



1 **Regime-aware retrieval of lower-tropospheric stability and a sensitivity-weighted**
2 **diagnostic of cross-site transferability over heterogeneous land surfaces**
3 Yufei Chu^{1,2}, Jun A. Zhang^{3,4*}, Guo Lin⁵, Hanqin Guo⁶, Lulin Xue⁷, Weiwei Li⁷, Hyeyum Hailey Shin⁷,
4 Min Deng⁸, Lan Jin⁹, and Flynn McGinnity²
5 ¹ Electrical and Computer Engineering, Rice University, Houston, TX, USA
6 ² School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, USA
7 ³ NOAA/AOML/Hurricane Research Division, Miami, FL, USA
8 ⁴ Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL, USA
9 ⁵ GFZ Helmholtz Centre for Geosciences, Telegrafenberg, Potsdam, Germany
10 ⁶ Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA
11 ⁷ NSF National Center for Atmospheric Research, Boulder, CO, USA
12 ⁸ Environmental Science and Technologies Department, Brookhaven National Laboratory, Upton, NY,
13 USA
14 ⁹ Department of Nuclear Engineering and Radiological Sciences, University of Michigan, Ann Arbor, MI,
15 USA
16 *Correspondence to: Jun A. Zhang (jun.zhang@noaa.gov)
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33



34

Abstract

35 Lower-tropospheric stability (LTS) regulates convective boundary-layer growth, low-cloud evolution,
 36 pollutant dispersion, and land–atmosphere coupling, yet continuous observations are scarce because
 37 thermodynamic profilers are sparse and expensive, and reanalysis products explain less than one-third of
 38 the observed variance in inversion strength even at well-instrumented sites. We retrieve continuous LTS at
 39 the Atmospheric Radiation Measurement Southern Great Plains observatory from routine surface-flux
 40 measurements and Doppler-lidar-derived mixed-layer height, using a regime-aware, stage-partitioned
 41 framework established in earlier work. At the central facility over May 2016–December 2020 the retrieval
 42 reproduces observed LTS with a coefficient of determination of 0.78 and a mean absolute error of 2.39 K,
 43 and captures the full diurnal evolution of stability. The principal contribution is a diagnostic for judging
 44 when such a retrieval can be transferred to another site. Predictor-distribution differences, quantified by the
 45 Kullback–Leibler divergence, are weighted by SHapley Additive exPlanations sensitivities to form a single
 46 metric, $\Delta\text{KL} \times |\text{SHAP}|$. Applied to two extended facilities, the metric shows that distributional differences
 47 degrade transferability only where the retrieval is sensitive to the shifted predictor, and identifies cloud-
 48 base height, mixed-layer height, and thermodynamic predictors as the dominant contributors at the poorer-
 49 performing site. Gaussian perturbation experiments provide an independent fingerprint: monotonic error
 50 growth indicates a transferable relationship limited by random observational uncertainty, whereas a non-
 51 monotonic response indicates structural thermodynamic mismatch. Requiring only predictor distributions
 52 and model sensitivities, the diagnostic applies to any data-driven atmospheric retrieval.

53 Short summary

54 Atmospheric stability near the ground controls clouds, pollution and mixing, but is rarely measured
 55 continuously. We show it can be estimated from routine surface and boundary-layer measurements, and —
 56 more importantly — we introduce a way to tell in advance whether such an estimate will still work at a
 57 different location. The method identifies which measurements are responsible when it fails, and whether
 58 the failure is due to random noise or a genuine difference in local conditions.

59 Hight light

- 60 1. Continuous lower-tropospheric stability retrieved from routine surface observations
- 61 2. $\Delta\text{KL} \times |\text{SHAP}|$ metric attributes cross-site transferability loss to predictors
- 62 3. Distribution shifts degrade skill only in predictors the model is sensitive to
- 63 4. Non-monotonic noise response flags structural mismatch, not random error

64 **Keywords:** Lower-tropospheric stability; Boundary-layer retrieval; Transferability; Covariate shift;
 65 Kullback–Leibler divergence; SHAP; Doppler lidar

66

67

68

69



70 1 Introduction

71 The planetary boundary layer (PBL) governs the exchange of heat, moisture, momentum, aerosols, and
 72 trace gases between the Earth's surface and the free troposphere and therefore plays a central role in weather,
 73 cloud evolution, air quality, and climate (Stull, 1988; Garratt, 1994). The thermodynamic structure of the
 74 lower troposphere strongly modulates convective boundary-layer growth, turbulent entrainment, cloud
 75 development, and pollutant dispersion (Bretherton and Wyant, 1997; Lothon et al., 2014; Su et al., 2018;
 76 Lin et al., 2023; Lin et al., 2024; Deng et al., 2025). Among the variables commonly used to characterize
 77 this structure, lower-tropospheric stability (LTS), typically defined as the potential-temperature difference
 78 between the surface and 700 hPa, provides a useful bulk measure of the resistance to vertical mixing (Klein
 79 and Hartmann, 1993; Wood and Bretherton, 2006). Strong LTS suppresses turbulent exchange and weakens
 80 coupling between the surface and free troposphere, whereas weak LTS promotes deeper convective mixing
 81 and stronger vertical transport of heat, moisture, and pollutants.

82 LTS has broad relevance across boundary-layer meteorology, air quality, cloud physics, and climate
 83 research. Over subtropical oceans, LTS and related inversion-strength metrics strongly influence marine
 84 low-cloud fraction and cloud radiative feedback (Klein and Hartmann, 1993; Wood and Bretherton, 2006;
 85 Qu et al., 2015). Over land, enhanced lower-tropospheric stability suppresses vertical dispersion and
 86 contributes to severe haze and PM_{2.5} accumulation episodes, particularly under weakly convective
 87 conditions (Li et al., 2017; Che and Zhao, 2021). LTS also affects entrainment processes, shallow-cloud
 88 evolution, and the transition between convective and stratified boundary-layer regimes (Park and Bretherton,
 89 2009). In numerical weather prediction and climate models, biases in LTS can propagate into errors in
 90 turbulent mixing, cloud radiative forcing, and land-atmosphere coupling (Holtslag et al., 2013; Shin et al.,
 91 2021; Li et al., 2025). Consequently, LTS is increasingly used as a diagnostic variable for evaluating PBL
 92 parameterizations and atmospheric model performance.

93 Despite its importance, continuous observations of LTS remain limited. Radiosondes provide accurate
 94 thermodynamic profiles but are typically launched only a few times per day and therefore do not fully
 95 resolve the rapid transitions associated with morning boundary-layer growth, afternoon convective mixing,
 96 and evening stabilization. Advanced remote-sensing systems, including Atmospheric Emitted Radiance
 97 Interferometers (AERIs), microwave radiometers, and Raman lidars, can provide continuous
 98 thermodynamic profiling, but these systems remain relatively expensive and geographically sparse (Turner
 99 and Löhnert, 2014; Wulfmeyer et al., 2015). Reanalysis and numerical-model products provide continuous
 100 coverage but often underestimate inversion sharpness and boundary-layer heterogeneity over land,
 101 particularly under stable conditions and heterogeneous surface forcing (Hersbach et al., 2020). There
 102 therefore remains a need for observationally constrained approaches capable of extending continuous LTS
 103 monitoring beyond locations equipped with dedicated thermodynamic profilers. These limitations are
 104 quantitative as well as practical. Even at the ARM SGP observatory, one of the most densely instrumented
 105 sites in the world, LTS derived from ERA5 explains only about 28 % of the variance in the inversion
 106 strength measured by high-resolution radiosondes, with a regression slope of approximately 0.26 K K⁻¹
 107 (Wang et al., 2023), reflecting the fact that the vertical spacing of reanalysis levels is generally coarser than
 108 the inversion layer itself. Reanalysis products therefore cannot substitute for observationally constrained,
 109 continuous estimates of LTS.

110 At the same time, long-term surface-flux measurements and Doppler-lidar observations are becoming
 111 increasingly available at atmospheric observatories and field campaigns. Surface sensible and latent heat
 112 fluxes, turbulent statistics, and boundary-layer depth collectively reflect the evolving coupling between the
 113 land surface and lower atmosphere. Previous studies have demonstrated that data-driven approaches can
 114 successfully retrieve boundary-layer quantities such as mixed-layer height (MLH) and turbulence
 115 characteristics from remote-sensing and surface observations (Dang et al., 2019; Krishnamurthy et al., 2021;
 116 Rieutord et al., 2021; Chu et al., 2022b; Zhang et al., 2025). However, most existing work has focused
 117 primarily on MLH retrieval itself, while comparatively little attention has been given to continuous retrieval
 118 of LTS or to the transferability of such retrievals across heterogeneous environments. The scale of this



observational asymmetry is substantial. ARM operated 15 coherent Doppler lidars across its sites as of early 2023, including a network of five instruments at SGP alone — one at each of the extended facilities E32, E37, E39 and E41, and the central facility (Newsom and Krishnamurthy, 2022) — while eddy-covariance flux systems have been deployed at seven SGP extended facilities, with records extending back to 2003 (Sullivan et al., 2025). Continuous AERI-based thermodynamic profiling is far more restricted: long-term AERI records exist at only a small number of ARM observatories (Gero and Hackel, 2025), and within the SGP domain the extended-facility retrievals used here span a substantially shorter window than the corresponding surface-flux and Doppler-lidar records. A retrieval that reconstructs LTS from surface and boundary-layer observations therefore has the potential to extend continuous thermodynamic monitoring both spatially, to facilities without a profiler, and temporally, to periods when profiling instruments were not operating.

The relationship between surface observations and LTS is also fundamentally nonstationary. The convective boundary layer evolves through distinct diurnal regimes, including the morning transition, rapid-growth period, mature convective stage, and evening decay stage, each governed by different balances among surface heating, entrainment, turbulent transport, and radiative cooling (Sun et al., 2012; Salesky and Anderson, 2018). As a result, the statistical relationship between surface forcing and lower-tropospheric thermodynamic structure changes substantially throughout the day. Retrieval frameworks that ignore this regime dependence may average across physically distinct atmospheric states and thereby reduce both interpretability and predictive skill. Importantly, the surface and boundary-layer variables used here should not be interpreted as direct physical drivers of LTS. Rather, they represent observational indicators of the evolving thermodynamic state of the lower atmosphere and therefore provide an observationally constrained proxy for LTS variability.

This study uses observations from the U.S. Department of Energy ARM SGP observatory to address two connected questions. The first is whether continuous LTS variability can be reconstructed from routinely available surface-flux and boundary-layer observations. For this we adopt the regime-aware, stage-partitioned retrieval framework developed and validated in our earlier work (Chu et al., 2025a, 2026a, 2026c); that framework is treated here as an established tool, and its formulation and evaluation are not repeated. The second question, which is the principal focus of the present study, is how far such a retrieval can be transferred to nearby sites with different land-surface and thermodynamic characteristics, and how the sources of transferability loss can be diagnosed physically rather than treated as unexplained statistical error. Observations from the central facility (C1) and two extended facilities (E37 and E39) during May 2016–December 2020 are used throughout.

