

A Physics-Based Ensemble Machine-Learning Approach to Identifying a Relationship Between Lightning Indices and Binary Lightning Hazard

1 **Andrew M. Thomas^{1*}, Stephen Noble¹**

2 ¹Atmospheric Technologies Group, Savannah River National Laboratory, Department of Energy,
3 Aiken, SC

4 *** Correspondence:**

5 Andrew Thomas

6 Andrew.thomas@srnl.doe.gov

7 **Keywords: lightning, machine-learning, numerical-weather-prediction, microphysics,**
8 **planetary-boundary-layer.**

9 **Abstract**

10 To convert lightning indices generated by numerical weather prediction experiments into binary
11 lightning hazard, a machine-learning tool was developed. This tool, consisting of parallel multilayer
12 perceptron classifiers, was trained on an ensemble of planetary boundary layer schemes and
13 microphysics parameterizations that generated four different lightning indices over one week. In a
14 subsequent week, the multi-physics ensemble was applied and the machine-learning tool was used to
15 evaluate the accuracy. Unintuitively, the machine-learning tool performed better on the testing
16 dataset than the training dataset. Much of the error may be attributed to mischaracterizing the
17 convection. The combination of the machine learning model and simulations could not differentiate
18 between cloud-to-cloud lightning and cloud-to-ground lightning, despite being trained on cloud-to-
19 ground lightning. It was found that the simulation most representative of the local operational model
20 was the most accurate simulation tested.

21 **1 Introduction**

22 While lightning risk has been steadily decreasing in the United States (Holle, 2016), exposure to
23 lightning is projected to increase in a warming climate (Romps et al., 2014;Romps, 2019). Lightning
24 indices, which quantify lightning hazards derived from numerical weather prediction models, are an
25 important means of alerting those exposed to the hazard (Price and Rind, 1994;McCaul et al.,
26 2009;Romps et al., 2014). The lightning potential index (LPI) (Yair et al., 2010) is one index that is
27 useful for cloud-resolving models and is well-correlated with lightning strikes (Lynn and Yair,
28 2010;Gharaylou et al., 2019). The LPI quantifies the charge separation within the charging zone

29 using model state variables via integrating the vertical velocity squared and a microphysical scaling
30 parameter through the charging zone (the vertical column between 0°C and -20°C). While there are
31 more sophisticated methods of predicting lightning (Lynn et al., 2012;Gharaylou et al., 2020), the
32 computational efficiency of the LPI makes it more useful in an operational setting. One alternative to
33 the LPI is the product of convective available potential energy (CAPE) and convective precipitation
34 (hereafter referred to as CAPE-P). CAPE-P relies under the assumption that if there is convective
35 precipitation, then lightning should be generated. This seems to perform better in some areas, such as
36 the boreal region (Mortelmans et al., 2022), and over land (Romps et al., 2018). Two more useful
37 indices are McCaul’s lightning threat index (McCaul et al., 2009) (LTI) and the index proposed by
38 Price and Rind (1994) (PR92W). The LTI combines the vertical movement of graupel at -15°C and
39 the vertical integral of cloud ice and hydrometeors. PR92W relies on correcting the maximum updraft
40 velocity for resolution. PR92W and CAPE-P were originally designed for climate models, where
41 resolution requirements forced parameterization of convection, and both indices were used to make
42 inferences for lightning. While the LPI has been shown to be a better indicator of lightning than some
43 of the other indices described (Saleh et al., 2023), the consideration of multiple lightning indices may
44 be more informative than any one lightning index.

45 Previous studies of lightning index skill focus on particular case studies. Malečić et al.
46 (2022) examined three cases with a large amount of hail and found systematic underprediction in the
47 amount of lightning predicted using an ensemble of microphysics and planetary boundary layer
48 (PBL) schemes by using the LPI. Lynn and Yair (2010) used two cases where there was a significant
49 amount of lightning and found that the time-averaged LPI was well correlated with accumulated
50 lightning flash density over space. The examination of the most significant lightning events may
51 favor model performance, whereas the “garden variety” thunderstorm may be overlooked. These
52 types of unorganized, non-severe storms tend to lead to the most lethal lightning incidents (Ashley
53 and Gilson, 2009). Some studies examine the global distribution of lightning through climate change
54 (Price and Rind, 1992;Finney et al., 2014;Romps et al., 2014).

55 These “garden variety” or “pulse thunderstorms” are regular occurrence to the southeastern
56 United States (Miller, 2017) with topography influencing the spatial variability (Miller and Mote,
57 2017). The observed positive feedback mechanism between the soil moisture and atmosphere
58 (Findell and Eltahir, 2003) resulting in high precipitation recycling (Dominguez et al., 2006) and a
59 subtropical climate promote these non-severe thunderstorms. Therefore, the southeastern United
60 States during the warm season is a suitable environment to study non-severe thunderstorms. This

61 study will establish relationships between the four previously mentioned lightning indices, as
62 computed by a physics-based ensemble of Weather Research and Forecasting (WRF) model
63 simulations and observations of lightning through the use of a multilayer perceptron classifier.

64 There is some regional and event-specific variation to the relationship between lightning
65 indices and lightning flashes (Yair et al., 2010). For example, while the LPI is clearly indicative of
66 lightning, there is currently no universal relationship between the two. One method for mitigating this
67 is to form a regression on the sorted values (Brisson et al., 2021; Mortelmans et al., 2022). This
68 disregards amplitude error, favoring smoothing out spatiotemporal errors associated with the model.
69 Other circumstances, such as the seasonality or event-specific nature of lightning, are also
70 outstanding factors. Lead time or model configuration may be influential in identifying the
71 predictability of lightning strikes in a future study. Disentangling the diagnostic capacity of lightning
72 indices from model error is a prerequisite for identifying a generalizable, albeit local, relationship for
73 predicting lightning.

74 We hypothesize that identifying a mean classification relationship from a physics-based
75 model ensemble should reduce the bias associated with individual simulations. To this end, we
76 trained a series of machine-learning based tool on a physics-based ensemble that will test different
77 configurations of the microphysics and PBL schemes. The tool should combine lightning indices to
78 identify a probability of cloud-to-ground lightning. This tool should not be reliant of model
79 configuration, as to compare the model performance. The microphysics schemes should address
80 variances in knowledge concerning microphysical development, while PBL schemes concern
81 variances in the development of convection and the supply of moisture. Both are important to the
82 development of convection, and especially lightning. This should lead to an ensemble average of
83 probabilities based on indices that had variations from microphysics and PBL parameterizations. The
84 effectiveness of this ensemble averaged probability will be tested on a subsequent period on each
85 model run to compare model configurations for optimal use. The metrics to be evaluated include
86 correlation coefficient, elements of confusion matrices, and the standard deviation normalized by the
87 observations.

88 **2 Methods**

89 Operational forecasts are issued by the Atmospheric Technologies Group (ATG) of the Savannah
90 River National Laboratory (SRNL) to protect workers at the Savannah River Site (SRS). SRNL and
91 SRS have a wealth of meteorological observations for the southeastern United States, which is a

92 region of increasing research interest regarding convection (Rasmussen, 2015;C Kuang et al.,
93 2023;Kosiba, 2023). SRS is a vast industrial facility with a large contingent of outdoor workers (over
94 1,000). ATG provides tailored hyper-local forecasts to assist in both weather-sensitive operations and
95 worker safety. ATG is interested in producing guidance regarding lightning forecasts. Daily
96 numerical weather model predictions from the Weather and Research Forecasting model (WRF)
97 (Skamarock et al., 2019), with 36 hours of lead time, are used to assist forecasters (Figure 1). The
98 nested domain (D02) (Figure 1) is our area of interest. To test the implementation of the lightning
99 indices into routine forecast operations, we have adopted the configuration described in Table 1. To
100 create our 12-member set of experiments, we adjusted the microphysics parameterization, the
101 boundary layer parameterization, and the surface layer parameterization (Table 2).

102 For some lightning indices, namely the LPI and LTI, the microphysical mixing ratios are crucial
103 for calculation. The microphysics parameterization modulates latent heating, which is a requirement
104 for moist convection. The PBL parameterization is also an important feature, as it modifies the
105 vertical transfer of heat and moisture that triggers convection. Vertical fluxes of moisture and heat
106 have an important role in the development of convection and supply moisture from the surface. This
107 is relevant, since the eastern US experiences a positive feedback mechanism between surface
108 moisture and precipitation (Findell and Eltahir, 2003).

