

**Interpretable and Risk-Aware Machine Learning for Environmental Suitability,
Exposure, and Decision Systems**

by

Liangliang Xu

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
Aug 8, 2026

Keywords: interpretable machine learning, risk-aware learning, environmental suitability,
environmental exposure, random forest, reinforcement learning

Copyright 2026 by Liangliang Xu

Approved by

Zongliang Yue(Co-chair), Assistant Research Professor, Harrison College of Pharmacy
Hao Chen(Co-chair), Assistant Professor of Forest Genomics
Xiao Qin, Professor of Computer Science and Software Engineering
Gerry Dozier, Professor of Computer Science and Software Engineering
Sathyanarayanan Aakur, Assistant Professor of Computer Science and Software
Engineering
Chen Ding(University Reader), Assistant Professor of Forest Ecophysiology

Abstract

Environmental and biological systems increasingly require prediction and decision-making under uncertainty. This dissertation develops and applies interpretable and risk-aware machine-learning approaches for environmental suitability, environmental exposure, and sequential decision systems. The central objective is to translate complex ecological, public-health, and decision-process data into useful predictions while making model interpretation, validation limits, and decision risk explicit.

Chapter 2 synthesizes related work on environmental machine learning, forest species distribution modeling, pollen-centered environmental health prediction, model interpretability, and risk-aware policy learning. It provides the shared conceptual framework for the three studies that follow: machine-learning outputs should be interpreted as decision-relevant evidence only when their data limitations, predictor correlations, validation design, and uncertainty are made explicit.

Chapter 3 develops a climate-regeneration-fire informatics framework for future suitability mapping of loblolly pine, shortleaf pine, slash pine, and longleaf pine in the southeastern United States. The framework integrated 79,349 adult and 22,969 seedling occurrence records compiled from the U.S. Department of Agriculture Forest Service Forest Inventory and Analysis database (FIADB), mapped fire-history intersection, four historical climate baselines, variance-inflation screening, seven species distribution modeling algorithms, Random Forest interpretation, future climate projections, percentile-envelope extrapolation screening, and management-oriented spatial summaries. Adult and seedling records overlapped broadly but were not spatially or climatically interchangeable, indicating that adult persistence and early regeneration should not be treated as the same suitability signal. Wildfire stratification was interpreted as a comparison between mapped

fire-history-intersected and non-intersected landscape strata rather than as a causal fire-effects analysis. XGBoost and Random Forest were the most consistent algorithms, and Random Forest interpretation showed that moisture availability, thermal energy, cold limitation, growing-season timing, evaporative demand, heat-moisture balance, and seasonality contributed to modeled suitability. Future projections were translated into management classes such as seedling-supported suitability, adult-only suitability, future gain, future loss, non-analog uncertainty, and aggregate mill-proximity consensus.

Chapter 4 applies ecological machine learning to environmental health by evaluating whether airborne pollen signatures provide meaningful environmental information for county-level lung cancer incidence and mortality prediction. The analysis used county-year observations for 11 U.S. counties from 2015 through 2021 and combined five daily-derived annual active-sum pollen variables with air pollutants, Air Quality Index summaries, humidity, lag-window exposure engineering, Random Forest prediction, permutation importance, SHAP interpretation, and county-grouped validation. Pollen was not treated as an established lung carcinogen; instead, it was evaluated as a climate-sensitive aeroallergen and environmental-regime marker that may be informative alongside air pollution and humidity. Pooled Random Forest models reproduced county-year patterns more strongly for incidence than mortality, but county-grouped validation showed poor geographic transfer. The chapter therefore interprets pollen-associated signals as exploratory, ecological, and hypothesis-generating rather than causal or ready for external prediction.

Chapter 5 develops TOPS, a transition-based volatility-controlled policy-search framework for risk-averse reinforcement learning. TOPS addresses a general problem exposed by the applied chapters: decisions based only on average predicted performance can be unsafe when outcomes are uncertain, variable, or difficult to validate externally. The framework controls reward volatility, learns from short trajectories or transition segments rather than requiring long uninterrupted rollouts, and provides global convergence guarantees under over-parameterized neural-network policy representations. This theoretical chapter

connects environmental prediction to risk-aware decision-making by showing how learning systems can optimize policies while accounting for variability and hazardous outcomes.

Chapter 6 outlines future work that links the three studies into a broader research program. Key directions include applying risk-averse policy learning to forest adaptation and environmental-health decision support, expanding pollen and lung cancer analysis with richer individual-level and spatial-temporal data, validating forest suitability classes with independent survival and recruitment data, improving uncertainty quantification and causal interpretation, and developing reproducible tools that communicate model uncertainty to stakeholders.

Artificial Intelligence (AI) Use Disclosure Statement

In the preparation of this dissertation, Artificial Intelligence (AI) tools were used to assist with document organization, language editing, LaTeX conversion, figure and table formatting, and dissertation review. The author acknowledges full responsibility for the intellectual content of this work and has ensured that all AI-assisted sections have been reviewed and revised for accuracy and appropriate academic style. All AI-assisted content was reviewed and validated for relevance, appropriateness, and accuracy before incorporation into the final document to maintain scholarly integrity of this research.

Digital Accessibility Disclosure Statement

In the preparation of this dissertation, digital accessibility tools and manual review procedures were used to support compliance with accessibility requirements. These procedures included review of document structure, captions, tables, figures, and PDF output where feasible. The author acknowledges full responsibility for the intellectual content of this work and has made a good-faith effort to comply with digital accessibility requirements in publishing, wherein the nature of the content does not significantly change in order to do so. Furthermore, the document should be reviewed and revised to meet final accessibility requirements prior to final publication.

Table of Contents

Abstract	2
Artificial Intelligence (AI) Use Disclosure Statement	5
Digital Accessibility Disclosure Statement	6
List of Tables	13
List of Figures	14
1 Introduction	17
1.1 Environmental prediction under uncertainty	17
1.2 A unified model interpretability framework	18
1.3 Risk-aware environmental modeling	19
1.4 Ecological suitability, regeneration, and disturbance	20
1.5 Environmental exposure signatures and health burden	21
1.6 From prediction to risk-aware decision-making	22
1.7 Dissertation objectives and organization	22
2 Related Work	24
2.1 Environmental prediction under uncertainty	24
2.2 Forest suitability modeling, regeneration, and disturbance context	25
2.3 Environmental exposure signatures and ecological health modeling	26
2.4 Model interpretability as diagnostic evidence	26
2.5 Risk-aware learning and short-trajectory policy optimization	27
2.6 Synthesis	28
3 Climate-regeneration-fire informatics for future suitability mapping of southern yellow pines	29
3.1 Background and objectives	29

3.2	Materials and methods	30
3.2.1	Study region, focal species, and conceptual framework	30
3.2.2	Occurrence records and life-stage classification	31
3.2.3	Wildfire-intersection stratification	31
3.2.4	Climate predictors and historical baseline windows	34
3.2.5	Background sampling and accessible area	34
3.2.6	Species distribution modeling and validation	35
3.2.7	Random Forest interpretation and climate-filter grouping	35
3.2.8	Adult-seedling climate-space and spatial agreement analyses	36
3.2.9	Future projections, non-analog uncertainty, and aggregate mill- proximity consensus	37
3.3	Results	41
3.3.1	Occurrence structure and wildfire stratification	41
3.3.2	Adult-seedling climate-space separation	43
3.3.3	Multi-algorithm validation and baseline-window sensitivity	44
3.3.4	Climate-filter importance and response curves	44
3.3.5	Adult versus seedling modeled suitability	49
3.3.6	Wildfire-conditioned climate filters	50
3.3.7	Future suitability, change classes, uncertainty, and mill proximity . .	51
3.4	Discussion	53
3.4.1	Adult persistence and seedling recruitment are different suitability signals	53
3.4.2	Species-specific climate filters explain differences in climatic sepa- rability	54
3.4.3	Wildfire intersection modifies realized climate associations but does not describe the full fire regime	55
3.4.4	Baseline sensitivity and future projection uncertainty	56

3.4.5	Implications for adaptation planning and forest management	56
3.4.6	Limitations and next steps	57
3.5	Conclusions	58
4	Airborne Pollen Signatures and County-Level Lung Cancer Burden	59
4.1	Background and objectives	59
4.1.1	Lung cancer burden and environmental context	59
4.1.2	Why focus on pollen?	60
4.1.3	Study rationale, hypothesis, and objectives	61
4.2	Materials and Methods	62
4.2.1	Study design and analytic unit	62
4.2.2	Outcomes	62
4.2.3	Pollen exposures	63
4.2.4	Air-quality and meteorological features	63
4.2.5	Daily pollen aggregation and illustrative weekly comparison	65
4.2.6	Windowed time-lagged exposure engineering	67
4.2.7	Descriptive correlation and collinearity assessment	67
4.2.8	Predictive modeling	69
4.2.9	Model interpretation	69
4.2.10	Humidity and interaction analysis	72
4.2.11	Validation strategy	74
4.2.12	Ethical considerations and reporting	74
4.3	Results	75
4.3.1	Lung cancer burden varied by county and year	75
4.3.2	Pollen levels showed geographically structured overlap with lung cancer outcomes	75
4.3.3	Illustrative weekly pollen comparisons showed seasonal structure	75

4.3.4	Lag-window and correlation analyses suggested temporally complex environmental signals	77
4.3.5	Random forest models identified both air-quality and pollen features as important predictors	77
4.3.6	SHAP analyses supported context-dependent pollen contributions .	78
4.3.7	Humidity changed pollen rankings but provided limited evidence for strong interactions	79
4.3.8	Absolute humidity analyses indicated sensitivity to pollen-meteorology context	80
4.3.9	Supplementary analyses reinforced heterogeneity and confounding concerns	80
4.4	Discussion	81
4.4.1	Principal findings	81
4.4.2	Interpretation of pollen-associated signal	82
4.4.3	Interpretation by pollen taxa	82
4.4.4	Pollen, pollution, and humidity as a combined exposome	83
4.4.5	Why incidence prediction was stronger than mortality prediction . . .	84
4.4.6	County-grouped and leave-one-county-out validation	84
4.4.7	Strengths	85
4.4.8	Limitations	85
4.4.9	Future directions	87
4.5	Summary	87
5	TOPS: Transition-based Volatility-controlled Policy Search and its Global Convergence	91
5.1	Background and problem formulation	91
5.2	Technical background	93
5.3	Algorithm Formulation	97

5.3.1	Augmented MDP	97
5.3.2	Proposed Algorithms	99
5.4	Theoretical Analysis	102
5.4.1	Assumptions	103
5.4.2	Major Theoretical results	104
5.5	Experiments	106
5.6	Discussion	109
5.6.1	Method comparison	109
5.6.2	Summary	110
6	Future Work	112
6.1	Apply risk-averse policy learning to environmental decision support	112
6.2	Expand the pollen-lung cancer analysis with richer exposure and health data	113
6.3	Strengthen forest suitability models with demographic, fire-regime, and ge- netic validation	113
6.4	Improve validation, uncertainty quantification, and causal interpretation . . .	114
6.5	Develop accessible, reproducible, and stakeholder-facing decision tools . .	115
6.6	Concluding synthesis: from interpretation to risk-aware action	115
	Bibliography	117
A	Supplementary Material for Forest Suitability Modeling	127
A.1	Supplementary plots, tables, and figure notes	127
A.2	Detailed captions for Chapter 3 main figures	161
B	Supplementary Material for Pollen and Lung Cancer Modeling	168
B.1	Supplementary plots and figure notes	168
B.2	Detailed captions for Chapter 4 main figures	185
C	Supplementary Material for TOPS Framework	188
C.1	Notation Systems	188
C.2	Algorithm Details	190

C.3	Experimental Details	192
C.3.1	Testbeds	193
C.3.2	Hyper-parameter Settings	193
C.3.3	Computing infrastructure	193
C.4	Theoretical Analysis Details	193
C.4.1	Overview	193
C.4.2	Additional Assumptions	195
C.4.3	Proof of Theorem 5.1	196
C.4.4	Proof of Theorem 5.2	209
C.4.5	Proof of Lemma 1	214
C.4.6	Proof of Lemma 2	215
C.5	Additional method comparisons	229
C.5.1	Global Optimality of Policy Search Methods	229
C.5.2	Over-parameterized Neural Networks in RL	230

List of Tables

3.1	Occurrence sample sizes by species, life stage, and wildfire status	42
3.2	Adult-seedling climate-space centroid distances	43
3.3	XGBoost and Random Forest performance by species	45
3.4	Future suitability-class percentages from ensemble maps	52
4.1	Environmental predictor domains and interpretation	63
4.2	Summary of key model-performance results	71
4.3	Pollen-centered interpretation of model results	76
A.1	Data provenance for occurrence, climate, and fire-history data streams	129
A.2	Model performance metrics for seven SDM algorithms	133

List of Figures

3.1	Adult-tree and seedling occurrence distributions for southern yellow pines . . .	32
3.2	Climate-space separation of adult-tree and seedling records by wildfire inter- section	36
3.3	Adult and seedling modeled climate suitability and agreement classes	38
3.4	Future suitability and stable/gain/loss change classes	39
3.5	Occurrence sample sizes by life stage and wildfire-intersection status	42
3.6	SDM algorithm performance across species and climate baselines	46
3.7	Random Forest climate-filter importance from SHAP and MDI	47
3.8	Random Forest response curves for selected climate predictors	48
3.9	Wildfire-stratified climate-filter importance by life stage	50
4.1	Spatial and temporal distribution of county-level lung cancer incidence and mortality	64
4.2	Pollen station levels overlaid on average lung cancer incidence and mortality .	64
4.3	Weekly pollen reconstruction in Pulaski and Sarasota counties	66
4.4	Lag-window correlations, predictor collinearity, and temporal/spatial correlation	68
4.5	Pooled random forest feature importance and predictive performance	70

4.6	SHAP interpretation and county-grouped validation	73
4.7	Humidity-index SHAP and interaction analysis	89
4.8	Absolute-humidity sensitivity, 3D pollen-humidity patterns, and SHAP interaction diagnostics	90
5.1	TOPS framework block diagram	98
5.2	Training progress of TOPS-NPG and baseline algorithms	107
5.3	Training progress of TOPS-PPO and baseline algorithms	107
A.1	VIF screening workflow and retained climate predictors across baselines . . .	131
A.2	ROC curves comparing SDM algorithms under the 1991-2020 baseline	134
A.3	MDI predictor-importance rankings across species	136
A.4	SHAP predictor-importance rankings across species	138
A.5	Response curves for retained predictors - loblolly pine	140
A.6	Response curves for retained predictors - shortleaf pine	141
A.7	Response curves for retained predictors - slash pine	142
A.8	Response curves for retained predictors - longleaf pine	143
A.9	Wildfire-stratified MDI rankings - loblolly pine	145
A.10	Wildfire-stratified MDI rankings - shortleaf pine	146
A.11	Wildfire-stratified MDI rankings - slash pine	147
A.12	Wildfire-stratified MDI rankings - longleaf pine	148

A.13 Wildfire-stratified SHAP rankings - loblolly pine	150
A.14 Wildfire-stratified SHAP rankings - shortleaf pine	151
A.15 Wildfire-stratified SHAP rankings - slash pine	152
A.16 Wildfire-stratified SHAP rankings - longleaf pine	153
A.17 Early-century suitability projections from the eight-GCM ensemble	155
A.18 Mid-century suitability projections from the eight-GCM ensemble	157
A.19 Non-analog climate uncertainty and dominant driver maps	159
B.1 Alternative outcome trend displays	169
B.2 Temporal correlations by outcome	171
B.3 Diagnostic environmental exposure heatmap	173
B.4 County-specific tree-based feature rankings	175
B.5 Held-out-county permutation importance	177
B.6 Exploratory future-projection concept	179
B.7 Sex- and race-stratified lung cancer outcome patterns	181
B.8 t-SNE and UMAP clustering of county profiles	183
C.1 Over-parameterized neural network block diagram	192
C.2 Flow chart of the theoretical analysis	194

Chapter 1

Introduction

Environmental and biological systems increasingly require decisions under uncertainty. Forest managers must anticipate whether long-lived species can persist and regenerate under changing climate, disturbance, and management conditions. Public-health researchers must evaluate complex environmental exposures, such as pollen, air pollution, and humidity, without overstating causal conclusions from ecological data. Decision scientists must design learning systems that perform well while avoiding high-variance or hazardous outcomes. These problems differ in subject matter, but they share a common challenge: machine-learning models can reveal useful patterns, yet their outputs are only useful when interpretation, validation, uncertainty, and risk are made explicit.

This dissertation addresses these challenges through three linked studies. The first applied study focuses on ecological suitability and forest adaptation. The second applied study focuses on environmental exposure signatures and county-level lung cancer burden. The third study develops a risk-averse learning method for sequential decision systems. Together, the chapters move from interpretable environmental prediction to risk-aware decision-making.