To address the second question we introduce a sensitivity-weighted representativeness diagnostic that combines predictor-distribution differences, quantified using the Kullback–Leibler (KL) divergence, with SHapley Additive exPlanations (SHAP)-based predictor sensitivity. The resulting metric, $\Delta\text{KL} \times |\text{SHAP}|$, isolates predictors that are simultaneously shifted between sites and influential for the retrieval, and therefore separates benign distributional variability from structurally important thermodynamic mismatch. Gaussian perturbation experiments are then used as an independent test that distinguishes random observational uncertainty from covariate shift. The objectives of this study are therefore to (1) demonstrate that continuous LTS can be retrieved from routine surface-flux and boundary-layer observations at the SGP central facility, (2) quantify how retrieval skill degrades when the retrieval is applied to two extended facilities within the same observational domain, and (3) introduce and demonstrate a physically interpretable diagnostic framework that identifies the dominant sources of cross-site transferability degradation and distinguishes them from random observational error. Objective (3) constitutes the principal methodological contribution and is intended to be applicable well beyond LTS.



2 Data and methods

2.1 ARM Southern Great Plains observational framework

This study uses observations collected at the ARM SGP observatory in north-central Oklahoma. The SGP site is one of the world's most extensively instrumented atmospheric observatories and provides long-term measurements of land-atmosphere interactions, boundary-layer structure, cloud processes, and surface energy exchange over heterogeneous continental terrain (Sisterson et al., 2016). The region contains substantial spatial variability in land cover, soil moisture, vegetation, and surface energy partitioning, making it well suited for investigating the transferability of LTS retrievals across different boundary-layer regimes.

Figure 1 summarizes the observational framework and the relative locations of the C1, E37, and E39 facilities within the SGP domain. The analysis focuses on the ARM central facility (C1; 36.61° N, 97.49° W) and two extended facilities, E37 (36.31° N, 97.93° W) and E39 (36.37° N, 97.07° W) (Fig. 1). The C1 site contains the most comprehensive instrumentation suite and is used as the primary training location for the retrieval framework, while E37 and E39 are used to evaluate spatial transferability within the SGP domain. Although all three sites are located within approximately 100 km of one another, they sample different local surface and thermodynamic environments associated with variations in vegetation cover, soil properties, and land-atmosphere coupling. The study period extends from May 2016 through December 2020 and encompasses multiple seasonal cycles and a wide range of synoptic and convective conditions.

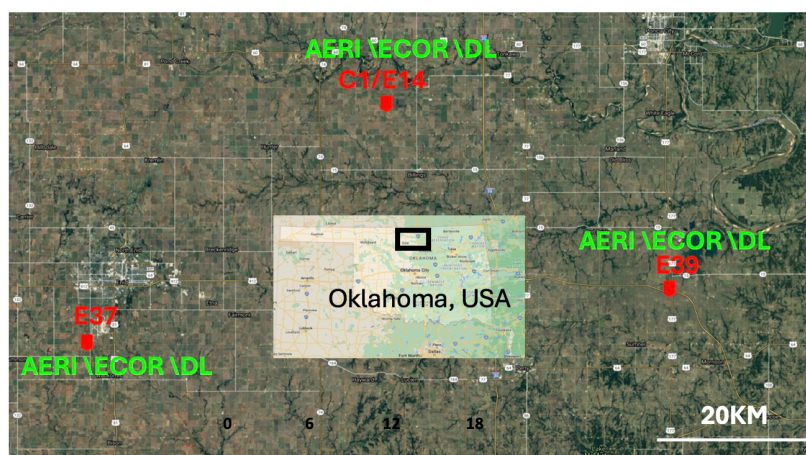


Figure 1. Locations of the ARM Southern Great Plains (SGP) central facility (C1) and extended facilities E37 and E39 used in this study (Imagery © 2026 NASA, Map data © 2026 Google). Surface-flux observations, Doppler-lidar-derived mixed-layer height (MLH), and Atmospheric Emitted Radiance Interferometer (AERI) thermodynamic profiles are combined to develop and evaluate the lower-tropospheric stability (LTS) retrieval framework. C1 is used for model training, while E37 and E39 are used for cross-site transferability evaluation.

Surface turbulent-flux measurements are obtained from the ARM Eddy Correlation Flux Measurement (ECOR) systems, which provide sensible heat flux, latent heat flux, friction velocity, and turbulence statistics at 30 min temporal resolution. The ECOR systems use the eddy-covariance technique to estimate turbulent exchange by correlating high-frequency fluctuations of vertical velocity with simultaneous fluctuations in temperature and water-vapor concentration (Cook, 2018). Standard ARM quality-control procedures, including coordinate rotation, despiking, and signal screening, are applied before use.



Boundary-layer structure is characterized using ARM Doppler lidars operating at 1.5 μm wavelength. The Doppler lidars provide continuous vertical profiles of radial velocity and attenuated backscatter and are widely used for monitoring boundary-layer turbulence and mixed-layer evolution (Manninen et al., 2018). The MLH is derived from the Doppler-lidar observations using the wavelet-covariance approach described by Chu et al. (2022a, 2022b). The method first identifies the dominant turbulent-eddy scales using wavelet analysis and subsequently determines the vertical-velocity-variance structure associated with the convective mixed layer. Dynamic thresholds are then applied to reconstruct the evolving variance field and estimate MLH. Compared with fixed-threshold approaches, this method better accounts for changing turbulent scales and reduces contamination from gravity waves and intermittent turbulence (Chu and Wang, 2025).

Thermodynamic profiles are obtained from the AERI, a ground-based Fourier-transform infrared spectrometer that continuously measures downwelling infrared radiance with high spectral resolution and radiometric accuracy (Turner, 2007). The AERI retrievals provide high-temporal-resolution temperature and moisture profiles in the lower troposphere and are used here to derive the reference LTS field. Radiosonde observations at the C1 site are additionally used to validate the AERI-derived LTS estimates.

2.2 Lower-tropospheric stability and observational derivation

The LTS is used in this study as a bulk thermodynamic measure of the resistance to vertical mixing within the lower atmosphere. Following Klein and Hartmann (1993), LTS is defined as

$$LTS = \theta_{700 \text{ hPa}} - \theta_{\text{surface}} \quad (1)$$

where $\theta_{700 \text{ hPa}}$ is the potential temperature at 700 hPa and θ_{surface} is the surface potential temperature. Positive LTS values indicate a more stable lower troposphere and reduced vertical coupling between the surface and free troposphere.

Continuous LTS observations are derived from AERI temperature profiles combined with surface thermodynamic measurements. Radiosondes launched at the C1 site are used to evaluate the quality of the AERI-based retrievals. Using 1245 collocated samples, the AERI-derived LTS exhibits strong agreement with radiosonde estimates, with a coefficient of determination of $R^2 = 0.96$, a root-mean-square error of 1.76 K, and a mean bias of 0.21 K. These results indicate that the AERI observations provide a reliable high-temporal-resolution reference dataset for evaluating the retrieval framework (Chu et al., 2026b).

2.3 Predictor selection and physical rationale

It is important to emphasize that the predictors used in this study are not interpreted as direct physical causes of LTS. LTS is fundamentally controlled by the thermodynamic structure of the lower troposphere, whereas surface-flux variables and MLH represent correlated indicators of the evolving boundary-layer state. The retrieval framework therefore aims to reconstruct LTS statistically from observationally related quantities while maintaining physical interpretability.

Fourteen predictor variables are used to characterize the thermodynamic and dynamical state of the boundary layer. These variables are grouped into four physically motivated categories: thermodynamic forcing, boundary-layer structure, turbulent dynamics, and temporal context. All predictors are averaged to hourly resolution to reduce high-frequency noise and to better represent the evolving boundary-layer state.

The thermodynamic-forcing variables include instantaneous sensible heat flux (I_SHF), instantaneous latent heat flux (I_LHF), cumulative sensible heat flux (C_SHF), and cumulative latent heat flux (C_LHF). The cumulative fluxes are computed by integrating the corresponding surface fluxes beginning at local sunrise until a specific time of day. These cumulative quantities represent the thermodynamic memory of the boundary layer and account for the progressive accumulation of surface energy input during the day.



241 Boundary-layer structure is represented primarily by MLH derived from the Doppler lidar and by cloud-
242 base height (CBH). MLH characterizes the geometric depth of the turbulent layer, while CBH provides
243 information about cloud development and moisture structure near the inversion layer.

244 Turbulent-dynamical predictors include friction velocity (u_*), turbulent kinetic energy (TKE), and vertical-
245 velocity skewness. These variables describe the intensity and asymmetry of turbulent motions within the
246 boundary layer and help characterize the transition between stable, weakly convective, and fully convective
247 regimes.

248 Finally, temporal context is represented using a normalized solar-time variable (SUNPERCENT), defined
249 as the fractional time between local sunrise and sunset. This variable provides information about the
250 evolving phase of the diurnal cycle and helps distinguish among boundary-layer stages with different
251 physical controls.

252 The selected predictors are not intended to provide a complete physical description of the lower troposphere.
253 Instead, they represent a set of routinely available observational quantities that collectively encode
254 information about the thermodynamic evolution of the boundary layer and its coupling to the lower
255 atmosphere.

256 **2.4 Regime-aware retrieval framework**

257 The relationship between surface observations and lower-tropospheric stability evolves substantially
258 throughout the diurnal cycle because the convective boundary layer passes through distinct physical
259 regimes (Acevedo et al., 2016). To account for this non stationarity, we employ a staged retrieval framework
260 in which separate machine-learning models are trained for four characteristic boundary-layer stages: (1)
261 morning transition, (2) rapid-growth period, (3) mature convective stage, and (4) evening decay stage.

262 The morning-transition stage is characterized by erosion of the nocturnal stable layer through surface
263 heating and turbulent entrainment. During the rapid-growth stage, mixed-layer depth increases quickly as
264 the residual layer becomes incorporated into the growing convective boundary layer. The mature convective
265 stage corresponds to the period of maximum daytime mixing and deepest MLH, while the evening-decay
266 stage represents the collapse of turbulence and redevelopment of stable stratification after sunset.

267 Separate models are trained for each stage using the Tree-based Pipeline Optimization Tool (TPOT), an
268 automated machine-learning framework based primarily on tree-based ensemble methods and linear
269 regressors. The formulation, hyperparameter treatment, and evaluation of this stage-partitioned framework
270 are described in detail by Chu et al. (2025a, 2026a, 2026c) and are not repeated here; the framework is used
271 in the present study as an established retrieval tool rather than as a contribution. We note explicitly that the
272 framework is not physics-informed in the strict sense used in the machine-learning literature: no physical
273 constraint is imposed on the loss function or on the model architecture. Physical knowledge instead enters
274 through the grouping of predictors into physically motivated categories and through the partitioning of the
275 retrieval by boundary-layer regime, so that each sub-model operates within a physically coherent state space.
276 The design is therefore inspired by, rather than an instance of, physics-informed learning.

277 To improve physical continuity across time, the predicted LTS from the previous time step is recursively
278 included as an additional predictor for the subsequent step. Model interpretability is evaluated using
279 SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017), which quantify the relative contribution
280 of each predictor to the retrievals and allow the dominant controls on LTS to be examined across seasons
281 and boundary-layer stages.