109 The specific parameterizations (Table 2) were chosen to describe commonly used PBL and
110 microphysics parameterizations according to a WRF Physics Use survey from August 2015. The
111 control run (CTRL) for this experiment most resembles the operational configuration, which uses the
112 Thompson microphysics scheme (Thompson et al., 2008) and the third order Mellor-Yamada-
113 Nakanishis-Niino (MYNN) scheme (Nakanishi and Niino, 2006;2009;Olson et al., 2019).. The other
114 microphysics schemes tested include the Morrison (Morrison et al., 2009) and WRF Single Moment
115 6-Class (WSM6) (Hong and Lim, 2006) schemes. The other PBL schemes examined include the
116 Yonsei University (YSU) (Hong et al., 2006), the Mellor-Yamada-Janjic (MYJ) (Mesinger,
117 1993;Janjić, 1994) and MYNN scheme (Nakanishi and Niino, 2006;2009;Olson et al., 2019),
118 including the 2.5 (denoted as MYN2) and 3rd order scheme (MYN3). The YSU scheme is a nonlocal
119 PBL scheme, while the MYJ is a 1.5-order turbulence closure model, and the MYN2 and MYN3 use
120 a 2.5 and 3rd order local turbulence closure. WSM6 is the only single moment (bulk mixing ratios
121 only) microphysics scheme considered, while the Thompson and Morrison are both double-moment
122 (mixing ratios and concentrations) microphysics parameterizations.

123 Two weeks of 36-hour simulations were conducted for each experiment, with the first
124 initialization on 7 July 2022 at 6 UTC and the last simulation initialized 20 June 2022 at 6 UTC. One
125 slight difference between the experimental setup described and the operational settings is that
126 operationally, nudging of in situ temperature, humidity, and wind observations at the SRS is
127 implemented. Although observations are available for the entirety of this study, incorporating
128 observations via nudging may reduce the differences between experiments, and thus dilute our result.
129 While routine simulations are conducted every 6 hours, only simulations initialized at 6 UTC were
130 conducted for this study. The reasoning for this is that a forecaster may place more weight on recent
131 initializations rather than previous initializations and 6 UTC runs are the most current to issue the
132 morning forecast. During this period, lightning was abundant and occurred on most days (Figure 2).
133 The National Lightning Detector Network (NLDN) (Murphy et al., 2021) was chosen since it
134 measures cloud-to-ground lightning strikes, which is more relevant to the safety of outside workers.
135 The number of lightning strikes varied during this period, with a peak exceeding 400 strikes, as well
136 as a single strike in a 15-minute period. There were some time periods where lightning was not
137 observed in the NLDN but observed by the Geostationary Lightning Mapper (GLM), namely 18 and
138 19 July 2022. That is, if lightning was observed by the GLM, which observes all lightning by
139 detecting flashes in the visual range on the Geostationary Operational Environmental Satellite, but
140 not reported by the NLDN, which observes cloud-to-ground lightning, then discrepancies between
141 the two may be caused by cloud-to-cloud or cloud-to-air lightning. While we primarily used the
142 NLDN for training the model, the GLM was useful for independent verification in the analysis.

143 We used four lightning indices to diagnose lightning. The Lightning Potential Index (LPI) was
144 calculated by WRF's internal implementation and is computed by integrating the product of the
145 vertical velocity squared against the ratio of the geometric mean to the arithmetic mean of liquid and
146 frozen hydrometeors. The product of CAPE and precipitation rate (CAPE-P) was computed by
147 multiplying the precipitation rate, which was calculated by applying centered finite differencing to
148 accumulated precipitation, by the CAPE as computed by WRF-Python (Ladwig, 2017). The PR92W
149 index uses the maximum vertical velocity, similar to that found by Price and Rind (1992), but scaled
150 by the resolution (Wong et al., 2013). The McCaul's Lightning Threat index (LTI) uses a
151 combination of vertical graupel flux and total hydrometeors. LTI, LPI, and PR92W all use the
152 vertical velocity, while CAPE-P simply uses CAPE as a proxy for the vertical velocity. LTI and LPI
153 directly use the 3D mixing ratios of microphysics, while CAPE-P uses the precipitation rate to infer
154 the convection being generated and PR92W does not consider the microphysical state. While both

155 LPI and PR92W can be computed internally within WRF, all calculations of the lightning indices
156 were performed offline.

157 Model errors in convective initiation were accounted for by both smoothing and sorting. First
158 a three-hour moving average was applied to both lightning indices and gridded binary NLDN
159 observations. Then the model and observed values were sorted at each timestep, retaining
160 chronological relevance. This is similar to Brisson et al. (2021) and Mortelmans et al. (2022) which
161 sorted all values in space and time, but is altered for operational value since each model run
162 comprises 36 hours of predictions. While this method of spatiotemporal reorganization may result in
163 less pristine relationships between lightning indices and lightning observations, it will prevent
164 associating lightning indices with observations potentially 36 hours prior. Moreover, it should be
165 more analogous to how lightning would be operationally evaluated than absolutely sorting time and
166 space. Indices are still sorted in space to account for discrepancies in location of model convection,
167 but smoothing and temporal sequencing retains the temporal relevance. This processing was only
168 applied to the data ingested by a machine-learned (ML) member.

169 Each experiment was divided into two datasets. Simulations initialized between 7 and 13 July
170 2022 were designated as a training dataset, while simulations initialized between 14 and 20 July 2022
171 were designated as a testing dataset. The testing dataset may then be a mostly independent dataset,
172 with the exceptions of being closely related in time as well as the 12-hour overlap between the
173 experiments initialized on the 13 July and 14 July. The lightning indices of the training dataset of
174 each experiment were converted into the standard score (subtracted by the mean and divide by the
175 standard deviation) by the Standard Scaler, and then underwent Principal Component Analysis
176 (PCA) via scikit-learn (Pedregosa et al., 2011). Each principal component is a linear combination of
177 the four lightning indices that is orthogonal to the other components. Orthogonality is important for
178 reducing the multicollinearity that arises when several indices are related on a similar physical
179 process- namely the generation of lightning. The first two components of the PCA were then used to
180 train a multilayer perceptron classifier-type of neural network on binary observations of lightning.
181 The pipeline of the StandardScaler, which divides the mean-subtracted variable by its trained
182 variance, to PCA to multilayer perceptron classifier (MLPClassifier) is denoted a ML member, with
183 each experiment associated with a ML member. The MLPClassifier can provide not just a
184 classification, but also a probability for each classification by applying a forward pass. Some
185 hyperparameters, shown in Table 3, were chosen via an exhaustive grid search with cross-validation
186 (GridSearchCV), maximizing the equitable threat score. The other hyperparameters, some of which

187 were dependent on the results of GridSearchCV, were found by testing randomly selected values on a
188 normal distribution (RandomizedSearchCV) to further optimize the equitable threat score. Details
189 regarding the final hyperparameters and the mean equitable threat score as applied to the (processed)
190 training dataset of the MLPClassifiers are found in Supplementary Table 1. Thus with 12 ML
191 members available, a composite mean of probabilities of lightning can be attained.

192 After a ML member was trained on each experiment during the training period, the ensemble
193 of ML members was applied to each experiment within the testing dataset. The average of the
194 probabilities was used, which ideally would make a composite estimator for determining model
195 performance, creating an impartial postprocessed binary lightning product for identifying model
196 performance. This process is depicted in Figure 3. One advantage to this is the generation of multiple
197 probabilities gives a possible range and confidence to the estimate, but we do not explore this aspect
198 in any detail. There are several weaknesses to this, such as biases in the training data leading to
199 biases in the estimator. While taking an average of the probabilities should lead to an impartial
200 probability and attempts to avoid overfitting, the biases in each trained dataset can lead to biases in
201 evaluating the testing dataset. Additionally, the composition of several neural networks and training
202 on a recurrent phenomenon limits the scope of the usability and portability of this composite of ML
203 members. One other downside of having different PCAs associated with different experiments brings
204 up an issue of not being able to intercompare the components of each PCA.

205

206 **3 Results**

207 **3.1 Fittedness of ML members on training data**

208 Table 4 shows the percentage of the contribution, which we define as the absolute value of the
209 coefficients normalized by the sum of the absolute values. Since the variables are standardized by
210 subtracting the mean and dividing by the standard deviation, the absolute values depict the total
211 contribution to each Principal Component (PC). For the first component, each of the indices have
212 similar contributions, though both CAPE-P and PR92W have slightly more of a contribution than
213 LPI or LTI. For PC2, the LPI is most influential in the Morrison and WSM6 microphysics schemes,
214 while the LTI is most influential in the Thompson microphysics scheme. The approximate weighting
215 of PC1 between all 4 indices, with slight weighting towards CAPE-P and PR92W suggests that PC1
216 represents the presence of convection. For PC2, the LPI and LTI both rely heavily on microphysics,
217 suggesting secondary importance on the microphysical state as compared to the presence and

218 activation of convection. In short, PC1 is indicative of the presence, or at least the conditions
219 conducive of, convection, while PC2 describes the sufficiency of the microphysics for generating
220 lightning. This assessment of the similarity in performance, but with weighting toward the inclusion
221 of microphysical processes, is consistent with the findings of Saleh et al. (2023).

