Check for updates

AGU ADVANCING EARTH AND SPACE SCIENCES

Geophysical Research Letters

RESEARCH LETTER
10.1029/2026GL123315

Uncertainty-Aware Machine Learning for Onset of Deep Convection: Under what Conditions Are Trigger Predictions More Reliable?

Ashish Bhattarai[1,2] **and Youtong Zheng**[1,2]

[1]Department of Earth and Atmospheric Science, University of Houston, Houston, TX, USA, [2]Institute of Climate and Atmospheric Science, University of Houston, Houston, TX, USA

Key Points:

- Uncertainty-aware ML provides a framework for diagnosing convection-triggering predictability across synoptic regimes
- Prediction reliability is lowest under weak-to-moderate ascent and modest CAPE generation, where convective and non-convective environments overlap
- Reliable predictions correspond to long-lived, near-equilibrium convection supported by stronger upper-level forcing and low-level moisture transport

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Zheng,
yzheng18@uh.edu

Citation:

Bhattarai, A., & Zheng, Y. (2026). Uncertainty-aware machine learning for onset of deep convection: Under what conditions are trigger predictions more reliable? *Geophysical Research Letters*, *53*, e2026GL123315. https://doi.org/10.1029/2026GL123315

Received 28 APR 2026
Accepted 15 JUL 2026

© 2026 The Author(s). Geophysical Research Letters published by Wiley Periodicals LLC on behalf of American Geophysical Union.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract Machine Learning (ML) models have emerged as a powerful tool for predicting deep convection triggering, yet the atmospheric conditions that systematically challenge these models in detecting deep convection remain poorly understood. To diagnose such ambiguous regimes, we trained a Controlled Abstention Neural Network (CAN) that separates high- and low-confidence predictions, enabling the physical characterization of uncertain environments over the Southern Great Plains during the warm season. Results show that low-confidence, abstained predictions cluster in environments with weak-to-moderate mid-level vertical velocity and low dynamic generation rate of Convective Available Potential Energy (CAPE), predominantly corresponding to short-lived convective episodes in locally forced, non-equilibrium regimes with weaker synoptic forcing. In contrast, confident predictions correspond to long-lived, equilibrium regimes driven by synoptic forcing of ascent and enhanced low-level southerly flow. Our findings provide physically grounded guidance for identifying forecast situations that require additional caution in operational ML-based convection triggering.

Plain Language Summary Deep convection can trigger powerful storms, including thunderstorms and heavy rainfall, that are difficult to forecast. While machine learning (ML) models are popular and reliable tools for predicting where and when these storms will ignite, a key question has largely gone unanswered: under what atmospheric conditions are these AI predictions trustworthy? To address this, we trained ML models on atmospheric data from the Southern Great Plains, designed to isolate ambiguous states where convection prediction remains highly uncertain. We find that models are most confident in predicting organized storms when air is rising strongly and when stronger winds transport moisture from the Gulf of America. However, the models struggle with short-lived, pop-up afternoon convection that forms locally and is less organized. This insight helps forecasters decide when to rely on AI-based storm prediction tools in the Southern Great Plains.

1. Introduction

Deep convection redistributes heat and moisture in the atmosphere, drives large-scale circulations (e.g., Emanuel, 1994; Houze Jr, 2004; Lin et al., 2022; Miyamoto et al., 2013; Retsch et al., 2022), and often induces hazardous weather, such as thunderstorms, heavy rainfall, and flooding (e.g., Doswell et al., 1996; Markowski & Richardson, 2010; Schumacher & Johnson, 2005, 2008). Most of these processes evolve on spatial and temporal scales too fine to be resolved by general circulation models (GCMs; Arakawa, 2004) and are represented through parametrization (e.g., Yano et al., 2013). However, these parametrizations remain highly uncertain in GCM simulations (e.g., Hourdin et al., 2017; Jakob, 2010; Kober et al., 2015; Randall et al., 2003; Rio et al., 2019), with errors in predicting the deep convection triggering accounting for a large part of the bias (Suhas & Zhang, 2014; Tang et al., 2022; Tao et al., 2024; Xie et al., 2002; Xie & Zhang, 2000).

The convection trigger function determines the onset of deep convection at each time step within a grid column using a deterministic approach, variously based on grid-scale vertical motion (Kain, 2004), boundary-layer moisture convergence (Tiedtke, 1989), and convective available potential energy (CAPE) measures (e.g., G. Zhang & McFarlane, 1995; G. J. Zhang, 2002). This formulation contributes to persistent model biases in simulating the frequency and amplitude of precipitation (Cui et al., 2021), and to inaccuracies in representing their variability across timescales from the diurnal cycle to interannual variation (Dai, 2006; Lee et al., 2008). One of the many efforts to reduce biases has focused on improving trigger function formulations. For example, the CAPE trigger was extended to include the large-scale atmospheric environment and to introduce a dynamic generation

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

rate of CAPE (dCAPE; Xie & Zhang, 2000) trigger. Subsequent improvements incorporated entrainment dilution (G. J. Zhang, 2009) and allowed parcel launch above the boundary layer, referred to as the Unrestricted Launch Level (ULL) trigger (Xie et al., 2019). Incorporating the dCAPE trigger formulations improved GCMs' performance across both tropical and midlatitude regions (Suhas & Zhang, 2014). ULL improved the simulation of the diurnal precipitation cycle; however, it still tends to overpredict the occurrence of light rainfall (1–10 mm day^{-1}; Wang et al., 2020), revealing persistent limitations.

Recent advances in Machine Learning (ML) models for identifying complex nonlinear relationships have gained traction in atmospheric sciences research (Bommer et al., 2024; Toms et al., 2020), including the development of trigger functions for identifying the onset of deep convection (Gentine et al., 2018). For example, T. Zhang et al. (2021) and Miller et al. (2025) have developed ML-based trigger functions that outperform traditional threshold-based triggers. Yet a critical limitation persists: most ML-based triggers remain essentially deterministic (Gentine et al., 2018), implicitly assuming that the resolved environmental state uniquely determines whether deep convection will occur. This assumption is not well-grounded physically. Because atmospheric motions exhibit a continuum of interacting scales and lack a clear spectral gap, resolved-scale variables cannot fully constrain subgrid convective processes, and deterministic closures will inevitably produce systematic errors and overconfidence. This limitation motivates a shift toward probabilistic and uncertainty-aware ML frameworks (Christensen et al., 2024), where the model output explicitly reflects uncertainty rather than providing a single best-guess prediction (Miller et al., 2025).

Beyond improving forecast skill, the probabilistic framework elucidates the physical basis of deep convection triggering certainty. By using uncertainty-aware ML as a proxy for intrinsic predictability, we systematically distinguish between convective environments that are physically well-constrained and those that are inherently more stochastic. In this way, uncertainty is treated not merely as a modeling limitation, but as a physically meaningful signal that provides new insight into the regime dependence of convection triggering predictability. To that end, we train an artificial neural network (ANN) and a controlled abstention neural network (CAN) (E. A. Barnes & Barnes, 2021b) using a compact, physically interpretable predictor set. In the ANN, confidence is diagnosed only after training by ranking the predicted probabilities, so low-confidence cases are identified post hoc. In contrast, CAN introduces an explicit abstention class and a NotWrong loss, allowing the network to say "I do not know" during training and to preferentially learn from the more confident samples while isolating ambiguous ones. Thus, CAN is not simply a confidence filter applied after prediction; it is a training framework designed to separate confidently predictable cases from intrinsically uncertain ones. Applying these models to both observations (Atmospheric Radiation Measurement (ARM) constrained variational analysis (VARANAL)) product (Tang & Zhang, 2015) and the convection-permitting model data sets (CONUS404 hydroclimate data set (Rasmussen et al., 2023)) in the United States Southern Great Plains (SGP) region, we assess how predictive skill depends on excluding low-confidence cases and under what conditions the trigger predictions are most uncertain.

