Environmental Research Letters

PAPER

Modeling cloud-to-ground lightning probability in Alaskan tundra through the integration of Weather Research and Forecast (WRF) model and machine learning method

Jiaying He and Tatiana V Loboda
Department of Geographical Sciences, University of Maryland, 2181 Samuel J. Lefrak Hall, 7251 Preinkert Drive, College Park, MD, United States of America
E-mail: hjy0608@terpmail.umd.edu

Keywords: Alaskan tundra, cloud-to-ground lightning, empirical-dynamic modeling, lightning-ignited wildfire, Weather Research and Forecast (WRF), random forest

Supplementary material for this article is available online

Abstract
Wildland fires exert substantial impacts on tundra ecosystems of the high northern latitudes (HNL), ranging from biogeochemical impact on climate system to habitat suitability for various species. Cloud-to-ground (CG) lightning is the primary ignition source of wildfires. It is critical to understand mechanisms and factors driving lightning strikes in this cold, treeless environment to support operational modeling and forecasting of fire activity. Existing studies on lightning strikes primarily focus on Alaskan and Canadian boreal forests where land-atmospheric interactions are different and, thus, not likely to represent tundra conditions. In this study, we designed an empirical-dynamical method integrating Weather Research and Forecast (WRF) simulation and machine learning algorithm to model the probability of lightning strikes across Alaskan tundra between 2001 and 2017. We recommended using Thompson 2-moment and Mellor–Yamada–Janjic schemes as microphysics and planetary boundary layer parameterizations for WRF simulations in the tundra. Our modeling and forecasting test results have shown a strong capability to predict CG lightning probability in Alaskan tundra, with the values of area under the receiver operator characteristics curves above 0.9. We found that parcel lifted index and vertical profiles of atmospheric variables, including geopotential height, dew point temperature, relative humidity, and velocity speed, important in predicting lightning occurrence, suggesting the key role of convection in lightning formation in the tundra. Our method can be applied to data-scarce regions and support future studies of fire potential in the HNL.

1. Introduction

Wildland fire is a dominant disturbance across boreal forest and tundra ecosystems in the high northern latitudes (HNL; Goetz et al 2005, Bond-Lamberty et al 2007, French et al 2015). Tundra fires can alter ecosystem functioning and drive environmental changes in carbon cycling and energy budget (Mack et al 2011, Bret-Harte et al 2013, Pearson et al 2013, Jones et al 2015, French et al 2016). Specifically, they can release large stores of carbon locked in organic soil and permafrost. For example, the 2007 Anaktuvuk River fire in Alaska burned 1039 km$^2$ and released $\sim$2.1 Tg carbon into the atmosphere (Mack et al 2011). Although much rarer and generally less severe than boreal fires, they are common, particularly in Alaska. Between 2001 and 2015, Alaskan tundra has burnt about 54% of 10 260 km$^2$ burned area in the tundra worldwide (He et al 2019). Additionally, post-fire recovery for critical components of the tundra ecosystem can last for several decades, which is comparable to that of the boreal forests (Racine et al 1987, Jandt et al 2008). Thus, tundra fires can impact long-term winter forage availability, particularly lichen, for caribou, which will subsequently influence the subsistence resources of local human communities (Gustine et al 2014, Bliss and Hu 2020).

Fires in the remote and generally inaccessible HNL are primarily ignited by cloud-to-ground (CG) lightning flashes (French et al 2015, Veraverbeke et al 2005). Understanding how lightning strikes are distributed in space and time is crucial in implementing effective response strategies towards potential wildfire events. The goal of this study is to design an empirical-dynamical model to model the probability of lightning strikes across Alaskan tundra between 2001 and 2017. We recommended using Thompson 2-moment and Mellor–Yamada–Janjic schemes as microphysics and planetary boundary layer parameterizations for WRF simulations. Our modeling and forecasting test results have shown a strong capability to predict CG lightning probability in Alaskan tundra, with the values of area under the receiver operator characteristics curves above 0.9. We found that parcel lifted index and vertical profiles of atmospheric variables, including geopotential height, dew point temperature, relative humidity, and velocity speed, important in predicting lightning occurrence, suggesting the key role of convection in lightning formation in the tundra. Our method can be applied to data-scarce regions and support future studies of fire potential in the HNL.
2017), also known as lightning strikes. Future projections indicate potential increase of lightning under climate warming (Price and Rind 1994b, Romps et al 2014), which will subsequently lead to more fire occurrences and larger burned areas in the HNL (Wotton et al 2010, Krause et al 2014, French et al 2015, Veraverbeke et al 2017). Numerous studies have explored the driving factors of lightning strikes and developed predictive models (Burrows et al 2005, Blouin et al 2016) for Alaskan and Canadian boreal forests. However, considerably less is known about factors driving CG lightning for the treeless tundra. The substantial differences in surface conditions between tree-dominated and treeless landscapes (Dissing and Verbyla 2003, Beringer et al 2005, Van Heerwaarden and Teuling 2014, Rivas Soriano et al 2019) imply that understanding lightning processes in the boreal forests is not necessarily readily transferable to the tundra. Therefore, tundra-focused studies are critical for enhancing the modeling capability of lightning and fire potential, and assisting wildfire management efforts in the future.

Typically, lightning formation is associated with atmospheric convection in cumulonimbus clouds (Anderson 1992). Lightning flashes are generated through buildup, separation, and transfer procedures of electric charges between cloud particles (Saunders 2008, Yair 2008). The occurrence and intensity of lightning are generally related to factors such as convective cloud development, cloud structure, and hydrometeor attributes (Price and Rind 1994b, Baker et al 1999, Buiat et al 2017). However, explicit simulation and prediction of the electrification processes can be computationally expensive (Zepka et al 2014). Further efforts are also required to comprehensively understand the detailed microphysical procedures contributing to the charge accumulation (Rakov and Uman 2003, Saunders 2008). Efforts on modeling lightning activity thus primarily rely on developing relationships with observed or model-resolved parameters related to convective activities and cloud microphysical properties.