1.1 Environmental prediction under uncertainty

Environmental prediction problems are rarely simple classification or regression tasks. Observations are often spatially structured, temporally uneven, and influenced by multiple interacting processes. In forest ecology, an occurrence record may reflect dispersal, management, planting history, disturbance, competition, and climate, rather than climate alone. In environmental health, county-level cancer outcomes may reflect smoking history, air pollution, demographic structure, health-care access, occupational exposure, radon,

screening, and unmeasured socioeconomic factors, as well as environmental signals such as pollen and humidity. In both settings, prediction models must be interpreted with caution because model importance does not automatically imply biological or causal effect.

Chapter 2 develops the shared literature and conceptual framework for the dissertation. It connects species distribution modeling, environmental-health exposure prediction, model interpretability, and risk-aware reinforcement learning. The remaining chapters then apply or extend this framework in three different settings.

1.2 A unified model interpretability framework

Interpretability plays a central role throughout the dissertation. I use response curves, mean decrease in impurity (MDI), permutation importance, SHAP values, grouped predictor summaries, feature rankings, sensitivity analyses, and validation diagnostics as complementary tools for understanding model behavior. These methods do not have the same meaning, and none of them establishes causality by itself.

Response curves visualize how a model prediction changes across a predictor gradient. In the forest-suitability chapter, response curves help show whether modeled suitability changes gradually or through threshold-like transitions along climate gradients. Their limitation is that they can become misleading when predictors are correlated or when the displayed range extends beyond well-supported training data.

MDI measures how much tree splits reduce impurity in Random Forest models. It is useful for summarizing repeated model runs, but it can be sensitive to predictor scale, split opportunities, and correlation structure [95]. Permutation importance measures the change in model performance after a feature is disrupted. It provides a direct performance-based sensitivity check, but correlated variables can substitute for one another, producing unstable or even negative importance values. In this dissertation, negative permutation values are interpreted as signs of instability or redundancy rather than as protective effects.

SHAP values summarize how features contribute to individual model predictions [66]. SHAP is useful for identifying heterogeneity, subgroup differences, and direction of model contribution, but it still explains the fitted model rather than the underlying causal process. Across the applied chapters, I therefore interpret response curves, MDI, permutation importance, and SHAP as model-diagnostic evidence. When the tools point to similar predictor groups, confidence in the predictive association increases. When they disagree, the disagreement is treated as evidence of collinearity, model sensitivity, or data-structure dependence.

1.3 Risk-aware environmental modeling

A second unifying theme is risk awareness. In the forest-suitability chapter, risk appears as uncertainty about regeneration, future climate extrapolation, adult-only suitability, and suitability loss. A map cell with high adult suitability is not necessarily low risk if seedling-supported suitability is weak or if future climate falls outside the historical calibration envelope. In the pollen and lung cancer chapter, risk appears as interpretive risk: pooled models can look accurate while county-grouped validation performs poorly, meaning that model outputs should not be used as external predictions or causal evidence. In the TOPS chapter, risk is formalized as reward volatility in sequential decision-making.

The dissertation therefore treats risk awareness broadly. In the applied chapters, risk-aware analysis means separating signals that are easy to overinterpret from signals that are robust enough for cautious decision support. In the methodological chapter, risk-aware learning means optimizing a policy while explicitly penalizing undesirable variability. This connection is important because environmental management and public-health decisions rarely depend only on average performance. They also depend on the probability of failure, the severity of poor outcomes, and the confidence that model behavior transfers to new conditions.

1.4 Ecological suitability, regeneration, and disturbance

Chapter 3 addresses future suitability mapping for southern yellow pines. Southern yellow pines are central to forest ecology, timber production, restoration, and private-land management in the southeastern United States. Climate change may affect these species through warming, altered precipitation, drought stress, evaporative demand, disturbance interactions, insects, pathogens, and management responses [73, 76]. For long-lived trees, however, climate stress may appear first as weak regeneration rather than adult mortality. Adult trees can persist for decades under conditions that no longer support reliable seedling establishment.

The first applied study separates adult-tree and seedling occurrence records before modeling. Adult records are interpreted as persistence and legacy signals, whereas seedling records are interpreted as early regeneration signals. The chapter also stratifies occurrences by mapped fire-history intersection. This stratification compares realized climate associations in mapped fire-intersected and non-intersected areas, but it is not a fire-effects analysis. The fire-history layer does not capture prescribed fire, burn severity, complete fire frequency, burn season, fuel condition, or exact time since fire for every observation.

The methodological contribution of Chapter 3 is an end-to-end climate-regeneration-fire informatics framework. It integrates occurrence processing, life-stage classification, mapped fire-history stratification, climate-predictor screening, multi-algorithm species distribution modeling, Random Forest interpretation, future projection, non-analog uncertainty screening, and management-oriented spatial summaries. The chapter demonstrates how machine learning can translate ecological data into adaptation classes such as seedling-supported suitability, adult-only suitability, future loss, future gain, non-analog uncertainty, and aggregate mill-proximity consensus.

1.5 Environmental exposure signatures and health burden

Chapter 4 shifts from forest suitability to environmental health. Lung cancer incidence and mortality vary across space and time, and established determinants include smoking, radon, occupational exposures, outdoor air pollution, demographic structure, and health-care access. Outdoor air pollution and particulate matter have strong external evidence for lung cancer risk [50, 45, 101]. Pollen is not an established lung carcinogen, but it is a climate-sensitive aeroallergen that can reflect vegetation, seasonality, humidity, geography, and respiratory inflammatory context [22, 24, 81].

The second applied study asks whether county-level pollen signatures provide meaningful environmental signal rather than random noise when evaluated alongside air pollution, AQI summaries, humidity, lag-window exposure features, and county structure. The study uses Random Forest models as the primary predictive tool, XGBoost as a supplementary sensitivity model, and SHAP, permutation importance, and feature-importance methods to interpret model behavior. Humidity is treated as a contextual factor that can change model explanations and feature rankings, but the chapter does not treat small interaction diagnostics as evidence of a strong pollen-humidity mechanism.

This chapter reinforces the importance of validation design. Pooled Random Forest models captured substantial county-year structure, especially for incidence, but county-grouped validation performed poorly. This contrast is scientifically valuable because it shows that strong pooled performance can reflect county baselines or repeated spatial structure rather than transferable exposure-response relationships. The environmental health chapter therefore complements the forest chapter: both studies use interpretable machine learning to detect environmental signals, and both emphasize that interpretation must be limited by validation design and data structure.

1.6 From prediction to risk-aware decision-making

Chapters 3 and 4 show that machine-learning predictions can support environmental interpretation, but they also show that prediction alone is not enough. A high suitability score may be uncertain under non-analog future climate. A pollen or air-quality feature may be model-informative but not causal. A model may perform well in pooled data yet fail under stricter validation. These limitations motivate a broader decision-science question: how should learning systems behave when uncertainty, outcome variability, and risk matter?

Chapter 5 addresses this question through risk-averse reinforcement learning. Reinforcement learning has achieved success across many domains, but many approaches remain risk-neutral and optimize expected return without explicitly controlling variability [96, 77, 31]. In applications such as finance, healthcare, robotics, and environmental management, this can be insufficient because hazardous outcomes, long-term shocks, or high volatility may be unacceptable even when average performance is high [42, 14, 68].

The third study develops TOPS, a transition-based volatility-controlled policy-search framework. TOPS is designed to control reward volatility, reduce dependence on long uninterrupted trajectories, and provide global convergence guarantees under over-parameterized neural networks. This chapter gives the dissertation a theoretical decision-learning component: the applied chapters show why uncertainty-aware interpretation is necessary, and the TOPS chapter formalizes a risk-aware approach for learning policies when variability and hazardous trajectories matter.

1.7 Dissertation objectives and organization

The overall objective of this dissertation is to develop and apply interpretable and risk-aware machine-learning methods for environmental prediction and decision support under uncertainty. This objective is pursued through four linked components.

Chapter 2 synthesizes the related work connecting environmental prediction, model interpretation, and risk-aware learning. Chapter 3 develops a climate-regeneration-fire informatics framework for life-stage-specific suitability mapping of southern yellow pines. Chapter 4 evaluates whether airborne pollen signatures contribute meaningful environmental information to county-level lung cancer incidence and mortality prediction. Chapter 5 develops a risk-averse policy-search method for decision-making under uncertain outcomes. Chapter 6 synthesizes future research directions that build on all three studies, including applying risk-averse policy learning to environmental decision problems, expanding environmental health data, strengthening forest-regeneration validation, improving uncertainty quantification, and developing reproducible decision-support tools.

Chapter 2

Related Work

This chapter places the dissertation within a shared literature on interpretable environmental machine learning, uncertainty-aware validation, and risk-aware decision systems. The three research chapters differ in domain—forest climate suitability, county-level environmental health, and risk-averse policy search—but they address a common modeling challenge. In each case, machine-learning outputs are useful only when they are interpreted with attention to predictor correlation, sampling structure, validation design, uncertainty, and the cost of overconfident decisions.

2.1 Environmental prediction under uncertainty

Environmental prediction problems often involve nonlinear relationships, spatial and temporal dependence, correlated covariates, and incomplete observation of the processes being modeled. Species occurrence, pollen exposure, air pollution, county-level disease burden, and sequential decision trajectories are all shaped by multiple interacting factors. A model may reproduce observed patterns while relying on sampling design, county or region effects, temporal trends, or correlated predictors rather than on a single causal mechanism. For this reason, the dissertation treats prediction, interpretation, validation, and risk communication as linked components of the same modeling workflow.

The applied chapters illustrate this principle in two domains. Chapter 3 uses machine learning to separate adult persistence, seedling regeneration, mapped fire-history strata, and future projection uncertainty in southern yellow pine suitability modeling. Chapter 4 evaluates whether airborne pollen provides meaningful environmental signal for county-level lung cancer prediction while emphasizing ecological-design limits, county-grouped validation, leave-one-county-out transfer, and geographic uncertainty. Chapter 5 then

moves from prediction to decision learning by developing a framework for controlling reward volatility when decisions must be made under uncertain outcomes.

2.2 Forest suitability modeling, regeneration, and disturbance context

Species distribution models (SDMs) are widely used to estimate climate suitability and potential range shifts for forest trees [44, 37]. Broad-scale products such as the Climate Change Tree Atlas provide valuable projections of potential suitable habitat for eastern U.S. tree species [51, 84]. However, occurrence-based SDMs estimate realized climatic suitability rather than fundamental physiological tolerance. Observed records can reflect dispersal limits, planting history, competition, soils, land use, disturbance, management, and sampling design. This distinction matters because a suitability map can be informative for planning while still being uncertain as a forecast of growth, survival, or regeneration.

For long-lived trees, life stage is a central source of uncertainty. Adult trees may persist for decades under conditions that no longer support reliable recruitment. Seedling records can provide a more contemporary early-regeneration signal, although seedling presence does not guarantee survival to sapling or canopy stages. Previous work has shown that adult distributions, fecundity, and recruitment can occupy different climate or habitat spaces [87]. Chapter 3 builds on this literature by treating adult-tree and seedling records as separate data streams before modeling.

Disturbance context adds a second layer of interpretation. Fire can influence canopy structure, hardwood competition, seedbed conditions, and regeneration opportunities in several southern pine systems [76, 114]. At the same time, mapped fire-history intersection is not a complete fire-regime descriptor. The source fire-history product maps burned areas associated with prescribed fire and wildfire activity, but the binary analytical classification does not distinguish fire type, fire severity, burn season, fire frequency, time since fire, fuel structure, stand age, or post-fire management. The forest-suitability chapter therefore uses mapped fire-history intersection as a landscape stratum for comparing

realized climate associations in fire-intersected and non-intersected areas, not as a causal fire-effects analysis.

2.3 Environmental exposure signatures and ecological health modeling

Environmental health prediction faces a related interpretive problem: an environmental feature may be predictive without being a direct cause. Lung cancer is shaped by smoking, radon, occupational exposures, outdoor air pollution, demographic structure, screening, and health-care access. Outdoor air pollution and particulate matter have the strongest external environmental evidence for lung cancer risk [50, 45, 101]. Pollen is not an established lung carcinogen, but it is a biologically active, climate-sensitive aeroallergen whose amount, timing, and spatial distribution are related to vegetation, humidity, land cover, seasonality, and geography [22, 24, 81].

The pollen-lung cancer chapter therefore frames pollen as an environmental-regime signal rather than as an isolated carcinogenic exposure. This framing asks whether pollen provides information beyond random noise when evaluated with air quality, humidity, lag windows, county structure, and model-interpretation diagnostics. The chapter remains ecological and hypothesis-generating because county-level model importance cannot establish individual exposure-response relationships. Its contrast between strong pooled prediction and weak county-grouped or leave-one-county-out transfer illustrates why validation design determines how strongly environmental-health findings can be interpreted.

2.4 Model interpretability as diagnostic evidence

Chapter 1 introduced the dissertation's unified interpretability framework. The relevant literature supports a cautious use of multiple diagnostics rather than reliance on a single importance score. Response curves can reveal threshold-like model behavior, but they may be unstable where predictors are sparse or correlated. Mean decrease in

impurity (MDI) is efficient for Random Forest summaries, but it can be biased by scale, split opportunities, and correlation structure [95]. Permutation importance evaluates performance loss after disrupting a feature, but correlated predictors can substitute for one another and produce small or negative values that indicate instability rather than protective effects. SHAP values provide local contribution summaries and can reveal heterogeneity, but they still explain the fitted model rather than a biological or epidemiologic mechanism [66].

Across the dissertation, these tools are used as complementary model diagnostics. Agreement among response curves, MDI, permutation importance, SHAP, grouped predictor summaries, and validation results strengthens confidence in a predictive association. Disagreement among them is not treated as a problem to hide; it is evidence of collinearity, model sensitivity, sampling structure, or domain-specific uncertainty. This interpretation is especially important for the forest and pollen chapters, where many predictors represent correlated environmental gradients rather than independent causal variables.

2.5 Risk-aware learning and short-trajectory policy optimization

Risk-aware learning becomes necessary when average performance is not enough. In forest adaptation, a strategy with high expected suitability may still be risky if it ignores regeneration weakness, future climate extrapolation, or projected suitability loss. In environmental health, a model with strong pooled performance may still be risky if it does not transfer to held-out counties. These applied examples motivate the broader decision-learning problem addressed in Chapter 5: how can a learning system pursue good performance while controlling variability and reducing exposure to poor outcomes?

Reinforcement learning provides a mathematical framework for sequential decision-making, but many standard approaches are risk-neutral and optimize expected return

[96, 77]. Risk-averse reinforcement learning introduces criteria that account for variability, hazardous states, or distributional outcomes [42, 14, 68]. The TOPS framework contributes to this literature by controlling reward volatility and by learning from short trajectories or transition segments rather than requiring long uninterrupted rollouts. This short-trajectory feature is important for safety-critical or data-limited settings because long experimental trajectories may be expensive, risky, or unavailable.

2.6 Synthesis

The related work supports a single dissertation-level argument. Machine learning can help translate complex environmental and decision-system data into useful predictions, but predictive accuracy alone is insufficient. Forest suitability maps must be interpreted through life stage, disturbance strata, future extrapolation, and management risk. Pollen-based health models must be interpreted through ecological design, predictor correlation, county-grouped validation, leave-one-county-out transfer, and geographic validation. Risk-averse policy learning formalizes the decision problem that follows from these applied chapters: when predictions are uncertain and poor outcomes matter, learning systems should account for variability rather than optimizing only average performance. The next three chapters develop this argument in sequence, moving from environmental suitability, to environmental health exposure prediction, to risk-aware policy search.

Chapter 3

Climate-regeneration-fire informatics for future suitability mapping of southern yellow pines

3.1 Background and objectives

This chapter applies the dissertation's interpretability and risk-aware modeling framework to southern yellow pine suitability. Loblolly pine, shortleaf pine, slash pine, and longleaf pine are ecologically and economically important in the southeastern United States, but they differ in climatic breadth, management history, regeneration ecology, and relationship to fire-maintained ecosystems. Climate change may alter the conditions under which these species persist, regenerate, and remain operationally relevant to forest management.

The chapter focuses on two distinctions that are easy to overlook in conventional occurrence-based species distribution modeling. First, adult-tree occurrence and seedling occurrence are treated as separate data streams. Adult records represent established populations and legacy persistence, whereas seedling records represent early regeneration and establishment. Seedling presence is not interpreted as proof of survival to sapling or canopy stages; it is used as an early indicator of regeneration-associated climate suitability.

Second, mapped fire-history intersection is used as a stratification variable, not as a direct fire-effects analysis. Records intersecting the mapped 1994-2023 fire-history product are compared with records without such mapped intersection to evaluate whether realized climate associations differ between mapped fire-history-intersected and non-intersected landscape strata. Because the fire-history layer does not fully describe prescribed fire, severity, seasonality, fuel structure, stand history, or the exact timing of

fire relative to every occurrence observation, wildfire-stratified results are interpreted as disturbance-conditioned climate associations rather than causal wildfire effects.