2. Data and Methods

2.1. Study Area and Data Sets

We focus our analysis on the SGP (Figure S1 in Supporting Information S1), a region characterized by a complex spectrum of warm-season convective forcing. Nocturnal convection over this region predominantly occurs when eastward-propagating mesoscale convective systems (MCSs) interact with southerly low-level jets from the Gulf of America, whereas boundary-layer destabilization due to diurnal solar heating mostly triggers afternoon convection (Pryor et al., 2023; Tao et al., 2021; Tian et al., 2005). To understand convective regimes in which ML models can confidently identify deep convective events and those in which predictive skill systematically degrades, we independently trained and evaluated our ML models on two distinct data sets: VARANAL and CONUS404. VARANAL is a continuous large-scale forcing data set in which the National Oceanic and Atmospheric Administration (NOAA) Rapid Update Cycle/Rapid Refresh (RUC/RAP) background analyses are constrained toward observations using a constrained variational analysis approach (Tang et al., 2019; Tang & Zhang, 2015). CONUS404 is the state-of-the-art Weather Research and Forecasting (WRF)-based hydroclimate data set over the conterminous United States (CONUS) at a 4 km spatial resolution (Rasmussen et al., 2023). At this convection-permitting scale, the model explicitly resolves large-scale convective events and does not necessarily rely on convection parameterization (Rasmussen et al., 2023). VARANAL provides single-column, vertically resolved domain-mean atmospheric conditions. In contrast, CONUS404 is a 3D gridded data set. To

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

enable a direct comparison, we horizontally domain-averaged CONUS404 to produce a single-column, vertically-resolved representation consistent with VARANAL (comparison statistics in Table S2 in Supporting Information S1). We analyzed both data sets for the summer months (JJA) from 2004 to 2018 at 1-hr intervals. Additionally, the fifth-generation European Center for Medium-Range Weather Forecasts Reanalysis (ERA5) data sets (Hersbach et al., 2023) provide a synoptic-scale context for the VARANAL analysis.

2.2. Selections of Predictors

To build an ML model that is both accurate and physically interpretable, we restrict our inputs to a compact set of nine predictors (Table S1 in Supporting Information S1), selected for their established physical relevance to the onset of deep convection and informed by prior ML-based studies of convective triggering (e.g., Miller et al., 2025; T. Zhang et al., 2021; I.-H. Chen et al., 2023; M. Chen et al., 2023). The predictors are: surface temperature (T_srf) and relative humidity (RH_srf), sensible heat flux (SH), latent heat flux (LH), precipitable water (PW), CAPE, dCAPE, mid-level vertical velocity (omega_mid, averaged over 300–700 hPa), and convective inhibition (CIN). Collectively, these variables represent three key ingredients of deep convection triggering: atmospheric instability, moisture availability, and lifting mechanisms. Surface temperature and humidity characterize the thermodynamic state of the near-surface environment (G. J. Zhang, 2003), while turbulent heat fluxes (SH and LH) determine mean profiles in the boundary layer (G. J. Zhang, 2003). CAPE quantifies the thermodynamic energy available to a buoyant parcel, while CIN represents the energetic barrier it must overcome to reach its level of free convection (Colby, 1984; Trier et al., 2004). Stronger vertical velocity promotes parcel lifting, reducing effective inhibition and increasing the likelihood of convective initiation (Schiro et al., 2018). Meanwhile, dCAPE represents synoptic-scale forcing of instability by considering large-scale temperature and moisture advection (Xie & Zhang, 2000). We also note that adding features such as moist static energy, wind shear, and humidity at different vertical levels had little to no impact on the model's skill in identifying deep convective events.

We considered periods when the SGP domain-mean precipitation exceeds 0.5 mm hr^{-1} as a proxy for deep convection (e.g., Schiro & Neelin, 2019; Song & Zhang, 2017; Suhas & Zhang, 2014). Because warm-season convection over the SGP is spatially intermittent, such a domain-mean threshold generally corresponds to substantially stronger local rainfall within the precipitating portions of the domain. Although this threshold may include stratiform precipitation and does not strictly isolate convective clouds, prior studies (Suhas & Zhang, 2014; T. Zhang et al., 2021) have demonstrated its robustness in capturing deep convection through sensitivity analyses, showing that moderate variations in the threshold value do not substantially impact the skills of the models in identifying deep convection.

2.3. Uncertainty-Aware Machine Learning Architectures

We trained two types of supervised neural network (NN) models: ANN and CAN to identify deep convection. The ANN (Gardner & Dorling, 1998) serves as the conventional baseline classifier, while CAN extends the same architecture by introducing an additional abstention class that allows the model to abstain from classifying less-confident samples when the atmospheric state does not uniquely constrain the outcome (E. A. Barnes & Barnes, 2021b). CAN employed a standard feed-forward architecture consisting of one input layer, three hidden layers, and one output layer with a softmax activation function. The classification problem was formulated as a binary prediction of deep convection occurrence. For the ANN model, parameters were optimized using the standard cross-entropy loss. For CAN, the standard cross-entropy loss was replaced with a customized "Not-Wrong" loss function (E. A. Barnes & Barnes, 2021b) defined as:

$$L_{NW}(x_j) = -\log(p_j + p_{k+1}) - \alpha \log q \quad (1)$$

here, p_j is the likelihood of the correct class, p_{k+1} is the likelihood of a designated "abstention" class, q is the likelihood of not abstaining, and α is a non-negative weight. The first term on the right-hand side of Equation 1 represents the likelihood of not getting the prediction wrong, which is the sum of the likelihoods of correct and abstention predictions, while the second term applies an abstention penalty weighted by α. This function prevents universal abstention but penalizes the network heavily for making an incorrect prediction (E. A. Barnes &

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Barnes, 2021b). Thus, the architectures of both models were otherwise similar, with CAN using a different loss function that can output one additional class.

A key advantage of CAN is its ability to focus learning on high-confidence samples while either enforcing a user-specified abstention rate or adaptively determining the abstention fraction, thereby assigning the least confident cases to a dedicated "abstention" class. The central rationale for employing CAN is to develop a more skillful model by training it on high-confidence samples while simultaneously isolating low-confidence samples to examine the characteristics of ambiguous cases. CAN was trained with prescribed abstention fractions ranging from 0% to 20% in 5% increments to evaluate the impact of abstention on model performance. In each experiment, the CAN isolated the least confident samples by assigning them to the abstention class during training. For a detailed comparison between confident and abstained samples, we focused on the 5% abstention configuration and compared the properties of both classes. A conceptual framework and implementation of CAN follow E. A. Barnes & Barnes (2021b, 2021a).

Both NN models were trained independently on each data set using 10 years of data, validated on the subsequent 2 years, and tested on the final 3 years. Input features were standardized to a zero mean and unit variance using statistics computed from the training set only. The hyperparameters were first optimized for the ANN using the optuna framework (Akiba et al., 2019), with the F1 score for the convective cases as an objective metric. The search space included network architecture, learning rate, L2 regularization penalty, batch size, and activation function, and all models were trained using stochastic gradient descent with Nesterov momentum and early stopping. The optimal configuration was a three-layer architecture with 128, 128, and 64 nodes, a learning rate of 0.0247, an L2 penalty of 0.003, and a batch size of 16, yielding a validation F1 score of 0.777. The same set of hyperparameters is used for the CAN, as independent tuning is impractical due to additional complexity introduced by the abstention mechanism.