Classificatory schemes or regression methods developed with convective indices or weather conditions were among the earliest lightning modeling attempts (Anderson 1991, Sly 1965, Reap and Foster 1979, Fuquay 1980, Reap and Macgorman 1989). With the development of numerical weather prediction (NWP) models, lightning schemes based on microphysics principles have been parameterized within models from regional to global scales (Price and Rind 1994b, Barthe et al 2010, Yair et al 2010, Lynn et al 2012, Wong et al 2013). Simple strategies were also designed with Weather Research and Forecast (WRF) simulations (Zepka et al 2014, Giannaros et al 2015). In addition to physical parameterizations, empirical-based methods such as logistic regression (Shafer and Fuelberg 2006, Bates et al 2018) and random forest (RF; Blouin et al 2016, Schön et al 2018) were adopted to model lightning based on meteorological conditions and thunderstorm characteristics. Opportunities for integrating dynamic NWP and statistical models have been explored to improve the modeling capability of lightning potential. Burrows et al (2005) trained tree-structured regression models to forecast lightning probability based on the Global Environmental Multiscale (GEM) model. Sousa et al (2013) and Gijben et al (2017) also combined WRF with logistic regression to develop a statistical-dynamical method for CG lightning prediction in different regions.

Due to the remoteness of tundra, meteorological observations are very sparsely distributed and thus unsuitable for describing the spatial variation of tundra conditions. Although reanalysis products ensure spatial-temporal consistency, their performances are limited by the coarse resolution, limited availability of observations, and uncertainty of diagnostic variables (Dee et al 2016). Therefore, purely empirical models trained with these data are inappropriate for lightning modeling in data-scarce regions like tundra. Although NWP has not been specifically adopted in tundra studies, existing research has demonstrated its suitability and effectiveness for modeling lightning potential (Reap 1991, Burrows et al 2005) and fire danger (Molders 2010, Di Giuseppe et al 2016) in boreal forests. In this study, we aim to understand what atmospheric factors affect the CG lightning probability in Alaskan tundra from 2001 to 2017 and to model the probability by integrating WRF and RF.

2. Study area

We defined Alaskan tundra using the Circumpolar Arctic Vegetation Map (CAVM; figure 1(a); Walker et al 2005). Fire regimes vary by year and across different tundra regions (figure 2). In Alaska, more than 99% of the lightning strikes occur from May to August (Reap 1991, Mcguiney et al 2005). Lightning activity shows a typical diurnal pattern from noon until midnight with a peak between 4 pm and 8 pm. Elevation and forest cover can affect the spatial variation of lightning strikes in Alaska by altering the convective activity (Dissing and Verbyla 2003). In particular, large-scale atmospheric instability and local convergence are the primary contributors to the formation thunderstorm of lightning (Reap 1991).

3. Materials and methods

We designed an empirical-dynamical method using WRF-simulated atmospheric variables to model CG lightning probability and understand its driving factors in Alaskan tundra. We first conducted a sensitivity analysis to identify optimal WRF parameterizations that best describes tundra meteorological conditions. We then ran WRF and...
developed RF models for predicting CG lightning probability.

3.1. CG lighting observations
We obtained lightning data from the Alaska Lightning Detection Network (ALDN; Fronterhouse 2012), with a detection efficiency better than 5 km and positional accuracy higher than 70% (Reap 1991, Dissinger and Verbyla 2003). The system has been updated to improve detection performance (Fronterhouse 2012). Devices employed before 2012, developed by Vaisala, Inc., recorded lightning flash with multiplicity (i.e., lightning strikes per flash). The new system, provided by TOA Systems, Inc., records lightning strikes instead (Fronterhouse 2012). To ensure data consistency between systems, we adopted lightning strikes rather than flashes in this study.

3.2. WRF setup and sensitivity analysis
We adopted the Advanced Research WRF version 4.0 (Skamarock et al. 2019) for simulation. The National
Table 1. WRF parameterization schemes used for the pan-Arctic region in existing literature.

| Components                  | Schemes                                  | References                                      |
|-----------------------------|------------------------------------------|-------------------------------------------------|
| Microphysics                | Morrison 2-moment                        | (Bieniek et al 2016)                           |
|                             | Thompson 2-moment                        | (Kim et al 2014; Mölders 2010, Mölders 2008)    |
|                             | WRF single-moment 6-class                | (Hines and Bromwich 2008)                      |
|                             | Goddard Cumulus Ensemble Model            | (Cassano et al 2011; Glisan et al 2013)        |
|                             | Rapid Radiative Transfer Model (RRTM) Longwave | (Bieniek et al 2016; Cai et al 2018; Glisan et al 2013) |
|                             | Community Atmospheric Model (CAM) Longwave | (Cassano et al 2011; Glisan et al 2013)        |
|                             | RRTM Shortwave                           | Cai et al 2018; Mölders 2010, Mölders 2008     |
|                             | Dubhia Shortwave                         | (Cai et al 2018; Mölders 2010, Mölders 2008)   |
| Shortwave radiation         | Goddard Shortwave                        | (Cassano et al 2011; Glisan et al 2013)        |
|                             | CAM Shortwave                            | (Bieniek et al 2016; Cassano et al 2011; Glisan et al 2013) |
|                             | Grell-Devenyi ensemble                    | (Cai et al 2018)                               |
| Cumulus                     | Kain-Fritsch                             | (Bieniek et al 2016; Cassano et al 2011; Glisan et al 2013) |
|                             | Mellor-Yamada-Janjic (MYJ)/Monin-Obukhov Eta (Eta) | (Cai et al 2018; Kim et al 2014; Mölders 2010, Mölders 2008) |
| PBL/ Surface layer          | Yonsei University (YSU)/Monin-Obukhov MM5 (MM5) | (Bieniek et al 2016; Cassano et al 2011; Glisan et al 2013; Hines et al 2011) |
|                             | Noah                                     | (Cai et al 2018; Kim et al 2014; Mölders 2010, Mölders 2008) |
| Land surface model          | Rapid Update Cycle                       |                                                 |

Table 2. Six candidates of parameterization combinations for sensitivity analysis.

| Combination notation | Microphysics | PBL + surface layer |
|----------------------|--------------|---------------------|
| Morrison_MYJ         | Morrison 2-moment | MYJ + Eta          |
| Thompson_MYJ         | Thompson 2-moment | MYJ + Eta          |
| WRF6_MYJ             | WRF6          | MYJ + Eta          |
| Morrison_YSU         | Morrison 2-moment | YSU + MM5         |
| Thompson_YSU         | Thompson 2-moment | YSU + MM5         |
| WRF6_YSU             | WRF6          | YSU + MM5          |

Centers for Environmental Prediction Final Operational Model Global Tropospheric Analysis data (NCEP FNL) at 1° × 1° resolution and 6-h interval were used for model initialization (NCEP 2000). We defined two domains with 25 km (Domain 1) and 5 km (Domain 2) grid spacings for two-way nested simulation (figure 1(b)). The vertical dimension was configured with 33 unevenly spaced full sigma levels with the model top defined at 50 hPa.