222 Taylor diagrams (Taylor 2001) can map out the accuracy of the composite mean of ML members
223 applied to each experiment, relative to the observed mean. On the radial axis is the normalized
224 standard deviation (relative to observations), while the azimuth contains the correlation coefficient.
225 The distance from a “Perfect” prediction is proportional to the root mean squared error. For a
226 timeseries, the angle along the arc connecting the observed mean to the perfect prediction (which will
227 be referred to as the critical arc) denotes phase error. The corollary for a spatiotemporal dataset is
228 differences in timing or space. By contrast, the radial distance from the arc is indicative of an under-
229 or over-amplified signal. As applied to our problem regarding lightning prognostication, the Taylor
230 diagram is advantageous as it may assess the probabilities of lightning to the boolean observations, as
231 compared to other diagrams that may require a threshold that could be considered arbitrary or
232 subjective based on allowable risk. To summarize, Taylor diagrams are a useful way of comparing
233 the variability (radial coordinate) and correlation (angular coordinate), relative to observations, for
234 members of an ensemble.

235 Figure 4 shows a Taylor Diagram for the spatial maximum probability from the composite mean
236 of the ML members, as applied to the unprocessed training dataset. The clustering of the experiments
237 suggest consensus. All of the experiments show overpredictions in the variability. The departure
238 from the “Observed mean” indicates some informedness regarding the timing of the lightning. The
239 departure from the critical arc (bolded) may be the result of errors in convection production (i.e.
240 lightning indices are high before or after lightning was produced), the result of the processing applied
241 to the training dataset prior to the fitting of the data, or the production of either cloud-to-ground or
242 cloud-to-air lightning within the model. Although this performance is drastically improved when
243 applied to the processed training dataset (not shown) applying this to the unprocessed dataset is more
244 representative of the accuracy a forecaster at SRS gauging lightning potential may experience.

245 **3.2 Performance of the composite mean of ML members and experiments.**

246 Figure 5 applies the same concept of the Taylor diagram to the spatial maximum probability of
247 the testing dataset. As expected, the composite mean probability of the ML members was not as
248 accurate for the testing dataset as the training dataset. The composite mean probability of the ML

249 members was far more variable for the testing dataset than the training dataset. This increase in
250 variability may be the result of forecast periods where all of the lightning generated was not cloud-to-
251 ground. This is reinforced by the ratio of cloud to ground lightning to total lightning, which can be
252 inferred by Figure 2.

253 In order to reduce the influence of spatial error, the maximum probability of lightning was
254 compared to the presence of lightning in the domain. Figure 6 shows a timeseries of what a forecaster
255 may expect out of each experimental configuration on 14 July 2022. All of the experiments were able
256 to show high probability of lightning for the period with long periods of cloud-to-ground lightning,
257 though with some timing issues. The MYN2 experiments with double moment microphysics
258 (MYN2_Thomp, MYN2_Morr) were the only experiments that produced high probabilities for the
259 lightning event near 15 July at 04 UTC, though those also had timing issues. The WSM6 experiments
260 also seem to be overly conservative with lightning probabilities in the YSU, MYN2, and MYN3
261 experiments leading to more underestimates of lightning.

262 Figure 7 shows a forecast period, initialized 18 July 2022 06 UTC, where no cloud-to-ground
263 lightning was observed, making all predictions of lightning false positives. Yet high probabilities of
264 cloud-to-cloud or cloud-to-air lightning were observed by the GLM. This suggests that the composite
265 mean of the ML members is incapable of discriminating cloud-to-ground lightning, which impacts
266 workers on the ground, from lightning that may not impact workers. Figure 7 also shows how much
267 erroneously predicted thunderstorms may impact forecast accuracy. All of the experiments indicate
268 high probabilities of lightning over a long period for a brief duration of observed lightning. By 19
269 July at 04 UTC, all of the experiments indicate that the lightning event, which observations suggest
270 was more succinct than the observations suggest, spurious convection then increased the probability
271 at approximately 19 July at 08 UTC. Here, WSM6 experiments predicted higher (more erroneous)
272 probabilities than the Morr or Thomp experiments. Towards the end of the forecast period, at near 16
273 UTC on 19 July, all the experiments except for MYN3 predict non-negligible probabilities of
274 lightning, leading to false positives.

275 If the probabilities are used to decide on lightning or no lightning, a confusion matrix may be
276 computed. We applied a threshold of 50% to the composite mean probability of the ML members as
277 applied to the testing dataset to make confusion diagrams for each experiment, as shown in Table 5.
278 Although intuitive, true negatives are the most frequently observed occurrence with probabilities
279 ranging from 63-71%. WSM6, which is the only representative of the single moment microphysics
280 schemes evaluated in our choice of parameterizations, has the lowest true positive rate, with

281 MYN3_WSM6 having more false positives than true positives. The CTRL (represented as
282 MYN3_Thomp) not only has more true negatives than any of the other experiments, CTRL has more
283 true negatives than some experiments have total correct predictions (such as MYJ_Morr or
284 MYN2_Thomp).

285 **4 Discussion**

286 Identifying the relationship between lightning indices and lightning is not a trivial task. Research
287 is available for guidance, but there is currently no known universal relationship for multiple lightning
288 indices. Previous research has generated possible relationships between the two, but those
289 relationships were formed using spatiotemporal sorting which presumes perfect matching of
290 distributions and disregards the potential for model error. Those studies also focus on cases where
291 vast quantities of lightning occur, which occur in scenarios where lightning fatalities may not be as
292 representative of lightning events that are likely to cause injuries. In “garden variety” thunderstorms,
293 lightning may still be present, despite weakly forced convection, resulting in lower lightning indices,
294 and more risk to outdoor workers.

295 There are also several indices that may be used to aid in lightning prediction. And yet these
296 indices are confounded with multicollinearity, suggesting that they diagnose a similar attribute,
297 namely convection. Within our training dataset, we identified a maximum variance inflation factor of
298 24 for one experiment. The use of Principal Component Analysis to form a regression relationship
299 (called Principal Component Regression or PCR) was therefore applied to reduce not just the
300 dimensionality, but also the multicollinearity of the indices. This PCR was used to train a
301 classification mechanism to determine the probability of lightning for each model. With a probability
302 of lightning attained, each model was evaluated according to the results of the probabilities generated
303 during the testing dataset, which were simulations conducted during the latter half of the period of
304 interest.

305 One reasoning behind the choice to examine the combination of microphysics and PBL
306 schemes was to identify possible changes to CTRL to optimize the prediction of lightning. The
307 results of this study suggest that no deviations regarding the microphysics or PBL scheme are
308 necessary but reinforce the initial choice of model settings. Moreover, this study provides guidance
309 regarding accuracy of this tool for lightning prediction. However, using methods to generate
310 ensembles that may attempt to change the timing or distribution of convection, such as perturbing the

311 initial and boundary conditions, may be more beneficial for training ML members than preprocessing
312 or physics-based ensembles.

313 Despite of the ability of the composite mean of ML members to produce probabilistic
314 estimates of lightning, false positives resulted from an overproduction of convection and/or the
315 generation of lightning that was not cloud-to-ground. One way to improve upon this is to apply
316 machine learning to spatial distributions of convection to try to discern between lightning trajectories.
317 This would require a study over a longer period. The idea would be to describe the ratio of cloud-to-
318 ground to total lightning through spatial distributions of convection.

319 Another avenue of research is to extend this analysis to timescales of a year or longer, which
320 will have a variety of mesoscale drivers for lightning. Training a ML member on an extended period
321 should provide a diverse set of environments, whereas this study focused on a consistent
322 environment. Testing the composite ML tool formed from this study may be a suitable control
323 compared to the machine learning tool formed from the proposed study. Besides the amount of data
324 required to be stored, another downside may be choosing separate training and testing datasets that
325 do not have a seasonal bias. This may be remedied by choosing systematic sampling, such as
326 conducted in this study.

327

328 **5 Conflict of Interest**

329 *The authors declare that the research was conducted in the absence of any commercial or financial*
330 *relationships that could be construed as a potential conflict of interest.*

331 **6 Author Contributions**

332 AT: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources,
333 Software, Validation, Visualization, Writing- original draft, Writing- review & editing; SN: Data
334 curation, Funding acquisition, Project administration, Supervision, Writing- review & editing.

335 **7 Funding**

336 This work was supported by the Laboratory Directed Research and Development (LDRD) program
337 within the Savannah River National Laboratory (SRNL). This document was prepared in conjunction
338 with work accomplished by Battelle Savannah River Alliance, LLC under Contract No.
339 89303321CEM000080 with the U.S. Department of Energy. Publisher acknowledges the U.S.
340 Government license to provide public access under the DOE Public Access Plan
341 (<http://energy.gov/downloads/doe-public-access-plan>).