The chapter develops an integrated climate-regeneration-fire informatics framework for the four focal pines. The framework separates adult-tree, seedling, fire-history-intersected, and non-intersected occurrence records; evaluates multi-algorithm model performance across historical climate baselines; interprets Random Forest models using response curves, mean decrease in impurity (MDI), and SHAP summaries; and translates future projections into management-relevant spatial classes and aggregate infrastructure-proximity summaries. The specific questions are: (1) do adult-tree and seedling records produce different spatial and climatic suitability signals; (2) are model results stable across algorithms and historical climate baselines; (3) which climate-filter groups contribute most to predicted suitability; (4) how do future suitability patterns and aggregate mill-proximity consensus vary across species and scenarios; and (5) does mapped fire-history intersection alter realized climate associations for adult persistence and seedling establishment?

3.2 Materials and methods

3.2.1 Study region, focal species, and conceptual framework

The study focused on the southeastern United States and adjacent portions of the broader southern pine region represented by the occurrence records. The four focal species - loblolly, shortleaf, slash, and longleaf pine - were selected because they are economically, ecologically, and restoration-relevant, and because they differ in climatic breadth, hydrologic association, disturbance ecology, regeneration strategy, and management history.

The analytical framework treated species suitability as a demographic and disturbance-conditioned process rather than as a single pooled occurrence pattern. Adult-tree records

were interpreted as a persistence and legacy signal. Seedling records were interpreted as a recruitment and early-establishment signal. Wildfire-intersected records were interpreted as a fire-associated landscape stratum rather than as evidence of a complete fire regime. These data streams were analyzed separately and then synthesized to distinguish adult-only suitability, seedling-supported suitability, future gain and loss, and wildfire-conditioned regeneration priorities.

3.2.2 Occurrence records and life-stage classification

Occurrence records were obtained from the U.S. Department of Agriculture Forest Service Forest Inventory and Analysis database (FIADB) and organized into adult-tree and seedling data streams. Multiple individuals of the same species and life stage within the same spatial unit were collapsed to a single presence record to reduce pseudoreplication. Records with missing species identity, missing life-stage information, invalid coordinates, or locations outside the study domain were removed. The occurrence maps retained 48,737 adult-tree and 16,733 seedling records for loblolly pine, 24,700 adult-tree and 2,689 seedling records for shortleaf pine, 14,314 adult-tree and 2,364 seedling records for slash pine, and 10,121 adult-tree and 1,183 seedling records for longleaf pine (Fig. 3.1).

Adult-tree records were used to model established populations and legacy persistence. Seedling records were used to model recent regeneration and early establishment. Because seedling presence does not ensure survival to sapling or canopy stages, seedling suitability was interpreted as an early recruitment signal rather than as confirmed population replacement.

3.2.3 Wildfire-intersection stratification

Adult-tree and seedling records were overlaid with mapped wildfire layers and classified as wildfire-intersected or non-wildfire-intersected. The terms wildfire-intersected and

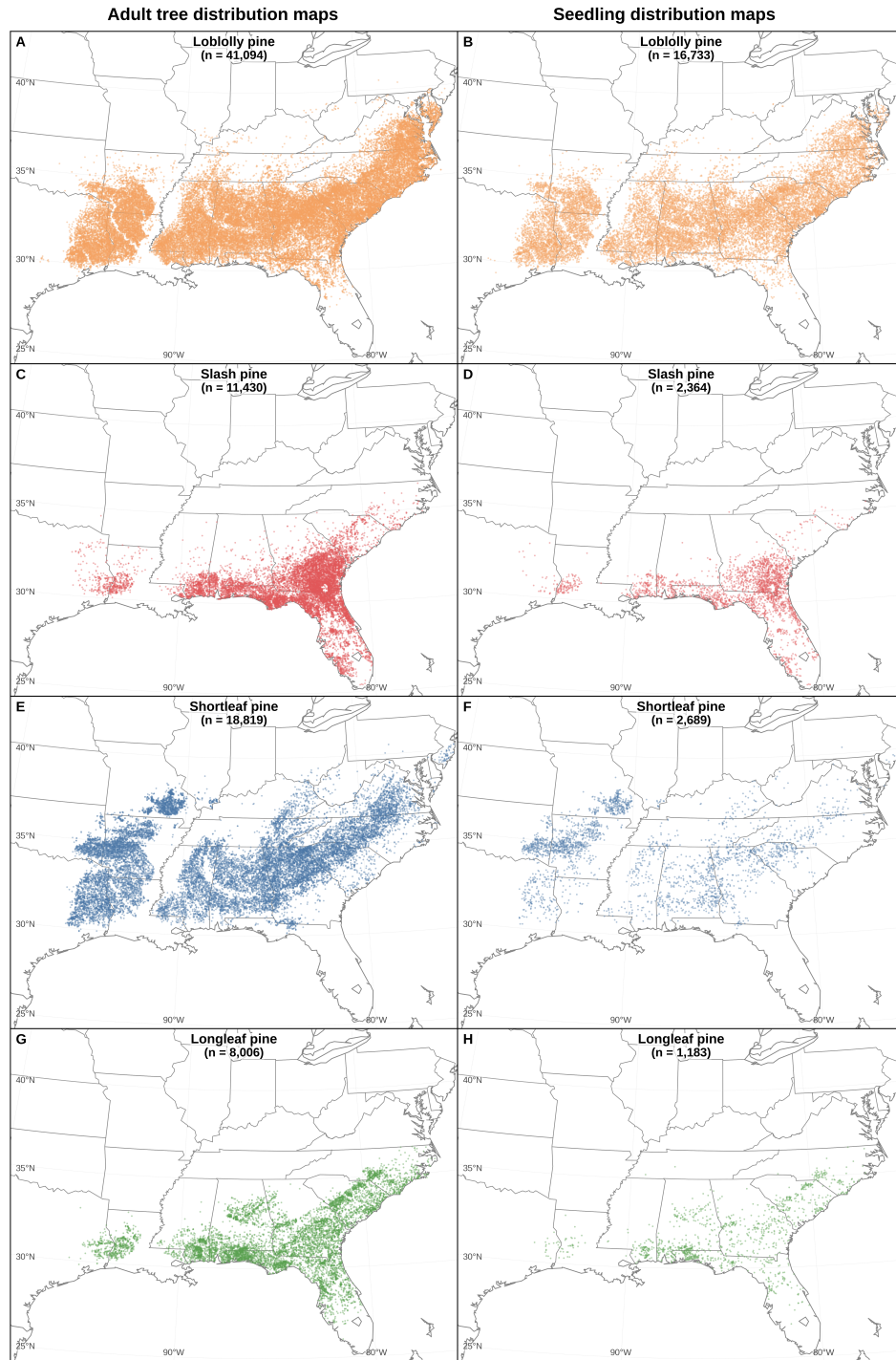


Figure 3.1: Spatial distributions of adult-tree and seedling occurrence records for four southern yellow pine species in the southeastern United States.

non-wildfire-intersected are used throughout this chapter because absence of a mapped wildfire intersection does not mean that a site was never burned, did not experience prescribed fire, or lacks fire history.

Wildfire intersection status was assigned using a fire-history point layer derived from mapped wildfire or burned-area polygons. The working fire-history dataset contained annual burn indicators from 1994 to 2023 and derived fire-history metrics, including fire frequency (FRQ), longest fire-free interval (LFFI), time since previous fire (TSPF), and year last burned (YLB). For each FIADB occurrence record, fire-history points were searched within a $\pm 0.001^\circ$ latitude-longitude window centered on the occurrence location, corresponding to an approximately 100 m spatial neighborhood in the southeastern United States. If one or more fire-history points occurred within this window, the occurrence record was classified as wildfire-intersected and assigned the mean FRQ value of the matched fire-history points. If no fire-history point was found, the record was assigned FRQ = 0 and classified as non-intersected. This procedure represents spatial intersection with mapped fire history over the available 1994-2023 fire-history period; no temporal matching between fire year and FIADB observation year was applied. Therefore, non-intersected records should be interpreted as having no mapped fire intersection in the source product, rather than definitive absence of fire history.

The wildfire-stratified 1991-2020 dataset produced four groups for each species: adult-tree records with wildfire intersection, adult-tree records without wildfire intersection, seedling records with wildfire intersection, and seedling records without wildfire intersection (Fig. 3.5). Wildfire-intersected seedling groups were consistently the smallest strata, whereas adult non-wildfire-intersected groups were consistently the largest. This imbalance was considered when interpreting wildfire-conditioned models.

3.2.4 Climate predictors and historical baseline windows

Climate predictors were assembled for four 30-year historical baselines: 1961-1990, 1971-2000, 1981-2010, and 1991-2020. These baselines were used to test calibration-window sensitivity rather than assuming that one climate normal fully represents the species-climate relationship. The 1991-2020 baseline was treated as the most contemporary historical climate window for interpreting recent adult-seedling and wildfire-conditioned patterns.

Candidate predictors represented biologically interpretable climate dimensions relevant to southern pine distribution and regeneration, including thermal energy, cold limitation, temperature extremes, growing-season or frost timing, moisture availability, evaporative demand or water deficit, heat-moisture balance, humidity, and seasonality or continentality. Variance inflation factor (VIF) screening was conducted separately for each baseline, following common collinearity-screening practice (Dormann et al., 2013). The screening procedure reduced candidate sets from 25 predictors to 15 for the 1961-1990, 1971-2000, and 1981-2010 baselines, and from 24 to 15 for the 1991-2020 baseline, where MAR was unavailable (Fig. A.1). Because retained predictors differed slightly among baselines and represented related climate gradients, interpretation emphasized climate-filter groups rather than treating individual predictors as independent causal drivers.

3.2.5 Background sampling and accessible area

For algorithms requiring absence, background, or pseudoabsence data, background points were sampled within species-specific accessible areas. Each accessible area represented the region in which a species could plausibly occur, disperse, or be sampled during the period represented by the records (Barve et al., 2011; Barbet-Massin et al., 2012). The same accessible area and background design were used for adult-tree and seedling models of a species, and for wildfire-intersected and non-wildfire-intersected comparisons,

so that differences among models reflected occurrence structure rather than different available climate spaces.

3.2.6 Species distribution modeling and validation

Seven SDM algorithms were compared for each species and historical baseline: support vector machine (SVM), generalized linear model (GLM), gradient boosting machine (GBM), Random Forest (RF), naive Bayes model (NBM), extreme gradient boosting (XGBoost), and MaxEnt (Phillips et al., 2006; Feng et al., 2022). Models used the same occurrence records, background or pseudoabsence design, climate predictors, and validation partitions within each species-baseline combination.

Model performance was evaluated using Accuracy, receiver operating characteristic area under the curve (ROC AUC), Cohen's Kappa and True Skill Statistic (TSS). ROC AUC and Kappa were used as primary comparison metrics, with Accuracy and TSS used as supporting information. XGBoost and Random Forest were the most consistently strong algorithms. Random Forest was retained as the main interpretation model because it combined strong predictive performance with response curves, MDI feature importance, and SHAP-based interpretation.

3.2.7 Random Forest interpretation and climate-filter grouping

Random Forest interpretation used response curves, MDI feature importance, and SHAP summaries (Strobl et al., 2007; Lundberg and Lee, 2017). Individual predictors were grouped into eight climate-filter categories: thermal energy, cold limitation, temperature extremes, growing-season or frost timing, moisture availability, evaporative demand or water deficit, heat-moisture balance, and seasonality or continentality. Grouped importance values were summarized for the 1991-2020 baseline and for wildfire-stratified life-stage models (Figs. 3.7 and 3.9).

MDI and SHAP values were interpreted as model-based indicators of predictive contribution, not as causal effects. Because climate predictors are correlated and occurrence data represent realized climate associations, response curves and importance rankings were treated as descriptive model interpretations rather than physiological tolerance thresholds.

3.2.8 Adult-seedling climate-space and spatial agreement analyses

Adult-tree and seedling records were projected into a shared principal component analysis (PCA) climate space using retained 1991-2020 predictors. Centroid distances between adult-tree and seedling records were summarized separately for non-wildfire-intersected and wildfire-intersected strata. The first two PCA axes explained 65.3% and 19.1% of total climate-space variation, respectively (Fig. 3.2).

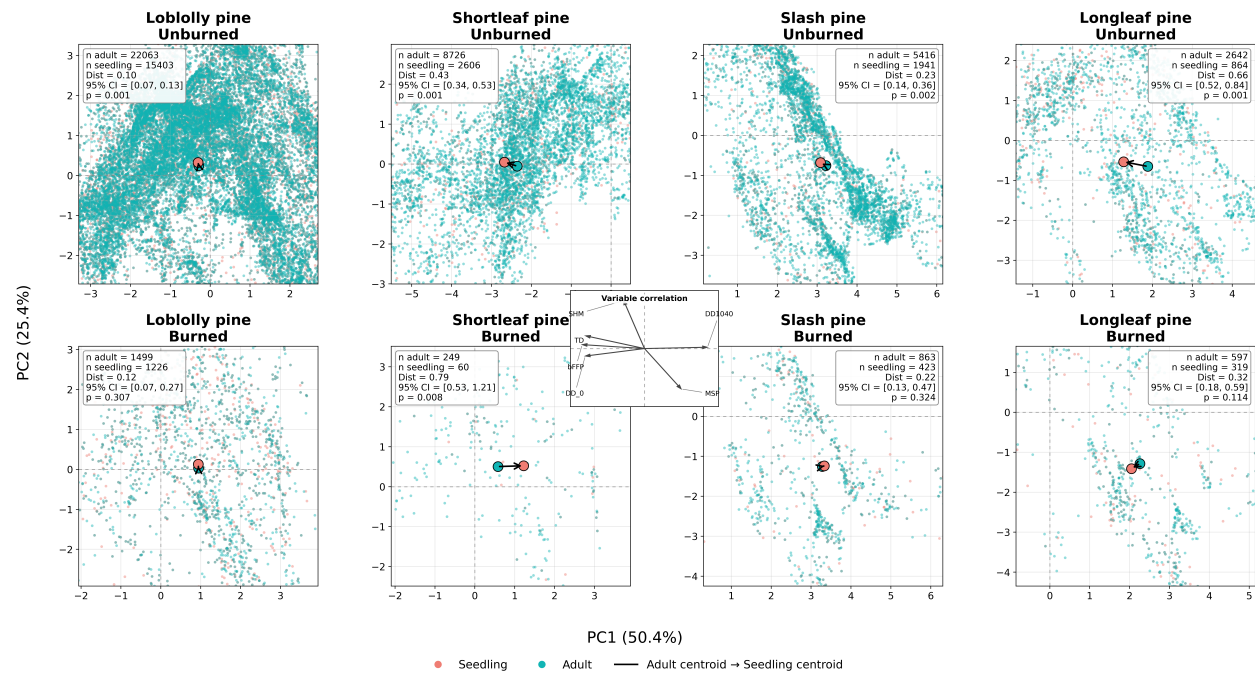


Figure 3.2: Adult-seedling climate-space separation by wildfire-intersection status for four southern yellow pine species during the 1991-2020 climate baseline.

Spatial adult–seedling agreement was evaluated by comparing gridded adult-tree and seedling suitability maps at 0.1° resolution under the 1991–2020 baseline (Fig. 3.3). The

0.1° value refers to the spatial resolution of the rasterized geographic grid, meaning that the study area was divided into regular latitude–longitude grid cells with boundaries spaced by 0.1° in both latitude and longitude. Each grid cell therefore represents one mapping unit for summarizing adult-tree suitability, seedling suitability, and their agreement or disagreement status. This grid size describes the mapping resolution of the suitability surfaces rather than the statistical accuracy of the model or a temperature unit; in ground distance, 0.1° latitude is approximately 11.1 km, while 0.1° longitude varies with latitude and becomes smaller at higher latitudes. After applying the selected suitability threshold to the adult-tree and seedling predictions, grid cells were classified as suitable for both adult trees and seedlings, adult-only suitable, seedling-only suitable, or unsuitable for both groups.

3.2.9 Future projections, non-analog uncertainty, and aggregate mill-proximity consensus

Calibrated SDMs were projected to future climate layers organized by species, historical baseline, climate source, emissions scenario, and future period. All future habitat-suitability projections were generated using Random Forest models for consistency across species, baseline periods, and climate scenarios, as Random Forest consistently performed well during baseline model evaluation. Future maps were generated for SSP126, SSP245, and SSP370 and summarized for early-century (2011-2040), mid-century (2041-2070), and other available future periods. The supplied figure set includes eight-GCM ensemble projections for early and mid-century periods and an average future suitability and change map based on the 1961-1990 baseline model (Fig. 3.4; Figs. A.17 and A.18).