WRF provides multiple parameterization schemes for major physics components (table 1), which can affect simulation results for a specific region given various assumptions and mechanisms. Since existing WRF applications in the HNL primarily focus on the boreal forests or the pan-Arctic region, their schemes may not be suitable for the tundra. Therefore, we conducted a sensitivity analysis to determine WRF settings that most closely match tundra meteorological observations.

We first identified a list of candidate schemes through literature review, the majority of which utilized the Noah, Rapid Radiative Transfer Model (RRTM), and Grell–Devenyi schemes for land surface model, radiation, and cumulus components, respectively (table 1). We adopted their updated versions in WRF version 4.0. Mixed-phased microphysics schemes are typically recommended for simulating icing or convective conditions at a horizontal resolution finer than 10 km (Skamarock et al 2019). We thus selected 2-moment schemes of Morrison, Thompson, and WRF 6-class (WRF6) for microphysics, and considered both Mellor–Yamada–Janjic (MYJ) and Yonsei University (YSU) schemes for planetary boundary layer (PBL) representation in our sensitivity analysis (table 2).
Remote Automated Weather Stations (RAWS; https://raws.nifc.gov) provide the densest network of near-surface weather observations across Alaska to date. Four variables, including air temperature (T), dew point temperature (Td), relative humidity (RH), and solar radiation, were utilized as ‘ground truth’ for sensitivity analysis. To consider different fire regimes (figure 2), we selected four cases for 24 h simulations in years of varying fire season intensity (2006—low, 2007—moderate, 2010 and 2015—intense, figure 2; table 3). We calculated three statistical metrics for each variable, including root-mean-square error (RMSE), mean absolute error (MAE) and Pearson’s r correlation, and ranked them of all candidates for each year from 1 (lowest) to 6 (highest). For each metric, the yearly rankings were summed up to create a single rank sum (ranging from 4 to 24), with the larger value representing better overall performance.

### Table 3. Summary of case studies for sensitivity analysis.

| Year | Number of RAWS stations | Simulation period (UTC time) |
|------|--------------------------|------------------------------|
| 2006 | 8                        | 08/16 00:00–08/17 00:00      |
| 2007 | 9                        | 05/01 00:00–05/02 00:00      |
| 2010 | 12                       | 07/01 00:00–07/02 00:00      |
| 2015 | 29                       | 06/18 00:00–06/19 00:00      |

3.3. Random forest modeling of CG lightning probability

We used the RF classification algorithm (Breiman 2001) to model CG lightning probability with WRF-simulated atmospheric variables as predictors. As an ensemble method, RF generates a large number of decision trees through permutation and integrates their results for a more stable modeling performance. Two separate modeling experiments referred to as ‘24 hr model’ and ‘48 hr model’ were designed to compare the consistency of modeling performance and variable importance with different simulation periods (0–24 h and 0–48 h, respectively; figure 3).

For each experiment, atmospheric variables were extracted from WRF outputs (Domain 2) simulated at 24 h or 48 h after initialization, respectively. Four groups of variables were summarized from literature (Reap 1991, Burrows et al 2005, Sousa et al 2013, Blouin et al 2016), including stability indices, cloud properties, weather conditions at multiple pressure levels (500 hPa, 700 hPa, 850 hPa, and 1000 hPa), and two lightning parameterizations from WRF (table 4). Then we used these predictors to model the presence and absence of lightning strikes during the following 24 h after the timing of variable extraction (figure 3). Lightning points from the ALDN dataset were labeled as presence, while random sample points within tundra with no lightning occurred were labeled as absence.

To ensure the representativeness of our models in describing tundra lightning, three lightning severity levels were identified based on the total number of daily lightning strikes for model training, following a similar method by Farukh et al (2011): >2000 strikes per day as severe, 500–2000 strikes per day as moderate, and 0–500 strikes per day as low. For each level, five cases were selected for WRF simulation using a stratified random sampling strategy (table 5). We then randomly selected 70% of presence and absence points for model training and reserved the rest 30% for validation. Training and testing points from all levels were combined for model development and accuracy assessment, respectively. We set the number of trees to 500 and the number of variables at each split as 8 in the RF algorithm.

We reported the out-of-bag (OOB) error rate to evaluate the overall accuracy of model training. For validation, we calculated statistical criteria based on the reserved dataset (tables 6 and 7), and generated receiver operating characteristics (ROC) curves and
Table 4. List of independent variables retrieved from WRF simulation output for RF modeling.