282 **2.5 Sensitivity-weighted transferability and robustness diagnostics**

283 A central objective of this study is to evaluate whether a retrieval trained at one location can generalize to
284 nearby sites with different surface and thermodynamic characteristics. To quantify the similarity of
285 predictor distributions between sites, we compute the KL divergence for each predictor:



$$D_{KL}(P \parallel Q) = \int p(x) \ln \left(\frac{p(x)}{q(x)} \right) dx \quad (2)$$

where P and Q denote the predictor distributions at the training and testing sites, respectively (Kullback and Leibler, 1951). Small KL divergence indicates similar predictor distributions and therefore greater expected transferability.

Distributional differences alone, however, do not determine retrieval degradation, because a predictor may differ strongly between sites while contributing little to the retrieval. We therefore combine the predictor-wise KL divergence with SHAP-based predictor importance to form a sensitivity-weighted representativeness metric, defined for each predictor as $\Delta KL \times |SHAP|$. Large values identify predictors that are simultaneously shifted between the training and testing environments and influential for the retrieval, and are interpreted as the dominant contributors to transferability degradation. This construction is the central methodological element of the present study: it converts an undifferentiated statement about covariate shift into an attribution of transferability loss to specific, physically meaningful observational quantities, and it requires only predictor distributions and model sensitivities, both of which are available for any data-driven retrieval.

Model robustness is further evaluated using noise-perturbation sensitivity experiments, following the general perturbation framework of Goodfellow et al. (2014) and Tsipras et al. (2019) but employing isotropic Gaussian perturbations rather than gradient-directed adversarial noise, since the objective here is to simulate random observational uncertainty rather than to maximally exploit model vulnerabilities. Gaussian perturbations with varying amplitudes are added to the predictor set to simulate measurement uncertainty and unresolved atmospheric variability:

$$x_{adv} = x + \epsilon \sigma_x, \epsilon \sim N(0, \eta^2) \quad (3)$$

where σ_x is the predictor standard deviation and η controls the noise amplitude (Rodgers, 2000). The resulting degradation of retrieval performance is used to distinguish random sensitivity from structural transferability error associated with covariate shift and heterogeneous boundary-layer regimes (Tsipras et al., 2019). The shape of the degradation curve, rather than its magnitude alone, is diagnostic: monotonic growth of error with perturbation amplitude is expected when the underlying predictor–target relationship is transferable and performance is limited by observational noise, whereas a non-monotonic response is consistent with a scenario in which that part of the error originates from a systematic offset in the predictor space that weak perturbations partially mask. The perturbation experiment therefore provides an independent test of the interpretation obtained from the $\Delta KL \times |SHAP|$ diagnostic.

3 Results

3.1 Overall retrieval performance and diurnal evolution

The staged retrieval framework reproduces the observed LTS variability at the ARM SGP central facility (C1) with high overall skill. Figure 2a compares the retrieved and observed LTS values for the independent test dataset. The framework achieves an overall coefficient of determination of $R^2 = 0.78$ with a mean absolute error (MAE) of 2.39 K across 1890 samples. The predicted values closely follow the observed distribution across both weakly stable and strongly stable conditions, indicating that the selected surface and boundary-layer predictors collectively contain substantial information about the thermodynamic structure of the lower troposphere.

The retrieval also successfully captures the characteristic diurnal evolution of LTS (Fig. 2b). Following sunrise, LTS decreases steadily as surface heating and turbulent entrainment deepen the convective boundary layer and weaken the lower-tropospheric inversion. Minimum LTS values occur during the mature convective period in the afternoon, coincident with maximum mixed-layer depth and strongest surface coupling. After sunset, stability rapidly increases as turbulent mixing collapses and a shallow stable layer reforms near the surface. The retrieved diurnal cycle reproduces both the amplitude and timing of



these transitions and remains within the 95 % bootstrap confidence intervals of the observations throughout most of the day.

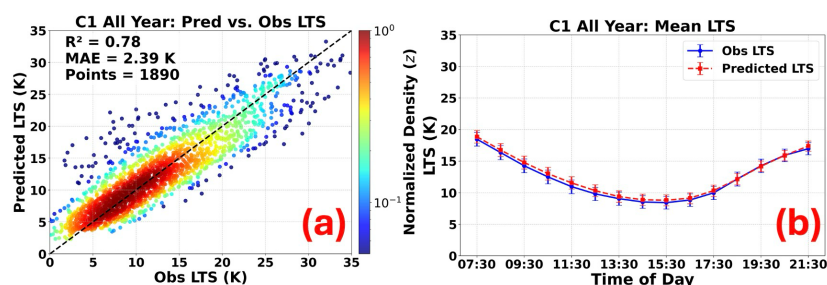


Figure 2. Overall retrieval performance of lower-tropospheric stability (LTS) at the ARM Southern Great Plains central facility (C1). (a) Retrieved versus observed LTS for the independent test dataset. The dashed line indicates the 1:1 relationship. (b) Mean diurnal evolution of observed and retrieved LTS with 95 % bootstrap confidence intervals. The retrieval captures the daytime destabilization and evening recovery of lower-tropospheric stability.

The largest retrieval discrepancies occur during the late-afternoon and evening transition period. Prior to sunset, the framework slightly overestimates LTS, whereas after the onset of evening stabilization it tends to underestimate the rapid redevelopment of the stable layer. This behavior may reflect the increasing influence of radiative cooling, residual-layer decoupling, and local heterogeneity during the evening transition, processes that are only partially represented by the surface-flux predictors. Nevertheless, the framework captures the primary thermodynamic evolution of the lower troposphere despite relying solely on surface observations and Doppler-lidar-derived boundary-layer structure.

Overall, the results demonstrate that continuous LTS variability can be reconstructed with useful accuracy from routinely available observations when the retrieval explicitly accounts for the evolving diurnal boundary-layer regime.

3.2 Seasonal variability of retrieval performance

The retrieval framework maintains robust skill across all four meteorological seasons, although both the magnitude of LTS and the retrieval performance vary substantially throughout the year (Fig. 3). The strongest performance occurs during autumn, when the framework achieves $R^2 = 0.82$, while the lowest skill occurs during winter, when stable-boundary-layer processes and synoptic forcing weaken the coupling between surface observations and lower-tropospheric thermodynamic structure. Spring and summer exhibit intermediate performance with consistently low retrieval errors.

The seasonal diurnal cycles reveal distinct thermodynamic regimes associated with variations in solar forcing, surface moisture availability, and boundary-layer coupling. During summer, LTS exhibits the weakest overall stability and the most coherent daytime destabilization cycle. Retrieved and observed values are nearly indistinguishable throughout most of the day, indicating strong coupling between surface thermodynamics and lower-tropospheric structure during strongly convective conditions. The narrow confidence intervals further suggest that summertime LTS evolution at SGP is relatively reproducible and dominated by local surface forcing.

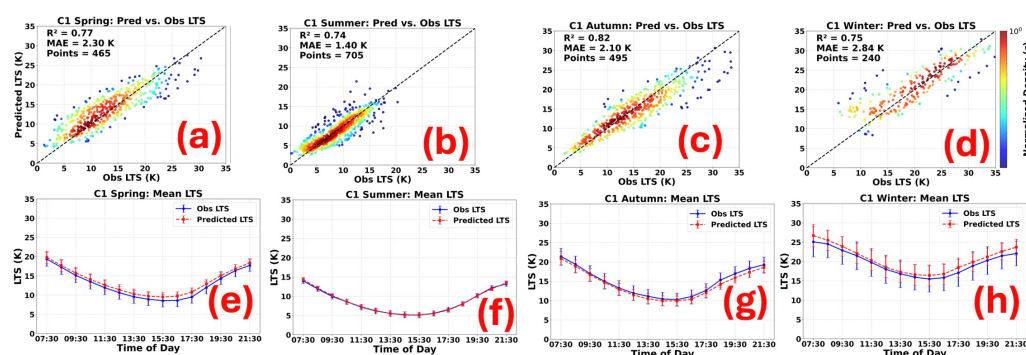


Figure 3. Seasonal retrieval performance of lower-tropospheric stability (LTS) at the ARM Southern Great Plains central facility (C1). (a–d) Retrieved versus observed LTS for spring, summer, autumn, and winter, respectively. The dashed lines indicate the 1:1 relationship. (e–h) Mean diurnal evolution of observed and retrieved LTS for spring, summer, autumn, and winter, respectively. Shading denotes the 95 % bootstrap confidence interval. The retrieval reproduces the seasonal evolution of daytime destabilization and nighttime stabilization, with strongest agreement during summer and autumn and larger variability

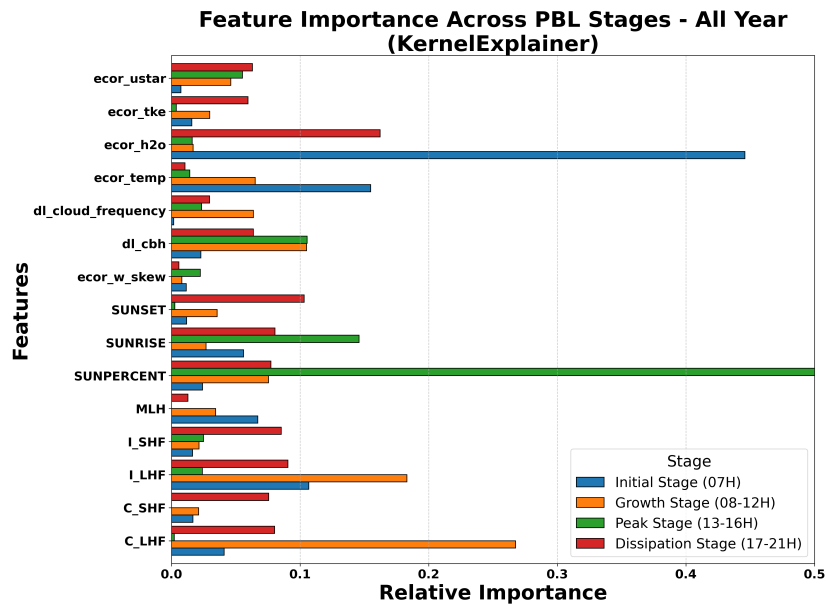
In contrast, winter exhibits substantially larger LTS values and broader variability. The wintertime boundary layer is more frequently influenced by synoptic advection, persistent stable stratification, and weak turbulent mixing, reducing the degree to which local surface observations constrain the free-tropospheric thermodynamic state. These conditions contribute to the larger retrieval uncertainty and broader confidence intervals observed during winter. The framework still reproduces the overall diurnal tendency toward daytime destabilization and nighttime stabilization, but with larger residual errors than during the warm season.

Spring and autumn occupy an intermediate regime characterized by stronger transitional variability. In spring, relatively moist soils and enhanced latent-heat partitioning suppress rapid convective growth during portions of the day, while autumn is characterized by drier surface conditions and stronger sensible-heating dominance. The autumn environment appears particularly favorable for the retrieval framework. Compared with winter, synoptic-scale disturbances are generally less frequent and intense in autumn, while the surface sensible-heat-dominated energy balance maintains a relatively stable coupling between surface forcing and lower-tropospheric thermodynamic structure, providing more favorable conditions for the retrieval approach.