Continuous suitability predictions were classified into unsuitable, low, medium, and high suitability classes with fixed suitability-index intervals (Feng et al., 2022). Specifically, unsuitable/very low suitability = 0-0.20, low suitability = 0.20-0.40, medium suitability

Adult versus seedling modeled climate suitability (1991–2020)
Grid size = 0.1°

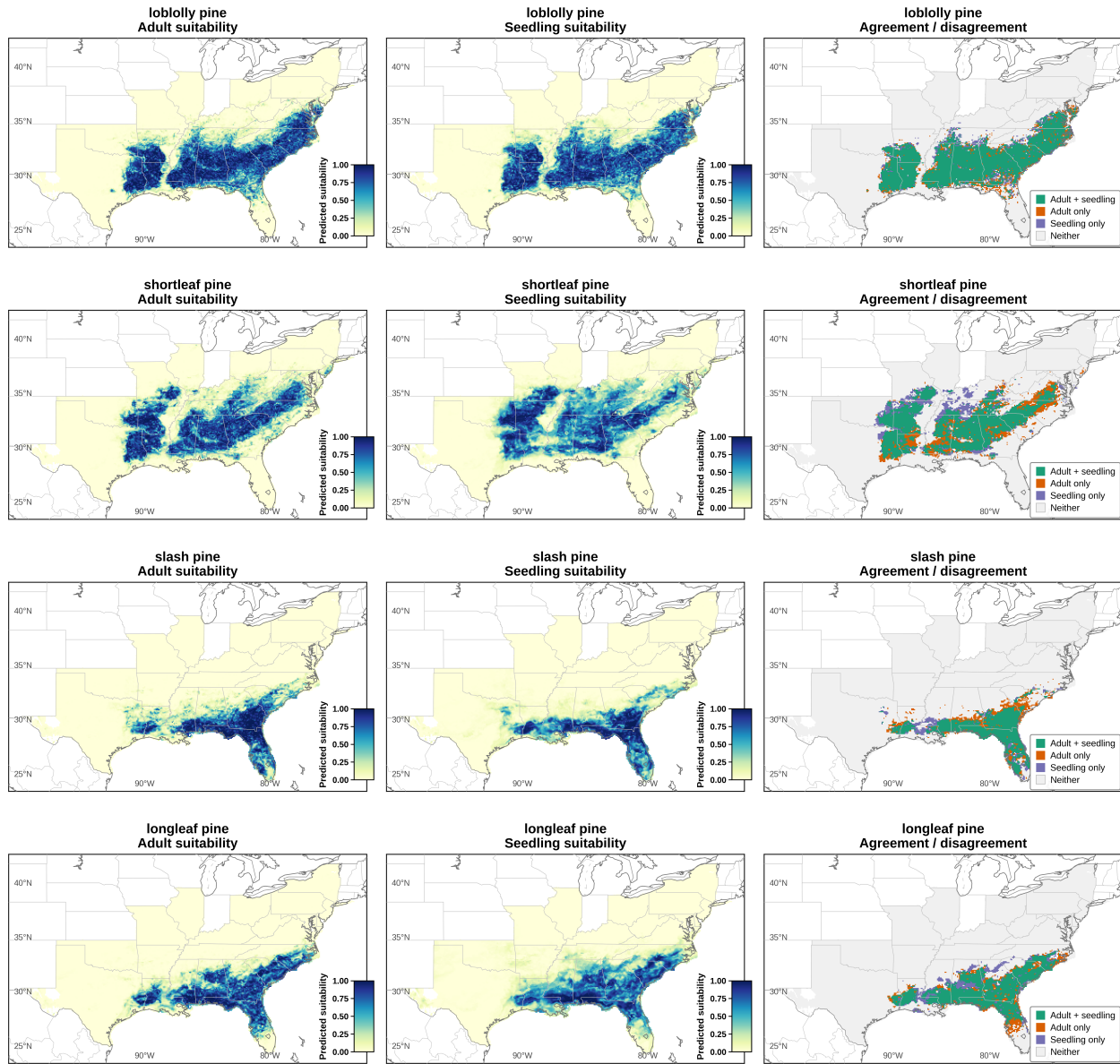


Figure 3.3: Adult versus seedling modeled climate suitability for four southern yellow pine species under the 1991-2020 climate baseline.

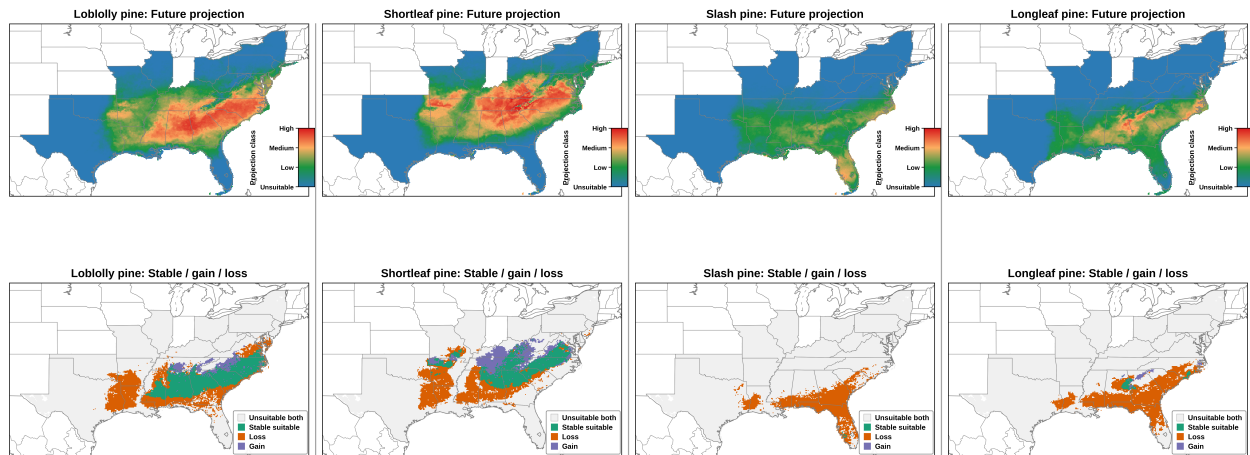


Figure 3.4: Average future climatic suitability and suitability change for four southern yellow pine species relative to the 1961-1990 baseline model.

= 0.40-0.60, and high suitability > 0.60 . This fixed-threshold approach was applied consistently across species, life stages, historical baselines, and future projections to ensure that adult-seedling comparisons and gain/loss classifications were based on the same suitability scale rather than on changing threshold values. Binary low-suitability maps were then derived using the 0.20 cutoff, so cells below 0.20 were treated as low/unsuitable and cells ≥ 0.20 were treated as having at least low suitability.

Binary change maps identified stable suitable, future gain, future loss, and unsuitable-both areas relative to the historical baseline model. A fixed binary suitability threshold of 0.50 was used. Grid cells with both historical and future suitability ≥ 0.50 were classified as stable suitable, cells with historical suitability < 0.50 and future suitability ≥ 0.50 as future gain, cells with historical suitability ≥ 0.50 and future suitability < 0.50 as future loss, and cells with both values < 0.50 as unsuitable-both areas.

Non-analog climate uncertainty was mapped by comparing future climate values with the historical calibration space and identifying dominant climate-filter groups responsible for extrapolation concern (Fig. A.19). The non-analog climate analysis was designed to address the area-of-applicability (AOA) problem in future species-distribution projections. Machine-learning models should not be assumed to provide equally reliable predictions in

all mapped areas, because cross-validation performance is only expected to apply where new prediction environments are sufficiently similar to the training environments (Meyer and Pebesma, 2021). Predictions made into novel predictor space may therefore represent extrapolation rather than true model-supported inference. In this study, this AOA concept was used as the basis for a practical screening approach to identify where future climate conditions fall outside the historical climate space used to support model training. For each species and historical baseline, the retained climate predictors were first obtained from the VIF-filtered predictor set, and species-specific historical reference conditions were defined from adult occurrence training records. For each retained predictor, the central historical envelope was calculated using the 2.5th and 97.5th percentiles. Future climate grids were then aggregated to 0.10° cells and compared against these species-specific historical envelopes. Grid cells were flagged when one or more future predictor values fell below or above the historical range, and the magnitude of departure was standardized by the historical predictor range. These variable-level departures were summarized as the number and proportion of extrapolated predictors, mean and summed extrapolation distance, and an overall extrapolation-risk score. Each grid cell was then assigned to no, low, moderate, or high non-analog risk, and the dominant extrapolating variable was grouped into broader climate-driver categories. This implementation provides a transparent and scalable realization of the AOA principle by identifying future projection areas where southern pine suitability predictions are based on climates that were poorly represented in the historical training domain.

Because the lumber-mill point dataset is restricted, the mill-proximity map is not reported in this dissertation. The mill dataset was used only to generate an aggregate infrastructure-proximity consensus summary. For each evaluated mill, future suitability outputs were screened to determine whether at least one focal southern pine species was projected as climatically suitable within a 50-mile surrounding area. Mills were assigned to consensus tiers based on the percentage of future model combinations that met this

condition, and results are reported only as non-spatial counts and percentages across tiers.

3.3 Results

3.3.1 Occurrence structure and wildfire stratification

Adult-tree and seedling records showed broad spatial overlap across the southeastern United States but were not identical in density, spatial continuity, or range position (Fig. 3.1). Loblolly pine had the broadest and most continuous footprint, spanning much of the southeastern pine region. Shortleaf pine was more interior and upland-centered, whereas slash and longleaf pine were concentrated more strongly in the Gulf Coastal Plain, Atlantic Coastal Plain, and adjacent warm southeastern regions. Seedling maps generally followed the adult occurrence footprint, but seedlings were locally sparser and more discontinuous, especially for shortleaf and longleaf pine.

The all-record occurrence maps retained the largest adult and seedling sample sizes for loblolly pine, followed by shortleaf, slash, and longleaf pine (Table 3.1). Wildfire stratification produced strong sample imbalance. In the 1991-2020 wildfire-stratified subset, adult non-wildfire-intersected records represented 67.1% of loblolly, 85.4% of shortleaf, 67.3% of slash, and 69.3% of longleaf records (Fig. 3.5). Seedling wildfire-intersected records represented only 2.8% of loblolly, 0.6% of shortleaf, 3.5% of slash, and 2.1% of longleaf records. These imbalances are important for interpreting wildfire-conditioned seedling models, which were based on smaller samples than adult or non-wildfire-intersected models.

Table 3.1: Occurrence-record sample sizes by species, life stage, and wildfire-intersection status.

Species	Adult total	Seedling total	Adult wildfire	Adult non-wildfire	Seedling wildfire	Seedling non-wildfire
Loblolly pine	48,737	16,733	4,667	43,408	1,730	14,899
Shortleaf pine	24,700	2,689	1,274	23,131	105	2,561
Slash pine	14,314	2,364	3,084	11,230	582	1,782
Longleaf pine	10,121	1,183	2,292	7,829	389	794

Note. All-record adult and seedling counts are from the occurrence maps in Figure 3.1. Wildfire-intersection counts are from the 1991-2020 wildfire-stratified summary in Figure 3.5. Small differences between all-record and wildfire-stratified totals reflect filtering and overlay-specific subsets in the supplied figure set.

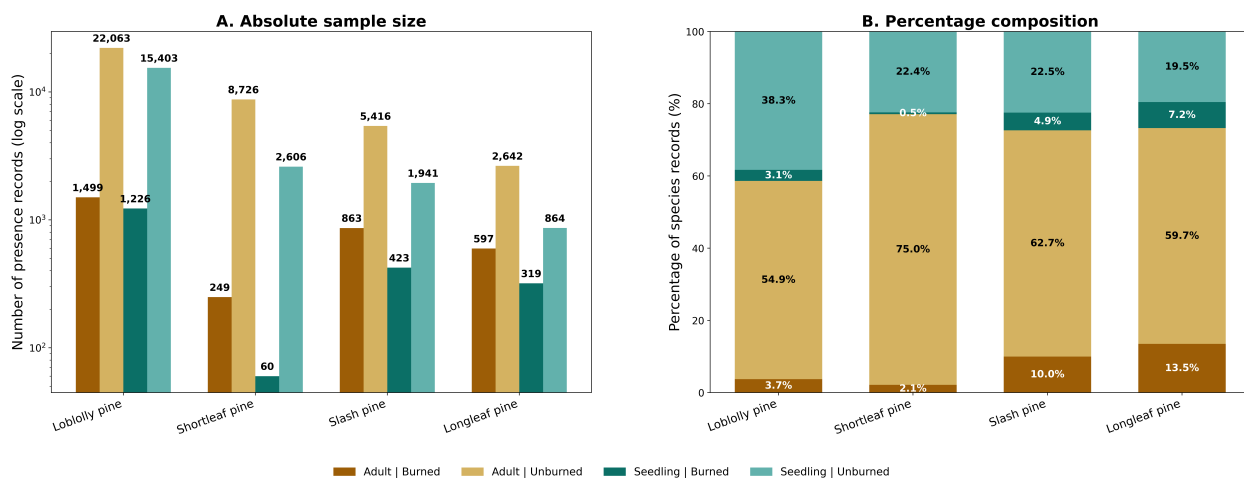


Figure 3.5: Sample size and wildfire stratification of adult-tree and seedling occurrence records for four southern yellow pine species during the 1991-2020 baseline period.

3.3.2 Adult-seedling climate-space separation

The shared PCA climate-space analysis showed that adult-tree and seedling records occupied related but not always identical climate spaces (Fig. 3.2; Table 3.2). In non-wildfire-intersected records, adult-seedling centroid distances were 0.10 for loblolly pine, 0.38 for shortleaf pine, 0.22 for slash pine, and 0.27 for longleaf pine. Distances were significant for loblolly, shortleaf, and longleaf pine ($p = 0.001$), whereas slash pine showed a weaker non-wildfire-intersected separation ($p = 0.129$).

Table 3.2: Adult-seedling climate-space centroid distances by species and wildfire status for the 1991-2020 baseline.

Species	Non-wildfire adult n	Non-wildfire seedling n	Non-wildfire distance (95% CI; p)	Wildfire adult n	Wildfire seedling n	Wildfire distance (95% CI; p)
Loblolly pine	43,408	14,899	0.10 [0.07, 0.12]; $p = 0.001$	4,667	1,730	0.07 [0.03, 0.14]; $p = 0.102$
Shortleaf pine	23,131	2,561	0.38 [0.31, 0.44]; $p = 0.001$	1,274	105	0.27 [0.13, 0.51]; $p = 0.082$
Slash pine	11,230	1,782	0.22 [0.16, 0.34]; $p = 0.129$	3,084	582	0.17 [0.08, 0.31]; $p = 0.031$
Longleaf pine	7,829	794	0.27 [0.20, 0.38]; $p = 0.001$	2,292	389	0.32 [0.20, 0.47]; $p = 0.001$

Note. Distances, confidence intervals, p values, and sample sizes are reported from Figure 3.2. The figure labels use burned and unburned; this chapter interprets those labels as wildfire-intersected and non-wildfire-intersected records.

In wildfire-intersected records, centroid distances were 0.07 for loblolly pine, 0.27 for shortleaf pine, 0.17 for slash pine, and 0.32 for longleaf pine. Slash pine and longleaf pine had significant wildfire-intersected adult-seedling separation ($p = 0.031$ and $p = 0.001$, respectively), whereas loblolly and shortleaf pine were weaker or not significant in the wildfire-intersected stratum ($p = 0.102$ and $p = 0.082$, respectively). These results indicate

that seedling recruitment is not simply a random subset of adult climate space and that the strength of adult-seedling separation varies by species and wildfire-intersection status.

3.3.3 Multi-algorithm validation and baseline-window sensitivity

Model validation showed consistently useful to strong discrimination across species, algorithms, and historical climate baselines (Fig. 3.6, Fig. A.2, and Table A.2). XGBoost and Random Forest were the strongest and most stable algorithms. Under the 1991-2020 baseline, XGBoost produced the highest Kappa for loblolly pine, shortleaf pine, slash pine, and longleaf pine, with Kappa values of 0.753, 0.726, 0.831, and 0.785, respectively. ROC AUC values were also highest or nearly highest for XGBoost and Random Forest. XGBoost achieved ROC AUC values of 0.935, 0.916, 0.966, and 0.938 across the four species, while Random Forest achieved ROC AUC values of 0.933, 0.918, 0.966, and 0.941, respectively (Table 3.3). These results indicate that Random Forest provided performance comparable to XGBoost while also supporting consistent interpretation and future projection across species.

Species differed in climatic separability. Slash pine had the strongest overall discrimination, consistent with its more restricted Coastal Plain climate association. Loblolly and longleaf pine also performed strongly and consistently across baselines. Shortleaf pine had lower relative discrimination, but its models remained useful, consistent with a broader and more heterogeneous realized climate space. Across baselines, top-level validation metrics changed little, but climate-variable interpretation changed more strongly, supporting the use of multiple climate normals for ecological interpretation rather than relying only on a single calibration window.

3.3.4 Climate-filter importance and response curves

Grouped Random Forest importance summaries showed that southern pine suitability was structured by multiple climate dimensions rather than by a single predictor (Fig. 3.7).

Table 3.3: Strongest 1991-2020 model performance by species for XGBoost and Random Forest.

Species	XGBoost ROC AUC	XGBoost Kappa	Random Forest ROC AUC	Random Forest Kappa	Interpreta- tion
Loblolly pine	0.935	0.753	0.933	0.744	Strong; broad realized climate space
Shortleaf pine	0.916	0.726	0.918	0.720	Useful but lower separability
Slash pine	0.966	0.831	0.966	0.829	Strongest climatic separability
Longleaf pine	0.938	0.785	0.941	0.778	Strong; climate and disturbance context important

Note. Values are from Table A.2. AUC is the area under the receiver operating characteristic curve.