2.4. Convective-Adjustment Time Scale Diagnostic

To interpret the physical origin of prediction uncertainty identified by the ML models, we introduce a convective adjustment timescale diagnostic that distinguishes between strongly forced, more predictable convection and locally driven, less predictable regimes (Done et al., 2006). Convection associated with mesoscale-to-synoptic forcing tends to approach a quasi-equilibrium state (Yano & Plant, 2012) because sustained large-scale ascent continuously generates CAPE, while convection rapidly removes instability through convective heating and vertical mixing, maintaining a balance between forcing and stabilization (Bechtold et al., 2014; Zimmer et al., 2011). In contrast, convection driven by local boundary-layer processes is inherently out of equilibrium, as instability builds intermittently under surface heating and must accumulate sufficiently to overcome CIN, leading to delayed, sporadic, and less predictable convective onset (Flack et al., 2016; Keil et al., 2014). To connect these contrasting convective behaviors to a measurable quantity, we use the convective adjustment timescale diagnostic, which quantifies how quickly convection removes instability relative to its generation. In this framework, near-equilibrium convection is associated with short adjustment timescales, reflecting rapid stabilization under strong large-scale forcing, whereas non-equilibrium convection exhibits longer timescales due to delayed and intermittent triggering. The convective adjustment timescale (τ_c) is defined as (Done et al., 2006):

$$\tau_c = \frac{\mathrm{CAPE}}{\frac{d(\mathrm{CAPE})}{dt}} \tag{2}$$

The rate of change of CAPE is derived from the column-integrated rate of latent heat release implied by precipitation rate as:

$$\tau_c = \frac{\mathrm{CAPE}}{\frac{d(\mathrm{CAPE})}{dt}} = b\,\frac{c_p\,\rho_0 T_0\,\mathrm{CAPE}}{L_v\,g\,P_r} \tag{3}$$

Here, P_r is the hourly precipitation rate (mm s^{-1}), L_v is the latent heat of evaporation (J kg^{-1}), c_p is the specific heat capacity of air at constant pressure (J kg^{-1} K^{-1}), g is the acceleration due to gravity, T_0 is the reference temperature (K), and ρ_0 is the reference density (kg m^{-3}). As in previous studies (Bechtold et al., 2014; I.-H. Chen et al., 2023; Matsunobu et al., 2024), we use an effective convective adjustment time scale by introducing an extra

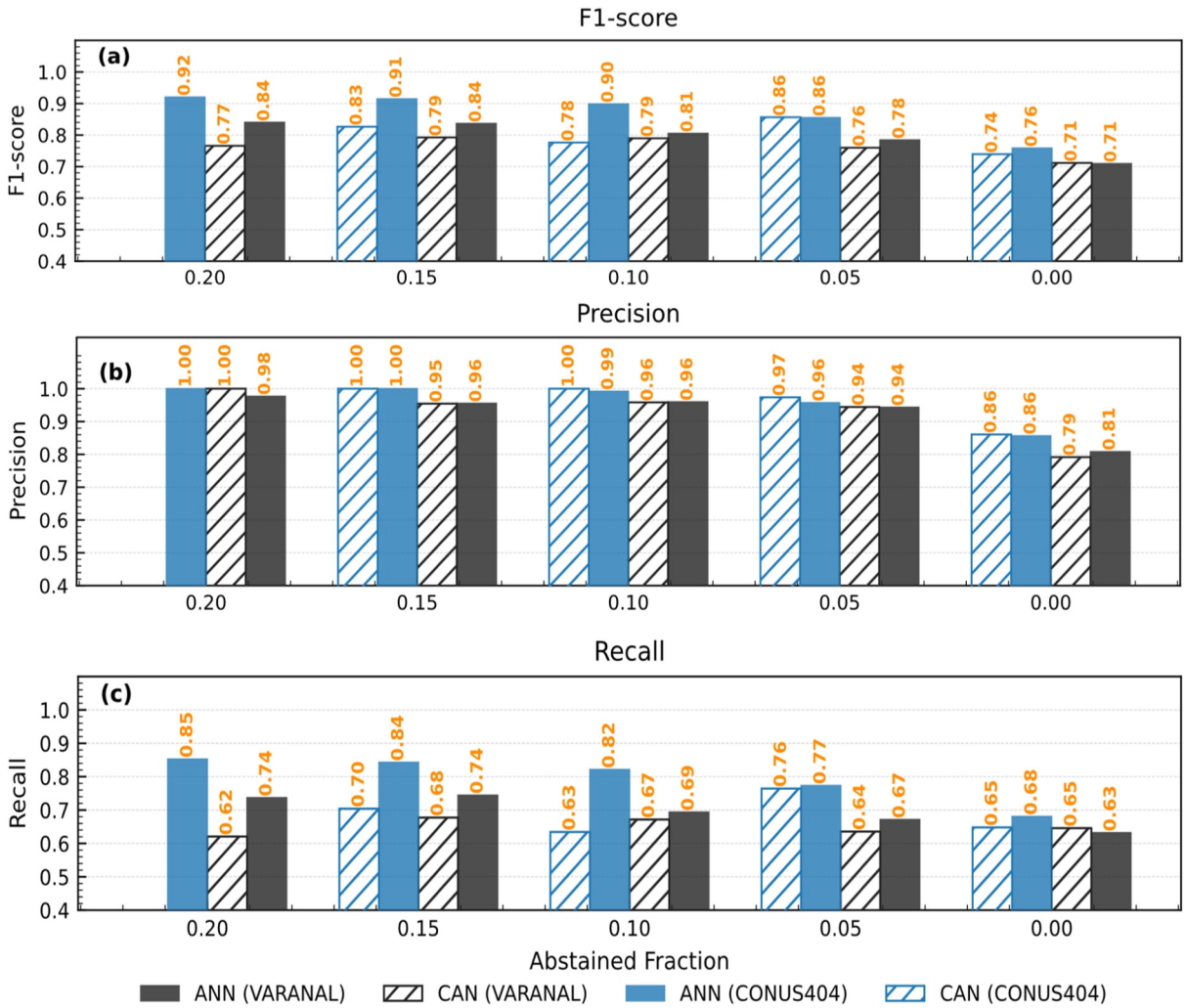


Figure 1. Model Evaluation using (a) F1-score, (b) Precision, and (c) Recall by varying the fraction of abstention samples in the VARANAL and CONUS404 data sets.

factor $b = 0.5$ in (3). This factor accounts for the reduction of CAPE by other aspects of convection, such as cooling and drying of the boundary layer.

3. Results

3.1. Evaluation of Model Performance

We first evaluated whether the uncertainty-aware framework provides useful information beyond overall classification skill. Model performance on identifying the deep convective events is assessed using F1, precision, and recall scores (Text S2 in Supporting Information S1) computed on the test sets after withholding a varying fraction of low-confidence samples. As shown in Figure 1, ANN and CAN exhibit similar baseline performance, with F1 scores of about 0.71 in VARANAL and about 0.75 in CONUS404 when all samples are included. Although these values are slightly lower than those reported in some previous studies (T. Zhang et al., 2021; Miller et al., 2025; M. Chen et al., 2023), our purpose is not to maximize prediction skill through a large predictor set or a highly complex model. Instead, we prioritize physical interpretability through reduced complexity over maximum accuracy.