Although the absolute magnitude of LTS varies substantially across seasons, the daytime reduction in stability from morning to afternoon remains comparatively consistent. This behavior contrasts with the much larger seasonal variability typically observed in mixed-layer height at SGP and suggests that LTS and MLH provide complementary rather than interchangeable information about boundary-layer structure. MLH primarily describes the geometric depth of turbulent mixing, whereas LTS characterizes the thermodynamic resistance to that mixing.

3.3 Diurnal-regime dependence of predictor contributions

The relative importance of the predictor variables changes substantially throughout the diurnal cycle, indicating that the physical controls on LTS vary strongly among boundary-layer regimes (Fig. 4). This stage dependence is considerably larger than the corresponding seasonal variability and provides direct justification for the staged retrieval framework.



397

398 **Figure 4.** Stage-dependent predictor contributions to lower-tropospheric stability (LTS) retrieval derived
399 from SHAP analysis for the morning-transition, rapid-growth, mature-convective, and evening-decay
400 stages. Contributions are normalized within each stage. The dominant controls on LTS vary substantially
401 throughout the diurnal cycle, with moisture-sensitive predictors dominating during the morning transition
402 and solar forcing and surface heat fluxes becoming more important during the mature convective stage.

403 During the morning-transition stage, surface moisture and thermodynamic variables exert the strongest
404 influence on the retrievals. Surface water-vapor concentration contributes prominently during this period,
405 reflecting the importance of residual-layer moisture, nocturnal stratification, and early-morning entrainment
406 processes. As the boundary layer transitions into the rapid-growth stage, cumulative sensible and latent heat
407 fluxes become increasingly important, indicating that the integrated daytime surface-energy input governs
408 the progressive weakening of lower-tropospheric stability.

409 The mature convective stage exhibits a distinctly different sensitivity structure. During this period, the
410 normalized solar-time variable (SUNPERCENT) becomes the dominant predictor, indicating that the
411 retrieval framework strongly associates the timing of peak solar forcing with the deepest convective mixing
412 and minimum LTS values. Instantaneous sensible and latent heat fluxes also contribute strongly during this
413 stage, consistent with vigorous daytime turbulent exchange and strong land-atmosphere coupling.

414 The evening-decay stage is characterized by a broader and more distributed predictor sensitivity. The
415 relative contributions of cumulative fluxes, moisture variables, and temporal context become more balanced,
416 reflecting the increasingly heterogeneous nature of the atmosphere during the collapse of turbulence and
417 redevelopment of stable stratification. Surface moisture regains importance during this stage, suggesting
418 that moisture retention and residual-layer thermodynamics modulate the rate at which lower-tropospheric
419 stability rebuilds after sunset.

420 In contrast, several dynamical variables exhibit consistently weak contributions throughout the day.
421 Turbulent kinetic energy, friction velocity, and vertical-velocity skewness remain comparatively
422 unimportant predictors relative to the thermodynamic variables and cumulative surface-flux terms. This
423 result is physically consistent with the definition of LTS as a bulk thermodynamic quantity governed
424 primarily by temperature and moisture structure rather than by local mechanical turbulence alone.



Overall, the strong variation in predictor contributions across boundary-layer stages demonstrates that the relationship between surface observations and LTS is fundamentally nonstationary. A single retrieval framework applied uniformly across the entire day would therefore average across physically distinct regimes and reduce interpretability. Beyond justifying the stage partitioning, which has been established previously, this result shows that predictor sensitivity is a state-dependent quantity with clear physical structure. This is a prerequisite for the transferability diagnostic developed in Sect. 4.4, in which SHAP-based sensitivities are used to weight cross-site distributional differences: those weights are meaningful precisely because they reflect the evolving physical controls on LTS rather than an arbitrary statistical ranking.

3.4 Thermodynamic controls during mature and decay stages

The SHAP beeswarm distributions for the mature convective and evening-decay stages further illustrate how the thermodynamic controls on LTS evolve during the latter half of the diurnal cycle (Fig. 5). During the mature convective period, the retrieval is dominated by solar forcing and instantaneous surface-flux variability. High values of sensible and latent heat flux are generally associated with negative SHAP contributions, indicating reduced LTS and stronger convective destabilization. The normalized solar-time variable also exerts a strong and consistent influence during this stage, reflecting the highly organized daytime convective regime.

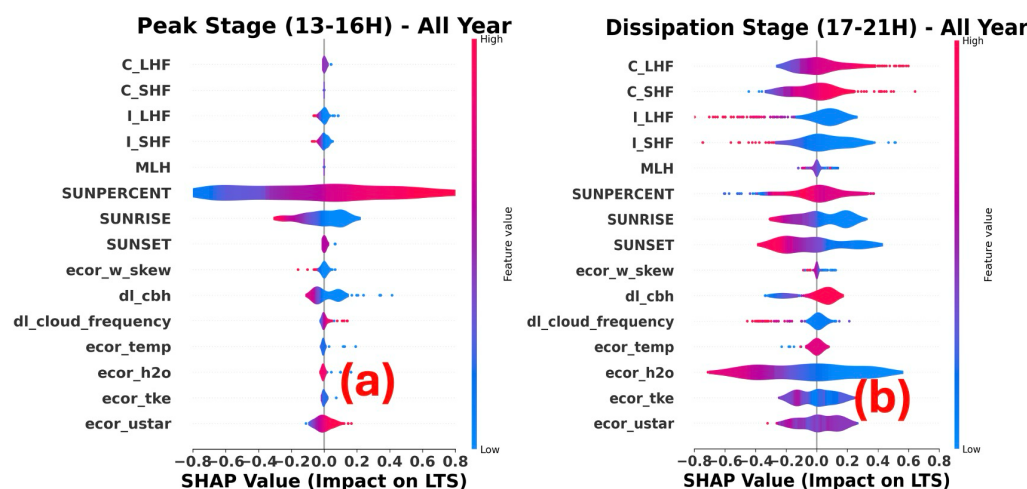


Figure 5. SHAP beeswarm distributions for the (a) mature-convective and (b) evening-decay stages. Each point represents an individual sample, with color indicating the predictor magnitude and horizontal position indicating the SHAP contribution to lower-tropospheric stability (LTS). The mature-convective stage is dominated by solar forcing and instantaneous surface heat fluxes, whereas moisture and cumulative-flux effects become increasingly important during the evening transition.

In contrast, the evening-decay stage exhibits substantially broader and more heterogeneous SHAP distributions. Predictor contributions become less tightly clustered and show greater sample-to-sample variability, indicating that the atmosphere becomes increasingly nonlinear and spatially heterogeneous as turbulence collapses. During this period, cumulative surface-flux variables become more influential than instantaneous fluxes, suggesting that the thermodynamic memory of daytime heating plays a central role in controlling the evening recovery of stability.

Surface moisture also exerts a stronger influence during the evening transition. High moisture conditions tend to suppress rapid increases in LTS, likely because retained near-surface moisture and residual-layer humidity slow radiative cooling and delay the redevelopment of strong stratification. This behavior



contrasts with the mature convective stage, during which the strong daytime turbulent mixing partially obscures the direct moisture sensitivity of LTS.

The widening of the SHAP distributions during the evening period highlights the difficulty of representing transitional boundary-layer regimes using a single stationary relationship. These results further support the need for a staged retrieval framework that explicitly accounts for the evolving physical structure of the boundary layer.

3.5 Seasonal versus diurnal-regime sensitivity

To evaluate whether retrieval sensitivity is controlled more strongly by season or by boundary-layer regime, SHAP-based predictor contributions were compared across both meteorological seasons and diurnal stages (Fig. 6). The analysis reveals that predictor variability associated with diurnal regime transitions is substantially larger than the corresponding seasonal variability.

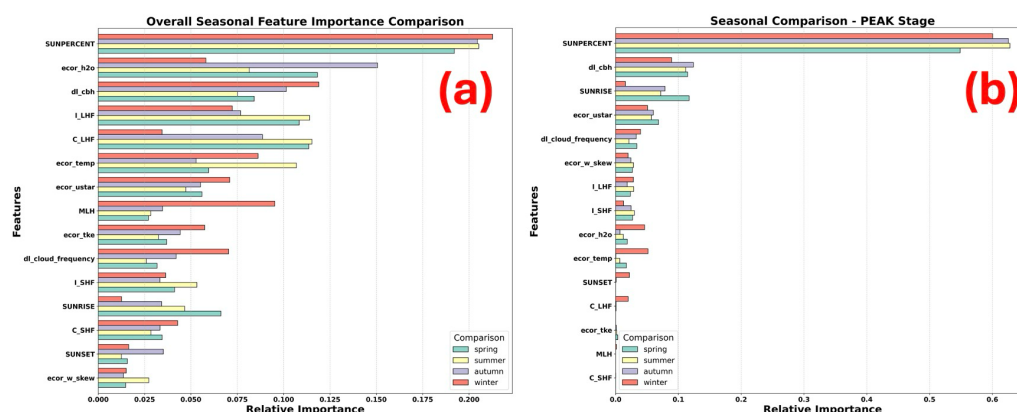


Figure 6. Comparison of predictor contributions across meteorological seasons and diurnal boundary-layer stages derived from SHAP analysis. (a) Seasonal variability of normalized predictor contributions. (b) Stage-dependent variability of normalized predictor contributions. Variability associated with boundary-layer stage is substantially larger than seasonal variability, indicating that the predictor–LTS relationship is governed primarily by diurnal boundary-layer evolution rather than by season alone.

Across seasons, the relative ranking of the dominant predictors remains comparatively stable. Surface moisture, sensible heat flux, cumulative fluxes, and temporal context remain consistently important throughout the year, with only moderate seasonal shifts in relative contribution. This result suggests that the fundamental physical relationships linking surface forcing and lower-tropospheric thermodynamics remain broadly similar across the SGP seasonal cycle.

In contrast, the transition among boundary-layer stages produces dramatic changes in predictor sensitivity. Variables such as surface moisture and solar timing exhibit strong shifts in contribution between the morning-transition, mature-convective, and evening-decay stages. These variations are substantially larger than the seasonal differences and indicate that the retrieval framework is responding primarily to the evolving physical state of the boundary layer rather than to climatological season alone.

The dominance of stage-dependent variability has important implications for observational retrievals of boundary-layer thermodynamics. It indicates that the primary source of nonstationarity in the predictor–LTS relationship arises from the internal evolution of the convective boundary layer during the day rather than from slower seasonal changes in the large-scale environment. This finding explains why the staged framework substantially improves retrieval interpretability and performance relative to a single global model.



The results also suggest that the SGP domain operates within a relatively coherent regional thermodynamic regime despite large seasonal variations in absolute LTS magnitude. Consequently, the principal challenge for generalization may arise less from seasonal concept drift and more from spatial heterogeneity and cross-site differences in land–atmosphere coupling, which are examined in the following discussion section.