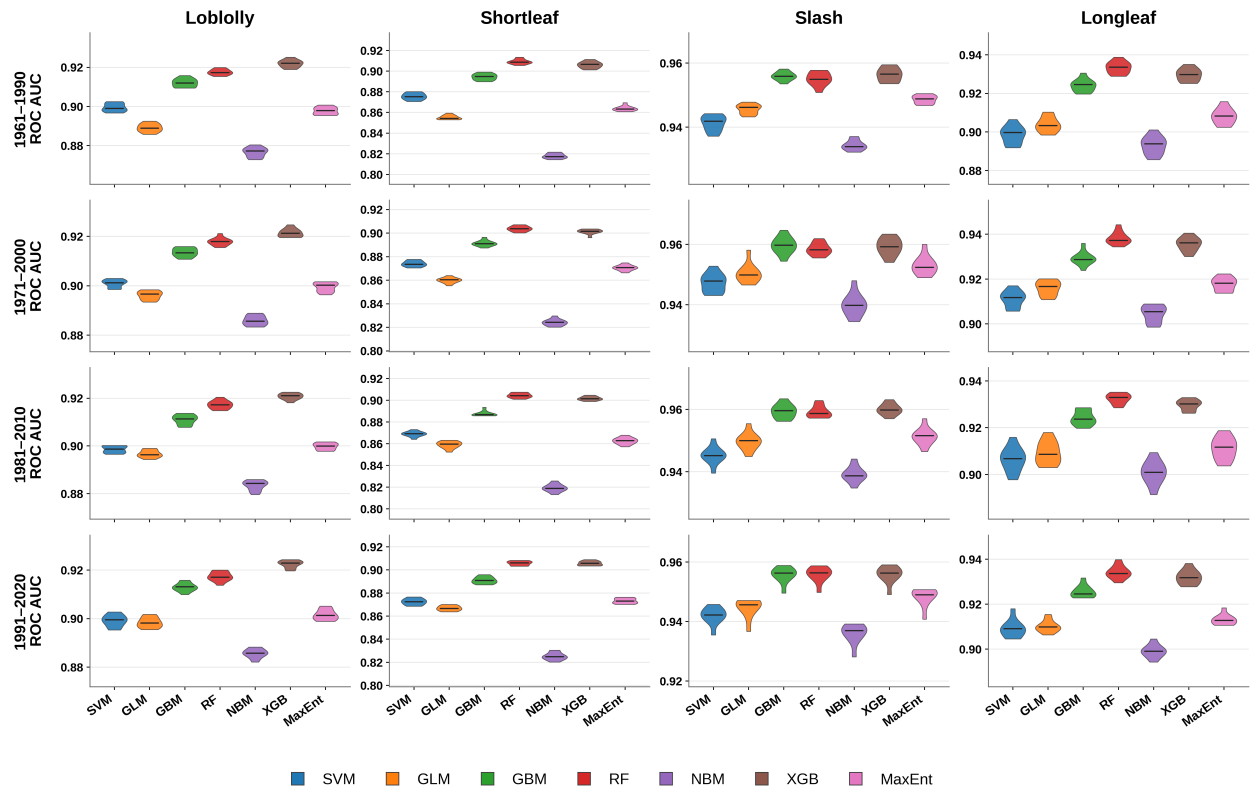


Figure 3.6: Multi-algorithm comparison of species distribution model performance across southern yellow pine species and historical climate baselines.

Moisture availability was the dominant climate-filter group for all four species under both SHAP and MDI summaries. SHAP-based moisture importance ranged from 0.19 for slash pine to 0.29 for shortleaf pine, while MDI-based moisture importance ranged from 0.17 for slash pine to 0.29 for shortleaf pine. Temperature extremes had consistently low relative importance, whereas thermal energy, cold limitation, growing-season or frost timing, evaporative demand or water deficit, heat-moisture balance, and seasonality contributed secondary but meaningful importance.

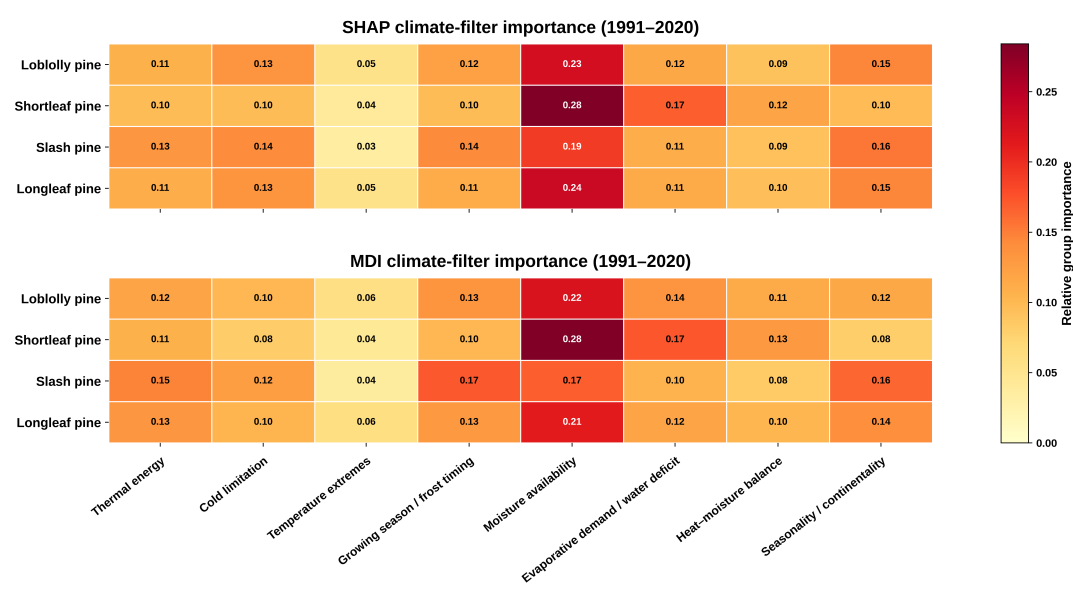


Figure 3.7: Relative importance of climate-filter groups in Random Forest species distribution models for four southern yellow pine species under the 1991-2020 climate baseline.

Response curves reinforced the species-specific interpretation (Fig. 3.8). Shortleaf pine showed strong threshold-like responses to mean annual precipitation (MAP) and reference evaporation or evaporative demand (Eref), as well as responses to temperature seasonality and cold limitation. Slash pine, longleaf pine, and loblolly pine showed stronger threshold responses to temperature difference or continentality (TD), degree-days below 0 degrees C (DD_0), degree-day accumulation (DD1040), precipitation as snow (PAS), and frost-free-period variables such as NFFD. These curves support a climate-filter

interpretation in which shortleaf pine is relatively more moisture and water-balance sensitive, whereas slash, longleaf, and loblolly pine are more strongly structured by thermal seasonality, cold limitation, heat accumulation, and growing-season constraints.

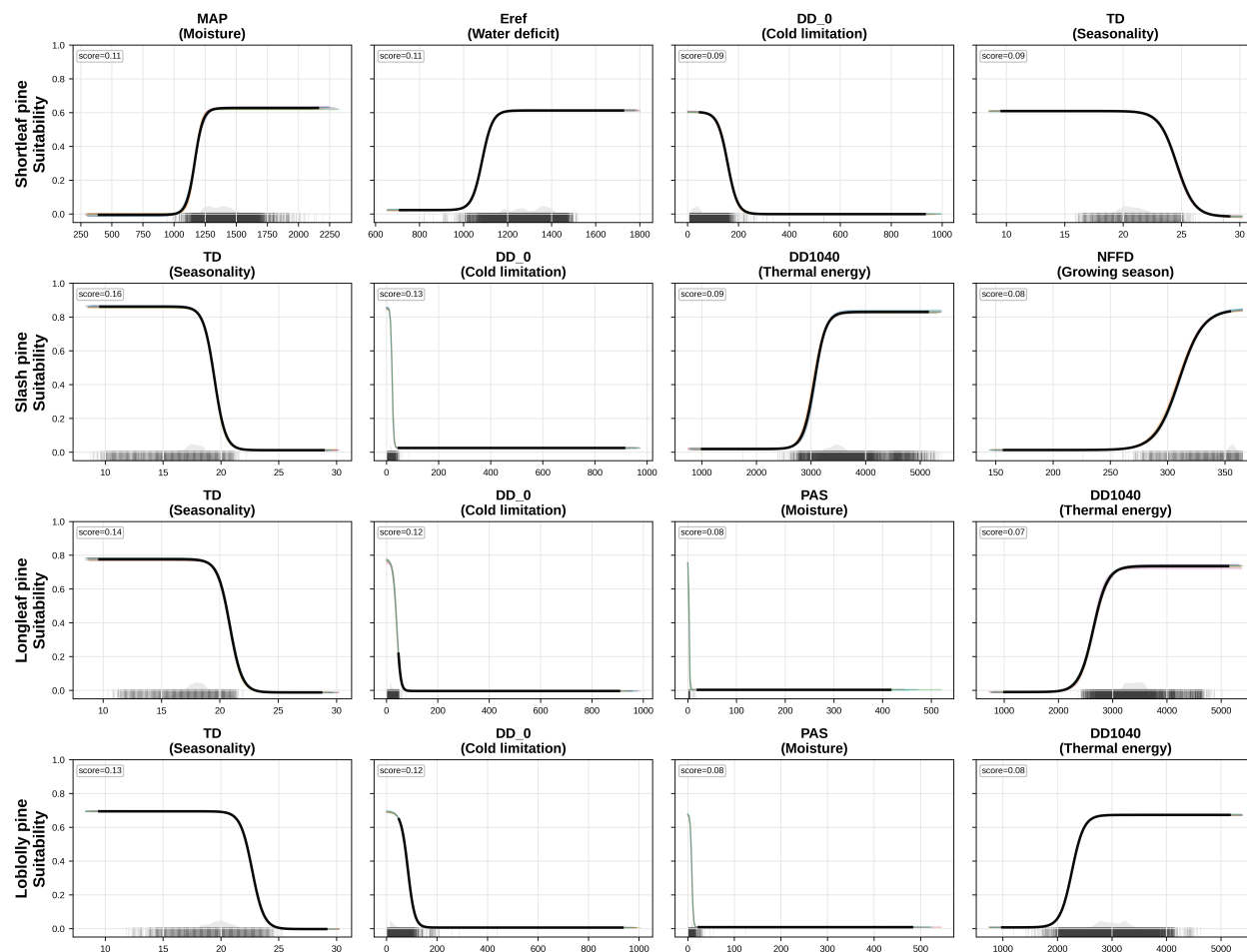


Figure 3.8: Random Forest response curves for selected climate predictors in southern yellow pine species distribution models under the 1991-2020 climate baseline.

Individual predictor rankings from MDI and SHAP were broadly consistent but not identical (Figs. A.3, A.4, and A.5–A.8). The most robust inference was therefore at the climate-filter level. For example, loblolly, slash, and longleaf pine were repeatedly associated with TD, DD_0, DD1040, NFFD, and PAS, while shortleaf pine repeatedly emphasized MAP and Eref. These patterns should be interpreted as realized climate associations rather than direct physiological thresholds.

3.3.5 Adult versus seedling modeled suitability

Adult and seedling suitability maps for 1991-2020 showed broad agreement but also important spatial mismatches (Fig. 3.3). Loblolly pine had extensive adult and seedling suitability across the central and eastern southern pine belt, but adult-only and seedling-only patches occurred around range margins and local discontinuities. Shortleaf pine showed a more complex mismatch, with adult suitability extending across interior and upland portions of the region and seedling suitability showing different local concentrations. Slash pine had adult and seedling suitability concentrated in the warmer southeastern Coastal Plain, with adult-only and seedling-only areas along the interior and coastal margins. Longleaf pine showed strong overlap across parts of the Gulf and Atlantic Coastal Plain, but adult-only areas were evident along some margins, indicating potential regeneration limitation or weaker seedling-supported suitability.

These adult-seedling maps translate the occurrence and climate-space results into management-relevant classes. Cells suitable for both adult trees and seedlings represent seedling-supported suitability and may be stronger candidates for persistence and regeneration-supported resilience. Adult-only suitable cells represent places where established trees may persist but modeled seedling suitability is weaker. Seedling-only cells represent areas where the seedling climate signal occurs outside the strongest adult suitability footprint.

These classes are intended as screening categories, not management prescriptions. Seedling-supported areas should be interpreted as places where modeled adult persistence and early-regeneration climate signals agree, whereas adult-only areas indicate locations where adult persistence is modeled more strongly than seedling suitability. Adult-only status does not prove regeneration failure, and seedling-supported status does not prove long-term recruitment success. Independent survival, sapling recruitment, growth, and management data are needed before translating these classes into site-level decisions.

3.3.6 Wildfire-conditioned climate filters

Wildfire-stratified Random Forest models showed that life stage and wildfire intersection altered climate-filter rankings, especially for seedling records (Fig. 3.9; Figs. A.9–A.12 and A.13–A.16). Across species, wildfire-intersected seedling models often gave greater relative importance to moisture availability, evaporative demand or water deficit, heat-moisture balance, or growing-season and frost-timing filters than corresponding non-wildfire-intersected models. The direction and strength of these shifts differed by species.

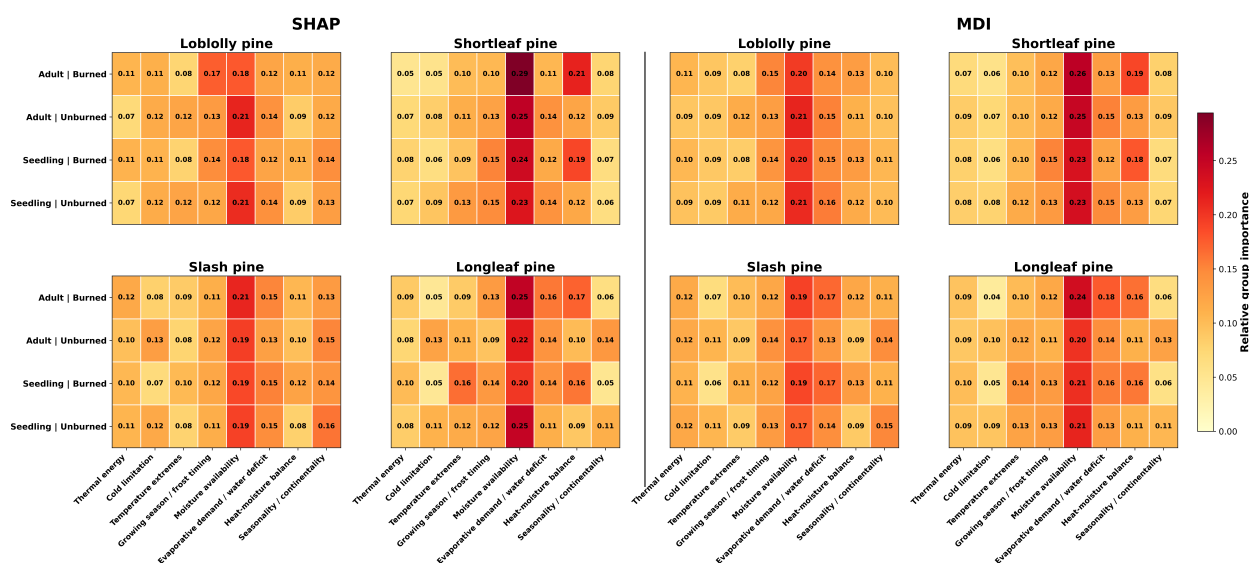


Figure 3.9: Relative importance of wildfire-stratified climate filters in Random Forest models for southern yellow pines.

Shortleaf pine wildfire-intersected records emphasized moisture availability and heat-moisture filters strongly. In the SHAP group summary, moisture availability reached 0.27 in seedling wildfire-intersected records and 0.33 in adult wildfire-intersected records. Slash pine showed a clearer contrast between non-wildfire-intersected thermal filters and wildfire-intersected combined thermal-moisture filters. Longleaf pine also showed wildfire-conditioned shifts, with wildfire-intersected seedlings emphasizing moisture availability, evaporative demand, and heat-moisture balance more strongly than some non-wildfire-intersected comparisons. Loblolly pine showed a more intermediate pattern

in which thermal and seasonal filters remained important, but seasonal moisture and heat-moisture filters gained importance in wildfire-intersected groups.

Because mapped fire-history intersection does not include complete fire severity, burn season, prescribed burning, fuel structure, stand history, or local post-fire management, these results should not be interpreted as causal effects of wildfire. They compare realized climate associations in records intersecting the mapped 1994-2023 fire-history product with records lacking such mapped intersection, with the strongest caution needed for small seedling strata.