A key result is that model skill generally increases when the least confident samples are excluded in most of the cases, except for CAN in the CONUS404 data set. The improvement is most evident at small abstention fractions, but performance degrades when too many samples are removed, particularly beyond about 15%, because an

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

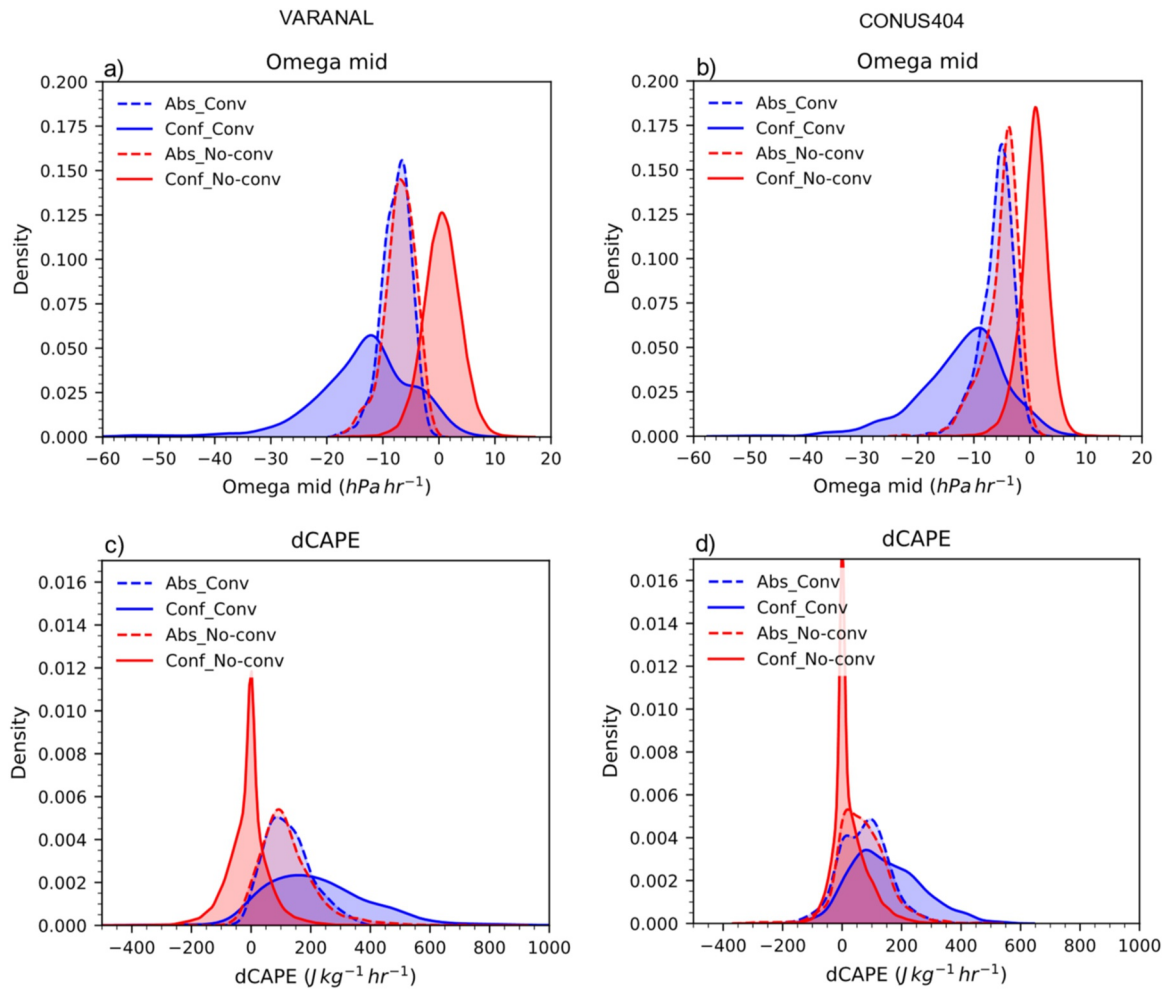


Figure 2. Probability density functions (PDFs) of (a, b) omega_mid and (c, d) dCAPE for the VARANAL and CONUS404 data sets. Colors denote confident convective (Conf_Conv), confident non-convective (Conf_No-conv), abstained convective (Abs_Conv), and abstained non-convective (Abs_No-conv) samples.

increasing number of convective events are assigned to the abstention class. This trade-off is reflected in the precision and recall scores: precision increases as low-confidence samples are removed, but recall decreases because more true convective events, which are relatively harder to predict, are assigned to the abstained class. The same mechanism likely explains inconsistent performance improvement in CONUS404 at higher abstention fractions. Overall, both models show higher precision than recall, indicating a tendency to miss some convective events. The NN exhibits higher skill when trained on CPMs output than on observations, likely because the model data set is more internally consistent and less noisy than the observation-constrained product.

3.2. Comparing the High and Low-Confident Regimes

Having shown that excluding low-confidence samples improves model skill, we now examine the physical environments associated with those samples. Using CAN, we isolated the lowest 5% of confidence samples (labeled "low-confidence") and compared them with the remaining high-confidence samples. After separating the samples by model confidence, we used the observed convection label to divide each group into convective and non-convective cases. The same analysis using ANN-based confidence yields similar results.

After the classification, we compared the environmental conditions between low-confidence (or abstained) and high-confidence samples (Figure 2), finding that the onset of deep convection is most predictable when the atmosphere is subject to strong upward motion and active generation of CAPE, and least predictable when these large-scale signals are weak. Figure 2 shows the probability density functions (PDFs) for mid-level omega and

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

dCAPE for both convective and non-convective samples across low- and high-confidence samples. In both data sets, low-confidence samples cluster in environments with weak-to-moderate ascent and modest positive dCAPE, where convective and non-convective cases overlap strongly (Figure 2). By contrast, high-confidence convective samples are associated with stronger ascent and larger positive dCAPE, while high-confidence non-convective samples are concentrated under weak ascent or subsidence and near-zero to negative dCAPE. We also examined other large-scale predictors, but they show weaker capability of distinction (Figures S2 and S3 in Supporting Information S1). These results indicate that ML uncertainty is concentrated in an intermediate synoptic-scale forcing regime, where the onset of convection is possible but not strongly constrained by the synoptic-scale environment.

To further understand the differences between confident and abstained convective regimes, we examine the time evolution of convective events and their environments using lead-lag composites centered on the time of deep convection occurrence ($t = 0$) (Figure 3). Here, $t = 0$ represents the time events at which domain-mean precipitation exceeds the threshold of 0.5 mm/hr. The confident and abstention samples peak at $t = 0$ by construction, but their temporal evolution differs. The confident samples show a broader curve with a gradual increase toward $t = 0$ and a similar gradual decrease afterward, whereas the abstention samples exhibit a sharper peak with a more rapid ramp-up immediately before $t = 0$ and a faster decay afterward. The variation of environmental predictors such as omega_mid (Figures 3c and 3d) and dCAPE (Figures 3e and 3f) shows a similar pattern. Confident samples are characterized by a strong ascent (more negative omega_mid) with a more pronounced local maximum at the time of the event ($t = 0$), whereas abstention samples show weaker variability and a flatter evolution in both data sets. This pattern is also apparent in dCAPE, with stronger dCAPE for confident samples compared with abstention samples. Although the variations in omega_mid and dCAPE are stronger in VARANAL than in CONUS404, the confident and abstention samples remain distinguishable in both data sets. This contrast indicates that confident cases are tied to more persistent and better-organized forcing (such as the MCS and the associated large-scale forcing), whereas low-confidence cases are associated with shorter-lived and weaker synoptic-scale perturbations (such as local convection driven by solar heating of the surface).

This interpretation is further supported by diurnal evolutions of τ_c (Figures 3g and 3h) across the confident and abstained samples. In both data sets, τ_c is systematically longer during the day than at night, consistent with the dominance of thermodynamically driven, non-equilibrium convection in the afternoon. Abstained samples consistently show larger τ_c, especially between local noon and late afternoon. The larger τ_c in abstained samples indicates that these events are associated with slower convective adjustment, where the atmosphere departs more from a convective quasi-equilibrium regime that is inherently harder to classify due to weaker coupling between large-scale forcing and convective response (Keil et al., 2014). Although no universal τ_c threshold has been established to definitively distinguish between equilibrium and non-equilibrium convection, the consistent offset between confident and abstained samples provides physically valid evidence that model uncertainty is higher for non-equilibrium short-lived convective regimes.

3.3. Synoptic Analyses

We have shown that high-confidence deep convection triggering occurs under stronger, large-scale forcing and with lower τ_c. To place these differences in a broader synoptic context, we composite the synoptic-scale circulation associated with the high- and low-confidence samples using 500-hPa geopotential height (GPH) and wind fields, together with 925-hPa moisture transport from ERA5 and CONUS404.