342 **8 Acknowledgments**

343 The authors would like to thank Dr. Steve Weinbeck for the helpful discussions, as well as Dr. David
344 Werth for proofreading and providing general guidance and Dr. Steven Chiswell for the initial
345 configuration of WRF. **Data Availability Statement**

346 The WRF model (available at <https://github.com/wrf-model/WRF>), combined with the NAM model
347 (available at <https://www.ncei.noaa.gov/data/north-american-mesoscale-model/access/analysis/>) was
348 used to generate the meteorological model data. The machine learning models were generated using
349 the scikit-learn module (available at <https://scikit-learn.org/stable/>). The NLDN data is a commercial
350 dataset, available at <https://www.vaisala.com/en/products/national-lightning-detection-network-nldn>.
351 GLM data may be found at
352 [https://www.avl.class.noaa.gov/saa/products/search?sub_id=0&datatype_family=GRGLMPROD&su](https://www.avl.class.noaa.gov/saa/products/search?sub_id=0&datatype_family=GRGLMPROD&submit.x=12&submit.y=8)
353 [bmit.x=12&submit.y=8](https://www.avl.class.noaa.gov/saa/products/search?sub_id=0&datatype_family=GRGLMPROD&submit.x=12&submit.y=8).

354 2 References

355

- 356 Ashley, W.S., and Gilson, C.W. (2009). A Reassessment of U.S. Lightning Mortality. *Bulletin of the*
357 *American Meteorological Society* 90, 1501-1518.
- 358 Brisson, E., Blahak, U., Lucas-Picher, P., Purr, C., and Ahrens, B. (2021). Contrasting lightning
359 projection using the lightning potential index adapted in a convection-permitting regional
360 climate model. *Climate Dynamics* 57, 2037-2051.
- 361 C Kuang, S.G., S Serbin, G Elsaesser, P Gentine, T Heus, M Oue, J Peters, J , Smith, A.S., A
362 Mccomiskey, M Jensen, a Sedlacek, P Kollias, a Vogelmann, H , and Morrison, M.P., D
363 Turner, (2023). "Science Plan for the Deployment of the Third ARM Mobile Facility to the
364 Southeastern United States at the Bankhead National Forest, Alabama (AMF3 BNF)".
365 (Richland, Washington).
- 366 Dominguez, F., Kumar, P., Liang, X.-Z., and Ting, M. (2006). Impact of Atmospheric Moisture
367 Storage on Precipitation Recycling. *Journal of Climate* 19, 1513-1530.
- 368 Findell, K.L., and Eltahir, E.a.B. (2003). Atmospheric Controls on Soil Moisture–Boundary Layer
369 Interactions. Part II: Feedbacks within the Continental United States. *Journal of*
370 *Hydrometeorology* 4, 570-583.
- 371 Finney, D.L., Doherty, R.M., Wild, O., Huntrieser, H., Pumphrey, H.C., and Blyth, A.M. (2014).
372 Using cloud ice flux to parametrise large-scale lightning. *Atmospheric Chemistry and Physics*
373 14, 12665-12682.
- 374 Forney, R.K., Debbage, N., Miller, P., and Uzquiano, J. (2022). Urban effects on weakly forced
375 thunderstorms observed in the Southeast United States. *Urban Climate* 43.
- 376 Gharaylou, M., Farahani, M.M., Hosseini, M., and Mahmoudian, A. (2019). Numerical study of
377 performance of two lightning prediction methods based on: Lightning Potential Index (LPI)
378 and electric POTential difference (POT) over Tehran area. *Journal of Atmospheric and Solar-*
379 *Terrestrial Physics* 193.
- 380 Gharaylou, M., Farahani, M.M., Mahmoudian, A., and Hosseini, M. (2020). Prediction of lightning
381 activity using WRF-ELEC model: Impact of initial and boundary conditions. *Journal of*
382 *Atmospheric and Solar-Terrestrial Physics* 210.

- 383 Holle, R.L. (2016). A Summary of Recent National-Scale Lightning Fatality Studies. *Weather,*
384 *Climate, and Society* 8, 35-42.
- 385 Hong, S.-Y., and Lim, J.-O.J. (2006). The WRF single-moment 6-class microphysics scheme
386 (WSM6). *Journal of the Korean Meteorological Society* 42, 129-151.
- 387 Hong, S.-Y., Noh, Y., and Dudhia, J. (2006). A New Vertical Diffusion Package with an Explicit
388 Treatment of Entrainment Processes. *Monthly Weather Review* 134, 2318-2341.
- 389 Iacono, M.J., Delamere, J.S., Mlawer, E.J., Shephard, M.W., Clough, S.A., and Collins, W.D. (2008).
390 Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative
391 transfer models. *Journal of Geophysical Research* 113.
- 392 Janjić, Z.I. (1994). The Step-Mountain Eta Coordinate Model: Further Developments of the
393 Convection, Viscous Sublayer, and Turbulence Closure Schemes. *Monthly Weather Review*
394 122, 927-945.
- 395 Kosiba, K.W., J; Trapp, J; Nesbitt, S; Parker, M (2023). "Overview of the PERiLS (Propagation,
396 Evolution and Rotation in Linear Storms) Project", in: *11th European Conference on Severe*
397 *Storms*. (Bucharest, Romania).
- 398 Ladwig, W. (2017). "wrf-python ()", in: *UCAR/NCAR Version 1.3. 2. ed.*
- 399 Lynn, B., and Yair, Y. (2010). Prediction of lightning flash density with the WRF model. *Advances*
400 *in Geosciences* 23, 11-16.
- 401 Lynn, B.H., Yair, Y., Price, C., Kelman, G., and Clark, A.J. (2012). Predicting Cloud-to-Ground and
402 Intracloud Lightning in Weather Forecast Models. *Weather and Forecasting* 27, 1470-1488.
- 403 Malečić, B., Telišman Prtenjak, M., Horvath, K., Jelić, D., Mikuš Jurković, P., Čorko, K., and
404 Mahović, N.S. (2022). Performance of HAILCAST and the Lightning Potential Index in
405 simulating hailstorms in Croatia in a mesoscale model – Sensitivity to the PBL and
406 microphysics parameterization schemes. *Atmospheric Research* 272.
- 407 Mccaul, E.W., Goodman, S.J., Lacasse, K.M., and Cecil, D.J. (2009). Forecasting Lightning Threat
408 Using Cloud-Resolving Model Simulations. *Weather and Forecasting* 24, 709-729.
- 409 Mesinger, F. (1993). Forecasting upper tropospheric turbulence within the framework of the Mellor-
410 Yamada 2.5 closure. *Res. Activ. Atmos. Oceanic Mod.*
- 411 Miller, P., and Debbage, N. (2023). Weakly Forced Thunderstorms in the Southeast US Are Stronger
412 Near Urban Areas. *Geophysical Research Letters* 50.
- 413 Miller, P.W., and Mote, T.L. (2017). A Climatology of Weakly Forced and Pulse Thunderstorms in
414 the Southeast United States. *Journal of Applied Meteorology and Climatology* 56, 3017-3033.
- 415 Miller, P.W.M., Thomas L. (2017). Standardizing the Definition of a “Pulse” Thunderstorm. *Bulletin*
416 *of the American Meteorological Society* 98, 905-913.
- 417 Morrison, H., Thompson, G., and Tatarskii, V. (2009). Impact of Cloud Microphysics on the
418 Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of
419 One- and Two-Moment Schemes. *Monthly Weather Review* 137, 991-1007.
- 420 Mortelmans, J., Bechtold, M., Brisson, E., Lynn, B., Kumar, S., and De Lannoy, G. (2022).
421 Lightning Over Central Canada: Skill Assessment for Various Land-Atmosphere Model
422 Configurations and Lightning Indices Over a Boreal Study Area. *Journal of Geophysical*
423 *Research: Atmospheres* 128.