3.3.7 Future suitability, change classes, uncertainty, and mill proximity

Future projections translated the fitted SDMs into spatial suitability and change classes (Fig. 3.4; Figs. A.17–A.19). In the early-century eight-GCM ensemble, loblolly and shortleaf pine retained broad medium-to-high suitability across the interior Southeast. Loblolly pine had 54-56% of mapped cells classified as suitable (low, medium, or high) across SSP126, SSP245, and SSP370, with 14-16% in the high-suitability class. Shortleaf pine had 59-61% suitable cells and 14-15% high suitability across the same scenarios. Slash and longleaf pine had more restricted high-suitability area: slash pine had 47-49% suitable cells with 0-1% high suitability, and longleaf pine had 46-47% suitable cells with 2% high suitability (Table 3.4).

By mid-century, loblolly and shortleaf pine still retained broad areas of medium-to-high suitability, but high-suitability area declined under higher-emissions scenarios. Loblolly high suitability decreased from 12% under SSP126 to 5% under SSP370, while shortleaf high suitability decreased from 12% to 9%. Slash pine and longleaf pine showed little or no high-suitability area by 2041-2070; slash remained mostly low to medium suitability, and longleaf shifted toward low suitability with declining medium-suitability area under higher emissions (Table 3.4).

Table 3.4: Future suitability-class percentages from the supplied early- and mid-century eight-GCM ensemble maps.

Species	Period	Scenario	Unsuitable (%)	Low (%)	Medium (%)	High (%)
Loblolly pine	2011-2040	SSP126	44	17	24	16
Loblolly pine	2011-2040	SSP245	45	16	24	14
Loblolly pine	2011-2040	SSP370	46	16	22	16
Shortleaf pine	2011-2040	SSP126	39	26	20	15
Shortleaf pine	2011-2040	SSP245	41	25	19	14
Shortleaf pine	2011-2040	SSP370	40	23	22	15
Slash pine	2011-2040	SSP126	51	29	19	1
Slash pine	2011-2040	SSP245	53	29	18	0
Slash pine	2011-2040	SSP370	53	27	20	1
Longleaf pine	2011-2040	SSP126	53	29	16	2
Longleaf pine	2011-2040	SSP245	54	32	13	2
Longleaf pine	2011-2040	SSP370	54	29	15	2
Loblolly pine	2041-2070	SSP126	43	20	26	12
Loblolly pine	2041-2070	SSP245	43	20	28	9
Loblolly pine	2041-2070	SSP370	44	21	30	5
Shortleaf pine	2041-2070	SSP126	41	26	21	12
Shortleaf pine	2041-2070	SSP245	43	27	20	10
Shortleaf pine	2041-2070	SSP370	44	27	20	9
Slash pine	2041-2070	SSP126	48	32	19	0
Slash pine	2041-2070	SSP245	49	35	15	0
Slash pine	2041-2070	SSP370	48	37	15	0
Longleaf pine	2041-2070	SSP126	50	40	10	0
Longleaf pine	2041-2070	SSP245	51	42	7	0
Longleaf pine	2041-2070	SSP370	51	44	5	0

Note. Percentages are the inset values shown in Figures A.17 and A.18. Rounding follows the supplied maps.

The average future change maps showed species-specific patterns of stable suitable area, gains, and losses (Fig. 3.4). Loblolly and shortleaf pine retained larger stable suitable cores, with gain and loss areas occurring around range margins and interior transitions. Slash and longleaf pine showed smaller stable cores and more extensive loss areas along portions of the Coastal Plain. Non-analog climate maps flagged areas where future predictions relied on climate conditions near or outside the historical calibration space, with dominant extrapolation drivers including temperature, growing-season variables, evaporative demand, and soil moisture (Fig. A.19).

The aggregate mill-proximity synthesis indicated that most evaluated lumber mills remained near future modeled suitability for at least one focal southern pine species. Of 1,327 mills, 1,141 mills (86.0%) fell in the high-consensus tier, meaning that more than 75% of the 180 future model-result files projected at least one suitable southern pine species within 50 miles. In contrast, 124 mills (9.3%) were in the low-consensus tier, 25 mills (1.9%) in the low-to-moderate tier, and 37 mills (2.8%) in the moderate tier. These results are reported only as aggregate counts and percentages because the underlying point layer is restricted.

3.4 Discussion

3.4.1 Adult persistence and seedling recruitment are different suitability signals

The central result of this study is that adult-tree and seedling records are related but not interchangeable. Adult records describe established populations and incorporate historical climate, planting, management, disturbance history, land-use legacy, and survivorship. Seedling records provide a more contemporary signal of early regeneration. The occurrence maps, PCA climate-space summaries, and 1991-2020 adult-seedling suitability maps all show that adult and seedling patterns can diverge. This means that adult-only

SDMs may overstate long-term population stability if current regeneration is weak or climatically displaced from adult persistence.

For management, this distinction is useful because adult-only and seedling-supported suitability imply different actions. Areas suitable for both adults and seedlings may be stronger candidates for persistence, conservation, monitoring, or restoration investment. Adult-only suitable areas may require management intervention, including planting, site preparation, thinning, competition control, prescribed fire, or seed-source selection. Seedling-only suitable areas may indicate emerging regeneration opportunities outside the strongest adult suitability footprint, although these areas require field validation before management interpretation.

3.4.2 Species-specific climate filters explain differences in climatic separability

The four species differed in climatic separability and dominant climate filters. Slash pine was the most climatically separable species, consistent with its concentration in warm, humid Coastal Plain environments. Its models emphasized thermal seasonality, cold limitation, heat accumulation, and frost-free-period variables. Loblolly pine had broad and consistently strong model performance, reflecting a wide realized climate space influenced by both natural occurrence and extensive planting. Longleaf pine showed strong thermal and seasonal signals, but its suitability must be interpreted alongside fire-maintained stand structure and regeneration ecology. Shortleaf pine had lower relative discrimination, likely reflecting a broader, more heterogeneous distribution and stronger influence of moisture availability, evaporative demand, soils, competition, land-use history, and stand structure.

The climate-filter grouping provides a more stable interpretation than single-variable rankings. Although TD, DD_0, DD1040, NFFD, PAS, MAP, Eref, CMD, and related predictors were frequently important, these variables represent correlated dimensions of thermal energy, cold limitation, moisture, and seasonality. Grouped SHAP and MDI summaries showed that moisture availability was the largest filter across all four species, while other

filters made substantial secondary contributions. This supports an interpretation of southern pine suitability as a multivariate climate-filter problem rather than a response to a single climatic threshold.

3.4.3 Wildfire intersection modifies realized climate associations but does not describe the full fire regime

Wildfire-conditioned models showed that records intersecting mapped wildfire landscapes often emphasized different climate filters from non-wildfire-intersected records. These differences were strongest in seedling models, suggesting that fire-associated recruitment may be more climate conditional than adult persistence. In several species, wildfire-intersected seedling records placed greater importance on moisture availability, heat-moisture balance, evaporative demand, or growing-season and frost-timing filters. This pattern is biologically plausible because seedling establishment after disturbance can be constrained by moisture stress, heat, competition, seedbed conditions, and post-fire management.

The wildfire results are especially relevant for longleaf and shortleaf pine, which are associated with fire-influenced ecosystems; recent shortleaf pine work has also highlighted regeneration sensitivity under current fire regimes (Chen et al., 2025). Fire can reduce hardwood competition and maintain open canopy conditions, but regeneration benefits depend on climate context and fire regime. Wildfire intersection alone cannot capture prescribed fire, severity, frequency, seasonality, time since fire, fuel structure, or stand history. Therefore, the results should be interpreted as disturbance-conditioned realized climate associations, not as causal wildfire effects. Future work should integrate prescribed-burn records, burn severity, fire-return intervals, and local stand conditions.

3.4.4 Baseline sensitivity and future projection uncertainty

Historical baseline choice had limited effect on top-level validation metrics but stronger effect on climate-variable interpretation. This finding is important because a model can maintain similar ROC AUC while relying on different climate gradients. Multi-baseline calibration therefore provides a more transparent assessment of ecological interpretation and future projection uncertainty than a single baseline alone.

Future projections showed broad retained suitability for loblolly and shortleaf pine but more restricted high-suitability patterns for slash and longleaf pine, particularly by mid-century. However, these projections are realized climatic suitability maps, not direct forecasts of growth, survival, establishment, timber supply, dispersal, planting feasibility, or physiological tolerance. Non-analog climate maps should be used alongside suitability maps to identify areas where model extrapolation is more uncertain. This is particularly important under higher-emissions scenarios and near the edges of the historical calibration space.

3.4.5 Implications for adaptation planning and forest management

The framework can support several management decisions. Stable high-suitability areas with adult and seedling agreement may be candidates for conservation, monitoring, seed-source preservation, or low-risk regeneration investment. Adult-only suitability can flag areas where established trees remain but seedling-supported suitability is weak, suggesting a need for regeneration assistance or site-specific diagnosis. Future-loss areas can guide monitoring and adaptation planning, whereas future-gain areas may identify potential opportunities for assisted regeneration or planting, subject to soils, ownership, dispersal, genetics, pests, and operational feasibility.

The aggregate mill-proximity analysis provides an additional industry-relevant synthesis without displaying the restricted point layer. Most evaluated mills had high consensus

for at least one suitable southern pine species within 50 miles across the future model-result ensemble, suggesting broad modeled climatic support for at least one focal species near much of the existing forest-products infrastructure. Lower-consensus tiers may warrant more detailed local assessment, including species diversification, regeneration monitoring, climate-risk screening, and supply-chain adaptation planning.

3.4.6 Limitations and next steps

Several limitations should be considered when interpreting the results and when extending this work. First, seedling presence is an early recruitment indicator, not proof of survival to sapling or canopy stages. Drought, competition, herbivory, pests, pathogens, fire timing, and management can all affect seedling survival. Future validation should use independent demographic information, including seedling survival, sapling recruitment, growth, mortality, and stand-development observations in adult-only, seedling-supported, future-gain, and future-loss areas. Second, wildfire intersection is incomplete as a fire-regime descriptor. Prescribed fire, severity, frequency, seasonality, and time since fire should be incorporated where available. Third, random validation splits can overestimate performance when occurrence data are spatially autocorrelated; spatial block cross-validation would strengthen evaluation (Roberts et al., 2017). Fourth, ROC AUC is sensitive to background extent and should be interpreted with Kappa, calibration diagnostics, and spatially independent validation where possible (Lobo et al., 2008). Fifth, public FIA coordinates are spatially protected for confidentiality, which can affect fine-scale overlays with wildfire perimeters or climate rasters (Bechtold and Patterson, 2005; USDA Forest Service, 2023).

Finally, MDI and SHAP should be treated as complementary model-interpretation tools rather than causal estimators. MDI can be biased by scale and correlation structure, while SHAP summaries describe contribution to model predictions within the fitted model space (Strobl et al., 2007; Lundberg and Lee, 2017). The most robust conclusions from

this study are therefore the separation of adult and seedling suitability, the species-specific climate-filter patterns, the wildfire-conditioned shifts in seedling climate associations, and the management value of translating SDM outputs into persistence, regeneration, change, and uncertainty classes.

3.5 Conclusions

Southern yellow pine climate suitability is best understood as a demographic and disturbance-conditioned process. Adult-tree and seedling records were spatially and climatically related but not interchangeable, confirming that adult-only SDMs can overlook regeneration limitations. The four species differed in climatic separability and dominant climate filters: slash pine was the most climatically separable; loblolly and longleaf pine showed strong but different thermal and moisture-associated signals; and shortleaf pine showed a broader, moisture-sensitive realized climate space. Wildfire-intersected records, especially seedlings, often emphasized different climate filters than non-wildfire-intersected records, indicating that fire-associated recruitment is climate conditional. Future projections and aggregate infrastructure-proximity summaries translated these model outputs into management-relevant categories, including stable suitability, future gain, future loss, adult-only suitability, seedling-supported suitability, non-analog uncertainty, and disturbance-conditioned regeneration priorities. For long-lived southern pines, future suitability is not only a question of where adult trees can persist, but also where seedlings can establish and where disturbance and management conditions support regeneration under changing climate.

This chapter demonstrates how interpretable machine learning can identify ecological suitability signals while also exposing uncertainty in regeneration, disturbance strata, and future projections. Chapter 4 applies the same broad logic to environmental health, where the central question becomes whether pollen adds meaningful environmental signal to county-level lung cancer prediction without overinterpreting ecological model outputs.

Chapter 4

Airborne Pollen Signatures and County-Level Lung Cancer Burden

Chapter 3 used interpretable machine learning to separate ecological suitability signals from regeneration and projection uncertainty. This chapter applies the same dissertation-level emphasis on cautious interpretation, validation, and risk awareness to environmental health, where the modeling objective is to determine whether pollen contributes meaningful county-level environmental signal without treating ecological prediction as causal evidence.

4.1 Background and objectives

4.1.1 Lung cancer burden and environmental context

Lung cancer remains a major public health burden in the United States and globally. Lung and bronchus cancer has high mortality relative to many other common cancers, and county-level rates can vary substantially because of differences in age structure, historical smoking patterns, radon exposure, occupational exposures, air pollution, socioeconomic conditions, screening, and access to diagnostic and therapeutic services [82, 23].

These known determinants must remain central when interpreting any environmental prediction model. A model that identifies an environmental predictor as important does not imply that the predictor replaces tobacco smoking, radon, or other established causes. Rather, it indicates that the predictor helps explain outcome variation in the available data and should be evaluated in the larger context of lung cancer epidemiology.

Outdoor air pollution provides the strongest established environmental rationale for the present study. The International Agency for Research on Cancer classified outdoor air pollution and particulate matter in outdoor air pollution as carcinogenic to humans, with

sufficient evidence for lung cancer [50]. Fine particulate matter is particularly important because small particles can be inhaled deeply into the respiratory tract and may contribute to systemic biological responses [104, 45, 101]. The Air Quality Index (AQI) is not a single exposure, but it provides a summary communication measure of short-term outdoor air-quality conditions and health concern [2]. Because AQI summaries correlate with pollutants and meteorology, they are useful for prediction but must be interpreted carefully.

4.1.2 Why focus on pollen?

Compared with air pollution, pollen has received much less attention in lung cancer research. This is understandable because pollen is not currently classified as a lung carcinogen. However, pollen is biologically active and environmentally structured. Pollen is an airborne allergen whose concentration, season timing, distribution, and allergenicity vary by plant species, climate, land cover, humidity, and geography [22, 24]. Climate change and urban environmental conditions can alter pollen exposure profiles by affecting pollen amount, season length, and plant distribution [22, 32]. These characteristics make pollen a potential marker of county-level environmental regimes, even if pollen is not an isolated carcinogenic exposure.

There are also mechanistic reasons to investigate pollen in respiratory-health prediction. Pollen exposure can trigger allergic rhinitis and asthma attacks and is linked with respiratory symptoms and airway inflammation [22, 24, 81]. Experimental evidence shows that pollen can generate reactive oxygen species through intrinsic NADPH oxidase activity and can augment allergic airway inflammation [15, 33]. Pollen-induced oxidative stress has also been shown to influence dendritic cell function and innate and adaptive immune responses [29]. Chronic inflammation, oxidative stress, and immune dysregulation are biologically relevant to cancer development and progression, although the present ecological analysis cannot determine whether pollen contributes causally to lung carcinogenesis [46, 70, 28].

Accordingly, pollen was evaluated as a marker of environmental and inflammatory conditions that may contribute to county-level lung cancer prediction when considered with PM2.5, AQI, other pollutants, humidity, and spatial-temporal structure.

4.1.3 Study rationale, hypothesis, and objectives

The current study was designed as an ecological machine-learning analysis rather than an individual-level etiologic study. The unit of analysis was the county-year. This design is useful for exploring spatial-temporal environmental signatures, but it cannot support individual-level risk conclusions. The study asked whether pollen variables provide meaningful environmental signal, rather than random noise, for prediction of annual lung cancer incidence and mortality across selected counties and whether that signal remains visible after evaluating air-quality covariates, humidity, lag windows, feature collinearity, model interpretability, and county-aware validation.

We hypothesized that pollen taxa with strong geographic and seasonal structure, especially *Ambrosia*, *Quercus*, *Pinaceae*, and grasses/*Poaceae*, would contribute to lung cancer outcome prediction as environmental-regime markers. We further hypothesized that pollen importance would be context-dependent because pollen co-varies with vegetation, climate, humidity, air quality, and county structure. Humidity was evaluated as a contextual factor for model explanation because it can affect pollen release, transport, fragmentation, particle behavior, and respiratory irritation, while also co-varying with seasonality and air-quality conditions. Finally, we expected incidence models to perform better than mortality models because mortality is influenced by downstream clinical, demographic, and health-system factors that were not fully represented in the environmental predictor set.

4.2 Materials and Methods

4.2.1 Study design and analytic unit

We performed a retrospective ecological prediction study using annual county-level observations. The analytic unit was the county-year rather than the individual patient. All associations and model explanations therefore apply to county-level environmental signatures and county-level lung cancer burden and do not represent individual exposure-response relationships.

The analysis included 11 counties observed from 2015 through 2021: Pulaski, Sarasota, Cobb, Madison, Chatham, Jefferson, Collin, McLennan, Denton, Harris, and Bexar. These counties form a geographically heterogeneous sample spanning Arkansas, Florida, Georgia, Alabama, and Texas. The small number of counties makes the analysis exploratory and increases the importance of sensitivity analyses and grouped validation.