| Categories                  | Abbreviation | Description                        |
|-----------------------------|--------------|------------------------------------|
| Atmospheric stability indices | CAPE         | Convective Available Potential Energy |
|                             | CIN          | Convective Inhibition              |
|                             | LCL          | Lifted Condensation Level          |
|                             | LFC          | Level of Free Convection           |
|                             | TT           | Total Totals                       |
|                             | KI           | K Index                            |
|                             | PLI          | Parcel Lifted Index (to 500 hPa)   |
|                             | BI           | Boyden Index                       |
|                             | SHOW         | Showalter Index                    |
|                             | CF_total     | Total cloud cover fraction         |
|                             | CF_high      | High-level cloud cover fraction    |
|                             | CF_mid       | Mid-level cloud cover fraction     |
|                             | CF_low       | Low-level cloud cover fraction     |
|                             | CTT          | Cloud top temperature             |
|                             | CTH          | Cloud top height                  |
|                             | CTP          | Cloud top pressure                |
| Cloud properties            | IWP          | Ice water path                     |
|                             | LWP          | Liquid water path                  |
|                             | ER_ice       | Effective radius of cloud ice      |
|                             | ER_water     | Effective radius of cloud water    |
|                             | Q_Cloud      | Cloud water mixing ratio           |
|                             | Q_ice        | Ice mixing ratio                   |
|                             | Q_rain       | Rain mixing ratio                  |
|                             | BT           | Brightness temperature             |
|                             | T            | Air temperature at surface and multiple pressure levels |
|                             | T_d          | Dewpoint temperature at surface and multiple pressure levels |
|                             | T_Td         | Temperature-dewpoint spread at multiple pressure levels |
|                             | RH           | Relative humidity at surface and multiple pressure levels |
|                             | GPZ          | Geopotential height at multiple pressure levels |
| Weather variables           | DZ           | Thickness between any two pressure layers |
|                             | W            | Vertical velocity at multiple pressure levels |
|                             | Helicity     | Helicity                           |
|                             | UH           | Updraft helicity                   |
|                             | Rain         | Total precipitation                |
|                             | PW           | Precipitable water                 |
|                             | SLP          | Sea level pressure                 |
| Lightning parameterizations | PR92         | Flash distribution of CG lightning with PR92 scheme |
|                             | LPI          | Lightning probability index        |

Table 5. Cases studies of WRF simulations for CG lightning modeling.

| Initialization of simulation (UTC) | Total CG lightning strokes during 0 ~ 24 h | Total CG lightning strokes during 24 ~ 48 h |
|-----------------------------------|--------------------------------------------|--------------------------------------------|
| 2003/06/23 00:00                  | 1305                                       | 1560                                       |
| 2005/06/11 00:00                  | 732                                        | 1027                                       |
| 2005/06/29 00:00                  | 3449                                       | 2151                                       |
| 2005/08/16 00:00                  | 5                                         | 15                                         |
| 2007/07/04 00:00                  | 4280                                       | 479                                        |
| 2007/07/11 00:00                  | 2321                                       | 2247                                       |
| 2008/06/25 00:00                  | 2030                                       | 295                                        |
| 2009/06/09 00:00                  | 335                                        | 64                                         |
| 2010/07/01 00:00                  | 1015                                       | 1991                                       |
| 2013/06/20 00:00                  | 3073                                       | 3641                                       |
| 2013/08/16 00:00                  | 60                                         | 10                                         |
| 2015/07/14 00:00                  | 847                                        | 4320                                       |
| 2015/06/21 00:00                  | 974                                        | 905                                        |
| 2015/07/23 00:00                  | 175                                        | 80                                         |
| 2016/07/11 00:00                  | 93                                         | 109                                        |

area under the curve (AUC). Moreover, we examined the contribution of predictors in modeling tundra lightning potential with variable importance quantified by mean decrease in accuracy (MDA).
and ROC curves 

were also generated to quantify the forecasting cap-
specifics schemes by meteorological variables and across

ability in the performance of different WRF

The results of sensitivity analysis indicate vari-

3.4. Forecasting capability assessment

In addition to empirical modeling, we tested the cap-

4. Results

4.1. Sensitivity analysis of WRF simulation

The results of sensitivity analysis indicate vari-

4.2. Accuracy assessment of RF models

The overall OOB error rate and the class errors sug-

Table 6. Contingency matrix of variables used to calculate statistical scores.

| Observed | Presence | Absence |
|----------|----------|---------|
| CG lightning event | a (hit) | b (false alarm) |
| Predicted | c (miss) | d (correct non-event) |

Table 7. Statistical criteria used for assessing modeling performance.

| Statistical scores | Abbreviation | Formula |
|--------------------|--------------|---------|
| Probability of Detection | POD | a / (a + c) |
| Critical Success Index | CSI | a / (a + b + c) |
| False Alarm Ratio | FAR | b / (a + c) |
| False Alarm Rate | F | b / (b + d) |

distribution of convective available potential energy (CAPE) as a representative. Considering that the

distribution of convective available potential energy (CAPE) as a representative. Considering that the

Table 7. Statistical criteria used for assessing modeling performance.

| Statistical scores | Abbreviation | Formula |
|--------------------|--------------|---------|
| Probability of Detection | POD | a / (a + c) |
| Critical Success Index | CSI | a / (a + b + c) |
| False Alarm Ratio | FAR | b / (a + c) |
| False Alarm Rate | F | b / (b + d) |

4.2. Accuracy assessment of RF models

The overall OOB error rate and the class errors sug-

Specifically, the absence of CG lightning has a class

Our validation results against the reserved dataset

Our validation results against the reserved dataset

In addition to the purely statistical accuracy

4.3. Forecasting performance

The forecasting test using the ‘24 hr model’ for the

distribution of convective available potential energy (CAPE) as a representative. Considering that the atmospheric soundings were very sparse in the tundra, and the spatial distribution of CAPE from the two schemes was fairly close, we chose the Thompson_MYJ scheme for further simulation, since it shows a more detailed distribution of CAPE values for describing convective activities (figure 5). The optimized WRF schemes were summarized in table 8.

The forecasting test using the ‘24 hr model’ for the

When separated by different severity levels of lightning days, the forecasting performance is consistent with previously reported results: the accuracy appears to be substantially higher for severe and moderate lightning conditions than low severity days.
Figure 4. Bar plots summarizing the ranks of six candidate schemes based on Pearson’s r correlation, MAE and RMSE for: (a) T, (b) $T_d$, (c) RH, and (d) solar radiation. The largest rank sum value represents the best overall performance.