In both ERA5 and CONUS404, the SGP lies downstream of an upper-level trough in both the confident and abstained samples (Figures 4a–4d), a circulation pattern favorable for ascent over the central United States. The key difference is that the confident samples are associated with a deeper, more amplified trough, as indicated by stronger negative GPH anomalies at 500 hPa. A stronger trough enhances large-scale ascent ahead of it and more effectively destabilizes the column, creating an environment favorable for deep convection triggering (Juckes & Smith, 2000). In contrast, the abstained samples are associated with a weaker, more diffuse GPH anomaly, indicating weaker dynamical forcing.

The low-level circulation shows a similarly coherent contrast (Figures 4e–4h). Confident samples are associated with stronger southerly flow and enhanced moisture transport into the SGP at 925 hPa, consistent with a stronger low-level jet. This stronger low-level flow supplies warm, moist air from the Gulf and supports greater moisture convergence and instability over the region (Bowman & Rapp, 2024). When combined with stronger ascent aloft,

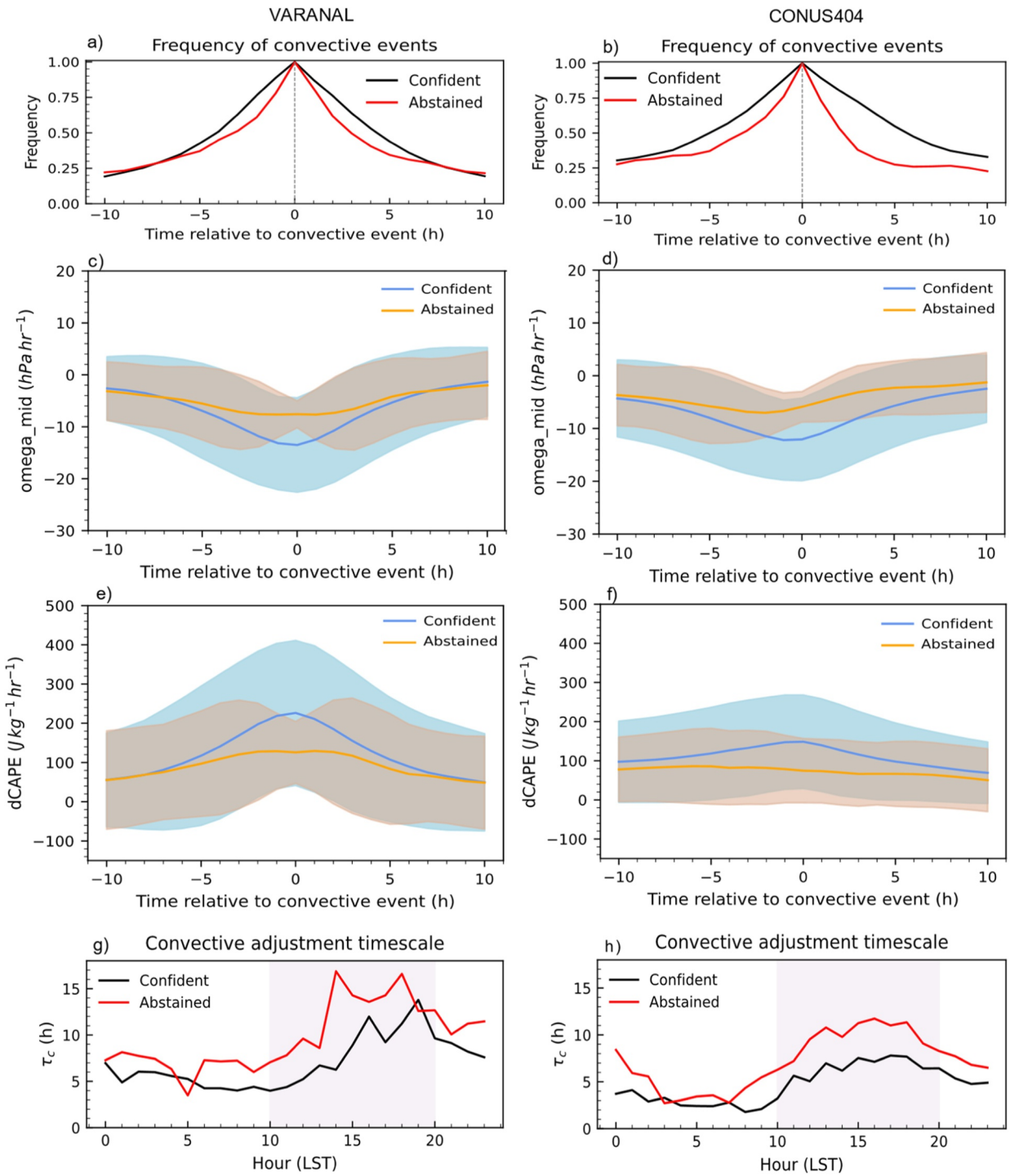


Figure 3. (a, b) Frequency of convective events; (c, d) time evolution of omega_mid; (e, f) time evolution of dCAPE; (g, h) convective-adjustment time scale, in the VARANAL and CONUS404 data sets. The solid lines in figures (c–f) represent the mean value of the quantities, and the shaded area represents ±1 S.D.

this vertical coupling between upper-level forcing and low-level moisture transport creates a more favorable and sustained environment for the onset of deep convection. In the abstained samples, both the low-level southerly flow (Figure S6 in Supporting Information S1) and moisture transport are weaker, suggesting that the environment remains capable of supporting convection, but with less large-scale support and a greater influence from local and short-lived processes.