- 424 Murphy, M.J., Cramer, J.A., and Said, R.K. (2021). Recent History of Upgrades to the U.S. National
425 Lightning Detection Network. *Journal of Atmospheric and Oceanic Technology* 38, 573-585.
- 426 Nakanishi, M., and Niino, H. (2006). An Improved Mellor–Yamada Level-3 Model: Its Numerical
427 Stability and Application to a Regional Prediction of Advection Fog. *Boundary-Layer
428 Meteorology* 119, 397-407.
- 429 Nakanishi, M., and Niino, H. (2009). Development of an Improved Turbulence Closure Model for
430 the Atmospheric Boundary Layer. *Journal of the Meteorological Society of Japan. Ser. II* 87,
431 895-912.
- 432 Olson, J.B., Kenyon, J.S., Angevine, W., Brown, J.M., Pagowski, M., and Sušelj, K. (2019). A
433 description of the MYNN-EDMF scheme and the coupling to other components in WRF–
434 ARW.
- 435 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,
436 Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in
437 Python. *the Journal of machine Learning research* 12, 2825-2830.
- 438 Price, C., and Rind, D. (1992). A simple lightning parameterization for calculating global lightning
439 distributions. *Journal of Geophysical Research: Atmospheres* 97, 9919-9933.
- 440 Price, C., and Rind, D. (1994). The Impact of a $2 \times \text{CO}_2$ Climate on Lightning-Caused Fires. *Journal
441 of Climate* 7, 1484-1494.
- 442 Rasmussen, E.N. (2015). "VORTEX-Southeast Program Overview". (National Severe Storms
443 Laboratory).
- 444 Romps, D.M. (2019). Evaluating the Future of Lightning in Cloud-Resolving Models. *Geophysical
445 Research Letters* 46, 14863-14871.
- 446 Romps, D.M., Charn, A.B., Holzworth, R.H., Lawrence, W.E., Molinari, J., and Vollaro, D. (2018).
447 CAPE Times P Explains Lightning Over Land But Not the Land-Ocean Contrast.
448 *Geophysical Research Letters* 45, 12,623-612,630.
- 449 Romps, D.M., Seeley, J.T., Vollaro, D., and Molinari, J. (2014). Climate change. Projected increase
450 in lightning strikes in the United States due to global warming. *Science* 346, 851-854.
- 451 Saleh, N., Gharaylou, M., Farahani, M.M., and Alizadeh, O. (2023). Performance of Lightning
452 Potential Index, Lightning Threat Index, and the Product of CAPE and Precipitation in the
453 WRF Model. *Earth and Space Science* 10.
- 454 Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G.,
455 Duda, M.G., and Barker, D.M. (2019). A description of the advanced research WRF model
456 version 4. *National Center for Atmospheric Research: Boulder, CO, USA* 145, 145.
- 457 Tewari, M., Chen, F., Wang, W., Dudhia, J., Lemone, M., Mitchell, K., Ek, M., Gayno, G., Wegiel,
458 J., and Cuenca, R. (Year). "Implementation and verification of the unified NOAA land
459 surface model in the WRF model", in: *20th conference on weather analysis and
460 forecasting/16th conference on numerical weather prediction*), 2165-2170.
- 461 Thompson, G., Field, P.R., Rasmussen, R.M., and Hall, W.D. (2008). Explicit Forecasts of Winter
462 Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a
463 New Snow Parameterization. *Monthly Weather Review* 136, 5095-5115.

- 464 Wong, J., Barth, M.C., and Noone, D. (2013). Evaluating a lightning parameterization based on
465 cloud-top height for mesoscale numerical model simulations. *Geoscientific Model*
466 *Development* 6, 429-443.
- 467 Yair, Y., Lynn, B., Price, C., Kotroni, V., Lagouvardos, K., Morin, E., Mugnai, A., and Llasat,
468 M.D.C. (2010). Predicting the potential for lightning activity in Mediterranean storms based
469 on the Weather Research and Forecasting (WRF) model dynamic and microphysical fields.
470 *Journal of Geophysical Research* 115.
471
472

473

474 **Table 1. Shared configuration of all experimental simulations.**

Setting	D01	D02	Reference
Resolution	4 km	1.3 km	N/A
Radiation	RRTMG		(Iacono et al., 2008)
Output time interval	15 minutes		N/A
Land-surface model	Unified NOAH Land Surface Model		(Tewari et al., 2004)
Initial and Boundary Conditions	North American Model		
Output interval	15 minutes		N/A

475

476

477 Table 2. Naming convention and physics parameterizations for experimental simulations.

		Microphysics Parameterizations		
		WSM6 (Hong and Lim, 2006)	Thompson (Thompson et al., 2008)	Morrison (Morrison et al., 2009)
PBL/Surface Layer	YSU (Hong et al., 2006)	YSU_WSM6	YSU_Thomp	YSU_Morr
	MYJ (Mesinger, 1993;Janjić, 1994)	MYJ_WSM6	MYJ_Thomp	MYJ_Morr
	2.5 order MYNN (option 5) (Nakanishi and Niino, 2006;2009;Olson et al., 2019)	MYN2_WSM6	MYN2_Thomp	MYN2_Morr
	3rd order MYNN (option 6) (Nakanishi and Niino, 2006;2009;Olson et al., 2019)	MYN3_WSM6	CTRL	MYN3_Morr

478

479 **Table 3. Choices for qualitative and unbounded hyperparameter selection**

	Activation function	Solver	Hidden layer sizes	L2 Regularization Parameter
Options	Identity	Stochastic Gradient Descent (sgd)	(100,) [1 hidden layer with 100 neurons]	10 ⁻⁵
				10 ⁻⁴
				10 ⁻³
	Logistic	Adam	(50,50) [two hidden layers with 50 neurons]	10 ⁻²
				10 ⁻¹
				0
	Hyperbolic Tangent (tanh)	Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (LBFGS)	(25,25,25,25) [four hidden layers with 25 neurons]	1
				10
	Rectified linear unit (relu)	Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (LBFGS)	(10,10,10,10,10,10,10,10,10,10) [ten hidden layers with ten neurons each]	10 ²
				10 ³

480

481 **Table 4. Absolute values of the coefficients for the Principal Component Analysis divided by the**
 482 **sum of the coefficients for each standardized lightning index.**

		Microphysics Parameterizations								
		WSM6			Thomp			Morr		
PBL/Surface Layer Parameterizations.	YSU		PC1 (%)	PC2 (%)		PC1 (%)	PC2 (%)		PC1 (%)	PC2 (%)
		LPI	23	49	LPI	24	39	LPI	23	50
		CAPE-P	26	3	CAPE-P	26	10	CAPE-P	26	2
		PR92W	26	13	PR92W	27	5	PR92W	27	9
		LTI	25	35	LTI	23	46	LTI	24	39
MYJ		PC1 (%)	PC2 (%)		PC1 (%)	PC2 (%)		PC1 (%)	PC2 (%)	
	LPI	23	49	LPI	23	43	LPI	23	51	
	CAPE-P	26	3	CAPE-P	27	7	CAPE-P	26	1	
	PR92W	26	16	PR92W	27	4	PR92W	27	8	
	LTI	25	33	LTI	23	46	LTI	24	40	
MYN2		PC1 (%)	PC2 (%)		PC1 (%)	PC2 (%)		PC1 (%)	PC2 (%)	
	LPI	23	48	LPI	23	42	LPI	23	50	
	CAPE-P	26	4	CAPE-P	26	8	CAPE-P	26	2	
	PR92W	27	13	PR92W	27	6	PR92W	26	11	
	LTI	24	35	LTI	23	44	LTI	25	37	
MYN3		PC1 (%)	PC2 (%)		PC1 (%)	PC2 (%)		PC1 (%)	PC2 (%)	
	LPI	23	53	LPI	23	42	LPI	23	52	
	CAPE-P	26	2	CAPE-P	26	8	CAPE-P	26	1	
	PR92W	26	17	PR92W	27	6	PR92W	27	11	
	LTI	25	29	LTI	23	45	LTI	24	36	

483

484

485 **Table 5. The confusion matrix (percentages of true and false positives and negatives) using the**
 486 **maximum probability in D02 for each experiment in the testing dataset, as compared to binary**
 487 **lightning observations. Incorrect predictions (false positives and false negatives) are shaded in**
 488 **brown.**

		Microphysics Parameterizations											
		WSM6				Thomp				Morr			
				Predicted				Predicted				Predicted	
True	False			True	False			True	False				
PBL/Surface Layer Parameterizations.	YJU	Observed	True	4.9	4.0	Observed	True	5.4	3.5	Observed	True	5.5	3.4
			False	24.7	66.3		False	21.7	69.4		False	26.1	64.9
	MYJ	Observed	True	4.9	4.0	Observed	True	6.0	3.0	Observed	True	4.8	4.1
			False	27.3	63.7		False	25.5	65.5		False	26.0	65.0
MYN2	Observed	True	4.8	4.1	Observed	True	5.1	3.8	Observed	True	4.8	4.1	
		False	24.5	66.5		False	25.8	65.2		False	25.6	65.4	
MYN3	Observed	True	4.4	4.5	Observed	True	5.0	3.9	Observed	True	4.5	4.4	
		False	21.4	69.7		False	19.8	71.2		False	23.8	67.2	

489

490

491

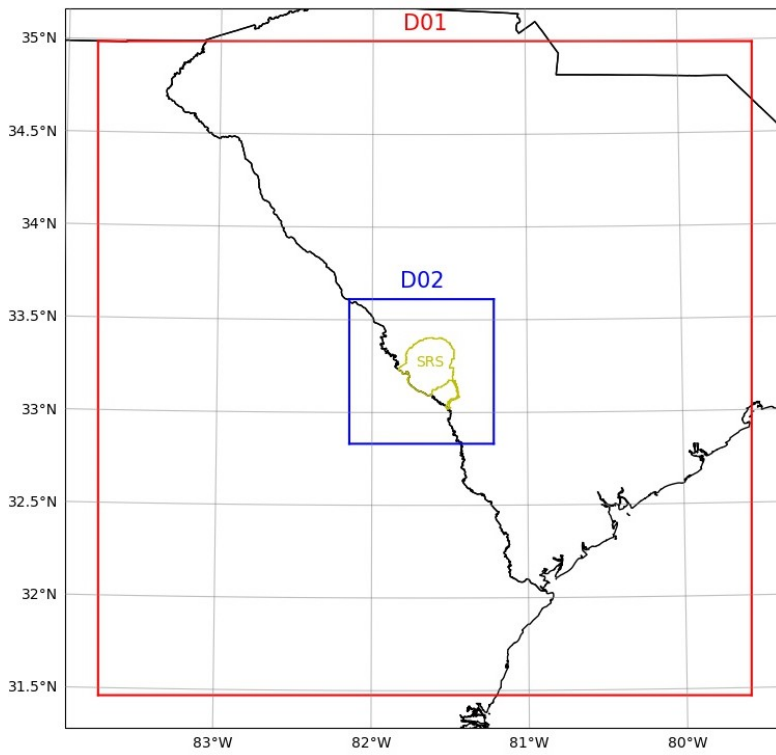
492

Supplementary Table 6: Table describing the selected hyperparameters and the resultant Mean Equitable Threat Score for each member of the machine learning model.