Table 8. Optimized combination of physical parameterizations for WRF.

| Physical component     | Parameterization scheme | Setting option in WRF |
|------------------------|-------------------------|-----------------------|
| Microphysics           | Thompson 2-moment       | mp_physics = 8        |
| Cumulus                | Grell-Freitas ensemble  | cu_physics = 3        |
| PBL                    | MYJ                     | bl_pbl_physics = 2    |
| Surface layer          | Eta                     | sf_sfcphys_physics = 2|
| Land surface model     | Noah                    | sf_surface_physics = 2|
| Longwave radiation     | RRTMG                   | ra_lw_physics = 4     |
| Shortwave radiation    | RRTMG                   | ra_sw_physics = 4     |

Similar patterns can be found from the ROC curves (figure 8), with the AUC values around 0.9 for the entire data and those from the moderate level cases (table 11).

4.4. Evaluation of variable importance

To understand the predictors’ roles in determining the CG lightning potential in the tundra, we examined the top 20 important variables ranked according to MDA from both the ‘24 hr model’ and the ‘48 hr model’ for comparison (figure 9). According to both models, the parcel lifted index (PLI) is found to be the most important variable in determining the accumulated lightning strikes among all the predictors. For the ‘24 hr model’, weather variables at multiple pressure levels, including geopotential height (GPZ), $T_d$, RH, and velocity speed, are also shown as highly important in
Figure 5. Comparison of CAPE simulation in Domain 2 between (a) Morrison_MYJ and (b) Thompson_MYJ for the 2010 case as an example.

Table 9. Confusion matrices for (a) '24 hr model' and (b) '48 hr model'.

(a) 24 h model

| Confusion matrix | CG lightning predictions | Absence | Presence | Class error |
|------------------|--------------------------|---------|----------|-------------|
| CG lightning observations | Absence | 7013 | 549 | 0.0726 |
| | Presence | 241 | 9239 | 0.0254 |

(b) 48 h model

| Confusion matrix | CG lightning predictions | Absence | Presence | Class error |
|------------------|--------------------------|---------|----------|-------------|
| CG lightning observations | Absence | 4872 | 612 | 0.1115 |
| | Presence | 243 | 6837 | 0.0343 |

Table 10. Statistical criteria calculated using the validation data for: (a) 24 h model; (b) 48 h model.

| Models | Metrics | All | Severe level | Moderate level | Low level |
|--------|---------|-----|--------------|----------------|-----------|
| 24 h model | POD | 0.9628 | 0.9821 | 0.9456 | 0.7885 |
| | CSI | 0.9065 | 0.9365 | 0.8569 | 0.7454 |
| | FAR | 0.0606 | 0.0472 | 0.0987 | 0.0682 |
| | F | 0.0858 | 0.1167 | 0.0794 | 0.0277 |
| | AUC | 0.9869 | 0.9898 | 0.9818 | 0.9743 |
| | POD | 0.9668 | 0.9701 | 0.9643 | 0.2273 |
| | CSI | 0.8815 | 0.8808 | 0.8833 | 0.1829 |
| 48 h model | FAR | 0.0909 | 0.0946 | 0.0868 | 0.5161 |
| | F | 0.1247 | 0.1291 | 0.1196 | 0.0366 |
| | AUC | 0.9810 | 0.9810 | 0.9809 | 0.8751 |

determining CG lightning occurrence in the tundra (figure 9(a)). Also, cloud fraction, sea level pressure, helicity, lifted condensation level, and atmospheric stability indices, such as the Total Totals and the Showalter Index (SHOW), are among the top 20 important predictors. Although the variable ranking order of the ‘48 hr model’ differs from that of the ‘24 hr model’, the majority of the top 20 important variables remain the same. Only layer thickness between 700 hPa and 850 hPa levels
and brightness temperature appear to play an important role in determining the lightning potential for the ‘48 hr model’ but not in the ‘24 hr model’.

5. Discussions

In general, this study successfully demonstrates the strong capability of the empirical-dynamical framework in representing meteorological conditions that support CG lightning well in Alaskan tundra. By integrating WRF and RF algorithm, our method shows excellent performance in both modeling CG lightning strikes with 24 h and 48 h WRF simulations, and forward forecasting of lightning probability. This supports the effectiveness of the framework in accurate prediction of future CG lightning potential in data-scarce regions like the HNL.

Our results identify the PLI to 500 hPa as the most important factor in modeling lightning in Alaskan tundra. Similarly, Farukh et al (2011) endorsed the lifted index for characterizing lightning in Alaska because of its sensitivity to evaluate upper air instability. This highlights the key role of the lift potential in providing sufficient convection to support lightning formation in the tundra, consistent with the results of Burrows et al (2005) in the far west and north region of North...
America dominated by tundra. While for boreal forests, SHOW and CAPE were typically recognized as the top-ranking indicators of lightning potential (Burrows et al. 2005, Blouin et al. 2016). Compared to static indices like SHOW, PLI can improve forecasting capability, especially during the afternoon (Galway 1956). Though commonly considered a superior measure of instability, CAPE was not found to be very important here. Unlike CAPE that represents integrated parcel buoyancy of the troposphere, PLI is more of a measure of actual ‘instability’ since it represents the potential buoyancy of a parcel at a certain level (Blanchard 1998). The various index performance in these studies may reflect different land-atmosphere energy exchange and convection between tundra and boreal forests, which also affect the boundary conditions in these ecosystems (Thompson et al. 2004, Beringer et al. 2005).

Additionally, GPZ at 500 hPa ranks top in determining CG lightning distribution, indicating potential connections between this pressure level and lightning activity in Alaskan tundra. While Burrows et al. (2005) suggested that the occurrence of lightning was influenced by the interaction between strong convection and precipitable water (PW) in the cloud, here we found much lower rankings of PW than instability indices like PLI and SHOW in both models. This suggests that convection plays a more critical role than PW in determining the lightning potential in the tundra, which is consistent with the findings of Reap (1991). We also found that the vertical profiles of variables like GPZ, RH, T_d, and layer thickness are highly important in modeling lightning activity in the tundra (figure 9). However, neither of the two internal WRF lightning parameterizations (PR92 and LPI) appears to represent lightning potential in the tundra at the regional scale well. This is not surprising as these parameterizations were not specifically developed for the HNL.