4 Discussion

4.1 Physical interpretation of the LTS retrieval

The results demonstrate that LTS can be reconstructed with useful accuracy from surface-flux observations and Doppler-lidar-derived MLH, provided that the evolving structure of the convective boundary layer is explicitly considered. This finding is physically significant because the retrieval does not rely on direct thermodynamic profiling above the surface. Instead, the framework exploits the coupled evolution of surface forcing, turbulent mixing, and boundary-layer growth that collectively encode information about the lower-tropospheric thermal structure.

A key result is that the predictor–LTS relationship is fundamentally nonstationary. The dominant controls on LTS vary substantially among the morning-transition, rapid-growth, mature-convective, and evening-decay stages, reflecting the evolving balance among surface heating, entrainment, moisture transport, and radiative cooling. During the daytime convective period, surface sensible heating and solar forcing dominate the retrieval, while moisture and cumulative surface-flux effects become increasingly important during the evening transition. These changes are consistent with the well-known hysteresis behavior of the convective boundary layer, in which the thermodynamic state depends not only on the instantaneous forcing but also on the accumulated surface-energy history (Wilde et al., 1985; Salesky and Anderson, 2018).

The staged framework therefore improves both retrieval performance and physical interpretability. Rather than imposing a single global relationship between surface variables and LTS, the framework resolves the distinct thermodynamic sensitivities associated with different boundary-layer regimes. The strong stage dependence identified by the SHAP analysis further suggests that much of the apparent nonlinearity in the retrieval problem arises from mixing physically distinct atmospheric states within a single model. Separating these regimes reduces internal variance and allows the dominant physical controls on LTS to emerge more clearly. We emphasise that the performance and interpretability benefits of the stage partitioning have been established in earlier work (Chu et al., 2025a, 2026a, 2026c) and are not claimed as a contribution of the present study; they are summarised here only insofar as they underpin the sensitivity weighting used in the transferability diagnostic below.

The results also clarify the relationship between LTS and MLH. Although the two quantities are dynamically linked, they describe different aspects of boundary-layer structure. MLH represents the geometric depth of turbulent mixing, whereas LTS characterizes the thermodynamic resistance to vertical exchange. The relatively modest contribution of MLH to the retrieval indicates that much of the information associated with lower-tropospheric thermodynamics is already encoded in the surface-flux variables and moisture structure. At the same time, the seasonal analysis demonstrates that MLH and LTS exhibit different variability characteristics, particularly during winter, emphasizing that the two quantities provide complementary rather than interchangeable information about the atmosphere.

4.2 Cross-site transferability and regional heterogeneity

The cross-site experiments reveal that retrieval performance depends strongly on regional representativeness and surface heterogeneity within the Southern Great Plains domain. Although the E37 and E39 facilities are both located within approximately 100 km of the central facility, the retrieval trained at C1 transfers substantially better to E39 than to E37. The C1-trained framework generalizes more successfully to E39 than to E37 (Figure 7), indicating that multiple thermodynamic regimes exist within the SGP domain despite the relatively small geographic separation among sites.

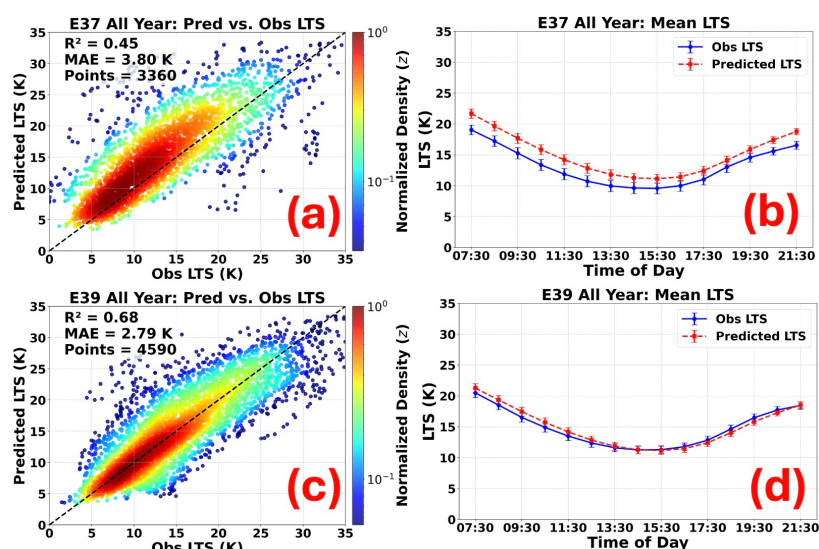


Figure 7. Cross-site transferability of the C1-trained lower-tropospheric stability (LTS) retrieval framework at the ARM Southern Great Plains extended facilities. (a, c) Retrieved versus observed LTS at E37 and E39, respectively. The dashed lines indicate the 1:1 relationship. (b, d) Mean diurnal evolution of observed and retrieved LTS at E37 and E39, respectively. The retrieval transfers more successfully to E39 than to E37, while E37 exhibits a systematic warm bias associated with regional thermodynamic and land–atmosphere heterogeneity.

The stronger transferability to E39 is physically consistent with its geographic relationship to C1. E39 lies southeast of the central facility and is more strongly influenced by the prevailing warm-season southerly flow over the SGP region. As a result, E39 and C1 are more likely to experience similar upstream surface-forcing histories, soil-moisture conditions, and boundary-layer evolution. In contrast, E37 samples a different portion of the regional land surface characterized by distinct vegetation and thermodynamic characteristics. These differences alter the partitioning between sensible and latent heat fluxes and ultimately modify the statistical relationship between surface observations and lower-tropospheric structure.

The transferability results therefore highlight an important limitation of observational retrieval frameworks over heterogeneous land surfaces. High skill at a training site does not guarantee regional generalization, even within a relatively compact observational domain. This issue is particularly important for boundary-layer applications because the thermodynamic coupling between the surface and free troposphere is strongly modulated by local land-surface conditions, vegetation state, soil moisture, and synoptic forcing.

At the same time, the cross-site analysis demonstrates that transferability degradation is physically interpretable rather than random. The systematic warm bias at E37 suggests that the retrieval framework encounters a structurally different thermodynamic regime rather than simply increased observational noise. The ability to diagnose this behavior using physically meaningful predictor differences represents one of the primary strengths of the framework.

4.3 Predictor distribution differences and regional representativeness

The cross-site transferability differences are closely linked to variations in the predictor distributions between the ARM SGP facilities. To quantify these differences, the KL divergence was computed for each predictor relative to the central facility (C1) (Fig. 8). Larger KL divergence indicates greater differences in the statistical distribution of a predictor between the training and testing sites and therefore reduced observational representativeness.

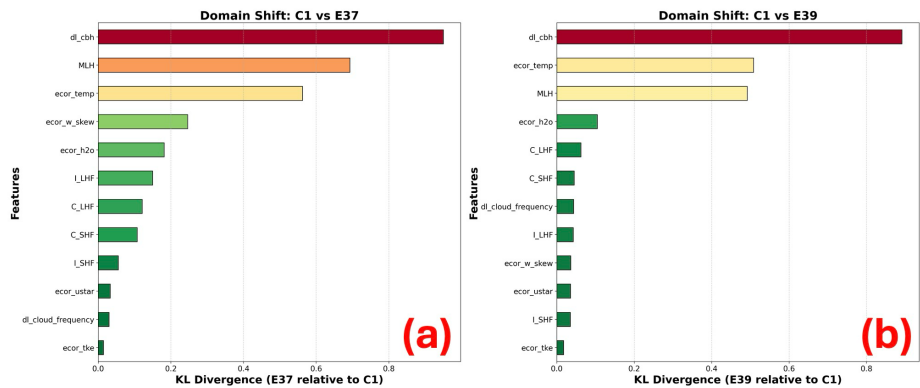


Figure 8. Predictor distribution differences between the ARM Southern Great Plains central facility (C1) and the E37 and E39 extended facilities quantified using Kullback–Leibler (KL) divergence. (a) C1–E37 comparison. (b) C1–E39 comparison. Larger KL divergence indicates stronger cross-site differences in predictor distributions. The E37 site exhibits larger thermodynamic and boundary-layer distribution differences relative to C1 than E39, particularly for cloud-base height, mixed-layer height, and temperature-related predictors.

Figure 8 shows that the E37 site exhibits substantially larger distributional differences relative to C1 than the E39 site. The strongest differences occur for CBH, MLH, and temperature-related predictors, indicating that the thermodynamic and boundary-layer environments sampled at E37 differ systematically from those represented in the C1 training dataset. In contrast, the E39 predictor distributions remain considerably closer to those at C1, consistent with the stronger cross-site retrieval performance shown in Fig. 7.

The large cloud-base-height divergence suggests substantial differences in cloud occurrence, boundary-layer moisture structure, and entrainment characteristics between the sites. Similarly, the enhanced MLH divergence at E37 indicates systematic differences in boundary-layer growth and turbulent mixing depth. Temperature and moisture variables also exhibit elevated divergence values, implying that the lower-tropospheric thermodynamic regimes sampled at E37 are not fully represented within the C1 training environment.

Importantly, large distributional differences alone do not necessarily imply retrieval degradation. Some predictors may exhibit strong cross-site variability while contributing only weakly to the retrieval framework. The results in Fig. 8 therefore establish the existence of regional thermodynamic heterogeneity but do not by themselves determine which differences dominate the transferability errors. To identify the physically important sources of retrieval degradation, the predictor distribution differences are combined with SHAP-based predictor sensitivity in the following section.

4.4 The sensitivity-weighted transferability diagnostic ($\Delta\text{KL} \times |\text{SHAP}|$)

Although Fig. 8 demonstrates that multiple predictors exhibit substantial cross-site distribution differences, not all distributional shifts contribute equally to retrieval degradation. To identify the physically important sources of transferability error, the predictor-wise KL divergence was combined with SHAP-based predictor sensitivity using the SHAP-weighted diagnostic shown in Fig. 9.

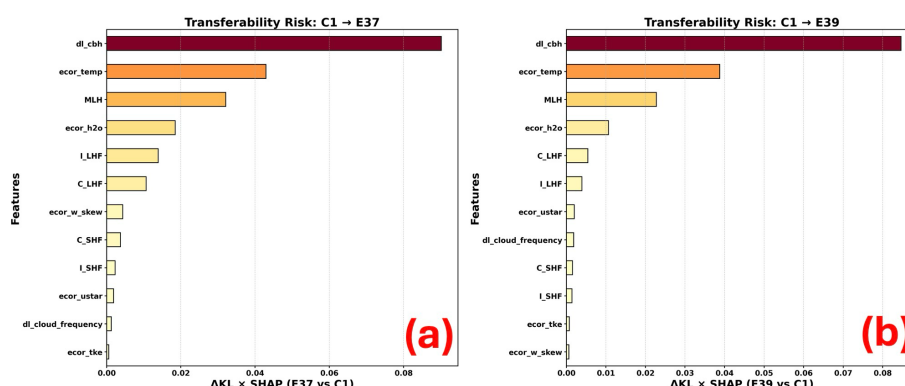


Figure 9. SHAP-weighted predictor distribution differences for the lower-tropospheric stability (LTS) retrieval framework at the ARM Southern Great Plains extended facilities. (a) C1–E37 comparison. (b) C1–E39 comparison. The diagnostics are expressed as $\Delta KL \times |SHAP|$, where larger values indicate predictors that exhibit both strong cross-site distributional differences and strong influence on retrieval performance. The E37 site exhibits larger thermodynamic and boundary-layer contributions to transferability degradation than E39, particularly for cloud-base height, mixed-layer height, and temperature-related predictors.