Name	Activation function	Strength of the L2 regularization term	Hidden Layers	Initial Learning Rate	Momentum	Solver	β_2	β_1	Mean Equitable Threat Score
YSU_WSM6	identity	0.01	(10,10,10,10,10,10,10,10,10,10,10)	0.296801	0.187721	sgd			0.122204
MYJ_WSM6	identity	0.00001	(25,25,25,25)	0.022933	0.577663	sgd			0.14127
MYN2_WSM6	identity	1000	(100,100)			lbfgs			0.11719
MYN3_WSM6	identity	0	(50,50)	0.111454	0.251245	sgd			0.043397
YSU_Thomp	identity	0	(50,50)	0.638878	0.493415	sgd			0.130055
MYJ_Thomp	identity	0	(50,50)	0.100827	0.383409	sgd			0.056118
MYN2_Thomp	identity	0.01	(10,10,10,10,10,10,10,10,10,10,10)	0.965158		adam	0.660844	0.892984	0.199087
CTRL	identity	0.0001	(50,50)	0.08507	0.807777	sgd			0.021351

YSU_Morr	identity	0.00001	(10,10,10,10, 10,10,10,10,1 0,10)	0.580919		adam	0.986501	0.10749	0.247515
MYJ_Morr	identity	0	(10,10,10,10, 10,10,10,10,1 0,10)	0.159681		adam	0.775013	0.624754	0.161312
MYN2_Morr	identity	0.01	(10,10,10,10, 10,10,10,10,1 0,10)	0.760532		adam	0.910097	0.749185	0.137888
MYN3_Morr	identity	0.01	(10,10,10,10, 10,10,10,10,1 0,10)	0.265113		adam	0.888929	0.713959	0.133163

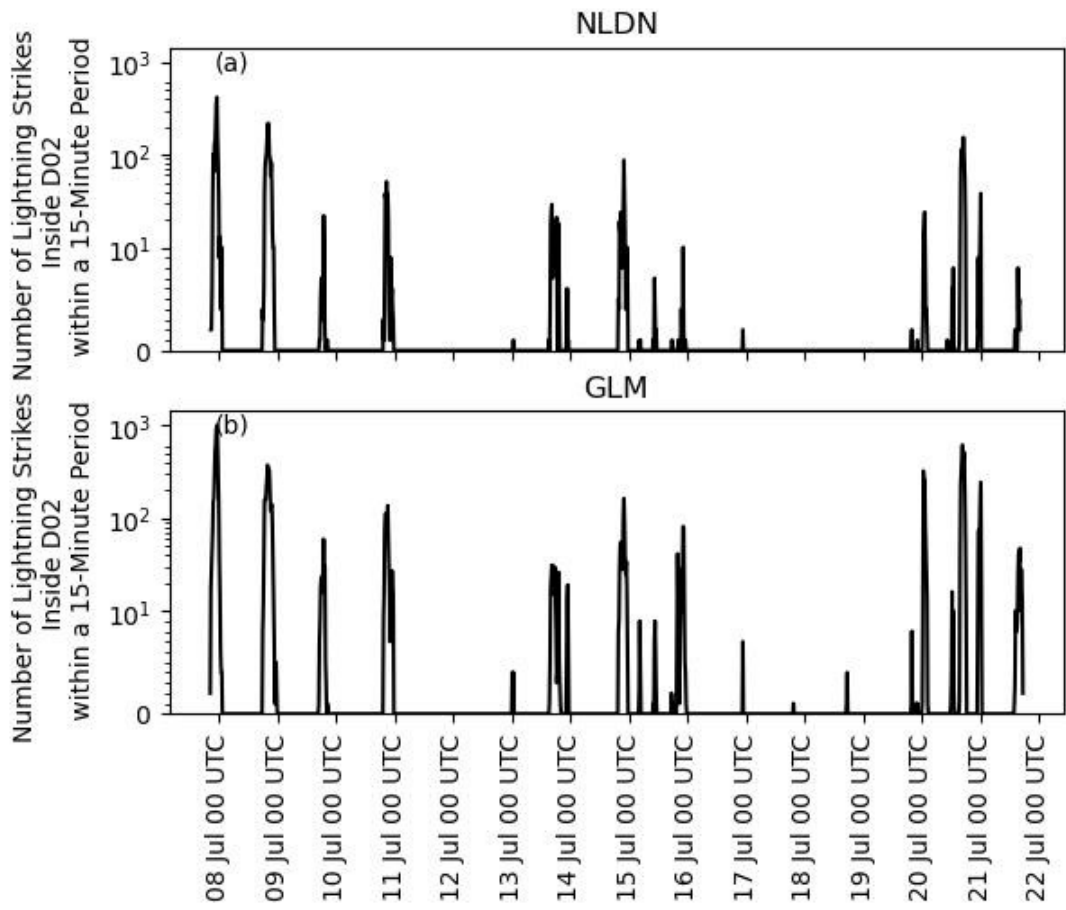
494



495

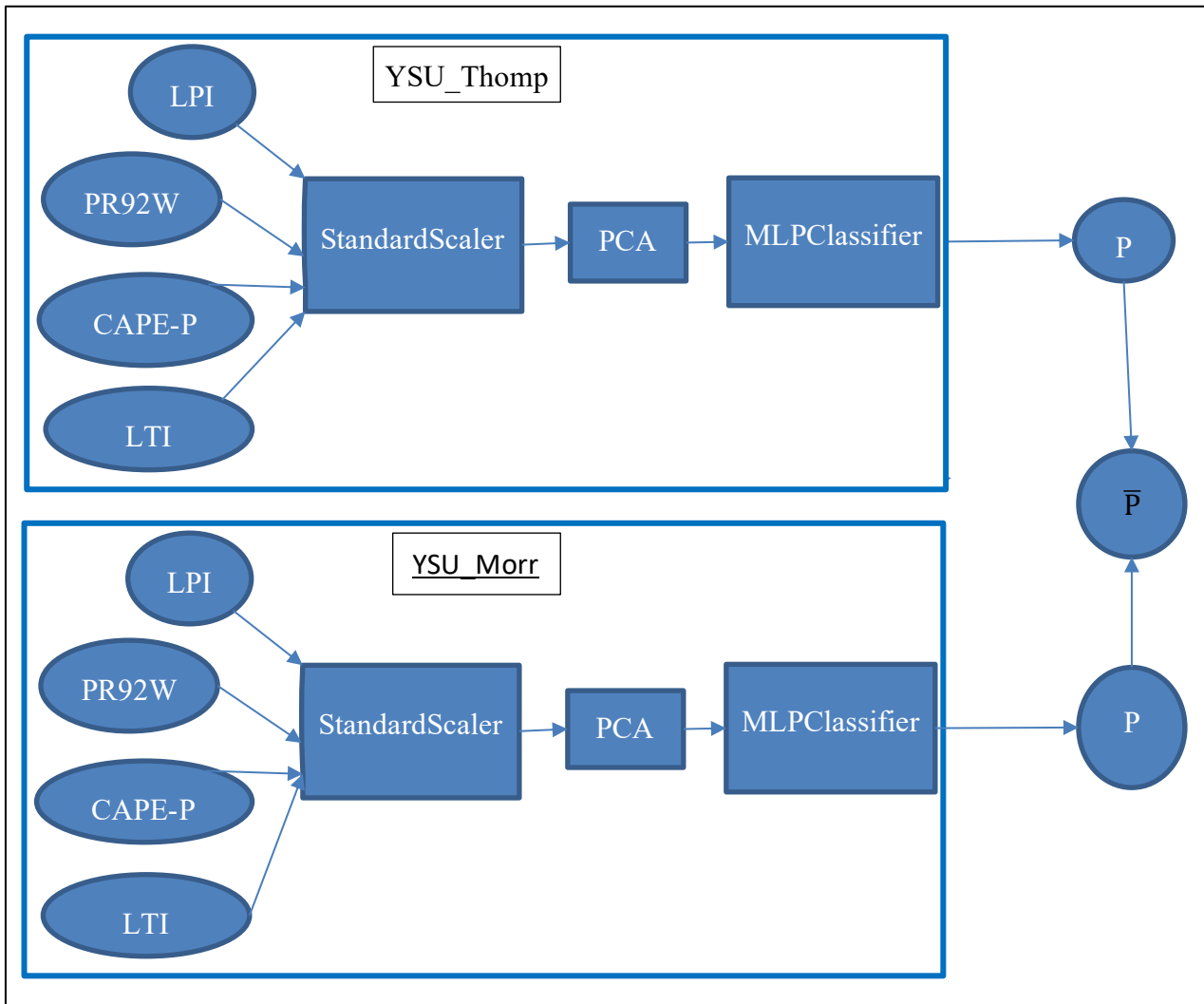
496 **Figure 1: Location of the parent domain (D01) and nested domain (D02) which makes up the**
497 **area of interest. The boundary of the Savannah River Site (SRS), located in the Southeastern**
498 **United States, is plotted in yellow.**