Our selection of Thompson 2-moment and MYJ schemes for WRF simulation in the tundra is consistent with the existing findings of WRF applications in lightning modeling. With a more detailed distribution of CAPE values for describing convective activities (figure 4), the Thompson scheme is also recommended by other studies given its detailed representation of ice-phase processes and improved simulation performance for convection related events like precipitation (Zepka et al. 2014, Giannaros et al. 2015). For lightning modeling purposes, the MYJ scheme is suitable for describing PBL conditions considering its optimal description of atmospheric conditions for triggering convection activities (Sousa et al. 2013, Giannaros et al. 2015).

Despite WRF’s strong capabilities in regional modeling and dynamic downscaling of atmospheric conditions, its performance is affected by the parameterization schemes’ assumptions and mechanisms for describing physical processes. Polar WRF is now under development to improve the representation of near-surface and atmospheric conditions for the Arctic regions, which can support lightning modeling for Alaska in the future (Hines et al. 2011, Wilson et al. 2011, Cai et al. 2018).

| Metrics | All | Severe level | Moderate level | Low level |
|---------|-----|--------------|----------------|-----------|
| POD     | 0.7063 | 0.6372       | 0.7707         | 0.1515    |
| CSI     | 0.6600 | 0.5898       | 0.7397         | 0.1397    |
| FAR     | 0.0788 | 0.1119       | 0.0516         | 0.2       |
| F       | 0.0780 | 0.1024       | 0.0547         | 0.0132    |
| AUC     | 0.9203 | 0.8766       | 0.9519         | 0.769     |

Figure 8. ROC curves of forecasting test on lightning days with different severity levels.
Although our modeling results show very impressive prediction capabilities, they are subject to uncertainties inherent in CG lightning observations detected from either the ground-based networks. Although satellite-based sensors such as Lightning Imaging Sensor or Geostationary Lightning Mapper monitor lightning at a large spatial scale, their datasets are not available for the HNL due to the limited spatial coverage and resolution (Nag et al. 2014, Matsangouras et al. 2016). Separating the CG and cloud-to-cloud lightning flashes from satellite observations can introduce errors as well (Nag et al. 2014). Although ground-based systems are constrained by position accuracy and detection efficiency, we found reporting probability a more effective way to describe CG lightning activity.

In addition to atmospheric factors that function as the key predictors for lightning, synoptic-scale dynamic forcings control meteorological mechanisms driving lightning and fire weather (Reap 1994, Flannigan and Wotton 2001, Santos et al. 2013). For example, Kochtubajda et al. (2019) found more frequent ridging and ridge displacements during the 2014 wildfire season in the Northwest Territories of Canada. Synoptic weather conditions should be explored and incorporated in future modeling efforts to improve our understanding of lightning and fire regimes in the tundra.

Since CG lightning is the primary ignition source of wildfire in boreal forests and tundra (French et al. 2015, Veraverbeke et al. 2017), our integration of NWP and machine learning is easily transferable to monitor and predict fire potential in other data-scarce regions. This provides opportunities for fire management efforts by enhancing our capability to assess future fire impacts on ecosystems and climate and improving fire prevention and suppression practices. Also, fire prediction can ultimately improve the safety and health of the human communities as an important tool in public health applications. Wildfires in the HNL frequently occur in remote and sparsely populated areas and thus pose little direct threat to life and property. However, emissions can be transported hundreds to thousands of kilometers away, exerting considerable impacts on local and regional air quality and leading to various health outcomes (Morris et al. 2006, Reisen et al. 2015).

6. Conclusions

This study makes the first effort in examining the factors driving lightning activity in Alaska tundra with our statistical-dynamical modeling method. We demonstrated the effectiveness of integrating WRF and machine learning for lightning modeling in Alaskan tundra at 5 km spatial resolution. Our results provide insights into understanding the mechanisms of lightning-ignited fires in the tundra. We found PLI and weather variables at multiple pressure levels the most important predictors for modeling lightning potential in the tundra, indicating the primary role of convection in the formation of thunderstorms and CG lightning. Moreover, applicable to other data-scarce regions, our method can further support lightning and fire prediction as well as fire management efforts in the HNL.

Acknowledgments

This work was supported by the NASA Terrestrial Ecology program grant NNX15AT79A.
Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Jiaying He  https://orcid.org/0000-0002-6394-5218
Tatiana V Loboda  https://orcid.org/0000-0002-2537-2447