The resulting metric highlights predictors that simultaneously exhibit strong cross-site distribution differences and strong influence on LTS retrieval performance. Larger values therefore identify the dominant contributors to transferability degradation. Figure 9 shows that the E37 site exhibits substantially larger SHAP-weighted shifts than E39, consistent with the poorer retrieval performance shown in Fig. 7.

Cloud-base height remains the dominant contributor at both sites, indicating that differences in cloud occurrence and boundary-layer moisture structure strongly affect retrieval transferability. However, the E37 site also exhibits substantially larger contributions from mixed-layer height, temperature, and moisture-related predictors. These results suggest that the thermodynamic and turbulent environments sampled at E37 differ more fundamentally from the C1 training environment than those at E39.

Importantly, the comparison between Figs. 8 and 9 demonstrates that large predictor distribution differences alone do not necessarily imply strong transferability degradation. Several predictors exhibit measurable KL divergence but relatively weak SHAP-weighted influence, indicating limited impact on retrieval performance. In contrast, thermodynamic and boundary-layer-structure variables simultaneously exhibit large distributional differences and strong retrieval sensitivity, making them the dominant controls on cross-site degradation.

The SHAP-weighted diagnostics therefore provide a physically interpretable framework for distinguishing between benign distributional variability and structurally important thermodynamic mismatch. More broadly, the results demonstrate that observational representativeness depends not only on the similarity of predictor distributions but also on the physical importance of those predictors to the retrieval framework itself. Because the metric is constructed only from predictor distributions and model sensitivities, it is not specific to LTS or to the present predictor set. It can be applied to any observational retrieval for which a trained model and its sensitivities are available, and it provides a quantitative, physically interpretable criterion for judging in advance whether a retrieval developed at one site is likely to transfer to another.

4.5 Perturbation sensitivity

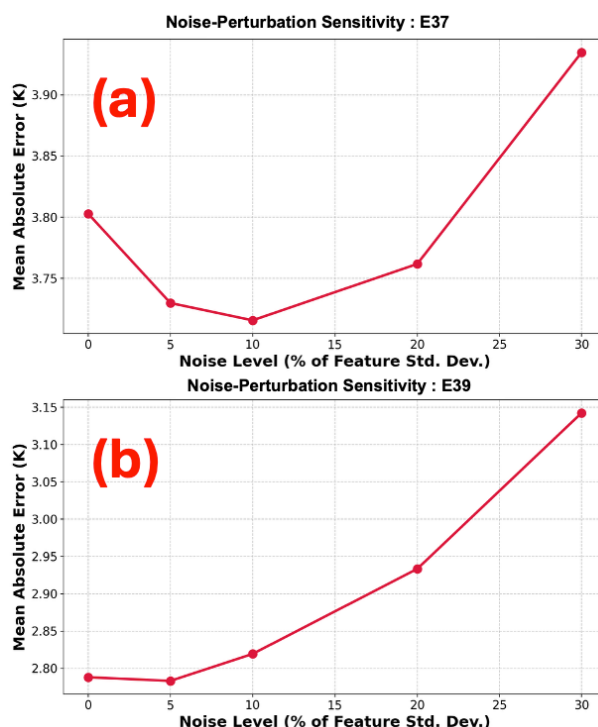
The noise-perturbation sensitivity experiments shown in Fig. 10 provide additional insight into the physical nature of the cross-site transferability degradation. Gaussian perturbations were applied to the predictor



variables to evaluate how retrieval performance responds to increasing distortion of the predictor space and to distinguish random observational sensitivity from structural thermodynamic mismatch.

At the E39 site, retrieval performance degrades approximately monotonically as perturbation amplitude increases (Fig. 10b). This behavior is consistent with a physically transferable predictor–LTS relationship subjected to progressively larger observational uncertainty. Although the perturbations reduce retrieval accuracy, the underlying relationship between the surface and boundary-layer predictors and lower-tropospheric stability remains broadly valid across the two sites. The E39 retrieval therefore appears primarily limited by random uncertainty rather than by systematic representativeness error.

In contrast, the E37 retrieval exhibits a distinctly non-monotonic response (Fig. 10a). Under weak perturbations, the mean absolute error initially decreases before increasing at larger perturbation amplitudes. This behavior is consistent with a scenario in which part of the retrieval degradation arises from structural mismatch rather than from observational uncertainty alone. One plausible interpretation is that small perturbations partially obscure systematic thermodynamic differences between the C1 training environment and the E37 predictor space, temporarily reducing the retrieval bias. As perturbation amplitude increases further, however, the physical information contained within the predictors becomes progressively degraded and retrieval performance deteriorates rapidly.



645

Figure 10. Sensitivity of the C1-trained lower-tropospheric stability (LTS) retrieval framework to Gaussian perturbations applied to the input predictors at the ARM Southern Great Plains extended facilities. Mean absolute error (MAE) is shown as a function of perturbation amplitude. (a) E37 exhibits a non-monotonic response, consistent with structural thermodynamic mismatch and partial covariate shift relative to C1. (b) E39 exhibits monotonic degradation consistent with increasing observational uncertainty and stronger cross-site transferability.

651



The contrasting responses between E37 and E39 strongly support the interpretation that the poorer E37 transferability is consistent with partial thermodynamic mismatch and covariate shift rather than from random noise alone. This result is consistent with the larger predictor distribution differences and stronger SHAP-weighted transferability contributions identified in Figs. 8 and 9. In particular, the enhanced divergence of cloud-base height, mixed-layer height, and thermodynamic predictors at E37 indicates that the site samples a boundary-layer regime that is insufficiently represented within the C1 training dataset.

More broadly, the noise-perturbation experiments suggest that retrieval uncertainty over heterogeneous land surfaces contains both random and structural components. Random uncertainty arises from observational noise and unresolved atmospheric variability, whereas structural uncertainty originates from incomplete representation of the target thermodynamic regime within the training dataset. Distinguishing between these two sources of uncertainty is important for evaluating observational representativeness and for assessing the reliability of retrieval frameworks applied across spatially heterogeneous atmospheric environments. Taken together with the $\Delta\text{KL} \times |\text{SHAP}|$ diagnostic, the shape of the perturbation-response curve constitutes a practical two-part test of transferability: the diagnostic identifies which observational quantities are responsible for degradation, while the perturbation response **may indicate** whether that degradation is structural or stochastic in origin. We suggest that this pairing is generally applicable to data-driven atmospheric retrievals deployed beyond their training environment.

4.6 Implications for boundary-layer observations and model evaluation

The results of this study demonstrate that LTS can be monitored continuously using widely available surface-flux observations and Doppler-lidar-derived boundary-layer structure when the evolving thermodynamic regime of the boundary layer is explicitly considered. This capability is important because continuous thermodynamic profiling observations remain geographically sparse, while long-term surface-flux and Doppler-lidar measurements are becoming increasingly common at atmospheric observatories and field campaigns.

The retrieval framework developed here therefore provides a practical pathway for extending thermodynamic monitoring beyond sites equipped with advanced profiling systems such as AERIs, microwave radiometers, or Raman lidars. Although these instruments remain essential for accurately resolving inversion structure and free-tropospheric thermodynamics, the present results indicate that surface and boundary-layer observations contain substantial information about the evolving lower-tropospheric thermal structure. The framework should therefore be viewed as a complementary observational tool capable of improving the spatial and temporal coverage of thermodynamic monitoring.

The strong stage dependence identified in this study also has important implications for boundary-layer process studies. The predictor–LTS relationship varies substantially among the morning-transition, rapid-growth, mature-convective, and evening-decay stages, indicating that the lower atmosphere cannot be treated as a statistically stationary system throughout the diurnal cycle. Retrieval frameworks that ignore this regime dependence may obscure physically distinct thermodynamic processes and reduce interpretability. More broadly, the results suggest that explicit representation of diurnal boundary-layer evolution may improve other observational retrievals involving turbulence, cloud structure, entrainment, or land–atmosphere coupling.

The transferability analysis further highlights the importance of observational representativeness over heterogeneous land surfaces. Even within the relatively compact ARM Southern Great Plains domain, substantial differences in cloud-base height, mixed-layer evolution, moisture structure, and surface thermodynamic forcing produce measurable degradation of cross-site retrieval performance. These findings suggest that regional heterogeneity may represent a major source of uncertainty for observational retrieval frameworks and data-driven atmospheric applications more broadly.

The combined KL divergence and SHAP-based diagnostics developed here provide a physically interpretable framework for identifying the dominant sources of transferability degradation. Rather than



treating cross-site uncertainty as purely statistical error, the framework links retrieval degradation directly to physically meaningful differences in thermodynamic and boundary-layer structure. This approach may therefore prove useful for evaluating observational representativeness in other boundary-layer retrieval problems, including mixed-layer height, cloud properties, turbulence characteristics, and air-quality-related variables.

Finally, the results have potential implications for atmospheric-model evaluation and development. LTS strongly influences turbulent mixing, entrainment, cloud development, and pollutant dispersion, all of which remain major sources of uncertainty in numerical weather prediction and climate models. Continuous retrievals of LTS could therefore complement existing evaluations based on mixed-layer height, surface fluxes, and cloud statistics and may provide additional constraints for planetary-boundary-layer parameterizations and land-atmosphere coupling processes.

4.7 Limitations and future work

Several limitations of the present study should be acknowledged. First, the analysis is restricted to three ARM SGP facilities located within a single climatic region. Although these sites sample meaningful variability in land-surface and thermodynamic conditions, they do not encompass the broader range of atmospheric environments associated with coastal regions, forests, mountainous terrain, arid landscapes, or strongly urbanized areas. Additional evaluation across other ARM observatories and field campaigns will therefore be necessary to determine the general applicability of the retrieval framework.

Second, the retrieval depends on the availability of high-quality surface-flux and Doppler-lidar observations. Retrieval performance may degrade under conditions with weak turbulence, intermittent precipitation, cloud contamination, or incomplete lidar coverage. Stable-boundary-layer conditions also remain particularly challenging because the thermodynamic coupling between the surface and lower troposphere weakens substantially during nighttime stratification and synoptic forcing becomes increasingly important. We note, however, that the framework is not tied to Doppler lidar specifically. Mixed-layer height enters the retrieval as one predictor among several and can equally be supplied by ceilometers, radar wind profilers, radiosondes, or other boundary-layer height estimates; the sensitivity analysis indicates that its contribution is modest relative to the thermodynamic and cumulative-flux predictors. The applicability of the framework is therefore governed primarily by the availability of surface-flux and near-surface thermodynamic observations rather than by any single remote-sensing instrument.