499

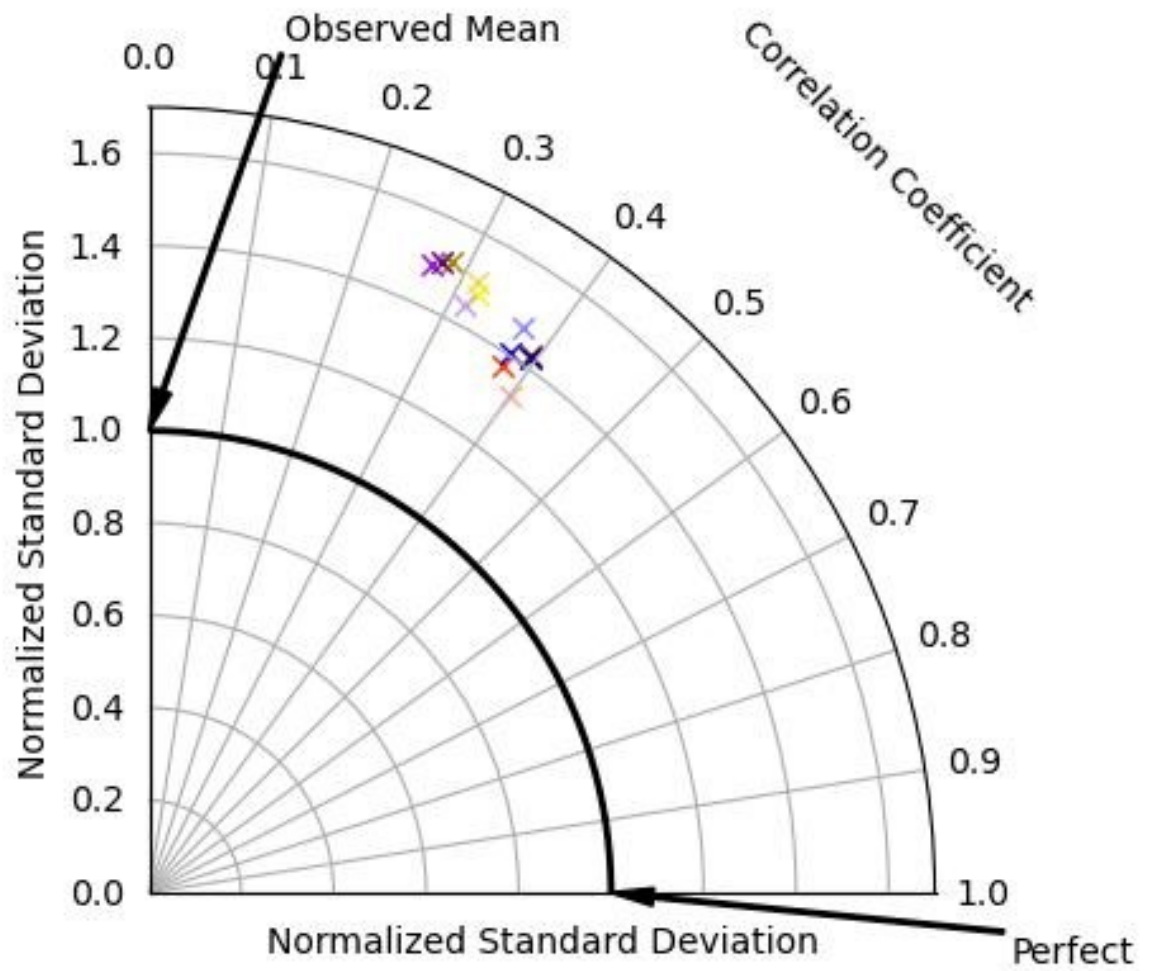


500

501 **Figure 2: Number of Lightning strikes within a 15-minute period inside D02 during the period**
 502 **of interest from the National Lightning Detection Network (NLDN) (a) and Geostationary**
 503 **Lightning Mapper (GLM) (b).**



505 **Figure 3. Flowchart showing part of the ensemble of machine learning tools in an abridged**
 506 **format (with only two variables and two members of the ensemble).**

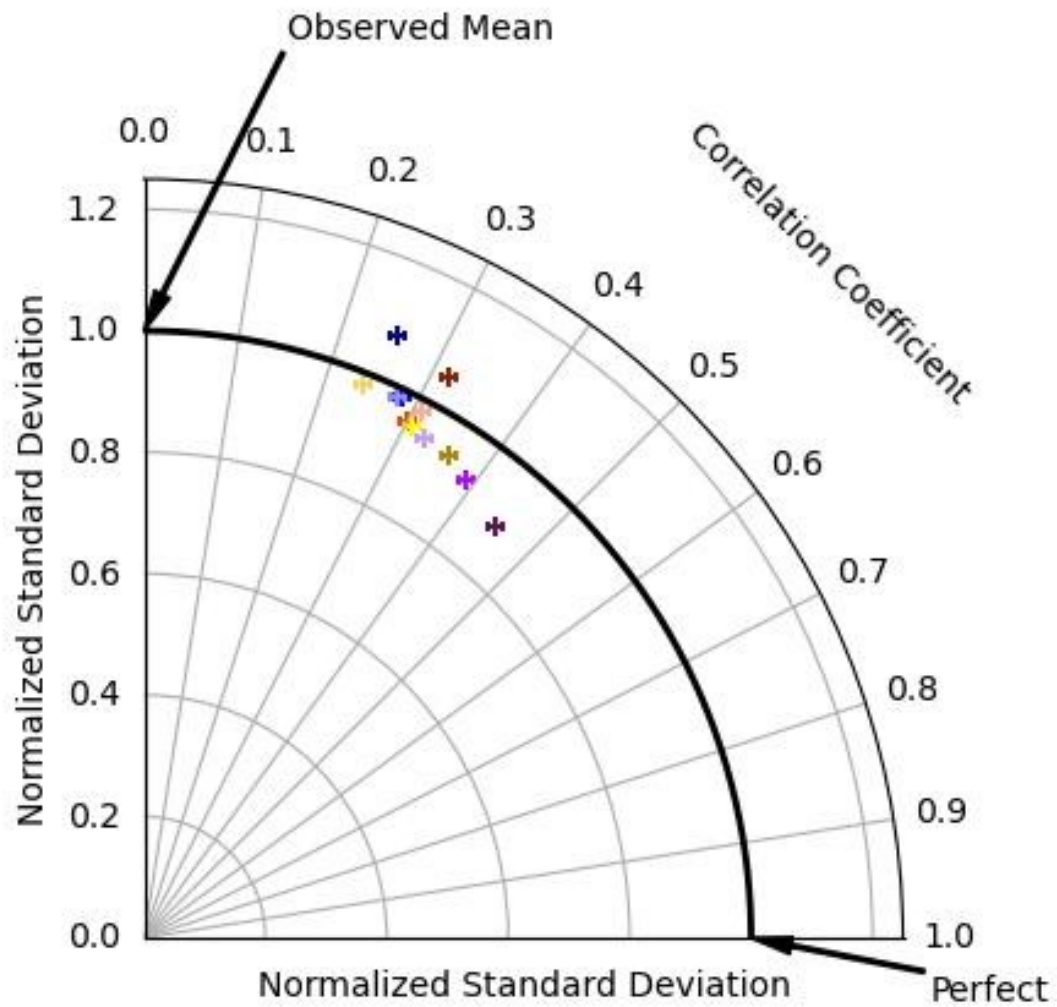


✕ YSU_WSM6	✕ YSU_Thomp	✕ YSU_Morr
✕ MYJ_WSM6	✕ MYJ_Thomp	✕ MYJ_Morr
✕ MYN2_WSM6	✕ MYN2_Thomp	✕ MYN2_Morr
✕ MYN3_WSM6	✕ CTRL	✕ MYN3_Morr