References

Anderson K R 1991 Models to predict lightning occurrence and frequency over Alberta MSc thesis University of Alberta
Anderson K 1992 Forecasting Lightning Occurrence and Frequency 1–18 Proc. 8th Central Region Fire Weather Committee Scientific and Technical Seminar Winnipeg, Manitoba Edmonton Forestry Canada
Baker M B, Blyth A M, Christian H J, Latham J, Miller K L and Gadain A M 1999 Relationships between lightning activity and various thundercloud parameters: satellite and modelling studies Atmos. Res. 51 221–36
Barthe C, Deierling W and Barth M C 2010 Estimation of total lightning from various storm parameters: a cloud-resolving model study J. Geophys. Res. Atmos. 115 1–17
Bates B C, Dowdy A J and Chandler R E 2018 Lightning prediction for Australia using multivariate analyses of large-scale atmospheric variables J. Appl. Meteorol. Climatol. 37 325–34
Berger J, Chapin F S, Thompson C C and Migueira A D 2005 Surface energy exchanges along a tundra-forest transition and feedbacks to climate Agric. For. Meteorol. 131 143–61
Bieniek P A, Bhatt U S, Walsh J E, Rupp T S, Zhang J, Krieger J R and Lader R 2016 Dynamical downscaling of ERA-interim temperature and precipitation for Alaska J. Appl. Meteorol. Climatol. 55 635–54
Blanchard D O 1998 Assessing the vertical distribution of convective available potential energy Weather Forecast 13 870–7
Bliss L C and Hu F S 2020 wwwbritannica.com/science/tundra Encyclopedia Britannica, inc.
Blouin K D, Flannigan M D, Wang X and Kochtubajda B 2016 Ensemble lightning prediction models for the province of Alberta, Canada Int. J. Wildl. Fire 25 421–32
Bond-Lamberty B, Peckham S D, Ahl D E and Gower S T 2007 Fire as the dominant driver of central Canadian boreal forest carbon balance Nature 450 89–92
Breiman L 2001 Random forests Mach. Learn. 45 5–32
Bret-Harte M S, Mack M C, Shaver G R, Huehner D C, Johnston M, Mojica C A, Pizano C and Reiskind J A 2013 The response of Arctic vegetation and soils following an unusually severe tundra fire Philos. Trans. R. Soc. B 368 20120490
Buia M, Porci F and Dietrich S 2017 Observing relationships between lightning and cloud profiles by means of a satellite-borne cloud radar Atmos. Meas. Tech. 10 221–30
Burrows W R, Price C and Wilson L J 2005 Warm season lightning probability prediction for Canada and the Northern United States Weather Forecast 20 971–88
Cai L, Alexeev V A, Arp C D, Jones B M, Liljedahl A K and Gadeke A 2018 The Polar WRF downscaled historical and projected twenty-first century climate for the coast and foothills of Arctic Alaska Front. Earth Sci. 5 1–15
Cassano J J, Higgins M E and Seefeldt M W 2011 Performance of the weather research and forecasting model for month-long pan-Arctic simulations Mon. Weather Rev. 139 3469–88
Dee D, Fasullo J, Shea D and Walsh J (ed) N.C. for A.R.S. 2016 The climate data guide: atmospheric reanalysis: overview & comparison tables
Di Giuseppe F, Pappenberger F, Wetterhall F, Krezminski B, Cannia A, Libertà G and Miguel J S 2016 The potential predictability of fire danger provided by numerical weather prediction J. Appl. Meteorol. Climatol. 55 2469–91
Dissing D and Verbyla D L 2003 Spatial patterns of lightning strikes in interior Alaska and their relations to elevation and vegetation Can. J. For. Res. 38 770–82
Farukhi M A, Hayasaka H and Kimura K 2011 Characterization of lightning occurrence in Alaska using various weather indices for lightning forecasting J. Disaster Res. 6 343–55
Flannigan M D and Wotton B M 2001 Climate, weather, and area burned Forest Fires (Amsterdam: Elsevier) pp 351–73
French N H F, Jenkins L K, Loboda T V, Flannigan M, Jandt R, Bourgeau-Chavez L L and Whitley M 2015 Fire in arctic tundra of Alaska: past fire activity, future fire potential, and significance for land management and ecology Int. J. Wildl. Fire 24 1045–61
French N H F, Whitley M A and Jenkins L K 2016 Fire disturbance effects on land surface albedo in Alaskan tundra Geophys. Res. Biogeosci. 121 841–54
Frontierhouse B A 2012 Alaska Lightning Detection Network (ALDN) briefing document Ft. Wainwright, AK Bur. L. Manag. Alaska Fire Serv 1–4
Fusquay D M 1980 Forecasting lightning activity level and associated weather USDA For. Serv. Res. Pap. INT-244
Galway J J 1956 The lifted index as a predictor of latent instability Bull. Am. Meteorol. Soc. 37 528–9
Giannaros T M, Kotroni V and Lagnavardos K 2015 Predicting lightning activity in Greece with the Weather Research and Forecasting (WRF) model Atmos. Res. 156 1–13
Gjibben M, Dyson L I and Lotots M T 2017 A statistical scheme to forecast the daily lightning threat over southern Africa using the Unified Model Atmos. Res. 194 78–88
Glisan J M, Gutowski W J, Cassano J J and Higgins M E 2013 Effects of spectral nudging in WRF on arctic temperature and precipitation simulations J. Clim. 26 3985–99
Goetz S J, Bunn A G, Fiske G J and Houghton R A 2005 Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance Proc. Natl Acad. Sci. USA 102 13521–5
Gustine D D, Brinkman T J, Lindgren M A, Schmidt J J, Rupp T S and Adams L G 2014 Climate-driven effects of fire on winter habitat for Caribou in the Alaskan-Yukon Arctic PLoS One 9 e100588
He J, Loboda T V, Jenkins L and Chen D 2019 Mapping fractional cover of major fuel type components across Alaskan tundra Remote Sens. Environ. 232 111324
Hines K M and Bromwich D H 2008 Development and testing of Polar Weather Research and Forecasting (WRF) model. Part I: Greenland ice sheet meteorology Mon. Weather Rev. 136 1971–89
Hines K M, Bromwich D H, Bai L-S, Barlage M and Slater A G 2011 Development and testing of Polar WRF. Part III: arctic land J. Clim. 24 26–48
Jandt R, Joly K, Meyers C R and Racine C 2008 Slow recovery of fire disturbance Factors Arctic, Antarctic, and Alpine Research 40 89–95
Jones B M, Grosse G, Arp C D, Miller E, Liu L, Hayes D J and Larsen C F 2015 Recent Arctic tundra fire initiates widespread thermokarst development Sci. Rep. 5 15865
Kim C K, Steuer M, Schmitt C G, Heysmied A and Thompson G 2014 Numerical modeling of ice fog in interior Alaska using the weather research and forecasting model Pure Appl. Geophys. 171 1963–82
Kochtubajda B, Stewart R E, Flannigan M D, Bonsal B R, Cuell C and Mooney C J 2019 An assessment of surface and atmospheric conditions associated with the extreme 2014 wildfire season in Canada’s Northwest territories Atmos.-Ocean 57 73–90
Krause A, Kloster S, Wilskenskeld S and Paeth H 2014 The sensitivity of global wildfires to simulated past, present, and