Third, LTS is itself a bulk thermodynamic metric and does not fully describe inversion sharpness, entrainment-zone structure, elevated stable layers, or the vertical distribution of moisture within the lower atmosphere. Consequently, retrieval skill for LTS does not necessarily imply accurate representation of all aspects of boundary-layer thermodynamic structure. Future studies should therefore investigate whether similar observationally constrained approaches can be extended to more detailed thermodynamic diagnostics, including estimated inversion strength, entrainment-related quantities, and vertically resolved stability structure.

The present framework also assumes that the dominant physical relationships between the predictor variables and LTS remain sufficiently stable within each boundary-layer stage. Although the staged approach substantially improves interpretability and retrieval performance, additional nonstationarity may still arise from synoptic variability, land-surface transitions, cloud-regime changes, and seasonal moisture anomalies. Future work could therefore explore adaptive or dynamically evolving stage classifications that account more explicitly for changing atmospheric regimes.

The transferability analysis presented here further suggests that observational representativeness is likely to become increasingly important as retrieval frameworks are extended across larger spatial domains. The combined KL divergence and SHAP-based diagnostics provide a useful first-order framework for identifying physically important covariate shifts, but additional methods may be needed to characterize more complex forms of thermodynamic regime mismatch and structural uncertainty. In particular, future



746 studies should investigate how transferability varies across strongly contrasting climatic regimes and
747 whether regional retraining or hybrid physics-informed approaches can improve generalization.

748 More broadly, the results suggest that physically interpretable machine-learning frameworks can provide
749 useful extensions of atmospheric observational capability when their limitations and representativeness
750 characteristics are explicitly quantified. Continued integration of observational diagnostics, boundary-layer
751 physics, and interpretable data-driven methods may therefore offer new opportunities for improving
752 thermodynamic monitoring and understanding land-atmosphere coupling across heterogeneous
753 environments.

754 5 Conclusions

755 This study developed an observationally constrained retrieval of lower-tropospheric stability at the ARM
756 SGP observatory and, more importantly, introduced a physically interpretable diagnostic framework for
757 evaluating and attributing the cross-site transferability of such retrievals. The retrieval itself is implemented
758 within the regime-aware, stage-partitioned framework established in our earlier work (Chu et al., 2025a,
759 2026a, 2026c), which is used here as an established tool; the methodological contribution of the present
760 study lies in the transferability diagnostics.

761 The retrieval successfully reproduces the observed variability and diurnal evolution of LTS at the SGP
762 central facility. The framework achieves an overall coefficient of determination of $R^2 = 0.78$ while
763 capturing the characteristic daytime destabilization and evening recovery of the lower troposphere.
764 Seasonal analysis shows the highest skill during autumn and the largest uncertainties during winter, when
765 stable-boundary-layer processes and synoptic forcing weaken the thermodynamic coupling between the
766 surface and free troposphere.

767 Predictor sensitivity was found to depend much more strongly on diurnal boundary-layer regime than on
768 season, with solar forcing and instantaneous surface heat fluxes dominating during the mature convective
769 stage and moisture and cumulative-flux effects becoming increasingly important during the evening
770 transition. This state dependence matters here because it establishes that SHAP-based sensitivities carry
771 physical meaning and can therefore legitimately be used to weight cross-site distributional differences.

772 The cross-site experiments further demonstrate that transferability across heterogeneous land surfaces
773 represents a major challenge for observational retrieval frameworks. A model trained at the central facility
774 transfers successfully to the E39 extended facility but exhibits systematic bias at E37, despite the relatively
775 small geographic separation among sites. This result suggests that multiple local thermodynamic regimes
776 exist within the SGP domain and that retrieval performance depends strongly on regional representativeness
777 and land-atmosphere coupling.

778 The principal contribution of this study is a sensitivity-weighted representativeness diagnostic,
779 $\Delta\text{KL} \times |\text{SHAP}|$, which combines predictor-distribution differences with predictor sensitivity to identify the
780 dominant sources of physically important covariate shift. The analysis showed that retrieval degradation
781 occurs specifically where thermodynamically influential predictors also exhibit strong cross-site
782 distributional differences, and that large distributional differences in low-sensitivity predictors are largely
783 benign. Gaussian perturbation experiments provided independent confirmation: the monotonic error growth
784 at E39 is consistent with random observational uncertainty acting on a transferable relationship, whereas
785 the non-monotonic response at E37 is consistent with structural thermodynamic mismatch. Together, the
786 two diagnostics form a practical test that both helps localize and classify transferability loss.

787 Overall, the results demonstrate that continuous monitoring of lower-tropospheric stability is feasible from
788 widely available surface and boundary-layer observations, and that the reliability of such monitoring
789 outside the training environment can be assessed quantitatively rather than assumed. Because the diagnostic
790 framework requires only predictor distributions and model sensitivities, it is directly applicable to other
791 observationally constrained retrievals, including mixed-layer height, cloud properties, turbulence
792 characteristics, and air-quality-related variables. The transferability diagnostics introduced here are



793 intended to establish the mechanistic understanding of spatial performance variability that must precede
 794 broader deployment, and to guide subsequent transfer-learning and domain-adaptation strategies over
 795 heterogeneous land surfaces.

796

797 **Data availability**

798 ARM data are available at <https://doi.org/10.2172/1253897> (ARM, 2016). The CBLH determination
 799 algorithm is accessible at <https://doi.org/10.1364/OE.451728> (Chu et al., 2022a). The MLH datasets are
 800 available at <https://doi.org/10.5439/2997130> (Chu and Wang, 2025).

801 **Author contributions**

802 Conceptualization: Y.C., J.A.Z. L.J. Methodology: Y.C., G.L., L.J. Data curation: Y.C., H.G., J.A.Z.
 803 Original draft: Y.C., J.A.Z., G.L., F.M. Review and editing: G.L., J.A.Z., M.D., L.X., W.L., H.H.S., L.J. ,
 804 Z.W. Funding acquisition: J.A.Z. All authors have read and agreed to the published version of the
 805 manuscript.

806 **Competing interests**

807 The authors declare that they have no conflict of interest.

808 **Acknowledgements**

809 This research was supported by the DOE Atmospheric System Research (ASR) program under grant DE-
 810 SC0020171 and the National Science Foundation (NSF) under grant AGS-1917693. J.Z. acknowledges
 811 support from NOAA under grants NA22OAR4590118 and NA22OAR4050669D, the Office of Naval
 812 Research (ONR) MURI under grant N00014-24-1-2554, and the National Science Foundation (NSF) under
 813 awards 2228299, 2211308 and 2532217. M.D. is supported by the Biological and Environmental Research
 814 under grant no. DE-AC05-76RL01830. Support for this project was provided by the Developmental Testbed
 815 Center (DTC). The DTC Visitor Program is funded by the National Oceanic and Atmospheric
 816 Administration, the National Science Foundation, and NSF National Center for Atmospheric Research.
 817 Data were obtained from the Atmospheric Radiation Measurement (ARM) user facility, a U.S. Department
 818 of Energy Office of Science user facility managed by the Biological and Environmental Research program.

819 **References**

820 Acevedo, O. C., Mahrt, L., Puhales, F. S., Costa, F. D., Medeiros, L. E., and Degrazia, G. A.: Contrasting structures
 821 between the decoupled and coupled states of the stable boundary layer, Q. J. R. Meteorol. Soc., 142, 693–702,
 822 <https://doi.org/10.1002/qj.2693>, 2016.



- 823 Bretherton, C. S. and Wyant, M. C.: Moisture transport, lower-tropospheric stability, and decoupling of cloud-topped
824 boundary layers, *J. Atmos. Sci.*, 54, 148–167, <https://doi.org/10.1175/1520->
825 0469(1997)054<0148:MTLTS>2.0.CO;2, 1997.
- 826 Che, J. and Zhao, P.: Characteristics of the summer atmospheric boundary layer height over the Tibetan Plateau and
827 influential factors, *Atmos. Chem. Phys.*, 21, 5253–5268, <https://doi.org/10.5194/acp-21-5253-2021>, 2021.
- 828 Chu, Y., Wang, Z., Xue, L., Li, Y., Wang, Q., and Zhang, Y.: Characterizing warm atmospheric boundary layer over
829 land by combining Raman and Doppler lidar measurements, *Opt. Express*, 30, 11892–11911,
830 <https://doi.org/10.1364/OE.451728>, 2022a.
- 831 Chu, Y., Wang, Z., Deng, M., Lin, G., Xue, L., Li, W., and Shin, H. H.: The spatial and temporal variability of the
832 planetary boundary layer at the ARM SGP supersite, in: *Proceedings of the AGU Fall Meeting, Chicago, IL,*
833 *USA, 12–16 December 2022, A35K-1601*, 2022b.
- 834 Chu, Y., Lin, G., Deng, M., Zhang, J. A., and Guo, H.: Characterizing seasonal variation of the atmospheric mixing
835 layer height using machine-learning approaches, *Remote Sens.*, 17, 1399, <https://doi.org/10.3390/rs17081399>,
836 2025a.
- 837 Chu, Y., Lin, G., Deng, M., and Wang, Z.: Characteristics of eddy dissipation rates in atmosphere boundary layer using
838 Doppler lidar, *Remote Sens.*, 17, 1652, <https://doi.org/10.3390/rs17091652>, 2025b.
- 839 Chu, Y., Guo, L., Deng, M., Xue, L., Li, W., Shin, H. H., Zhang, J. A., Guo, H., and Wang, Z.: Machine-learning model
840 for inverting convective boundary layer height with implicit physical constraints and its multi-site applicability,
841 *EGUsphere [preprint]*, <https://doi.org/10.5194/egusphere-2025-2490>, 2025c.
- 842 Chu, Y. and Wang, Z.: ARM PI Data: ARM SGP PBLH and MLH datasets from Raman lidar and Doppler lidar [data
843 set], <https://doi.org/10.5439/2997130>, 2025.
- 844 Chu, Y., Lin, G., Deng, M., Xue, L., Li, W., Shin, H. H., Zhang, J. A., and Wang, Z.: Thermodynamics-guided machine-
845 learning model for predicting convective boundary layer height and its multi-site applicability, *Atmos. Chem.*
846 *Phys.*, 26, 1415–1434, <https://doi.org/10.5194/acp-26-1415-2026>, 2026a.
- 847 Chu, Y., Wang, Z., Deng, M., Lin, G., Xue, L., Li, W., Shin, H. H., and Brabec, C. M.: The spatial and seasonal
848 variability of the mixing layer height at SGP sites, *Adv. Atmos. Sci.*, 2026, <https://doi.org/10.1007/s00376-025->
849 5575-2, 2026b.
- 850 Chu, Y., Zhang, J. A., Lin, G., Xue, L., Li, W., Shin, H. H., Deng, M., Zhang D., Guo, H., Li, Zi, and Jin L.: A
851 Physically-Guided and Stage-Wise Machine Learning Framework for Mixing Layer Height Retrieval
852 Based on Doppler Lidar Observations, *Adv. Atmos. Sci.*, accepted, 2026c.
- 853 Cook, D. R.: Eddy correlation flux measurement system (ECOR) instrument handbook, Report DOE/SC-ARM-TR-
854 052, U.S. Department of Energy, <https://doi.org/10.2172/1467448>, 2018.
- 855 Dang, R., Yang, Y., Hu, X., Wang, Z., and Zhang, S.: A review of techniques for diagnosing the atmospheric boundary
856 layer height (ABLH) using aerosol lidar data, *Remote Sens.*, 11, 1590, <https://doi.org/10.3390/rs11131590>,
857 2019.
- 858 Deng, M., Jensen, M. P., Giangrande, S. E., Johnson, K., Wang, D., Chu, Y., Subba, T., Peña, J. C., Walter, P., Flynn,
859 J., and Griggs, T.: A closed bay-breeze circulation and its lifecycle from TRACER with a new orienteering tape
860 recorder diagram, *J. Geophys. Res. Atmos.*, 130, e2024JD043187, <https://doi.org/10.1029/2024JD043187>,
861 2025.
- 862 Garratt, J. R.: *The Atmospheric Boundary Layer*, Cambridge Univ. Press, Cambridge, UK,
863 [https://doi.org/10.1016/0012-8252\(94\)90026-4](https://doi.org/10.1016/0012-8252(94)90026-4), 1994.
- 864 Gero, J. and Hackel, D.: Atmospheric Emitted Radiance Interferometer (AERI) instrument handbook,
865 Report DOE/SC-ARM-TR-054, U.S. Department of Energy, <https://doi.org/10.2172/1020273>, 2025.