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2026GL123315

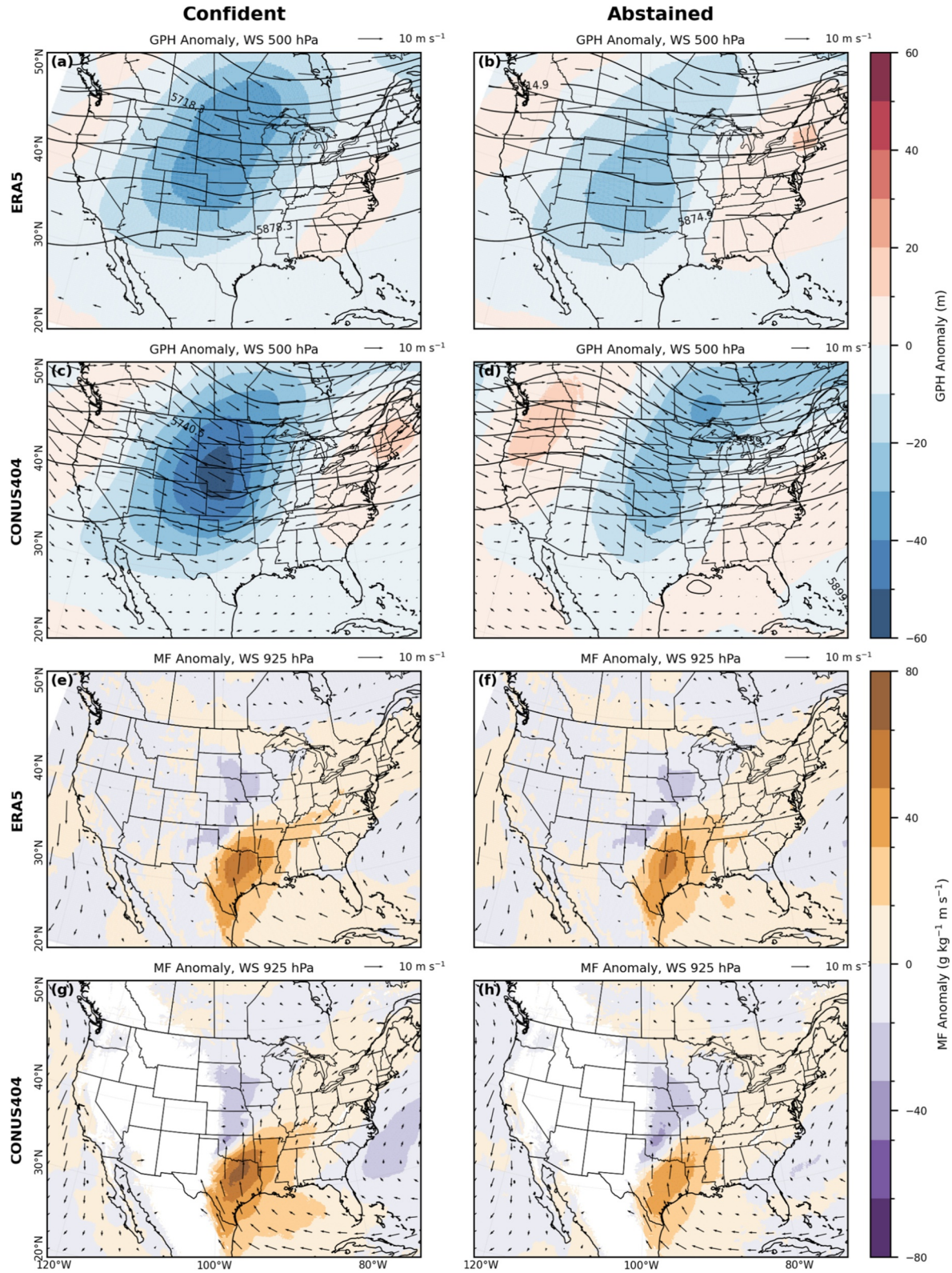


Figure 4. Synoptic-scale composites for ML model confident (left column) and abstained (right column) deep convection trigger predictions. (a, b) GPH contours, anomalies (shaded), and wind vectors at 500 hPa from ERA5; (c–d) same as (a–b) but from CONUS404; (e–f) moisture flux (MF) anomalies (shaded) and wind vectors at 925 hPa from ERA5; (g–h) same as (e–f) but from CONUS404. Anomalies are computed by subtracting the confident or abstained composite from the full training sample composite. The reference wind vector (10 m s^{-1}) is shown in the upper right of each panel.

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The synoptic composites lend physical support for the ML-based finding that onset of deep convection is more predictable when it is more strongly forced by the large-scale environment. This interpretation is consistent with earlier studies showing that precipitation and convection are more predictable when synoptic forcing is strong, but less predictable when outcomes depend more on local processes and internal variability (Keil et al., 2014, 2020). It is also consistent with Great Plains studies showing that organized convection is favored when low-level jets provide efficient moisture transport and convergence, helping to anchor onset of deep convection within a more coherent large-scale environment (Gebauer et al., 2018; Tang et al., 2017).

4. Conclusions

This study uses uncertainty-aware ML as a framework for diagnosing the predictability of deep convection triggering over the Southern Great Plains. By training conventional and abstention-based neural networks on both observationally constrained (VARANAL) and convection-permitting model hydroclimate (CONUS404) data sets, we show that prediction uncertainty is not simply a technical artifact of the ML model. Instead, it reflects how strongly the resolved atmospheric state constrains convective outcomes.

Both ML models achieved comparable baseline performance (F1 $\sim$ 0.71–0.75) using a compact set of nine physically interpretable predictors. Model skill improved as the least-confident samples were progressively excluded, demonstrating that uncertainty is not uniformly distributed but physically organized.

The central result is that ML-based deep convection predictions are most reliable when convection is organized by coherent large-scale forcing. These reliable cases are characterized by stronger mid-level ascent and larger dynamic CAPE generation. Lead-lag temporal composites further reveal that confident convective events are long-lived, with gradually intensifying signals consistent with synoptically forced, organized convection. Synoptic composites further explain this reliability: these cases tend to occur downstream of a more amplified upper-level trough, with stronger low-level southerly moisture transport into the Southern Great Plains. This vertically coupled synoptic environment supports organized, longer-lived convection (characterized by shorter convective-adjustment timescales) and helps explain why the ML model identifies these cases more confidently.

In contrast, prediction ambiguity is concentrated in weakly forced transition regimes. These cases occur in a transition regime characterized by weak-to-moderate ascent and modest dCAPE, where convective and non-convective environmental distributions overlap substantially. They are also associated with shorter-lived events, stronger daytime preference, and longer convective-adjustment timescales, consistent with locally forced, non-equilibrium convection. In such regimes, the convective outcome depends more strongly on local perturbations that are not fully captured by the predictor set.

Thus, the results provide physically grounded guidance for the operational use of ML-based convection triggering models over the SGP during the warm season. During convective regimes dominated by weak synoptic forcing, non-equilibrium, and afternoon diurnal heating, heightened caution should be followed, as these conditions may yield ambiguous outcomes. The consistency of these findings across the VARANAL and CONUS404 data sets strengthens confidence in their generality. Future work will extend this uncertainty-aware framework to spatially resolved and gridded data sets to better capture mesoscale and local-scale processes. Additionally, applying this framework across geographically diverse convective environments beyond the SGP would further test its generality and reveal whether the regime-dependent uncertainty structures identified here are regionally specific or represent a more universal characteristic of ML-based convection triggering models.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The ARM variational analysis data sets used in this study are available at https://www.arm.gov/data/science-data-products/vaps/varanal. The CONUS404 data sets are publicly available at https://www.usgs.gov/data/conus404-four-kilometer-long-term-regional-hydroclimate-reanalysis-over-conterminous-united. The ERA5 data is publicly available from the Copernicus Climate Data Store (CDS) (Hersbach et al., 2023), and can be accessed by selecting the variables, spatial domain, and time period described in this study. CAN codes are publicly available

at https://zenodo.org/records/5750222 (E. Barnes & Barnes, 2021). The post-processed data for model training, testing, and synoptic analysis, along with the code used in this study, are available at https://zenodo.org/records/19674393 (Bhattarai, 2026).

Acknowledgments
This work is supported by the University of Houston Presidential Frontier Faculty startup fund and DOE Early Career grant (DE-SC0024185). We acknowledge the resources provided by the Research Computing Data Core at the University of Houston. We thank the two anonymous reviewers for their constructive comments.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 2623–2631). Association for Computing Machinery. New York, NY, USA. https://doi.org/10.1145/3292500.3330701

Arakawa, A. (2004). The cumulus parameterization problem: Past, present, and future. *Journal of Climate*, *17*(13), 2493–2525. https://doi.org/10.1175/1520-0442(2004)017<2493:ratcpp>2.0.co;2

Barnes, E., & Barnes, R. J. (2021). eabarnes1010/controlled_abstention_networks: (version v1.0.1). *Zenodo*. https://doi.org/10.5281/ZENODO.5750222

Barnes, E. A., & Barnes, R. J. (2021a). Controlled abstention neural networks for identifying skillful predictions for regression problems. *Journal of Advances in Modeling Earth Systems*, *13*(12), e2021MS002575. https://doi.org/10.1029/2021MS002575

Barnes, E. A., & Barnes, R. J. (2021b). Controlled abstention neural networks for identifying skillful predictions for classification problems. *Journal of Advances in Modeling Earth Systems*, *13*(12), e2021MS002573. https://doi.org/10.1029/2021MS002573

Bechtold, P., Semane, N., Lopez, P., Chaboureau, J.-P., Beljaars, A., & Bormann, N. (2014). Representing equilibrium and nonequilibrium convection in large-scale models. *Journal of the Atmospheric Sciences*, *71*(2), 734–753. https://doi.org/10.1175/JAS-D-13-0163.1

Bhattarai, A. (2026). Supporting data for "Uncertainty-Aware machine learning for deep convection initiation: Under what conditions are trigger predictions more reliable? *Zenodo*. [Dataset]. https://doi.org/10.5281/ZENODO.19674393