507

508 **Figure 4: Taylor diagram of the composite machine learning model as applied to the training**
 509 **dataset.**

510

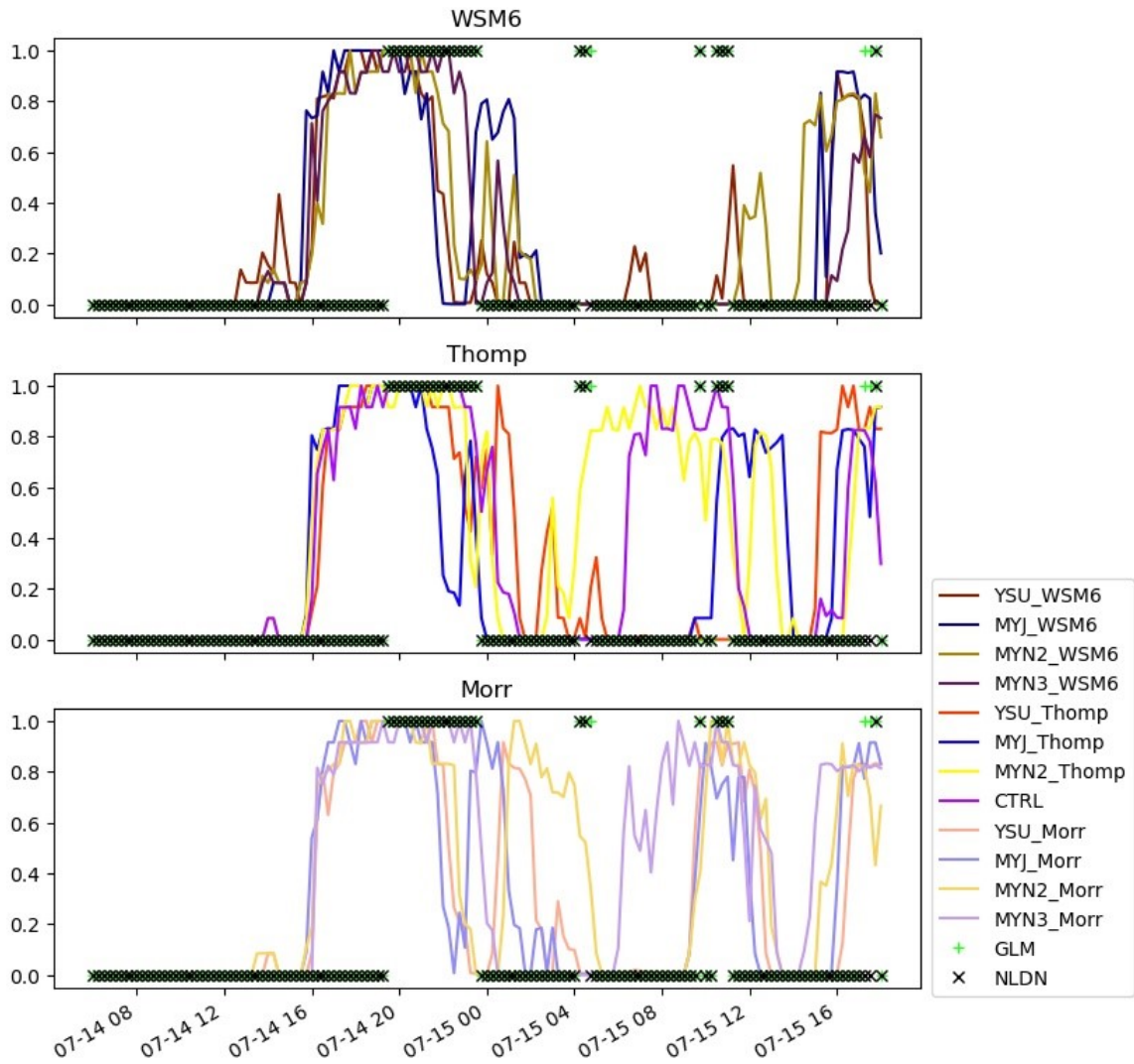


+	YSU_WSM6	+	YSU_Thomp	+	YSU_Morr
+	MYJ_WSM6	+	MYJ_Thomp	+	MYJ_Morr
+	MYN2_WSM6	+	MYN2_Thomp	+	MYN2_Morr
+	MYN3_WSM6	+	CTRL	+	MYN3_Morr

511

512 **Figure 5: Taylor diagram of the machine-learning tool as applied to the testing dataset of each**
 513 **experiment.**

514

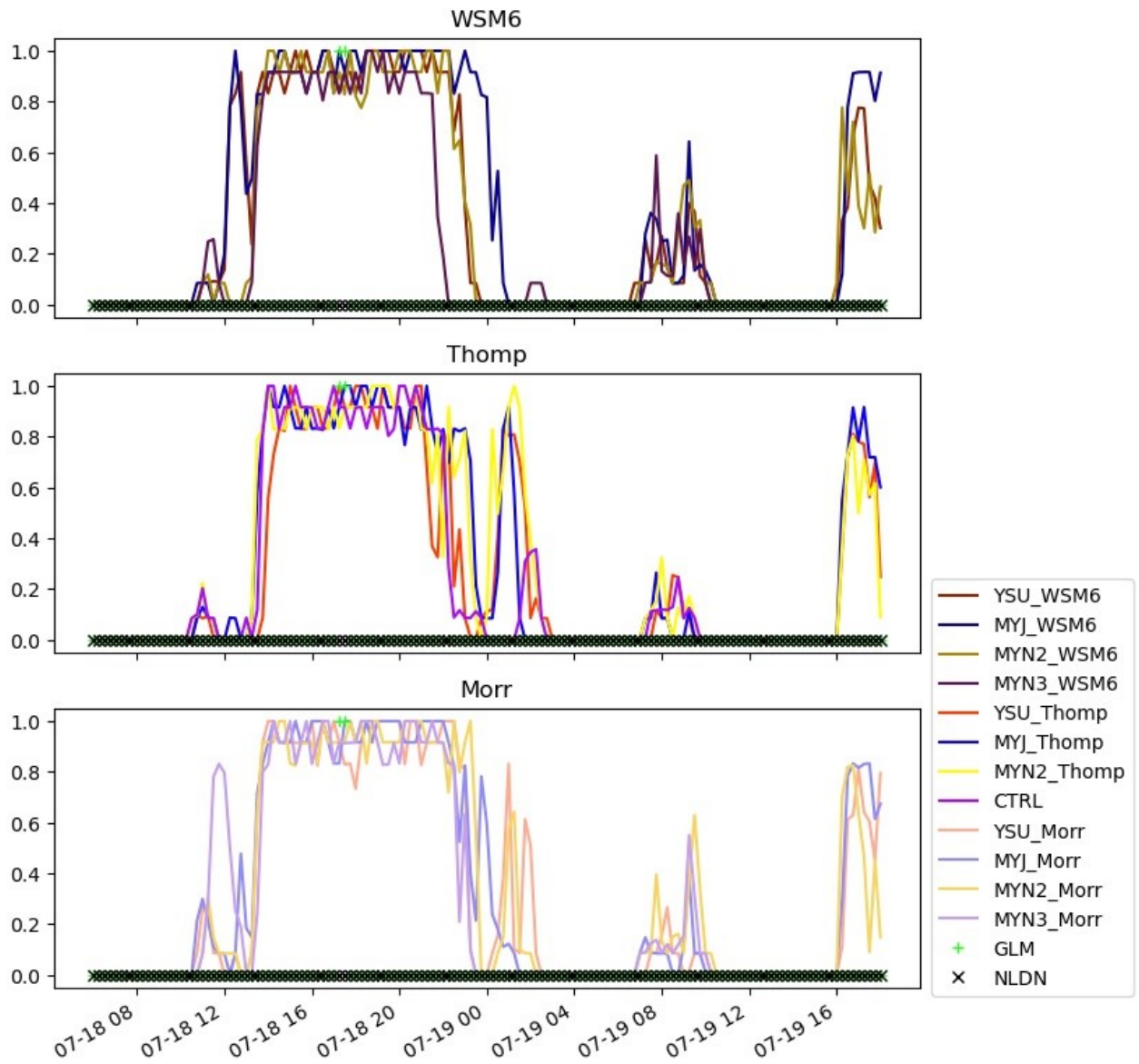


515

516 **Figure 6: Maximum probability of lightning in the model domain for the simulations initialized**
 517 **14 July 2022.**

518

519



520

521 **Figure 7: Similar to Figure 6, but for 18 July 2022.**