4.2.2 Outcomes

The primary outcomes were annual county-level lung cancer incidence and mortality rates from 2015 to 2021. Incidence and mortality were modeled separately because they represent related but distinct epidemiologic endpoints. Incidence is closer to disease occurrence, whereas mortality reflects both disease occurrence and downstream processes such as stage at diagnosis, treatment availability, survival, comorbidity, demographic composition, and health-care access.

Before predictive modeling, descriptive maps and county-by-year heatmaps were used to evaluate spatial heterogeneity and time trends, because a model can appear accurate if it learns stable county baselines or broad time trends rather than biologically meaningful exposure variation (Figure 4.1).

4.2.3 Pollen exposures

The principal exposures of interest were airborne pollen variables. The analysis included five pollen taxa or taxonomic groups: Ambrosia, Gramineae/Poaceae, Cupressaceae, Pinaceae, and Quercus. These pollen groups were selected because they represent common aeroallergen or vegetation signals and because they displayed measurable spatial and seasonal variation in the available data (Figure 4.2).

Pollen constituted the primary environmental exposure domain but was not analyzed in isolation. It was evaluated as part of a broader environmental exposome that included vegetation, climate, humidity, particulate pollution, gaseous pollutants, and AQI summaries (Table 4.1).

Table 4.1: Environmental predictor domains and interpretation.

Domain	Variables	Primary interpretive role
Pollen	Ambrosia; Gramineae/Poaceae; Cupressaceae; Pinaceae; Quercus	Main hypothesis-generating exposures; interpreted as aeroallergen, vegetation, seasonality, climate, airway inflammation, and geographic signals.
Air pollution	PM2.5; PM10; NO2; ozone	Established or plausible inhaled environmental exposures. PM2.5 carries the strongest external lung cancer evidence.
AQI summaries	Max AQI; median AQI; 90th percentile AQI	Integrated air-quality indicators that are useful for prediction but potentially collinear with pollutants.
Meteorology	Humidity index; absolute humidity	Contextual variables related to pollen dynamics, particle behavior, seasonality, and respiratory irritation.
Space and time	County; year; lag-window summaries	Capture geographic heterogeneity, temporal trends, and historical exposure-window alignment.

4.2.4 Air-quality and meteorological features

Co-exposures included PM2.5, PM10, nitrogen dioxide (NO2), ozone, Max AQI, median AQI, and 90th percentile AQI. PM2.5 was treated as the strongest externally supported air-pollution feature because of prior evidence linking fine particulate matter and

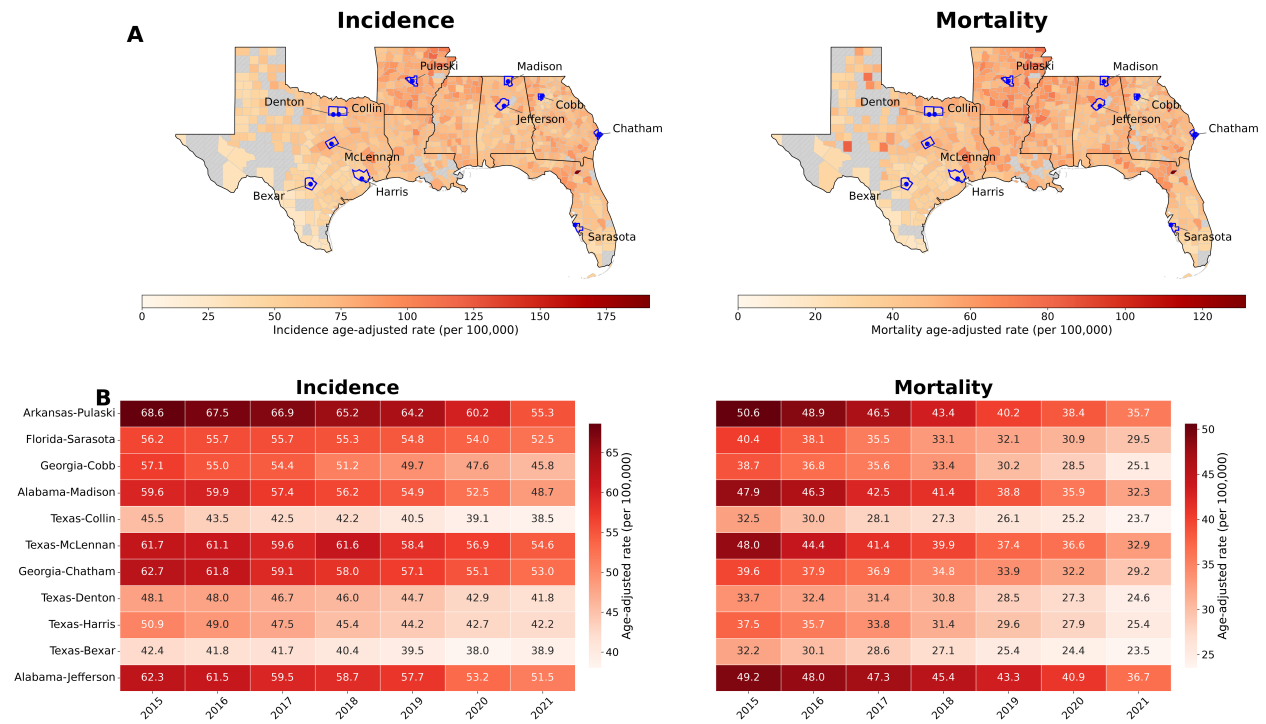


Figure 4.1: Spatial and temporal distribution of county-level lung cancer incidence and mortality.

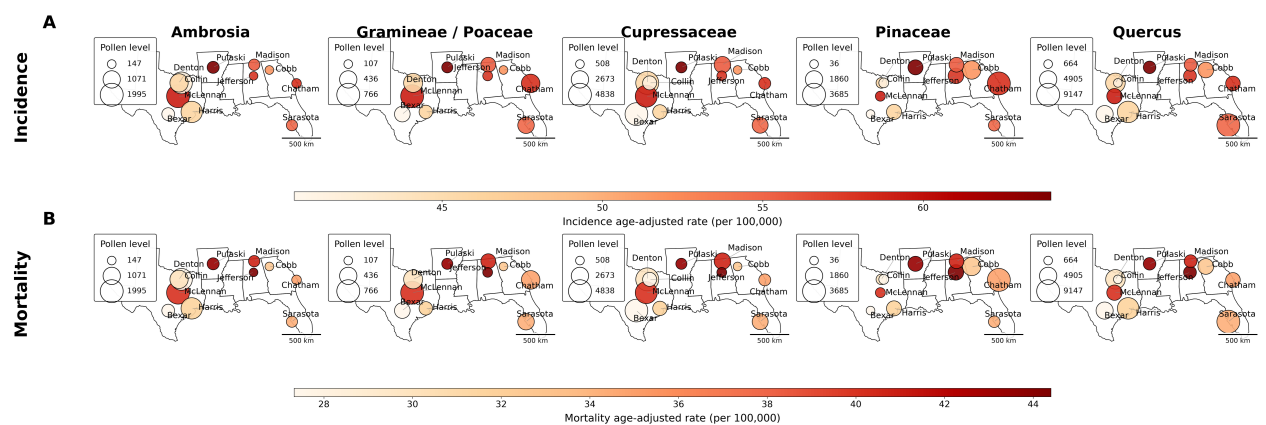


Figure 4.2: Pollen station levels overlaid on average lung cancer incidence and mortality.

outdoor air pollution with lung cancer risk [50, 104, 45, 101]. AQI summaries were retained because they may capture episodic and integrated air-quality conditions relevant to respiratory burden, although their correlation with underlying pollutants required careful interpretation (Table 4.1).

Humidity variables included a humidity index and absolute humidity. Humidity was evaluated because it can influence pollen release, pollen transport, pollen fragmentation, particle behavior, airway irritation, and pollutant chemistry. Absolute humidity (AH) was calculated as $AH = 216.7 \times e_{dew} / (T_{air} + 273.15)$, where $e_{dew} = 6.112 \times \exp[17.67 \times T_{dew} / (T_{dew} + 243.5)]$. T_{air} and T_{dew} were expressed in degrees Celsius, and AH was expressed in grams of water vapor per cubic meter of air (g/m³).

4.2.5 Daily pollen aggregation and illustrative weekly comparison

Daily predicted pollen values were used to construct the model predictors. For Ambrosia, Gramineae/Poaceae, Cupressaceae, Pinaceae, and Quercus, values at or above taxon-specific activity thresholds of 10, 5, 15, 15, and 15 in the source pollen units, respectively, were summed within each calendar year to create annual active-sum variables. These annual variables, rather than weekly values, were carried forward into the lag-window and predictive analyses.

Observed-versus-predicted weekly pollen plots for Sarasota County, Florida, and Pulaski County, Arkansas, were included only to illustrate seasonal agreement and taxon-specific peak timing (Figure 4.3); weekly values were not used as model predictors. Before lag-window construction, required annual Gramineae/Poaceae values for Collin County were completed using the year-matched arithmetic mean from the other 10 counties without using cancer outcomes. No additional predictor imputation was performed during model fitting.

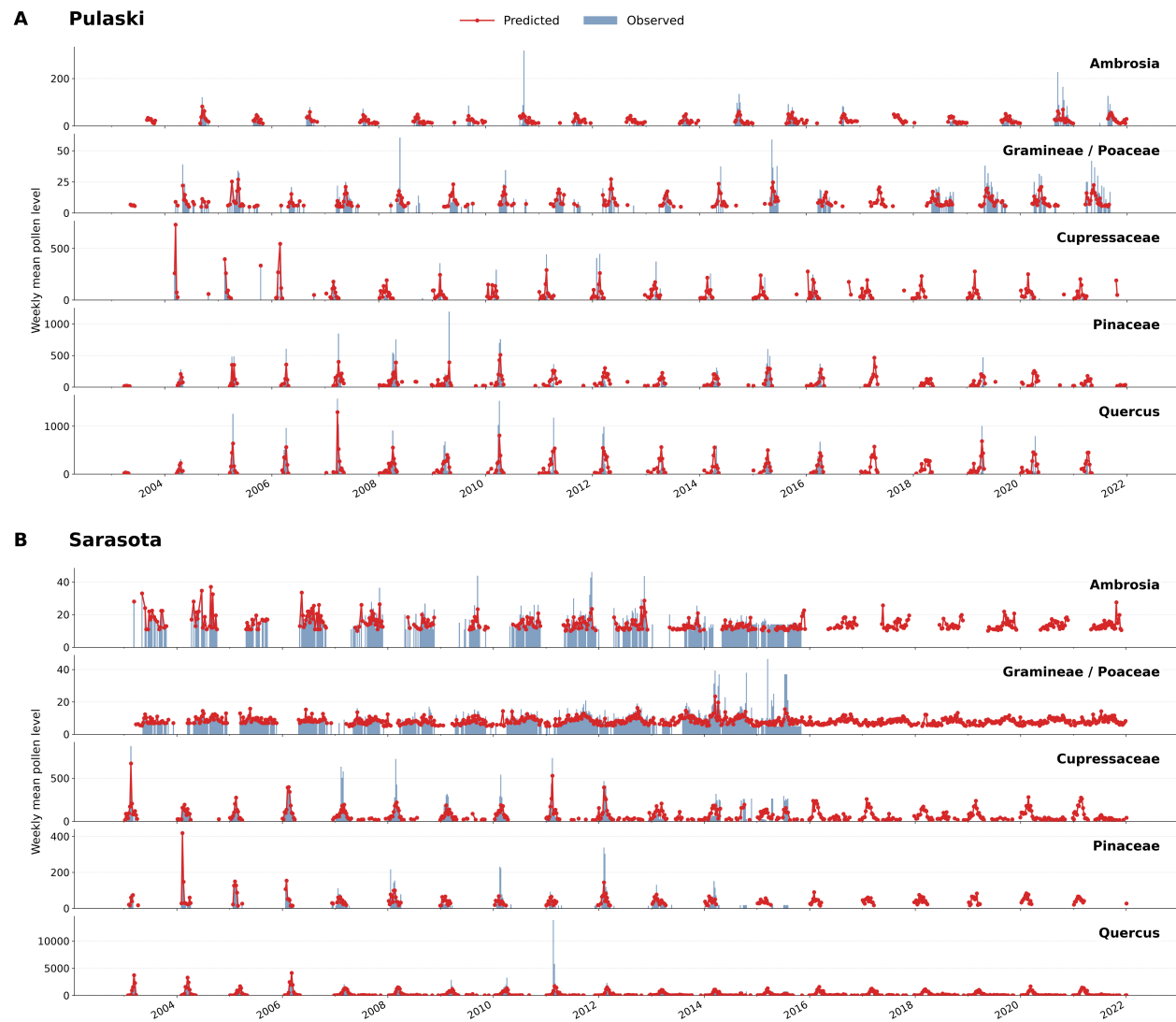


Figure 4.3: Weekly pollen reconstruction in Pulaski and Sarasota counties.

4.2.6 Windowed time-lagged exposure engineering

Because lung cancer can develop over long time scales, the analysis did not assume that same-year environmental conditions represented the relevant exposure period. For response year y , lag l , and window length w , each windowed time-lagged exposure was calculated as the arithmetic mean from year $y - l - w + 1$ through year $y - l$, inclusive. Pollen candidates used lags of 1-5 years and windows of 5-8 years. Pollutant and AQI candidates were restricted to complete historical windows: lag 1 with windows 5-8, lag 2 with windows 5-7, lag 3 with windows 5-6, and lag 4 with window 5 (Figure 4.4).

Within each county, feature, and outcome, the complete candidate with the highest signed Pearson correlation with the seven annual outcome observations was selected for the modeling dataset. This procedure was used for data-driven temporal alignment and was not interpreted as a definitive biological latency estimate. Because lag-window selection used the full outcome series before cross-validation, it may contribute to optimistic pooled performance and was considered in the limitations.

4.2.7 Descriptive correlation and collinearity assessment

Temporal and spatial correlations were examined to distinguish features that varied primarily across time from features that varied primarily across counties. This distinction was important because a feature may predict well by tracking a shared time trend, by marking county identity, or by capturing both spatial and temporal exposure patterns.

A predictor correlation matrix and an alternative circular correlation display were used to identify multicollinearity among pollen, pollutant, AQI, and humidity features. Correlation among predictors does not invalidate prediction, but it complicates interpretation because correlated features can substitute for one another in tree-based models.

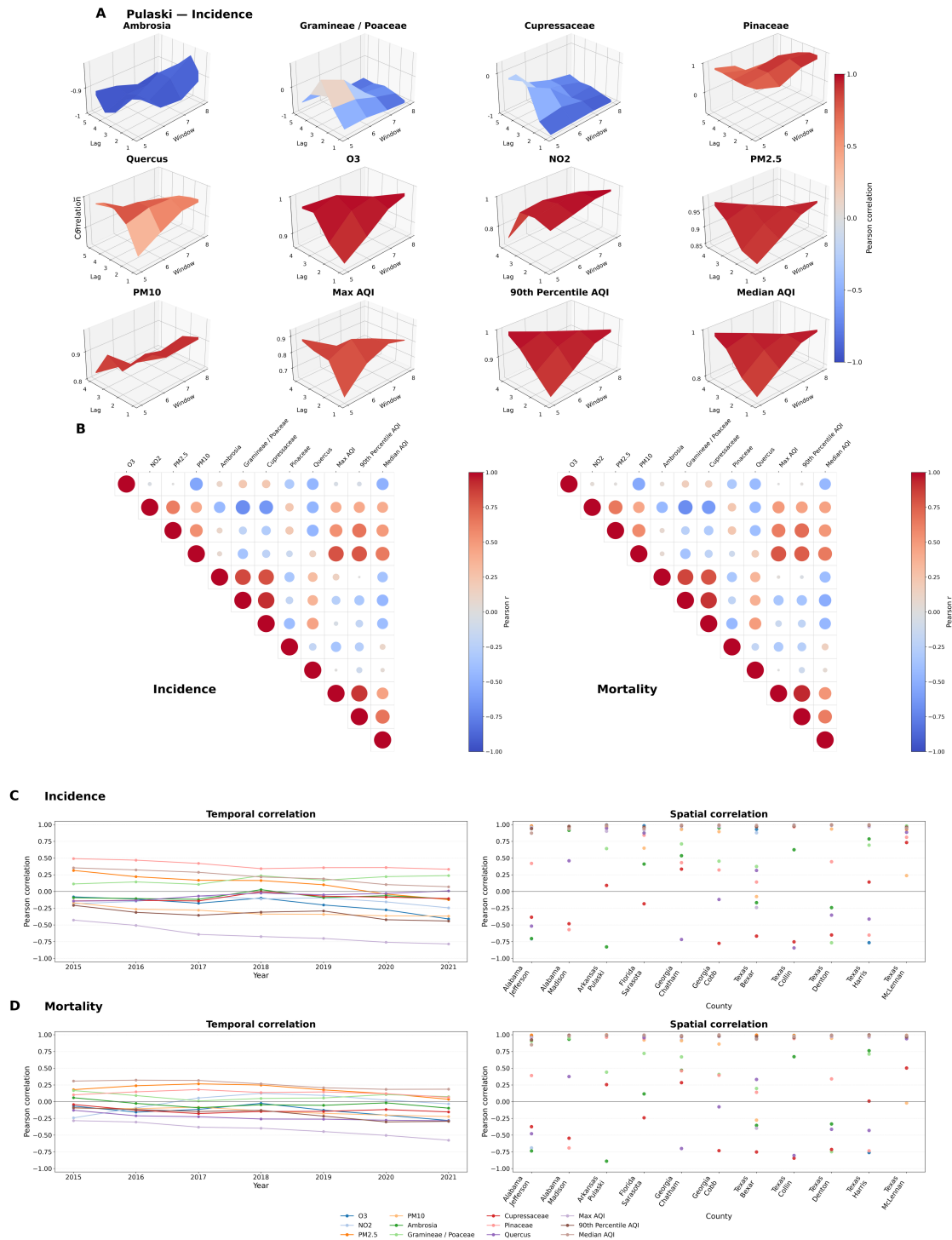


Figure 4.4: Lag-window correlations, predictor collinearity, and temporal/spatial correlation.