Bommer, P. L., Kretschmer, M., Hedström, A., Bareeva, D., & Höhne, M. M.-C. (2024). Finding the right XAI method—A guide for the evaluation and ranking of explainable AI methods in climate science. *Artificial Intelligence for the Earth Systems*, *3*(3). https://doi.org/10.1175/AIES-D-23-0074.1

Bowman, K. P., & Rapp, A. D. (2024). Environmental controls on deep and overshooting convection over the contiguous U.S. *Journal of Geophysical Research: Atmospheres*, *129*(22), e2024JD041841. https://doi.org/10.1029/2024JD041841

Chen, I.-H., Berner, J., Keil, C., Kuo, Y.-H., & Craig, G. (2023). Classification of warm-season precipitation in high-resolution rapid refresh (HRRR) model forecasts over the contiguous United States. *Monthly Weather Review*, *152*(1), 187–201. https://doi.org/10.1175/MWR-D-23-0108.1

Chen, M., Fu, H., Zhang, T., & Wang, L. (2023). ResU-Deep: Improving the trigger function of deep convection in tropical regions with deep learning. *Journal of Advances in Modeling Earth Systems*, *15*(11), e2022MS003521. https://doi.org/10.1029/2022MS003521

Christensen, H. M., Kouhen, S., Miller, G., & Parthipan, R. (2024). Machine learning for stochastic parametrization. *Environmental Data Science*, *3*, e38. https://doi.org/10.1017/eds.2024.45

Colby, F. P. (1984). Convective inhibition as a predictor of convection during AVE-SESAME II. *Monthly Weather Review*, *112*(11), 2239–2252. https://doi.org/10.1175/1520-0493(1984)112<2239:ciaapo>2.0.co;2

Cui, Z., Zhang, G. J., Wang, Y., & Xie, S. (2021). Understanding the roles of convective trigger functions in the diurnal cycle of precipitation in the NCAR CAM5. *Journal of Climate*, *34*(15), 6473–6489. https://doi.org/10.1175/JCLI-D-20-0699.1

Dai, A. (2006). Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, *19*(18), 4605–4630. https://doi.org/10.1175/JCLI3884.1

Done, J. M., Craig, G. C., Gray, S. L., Clark, P. A., & Gray, M. E. B. (2006). Mesoscale simulations of organized convection: Importance of convective equilibrium. *Quarterly Journal of the Royal Meteorological Society*, *132*(616), 737–756. https://doi.org/10.1256/qj.04.84

Doswell, C. A., Brooks, H. E., & Maddox, R. A. (1996). Flash flood forecasting: An ingredients-based methodology. *Weather and Forecasting*, *11*(4), 560–581. https://doi.org/10.1175/1520-0434(1996)011<0560:fffaib>2.0.co;2

Emanuel, K. A. (1994). *Atmospheric convection*. Oxford University Press.

Flack, D. L. A., Plant, R. S., Gray, S. L., Lean, H. W., Keil, C., & Craig, G. C. (2016). Characterisation of convective regimes over the British Isles. *Quarterly Journal of the Royal Meteorological Society*, *142*(696), 1541–1553. https://doi.org/10.1002/qj.2758

Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmospheric Environment*, *32*(14), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0

Gebauer, J. G., Shapiro, A., Fedorovich, E., & Klein, P. (2018). Convection initiation caused by heterogeneous low-level jets over the great plains. *Monthly Weather Review*, *146*(8), 2615–2637. https://doi.org/10.1175/MWR-D-18-0002.1

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*(11), 5742–5751. https://doi.org/10.1029/2018GL078202

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023). ERA5 hourly data on pressure levels from 1940 to present. *Copernicus Climate Change Service (C3s) Climate Data Store (Cds)*, *10*, 24381. https://doi.org/10.24381/cds.bd0915c6

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., et al. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602. https://doi.org/10.1175/BAMS-D-15-00135.1

Houze, R. A., Jr. (2004). Mesoscale convective systems. *Reviews of Geophysics*, *42*(4). https://doi.org/10.1029/2004RG000150

Jakob, C. (2010). Accelerating progress in global atmospheric model development through improved parameterizations: Challenges, opportunities, and strategies. *Bulletin of the American Meteorological Society*, *91*(7), 869–876. https://doi.org/10.1175/2009BAMS2898.1

Juckes, M., & Smith, R. K. (2000). Convective destabilization by upper-level troughs. *Quarterly Journal of the Royal Meteorological Society*, *126*(562), 111–123. https://doi.org/10.1002/qj.49712656206

Kain, J. S. (2004). The kain–fritsch convective parameterization: An update. *Journal of Applied Meteorology and Climatology*, *43*(1), 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:tkcpau>2.0.co;2

Keil, C., Chabert, L., Nuissier, O., & Raynaud, L. (2020). Dependence of predictability of precipitation in the northwestern Mediterranean coastal region on the strength of synoptic control. *Atmospheric Chemistry and Physics*, *20*(24), 15851–15865. https://doi.org/10.5194/acp-20-15851-2020

Keil, C., Heinlein, F., & Craig, G. C. (2014). The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quarterly Journal of the Royal Meteorological Society*, *140*(679), 480–490. https://doi.org/10.1002/qj.2143

Kober, K., Foerster, A. M., & Craig, G. C. (2015). Examination of a stochastic and deterministic convection parameterization in the COSMO model. *Monthly Weather Review*, *143*(10), 4088–4103. https://doi.org/10.1175/MWR-D-15-0012.1

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lee, M.-I., Schubert, S. D., Suarez, M. J., Schemm, J.-K. E., Pan, H.-L., Han, J., & Yoo, S.-H. (2008). Role of convection triggers in the simulation of the diurnal cycle of precipitation over the United States Great Plains in a general circulation model. *Journal of Geophysical Research*, *113*(D2). https://doi.org/10.1029/2007JD008984

Lin, J., Qian, T., Bechtold, P., Grell, G., Zhang, G. J., Zhu, P., et al. (2022). Atmospheric convection. *Atmosphere-Ocean*, *60*(3–4), 422–476. https://doi.org/10.1080/07055900.2022.2082915

Markowski, P., & Richardson, Y. (2010). *Mesoscale meteorology in midlatitudes*. Wiley-Blackwell. Chichester, West Sussex, UK ; Hoboken, NJ.

Matsunobu, T., Puh, M., & Keil, C. (2024). Flow- and scale-dependent spatial predictability of convective precipitation combining different model uncertainty representations. *Quarterly Journal of the Royal Meteorological Society*, *150*(761), 2364–2381. https://doi.org/10.1002/qj.4713

Miller, G. A., Stier, P., & Christensen, H. M. (2025). Characterizing uncertainty in deep convection triggering using explainable machine learning. *Journal of the Atmospheric Sciences*, *82*(6), 1093–1111. https://doi.org/10.1175/JAS-D-24-0085.1

Miyamoto, Y., Kajikawa, Y., Yoshida, R., Yamaura, T., Yashiro, H., & Tomita, H. (2013). Deep moist atmospheric convection in a subkilometer global simulation. *Geophysical Research Letters*, *40*(18), 4922–4926. https://doi.org/10.1002/grl.50944

Pryor, S. C., Letson, F., Shepherd, T., & Barthelmie, R. J. (2023). Evaluation of WRF simulation of deep convection in the U.S. Southern Great Plains. *Journal of Applied Meteorology and Climatology*, *62*(1), 41–62. https://doi.org/10.1175/JAMC-D-22-0090.1

Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*(11), 1547–1564. https://doi.org/10.1175/BAMS-84-11-1547

Rasmussen, R. M., Chen, F., Liu, C. H., Ikeda, K., Prein, A., Kim, J., et al. (2023). CONUS404: The NCAR–USGS 4-km long-term regional hydroclimate reanalysis over the CONUS. *Bulletin of the American Meteorological Society*, *104*(8), E1382–E1408. https://doi.org/10.1175/BAMS-D-21-0326.1