- 866 Goodfellow, I. J., Shlens, J., and Szegedy, C.: Explaining and harnessing adversarial examples, arXiv [preprint],
867 arXiv:1412.6572, <https://doi.org/10.48550/arXiv.1412.6572>, 2014.
- 868 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al.: The ERA5 global reanalysis,
869 Q. J. R. Meteorol. Soc., 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- 870 Holtslag, A. A. M., Svensson, G., Baas, P., Basu, S., Beare, B., Beljaars, A. C. M., Bosveld, F. C., Cuxart, J., Lindvall,
871 J., Steeneveld, G. J., et al.: Stable atmospheric boundary layers and diurnal cycles: challenges for weather and
872 climate models, Bull. Amer. Meteorol. Soc., 94, 1691–1706, <https://doi.org/10.1175/BAMS-D-11-00187.1>,
873 2013.
- 874 Klein, S. A. and Hartmann, D. L.: The seasonal cycle of low stratiform clouds, J. Climate, 6, 1587–1606,
875 [https://doi.org/10.1175/1520-0442\(1993\)006<1587:TSCOLS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2), 1993.
- 876 Krishnamurthy, R., Newsom, R. K., Berg, L. K., Xiao, H., Ma, P.-L., and Turner, D. D.: On the estimation of boundary
877 layer heights: a machine-learning approach, Atmos. Meas. Tech., 14, 4403–4424, [https://doi.org/10.5194/amt-](https://doi.org/10.5194/amt-14-4403-2021)
878 14-4403-2021, 2021.
- 879 Kullback, S. and Leibler, R. A.: On information and sufficiency, Ann. Math. Stat., 22, 79–86,
880 <https://doi.org/10.1214/aoms/1177729694>, 1951.
- 881 Li, W., D'Amico, D., Bernardet, L., Xue, L., Dudhia, J., Shin, H. H., Firl, G., Sun, S., Harrold, M., Nance, L. B., Ek,
882 M., and Chu, Y.: Demonstrating hierarchical system development with the Common Community Physics
883 Package Single-Column Model: a case study over the Southern Great Plains, Meteorol. Appl., 32, e70073,
884 <https://doi.org/10.1002/met.70073>, 2025.
- 885 Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., Sun, Y., Wang, T., Xue, H., Zhang, H., and Zhu, B.: Aerosol and boundary-
886 layer interactions and impact on air quality, Natl. Sci. Rev., 4, 810–833, <https://doi.org/10.1093/nsr/nwx117>,
887 2017.
- 888 Lin, G., Wang, Z., Ziegler, C., Hu, X.-M., Xue, M., Geerts, B., and Chu, Y.: A comparison of convective storm inflow
889 moisture variability between the Great Plains and the southeastern United States using multiplatform field
890 campaign observations, J. Atmos. Ocean. Technol., 40, 539–556, <https://doi.org/10.1175/JTECH-D-22-0037.1>,
891 2023.
- 892 Lin, G., Wang, Z., Chu, Y., Ziegler, C. L., Hu, X.-M., Xue, M., et al.: Airborne measurements of scale-dependent latent
893 heat flux impacted by water vapor and vertical velocity over heterogeneous land surfaces during the
894 CHEESEHEAD19 campaign, J. Geophys. Res. Atmos., 129, e2023JD039586,
895 <https://doi.org/10.1029/2023JD039586>, 2024.
- 896 Lothon, M., Lohou, F., Pino, D., Couvreux, F., Pardyjak, E. R., Reuder, J., et al.: The BLLAST field experiment:
897 boundary-layer late afternoon and sunset turbulence, Atmos. Chem. Phys., 14, 10931–10960,
898 <https://doi.org/10.5194/acp-14-10931-2014>, 2014.
- 899 Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst.,
900 30, <https://doi.org/10.48550/arXiv.1705.07874>, 2017.
- 901 Manninen, A. J., Marke, T., Tuononen, M., and O'Connor, E. J.: Atmospheric boundary layer classification with
902 Doppler lidar, J. Geophys. Res. Atmos., 123, 8172–8189, <https://doi.org/10.1029/2017JD028169>, 2018.
- 903 Newsom, R. K. and Krishnamurthy, R.: Doppler lidar (DL) instrument handbook, Report DOE/SC-ARM-
904 TR-101, U.S. Department of Energy, <https://doi.org/10.2172/1034640>, 2022.
- 905 Qu, X., Hall, A., Klein, S. A., and Caldwell, P. M.: The strength of the tropical inversion and its response to climate
906 change in 18 CMIP5 models, Clim. Dyn., 45, 375–396, <https://doi.org/10.1007/s00382-014-2441-9>, 2015.
- 907 Ricutord, T., Aubert, S., and Bellanger, M.: Deriving boundary layer height from aerosol lidar using machine learning:
908 KABL and ADABL algorithms, Atmos. Meas. Tech., 14, 4335–4353, [https://doi.org/10.5194/amt-14-4335-](https://doi.org/10.5194/amt-14-4335-2021)
909 2021, 2021.



- 910 Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific, Singapore, 400
911 pp., <https://doi.org/10.1142/3171>, 2000.
- 912 Salesky, S. T. and Anderson, W.: Buoyancy effects on large-scale motions in convective atmospheric boundary layers,
913 J. Fluid Mech., 856, 135–168, <https://doi.org/10.1017/jfm.2018.711>, 2018.
- 914 Shin, H. H., Xue, L., Li, W., Firl, G., D'Amico, D. F., Muñoz-Esparza, D., Ek, M. B., Chu, Y., Wang, Z., Gustafson Jr.,
915 W. I., and Vogelmann, A. M.: Large-scale forcing impact on the development of shallow convective clouds
916 revealed from LASSO large-eddy simulations, J. Geophys. Res. Atmos., 126, e2021JD035208,
917 <https://doi.org/10.1029/2021JD035208>, 2021.
- 918 Sisterson, D. L., Peppler, R. A., Cress, T. S., Lamb, P. J., and Turner, D. D.: The ARM Southern Great Plains (SGP)
919 site, Meteorol. Monogr., 57, 6.1–6.14, <https://doi.org/10.1175/AMSMONOGRAPH5-D-16-0004.1>, 2016.
- 920 Stull, R. B.: An Introduction to Boundary Layer Meteorology, Springer, Dordrecht, the Netherlands,
921 <https://doi.org/10.1007/978-94-009-3027-8>, 1988.
- 922 Su, T., Li, Z., and Kahn, R.: Relationships between the planetary boundary layer height and surface pollutants derived
923 from lidar observations over China: regional pattern and influencing factors, Atmos. Chem. Phys., 18, 15921–
924 15935, <https://doi.org/10.5194/acp-18-15921-2018>, 2018.
- 925 Sullivan, R. C., et al.: Over three decades, and counting, of near-surface turbulent flux measurements from
926 the Atmospheric Radiation Measurement (ARM) user facility, Earth Syst. Sci. Data, 17, 5007–5030,
927 <https://doi.org/10.5194/essd-17-5007-2025>, 2025.
- 928 Sun, J., Mahrt, L., Banta, R. M., and Pichugina, Y. L.: Turbulence regimes and turbulence intermittency in the stable
929 boundary layer during CASES-99, J. Atmos. Sci., 69, 338–351, <https://doi.org/10.1175/JAS-D-11-082.1>, 2012.
- 930 Tsipras, D., Santurkar, S., Engstrom, L., Turner, A., and Madry, A.: Robustness may be at odds with accuracy, in: Proc.
931 Int. Conf. Learn. Represent., <https://doi.org/10.48550/arXiv.1805.12152>, 2019.
- 932 Turner, D. D.: Improved ground-based liquid-water-path retrievals using a combined infrared brightness temperature
933 and millimeter-wavelength radar dataset, J. Geophys. Res. Atmos., 112, D15204,
934 <https://doi.org/10.1029/2007JD008530>, 2007.
- 935 Turner, D. D. and Löhnert, U.: Information content and uncertainties in thermodynamic profiles and liquid-cloud
936 properties retrieved from the ground-based Atmospheric Emitted Radiance Interferometer (AERI), J. Appl.
937 Meteorol. Climatol., 53, 752–771, <https://doi.org/10.1175/JAMC-D-13-0126.1>, 2014.
- 938 Wang, Z., Yuan, J., Wood, R., Chen, Y., and Tong, T.: Profile-based estimated inversion strength, Atmos.
939 Chem. Phys., 23, 3247–3266, <https://doi.org/10.5194/acp-23-3247-2023>, 2023.
- 940 Wilde, N. P., Stull, R. B., and Eloranta, E. W.: The LCL zone and cumulus onset, J. Climate Appl. Meteorol., 24, 640–
941 657, [https://doi.org/10.1175/1520-0450\(1985\)024<0640:TLZACO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1985)024<0640:TLZACO>2.0.CO;2), 1985.
- 942 Wood, R. and Bretherton, C. S.: On the relationship between stratocumulus cloud amount and lower-tropospheric
943 stability, J. Climate, 19, 6425–6432, <https://doi.org/10.1175/JCLI3988.1>, 2006.
- 944 Wulfmeyer, V., Hardesty, R. M., Turner, D. D., Behrendt, A., Cadeddu, M. P., Di Girolamo, P., Schlüssel, P., Van
945 Baelen, J., and Zus, F.: A review of the remote sensing of lower-tropospheric thermodynamic profiles and its
946 indispensable role for the understanding and the simulation of water and energy cycles, Rev. Geophys., 53,
947 819–895, <https://doi.org/10.1002/2014RG000476>, 2015.
- 948 Zhang, D., Comstock, J., Sivaraman, C., Mo, K., Krishnamurthy, R., Tian, J., Su, T., Li, Z., and Roldán-Henao, N.:
949 Best estimate of the planetary boundary layer height from multiple remote sensing measurements, Atmos. Meas.
950 Tech., 18, 3453–3475, <https://doi.org/10.5194/amt-18-3453-2025>, 2025.