13
future lightning frequency J. Geophys. Res. Biogeosci. 119 312–22
Lynn B H, Yair Y, Price C, Kelman G and Clark A J 2012 Predicting cloud-to-ground and intracloud lightning in weather forecast models Weather Forecast 27 1470–8
Mack M C, Bret-Harte M S, Hollingsworth T N, Jandt R R, Schuur E A G, Shaver G R and Verbyla D L. 2011 Carbon loss from an unprecedented Arctic tundra wildfire Nature 475 489–92
Matsangouras I T, Nastos P T and Kapsomenakis J 2016 Cloud-to-ground lightning activity over Greece: spatio-temporal analysis and impacts Atmos. Res. 169 485–96
Mcuiney E, Shukliski M and Wendler G 2005 Alaska lightning climatology and application to wildfire science: Proc. of the 85th American Meteorological Society Annual Meeting
Mölders N 2008 Suitability of the Weather Research and Forecasting (WRF) model to predict the june 2005 fire weather for interior Alaska Alaska Weather Forecast 23 953–73
Mölders N 2010 Comparison of Canadian forest fire danger rating system and national fire danger rating system fire indices derived from Weather Research and Forecasting (WRF) model data for the June 2005 Interior Alaska wildfires Atmos. Res. 95 290–306
Morris G A et al 2006 Alaskan and Canadian forest fires exacerbate ozone pollution over Houston, Texas, on 19 and 20 July 2004 J. Geophys. Res. Atmos. 111 D24S05
Nag A, Murphy M J, Schulz W and Cummins K L. 2015 Lightning locating systems: characteristics and validation techniques Earth Space Sci. 2 65–93
National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce 2000 NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999
Pearson R G, Phillips S J, Loranty M M, Beck P S A, Damoulas T, Knight S J and Goetz S J 2013 Shifts in Arctic vegetation and associated feedbacks under climate change Nat. Clim. Chang. 3 673–7
Price C and Rind D 1994a Possible implications of global climate change on global lightning distributions and frequencies J. Geophys. Res. 99 10823–31
Price C and Rind D 1994b Modeling global lightning distributions in a general circulation model Mon. Weather Rev. 122 1930–9
Racine C H, Johnson L A and Viereck L A 1987 Patterns of vegetation recovery after Tundra fires in Northwestern Alaska, U.S.A. Arct. Alp. Res. 19 461–9
Rakov V A and Uman M A 2003 Lightning: Physics and Effects (Cambridge: Cambridge University Press)
Reap R M, 1991 Climatological characteristics and objective prediction of thunderstorms over Alaska Alaska Weather Forecast 309–19
Reap R M and Foster D S 1979 Automated 12–36 hour probability forecasts of thunderstorms and severe local storms J. Appl. Meteorol. 18 1304–15
Reap R M and Macgorman D R 1989 Cloud-to-ground lightning: climatological characteristics and relationships to model fields, rai a and Del Carmen Huasat M 2010 Predicting the potential for lightning activity in Mediterranean storms based on the Weather Research and Forecasting (WRF) model dynamic and microphysical fields J. Geophys. Res. Atmos. 115 D04205
Reisen F, Duran S M, Flannigan M, Elliott C and Rideout K 2015 Wildfire smoke and public health risk Int. J. Wildl. Fire 24 1029–44
Rivas Soriano L, Sánchez Llorente J M, González Zamora A and de Pablo Dávila F 2019 Influence of land cover on lightning and convective precipitation over the European continent Prog. Phys. Geogr. 43 352–64
Romp D M, Seeley J T, Vollaro D and Molinari J 2014 Projected increase in lightning strikes in the United States due to global warming Science 346 851–4
Santos J A, Reis M A, De Pablo F, Rivas-Soriano L and Leite S M 2013 Forcing factors of cloud-to-ground lightning over Iberia: regional-scale assessments Nat. Hazards Earth Syst. Sci. 13 1745–58
Saunders C 2008 Charge separation mechanisms in clouds Space Sci. Rev. 137 335–53
Schön C, Dittrich J and Müller R 2018 The error is the feature: how to forecast lightning using a model prediction error 2979–88
Shafer P E and Fuelberg H E 2006 A statistical procedure to forecast warm season lightning over portions of the Florida peninsula Weather Forecast 21 851–68
Skamarock W C et al 2019 A description of the advanced research WRF model version 4 NCAR technical note Nat. Cent. Atmos. Res. 145
Sly W K 1965 A convective index in relation to lightning strikes in Northwestern Alberta, CIR 4220 TEC 566 Dept. of Transport, Canada 11 pp
Souza J F, Fontes M, Mendes S, Corte-Real J and Santos J A 2013 Statistical-dynamical modelling of the cloud-to-ground lightning activity in Portugal Atmos. Res. 132–133 46–64
Thompson C, Beringer J, Chapin F S and Mcguire A D 2004 Structural complexity and land-surface energy exchange along a gradient from arctic tundra to boreal forest J. Veg. Sci. 15 397–406
Van Heerwaarden C C and Teuling A J 2014 Disentangling the response of forest and grassland energy exchange to heatwaves under idealized land-atmosphere coupling Biogeosciences 11 6159–71
Veraverbeke S, Rogers B M, Goulden M L, Jandt R R, Miller C E, Wiggins E B and Randerson J T 2017 Lightning as a major driver of recent large fire years in North American boreal forests Nat. Clim. Chang. 7 529–34
Walker D A and et al (The other members of the CAVM team) 2005 The Circumpolar Arctic vegetation map J. Veg. Sci. 16 267–82
Wilson A B, Bromwich D H and Hines K M 2011 Evaluation of polar WRF forecasts on the arctic system reanalysis domain: Surface and upper air analysis J. Geophys. Res. Atmos. 116 D11112
Wong J, Barth M C and Noone D 2013 Evaluating a lightning parameterization based on cloud-top height for mesoscale numerical model simulations Geosci. Model Dev. 6 429–43 https://gmd.copernicus.org/articles/6/429/2013/
Wotton B M, Nock C A and Flannigan M D 2010 Forest fire occurrence and climate change in Canada Int. J. Wildl. Fire 19 253–71
Yair Y 2008 Charge generation and separation processes Space Sci. Rev. 137 119–31
Yair Y, Lynn B, Price C, Kotroni V, Lagouvardos K, Morin E, Mugnai A and Del Carmen Huasat M 2010 Predicting the potential for lightning activity in Mediterranean storms based on the Weather Research and Forecasting (WRF) model dynamic and microphysical fields J. Geophys. Res. Atmos. 115 D04205
Zepka G S, Pinto O and Saraiva A C V 2014 Lightning forecasting in southeastern Brazil using the WRF model Aeros. Res. 135–136 344–62