Retsch, M. H., Jakob, C., & Singh, M. S. (2022). Identifying relations between deep convection and the large-scale atmosphere using explainable artificial intelligence. *Journal of Geophysical Research: Atmospheres*, *127*(3), e2021JD035388. https://doi.org/10.1029/2021JD035388

Rio, C., Del Genio, A. D., & Hourdin, F. (2019). Ongoing breakthroughs in convective parameterization. *Current Climate Change Reports*, *5*(2), 95–111. https://doi.org/10.1007/s40641-019-00127-w

Schiro, K. A., Ahmed, F., Giangrande, S. E., & Neelin, J. D. (2018). GoAmazon2014/5 campaign points to deep-inflow approach to deep convection across scales. *Proceedings of the National Academy of Sciences*, *115*(18), 4577–4582. https://doi.org/10.1073/pnas.1719842115

Schiro, K. A., & Neelin, J. D. (2019). Deep convective organization, moisture vertical structure, and convective transition using deep-inflow mixing. *Journal of the Atmospheric Sciences*, *76*(4), 965–987. https://doi.org/10.1175/JAS-D-18-0122.1

Schumacher, R. S., & Johnson, R. H. (2005). Organization and environmental properties of extreme-rain-producing mesoscale convective systems. *Monthly Weather Review*, *133*(4), 961–976. https://doi.org/10.1175/MWR2899.1

Schumacher, R. S., & Johnson, R. H. (2008). Mesoscale processes contributing to extreme rainfall in a midlatitude warm-season flash flood. *Monthly Weather Review*, *136*(10), 3964–3986. https://doi.org/10.1175/2008MWR2471.1

Song, F., & Zhang, G. J. (2017). Improving trigger functions for convective parameterization schemes using GOAmazon observations. *Journal of Climate*, *30*(21), 8711–8726. https://doi.org/10.1175/JCLI-D-17-0042.1

Suhas, E., & Zhang, G. J. (2014). Evaluation of trigger functions for convective parameterization schemes using observations. *Journal of Climate*, *27*(20), 7647–7666. https://doi.org/10.1175/JCLI-D-13-00718.1

Tang, S., Tao, C., Xie, S., & Zhang, M. (2019). Description of the ARM large-scale forcing data from the constrained variational analysis (VARANAL) version 2.

Tang, S., Xie, S., Guo, Z., Hong, S.-Y., Khouider, B., Klocke, D., et al. (2022). Long-term single-column model intercomparison of diurnal cycle of precipitation over midlatitude and tropical land. *Quarterly Journal of the Royal Meteorological Society*, *148*(743), 641–669. https://doi.org/10.1002/qj.4222

Tang, S., & Zhang, M. (2015). Three-dimensional constrained variational analysis: Approach and application to analysis of atmospheric diabatic heating and derivative fields during an ARM SGP intensive observational period. *Journal of Geophysical Research: Atmospheres*, *120*(15), 7283–7299. https://doi.org/10.1002/2015JD023621

Tang, S., Zhang, M., & Xie, S. (2017). Investigating the dependence of SCM simulated precipitation and clouds on the spatial scale of large-scale forcing at SGP. *Journal of Geophysical Research: Atmospheres*, *122*(16), 8724–8738. https://doi.org/10.1002/2017JD026565

Tao, C., Xie, S., Ma, H.-Y., Bechtold, P., Cui, Z., Vaillancourt, P. A., et al. (2024). Diurnal cycle of precipitation over the tropics and central United States: Intercomparison of general circulation models. *Quarterly Journal of the Royal Meteorological Society*, *150*(759), 911–936. https://doi.org/10.1002/qj.4629

Tao, C., Zhang, Y., Tang, Q., Ma, H.-Y., Ghate, V. P., Tang, S., et al. (2021). Land–Atmosphere coupling at the U.S. Southern Great Plains: A comparison on local convective regimes between ARM observations, reanalysis, and climate model simulations. *Journal of Hydrometeorology*, *22*(2), 463–481. https://doi.org/10.1175/JHM-D-20-0078.1

Tian, B., Held, I. M., Lau, N.-C., & Soden, B. J. (2005). Diurnal cycle of summertime deep convection over North America: A satellite perspective. *Journal of Geophysical Research*, *110*(D8). https://doi.org/10.1029/2004JD005275

Tiedtke, M. (1989). A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Retrieved from https://journals.ametsoc.org/view/journals/mwre/117/8/1520-0493_1989_117_1779_acmfsf_2_0_co_2.xml

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically interpretable neural networks for the geosciences: Applications to Earth system variability. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS002002. https://doi.org/10.1029/2019MS002002

Trier, S. B., Chen, F., & Manning, K. W. (2004). A study of convection initiation in a mesoscale model using high-resolution land surface initial conditions. *Monthly Weather Review*, *132*(12), 2954–2976. https://doi.org/10.1175/MWR2839.1

Wang, Y.-C., Xie, S., Tang, S., & Lin, W. (2020). Evaluation of an improved convective triggering function: Observational evidence and SCM tests. *Journal of Geophysical Research: Atmospheres*, *125*(11), e2019JD031651. https://doi.org/10.1029/2019JD031651

Xie, S., Wang, Y.-C., Lin, W., Ma, H.-Y., Tang, Q., Tang, S., et al. (2019). Improved diurnal cycle of precipitation in E3SM with a revised convective triggering function. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2290–2310. https://doi.org/10.1029/2019MS001702

Xie, S., Xu, K.-M., Cederwall, R. T., Bechtold, P., Genio, A. D. D., Klein, S. A., et al. (2002). Intercomparison and evaluation of cumulus parametrizations under summertime midlatitude continental conditions. *Quarterly Journal of the Royal Meteorological Society*, *128*(582), 1095–1135. https://doi.org/10.1256/003590002320373229

Xie, S., & Zhang, M. (2000). Impact of the convection triggering function on single-column model simulations. *Journal of Geophysical Research*, *105*(D11), 14983–14996. https://doi.org/10.1029/2000JD900170

Yano, J.-I., Bister, M., Fuchs, Ž., Gerard, L., Phillips, V. T. J., Barkidija, S., & Piriou, J.-M. (2013). Phenomenology of convection-parameterization closure. *Atmospheric Chemistry and Physics*, *13*(8), 4111–4131. https://doi.org/10.5194/acp-13-4111-2013

Yano, J.-I., & Plant, R. S. (2012). Convective quasi-equilibrium. *Reviews of Geophysics*, *50*(4). https://doi.org/10.1029/2011RG000378

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Zhang, G. J. (2002). Convective quasi-equilibrium in midlatitude continental environment and its effect on convective parameterization. *Journal of Geophysical Research*, *107*(D14). ACL 12-1-ACL 12-16. https://doi.org/10.1029/2001JD001005

Zhang, G. J. (2003). Roles of tropospheric and boundary layer forcing in the diurnal cycle of convection in the U.S. southern Great Plains. *Geophysical Research Letters*, *30*(24). https://doi.org/10.1029/2003GL018554

Zhang, G. J. (2009). Effects of entrainment on convective available potential energy and closure assumptions in convection parameterization. *Journal of Geophysical Research*, *114*(D7). https://doi.org/10.1029/2008JD010976

Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model. *Atmosphere-Ocean*, *33*(3), 407–446. https://doi.org/10.1080/07055900.1995.9649539

Zhang, T., Lin, W., Vogelmann, A. M., Zhang, M., Xie, S., Qin, Y., & Golaz, J. (2021). Improving convection trigger functions in deep convective parameterization schemes using machine learning. *Journal of Advances in Modeling Earth Systems*, *13*(5), e2020MS002365. https://doi.org/10.1029/2020MS002365

Zimmer, M., Craig, G. C., Keil, C., & Wernli, H. (2011). Classification of precipitation events with a convective response timescale and their forecasting characteristics. *Geophysical Research Letters*, *38*(5). https://doi.org/10.1029/2010GL046199

19448007, 2026, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2026GL123315 by University Of Houston, Wiley Online Library on [19/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License