, Boundary-Layer Meteorol., 136, 253–284, https://doi.org/10.1007/s10546-010-9494-z, 2010.

Yao, L., Shen, D., Sun, X., Wang, D., Cao, X., Wang, J., Wang, D., Zhang, C., and Guo, Q.: The Beidou navigation radiosonde observation experiment and data evaluation, SSRN [preprint], https://doi.org/10.2139/ssrn.5085235, 2025.

Zhang, M., Li, J., Li, N., et al.: Spatiotemporal characteristics and associated circulation patterns of warm-season precipitation in a complex terrain region of Southwest China, Clim. Dyn., 63, 103, https://doi.org/10.1007/s00382-024-07581-9, 2025a.

Zhang, P., Li, R., Zhao, K., Wang, D., Wang, J., Lei, Y., Xia, X., Xian, D., Chen, Y., Wu, L., Zhao, P., Guo, Q., Yang, R., Li, C., Yan, P., Liu, C., Wang, X., Gui, H., Liu, L., Guan, M., Huang, H., He, J., Liu, L., Wang, B., Sun, Y., Wang, Y., Huang, Y., Hu, W., Li, B., Wang, Z., Ma, Y., Liu, J., and Fu, Y.: Development of an integrated meteorological observation system in China, Acta Meteorol. Sin., 83, 729–760, https://doi.org/10.11676/qxxb2025.20240193, 2025c.

Zhang, Q., Chen, T., Guo, J., Wu, Y., Deng, B., and Yan, J.: Retrieving thermodynamic profiles using TCKF1D-Var, Geosci. Model Dev., 19, 505–522, https://doi.org/10.5194/gmd-19-505-2026, 2026a.

Zhang, Q., Chen, T., and Guo, J.: Nocturnal boundary layer Mie–Raman lidar water vapor profiles retrieved by extended TCKF1D-Var framework, Zenodo [data set], https://doi.org/10.5281/zenodo.19605193, 2026b.

https://doi.org/10.5194/egusphere-2026-2184
Preprint. Discussion started: 23 April 2026

Zhang, Q., Deng, B., Wang, S., Dong, F., and Shao, M.: Multi-source retrieval of thermodynamic profiles using EnKF1D-Var, Remote Sens., 17, 3133, https://doi.org/10.3390/rs17183133, 2025b.

Ångström, A.: On the atmospheric transmission of sun radiation and on dust in the air, Geogr. Ann., 11, 156–166, https://doi.org/10.1080/20014422.1929.11880498, 1929.