4.2.8 Predictive modeling

Random forest models were used as the primary predictive approach because they can accommodate nonlinear relationships, feature interactions, and correlated predictors without requiring a prespecified linear form [16]. Separate models were fitted for lung cancer incidence and mortality. Each primary random forest contained 500 trees and was evaluated using 100 repetitions of shuffled 10-fold cross-validation. The model-stage pipeline required complete, finite predictor values and did not perform imputation. XGBoost was used as a supplementary tree-based method for county-specific feature-ranking sensitivity [26] (Figure B.4).

Model performance was summarized using mean absolute error, mean squared error, and coefficient of determination. Pooled prediction performance was evaluated with predicted-versus-actual plots (Figure 4.5). Predictions for each county-year were averaged across the repeated held-out fold predictions before MAE, MSE, and R-squared were calculated. These pooled metrics quantify reproduction of patterns within the available county-year data and do not represent external geographic validation (Table 4.2).

4.2.9 Model interpretation

Model interpretation used several complementary approaches. Average random forest feature importance across repeated runs was used to summarize stable pooled feature contributions. Held-out-county permutation importance assessed whether reliance on selected pollen and air-quality predictors transferred to individual counties (Figure B.5). Six predictors were retained from the consensus analysis: Pinaceae, PM_{2.5}, the AQI 90th percentile, Ambrosia, Quercus, and ozone. For each outcome and held-out county, random forest models were trained on the remaining 10 counties, corresponding to 70 county-year observations, and evaluated on the seven annual observations from the held-out county. The analysis used 50 repeated model fits, and permutation importance was summarized as the median percent increase in held-out mean squared error with interquartile ranges.

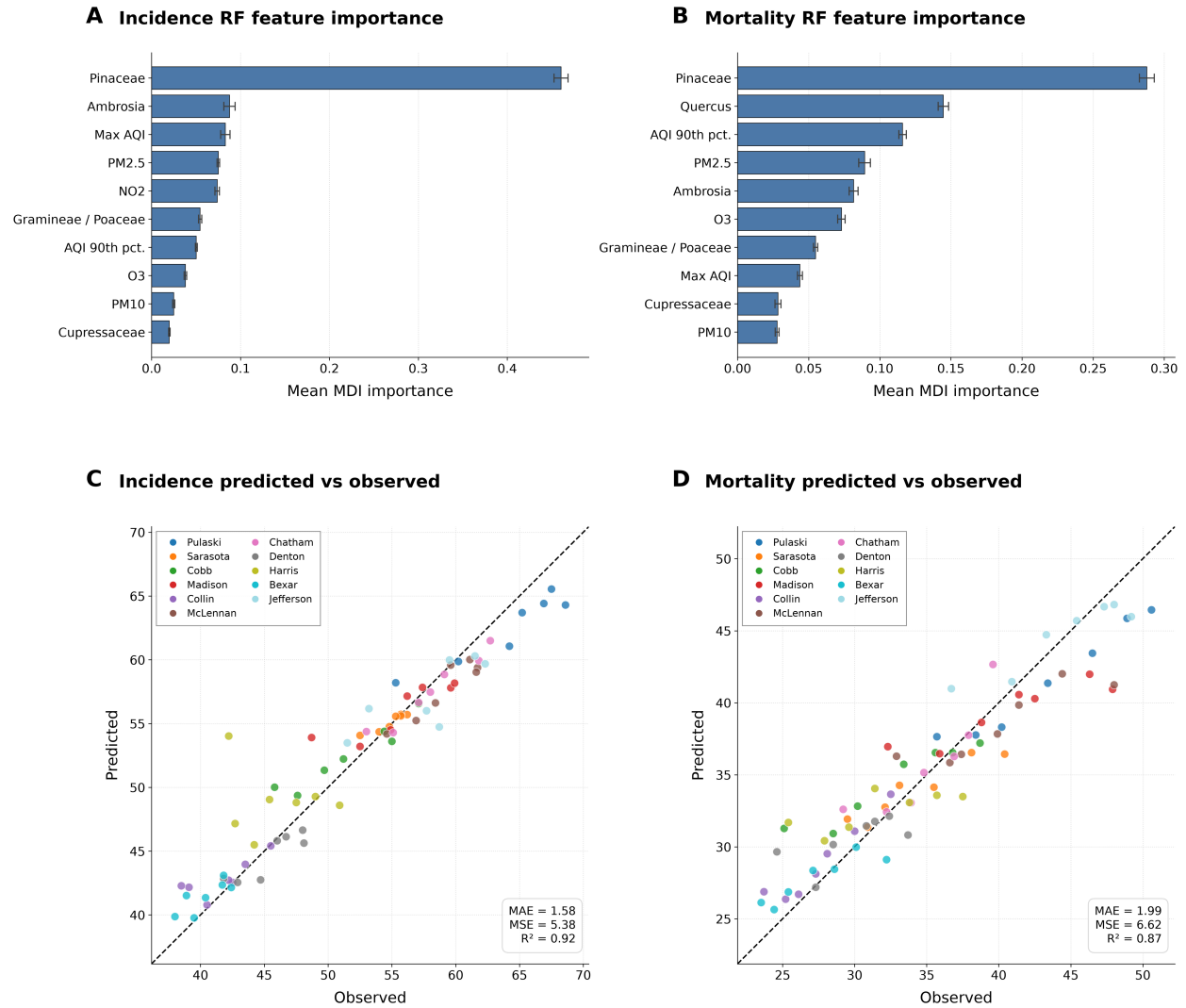


Figure 4.5: Pooled random forest feature importance and predictive performance.

Table 4.2: Summary of key model-performance results.

Model / evaluation view	Outcome	Reported MAE	Reported MSE	Reported R-squared	Interpretation
Averaged repeated-run pooled prediction	Incidence	1.58	5.38	0.92	Strong within-dataset fit to available county-year incidence patterns; not evidence of geographic generalizability.
Averaged repeated-run pooled prediction	Mortality	1.99	6.62	0.87	Strong within-dataset fit, but weaker than incidence and not evidence of geographic generalizability.
Reduced-feature county-grouped validation	Incidence	6.29	61.43	0.05	Near-zero geographic transfer and substantially poorer performance than pooled evaluation.
Reduced-feature county-grouped validation	Mortality	6.27	59.93	-0.19	Negative R-squared indicates poor geographic transfer; findings remain exploratory.

SHAP values were used to explain how individual features contributed to model predictions [67] (Figure 4.6). For Figures 4.7 and 4.8, SHAP and interaction summaries were obtained from random forests refitted to the complete 77-row model dataset after the corresponding repeated cross-validation analyses.

These methods explain model behavior, not causal effects. A high feature-importance score or SHAP value indicates that the model used the feature to make predictions, but it does not mean that intervening on that feature would necessarily change lung cancer incidence or mortality. In the held-out-county analysis, negative permutation-importance values indicated that permuting a predictor did not worsen, and sometimes improved, prediction; these values were interpreted as evidence of predictor redundancy or model instability rather than protective associations.

4.2.10 Humidity and interaction analysis

Humidity analyses were conducted to evaluate whether pollen importance was robust to meteorological context. Two humidity representations were considered: a humidity index and absolute humidity. Each humidity variable was summarized using candidate lags of 1-3 years and windows of 2-5 years; the complete candidate with the highest signed Pearson correlation was selected separately by county and outcome. In the absolute-humidity analysis, pollen $\times$ AH products and pollen/AH ratios were calculated after the pollen and AH lag-window variables had been constructed (Figures 4.7 and 4.8).

Friedman H-statistics were used to quantify interaction strength [40, 78]. In these analyses, H-statistics were generally small, so humidity was interpreted mainly as a context-shaping feature that redistributed importance rather than as evidence of strong pollen-humidity interaction.

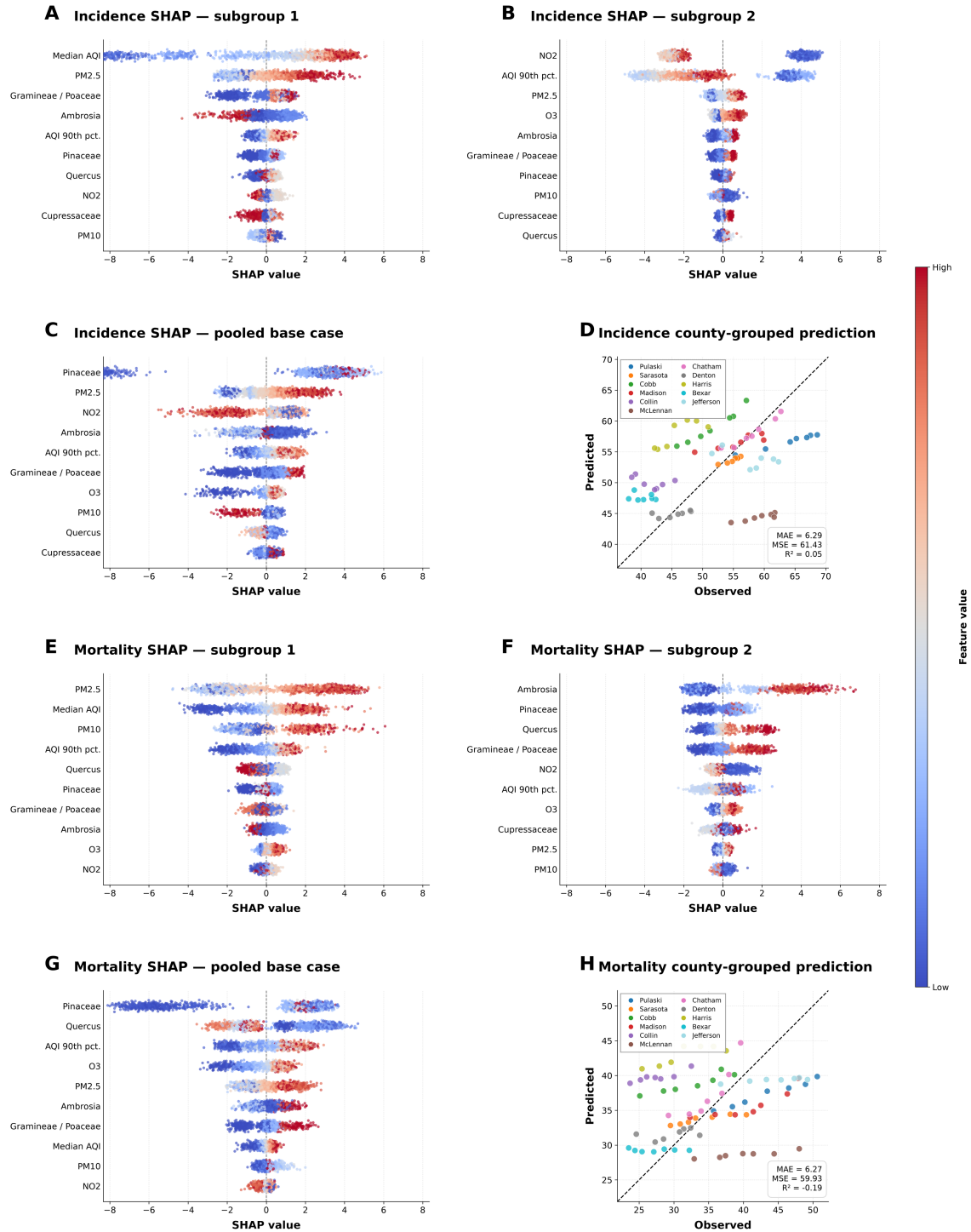


Figure 4.6: SHAP interpretation and county-grouped validation.

4.2.11 Validation strategy

Three validation perspectives were used. First, the pooled evaluation used shuffled row-level 10-fold cross-validation to assess how well random forest models reproduced patterns in the available county-year data. Second, county-grouped validation used complete county separation to assess transfer to held-out county groups. Third, leave-one-county-out evaluation assessed transfer to each individual county (Figure B.5).

The reduced-feature county-grouped experiment retained the five pollen variables, Max AQI, and PM2.5. It used 100 group-shuffled repetitions with three complete counties held out in each split. Figure 4.6 metrics were calculated after averaging predictions for each county-year across the splits in which that observation appeared in the test set. Performance was poor, with R-squared 0.05 for incidence and -0.19 for mortality (Table 4.2).

A complementary leave-one-county-out evaluation trained models on 10 counties and tested them on the remaining county, yielding 70 training observations and seven test observations per county. Fifty repeated fits were used to summarize county-specific MAE, RMSE, R-squared, and held-out permutation importance (Figure B.5).

4.2.12 Ethical considerations and reporting

The analysis used county-level aggregated data and did not include individual patient identifiers. Reporting was guided by relevant elements of observational-study and prediction-model reporting frameworks, including transparent description of data sources, predictors, missing-data handling, validation design, model tuning, and limitations [108, 27].

4.3 Results

4.3.1 Lung cancer burden varied by county and year

The descriptive maps and heatmaps showed clear spatial and temporal heterogeneity in lung cancer burden. The selected counties did not have uniform incidence or mortality rates: Pulaski County, Arkansas, consistently appeared as a high-burden county, whereas several Texas counties generally had lower rates, and incidence and mortality tended to decline between 2015 and 2021 in most counties.

The observed heterogeneity supported spatial-temporal modeling and indicated that county identity and calendar time could influence model performance independently of environmental exposures (Figure B.1).

4.3.2 Pollen levels showed geographically structured overlap with lung cancer outcomes

Pollen overlay maps demonstrated that pollen taxa were spatially structured and did not distribute uniformly across the study region. Ambrosia, grasses/Poaceae, Cupressaceae, Pinaceae, and Quercus had distinct geographic patterns relative to county-level incidence and mortality.

These descriptive patterns motivated evaluation of whether pollen signatures added predictive information about county-level lung cancer burden when considered with air quality, humidity, spatial structure, and lag windows (Table 4.3).

4.3.3 Illustrative weekly pollen comparisons showed seasonal structure

Observed-versus-predicted weekly pollen plots for Sarasota and Pulaski showed broad seasonal agreement and taxon-specific peak timing. The plots were included as descriptive illustrations of the underlying pollen series. Weekly values did not enter the

Table 4.3: Pollen-centered interpretation of model results.

Pollen-centered result	Interpretation for this chapter
Pollen overlay maps show geographic variation relative to lung cancer burden.	Use as descriptive motivation for pollen-focused modeling, not as causal evidence.
Weekly observed-versus-predicted pollen plots for Sarasota and Pulaski show seasonal structure.	Descriptive illustration only; the predictive models used threshold-based annual active-sum pollen variables rather than weekly predictors.
Lag-window surfaces show taxon- and county-specific correlations; complete candidate windows were selected separately by county, feature, and outcome using the highest signed Pearson correlation.	The environmental signal was temporally structured, but outcome-informed window selection may contribute to optimistic pooled performance.
Pinaceae was the leading pooled feature for both outcomes; Ambrosia was prominent for incidence and Quercus for mortality.	Pollen variables were model-informative alongside PM2.5 and AQI, but rankings varied by endpoint and county subgroup.
Pinaceae remained highly ranked after humidity adjustment, while humidity and engineered pollen-AH terms redistributed feature rankings.	Pollen importance was context-dependent, but the small interaction statistics did not support strong two-way interactions.
County-grouped and leave-one-county-out evaluations showed limited geographic transfer. All incidence leave-one-county-out median R-squared values were negative, mortality median R-squared was positive in only five of 11 counties, and held-out predictor importance varied substantially among counties.	Pooled pollen importance was not geographically stable. Findings remain exploratory and ecological, and larger external validation is necessary.

predictive models. The model analyses used threshold-based annual active-sum pollen variables and their selected lag-window summaries.

4.3.4 Lag-window and correlation analyses suggested temporally complex environmental signals

Windowed lag-correlation surfaces indicated that environmental features had different temporal signatures. PM-related and AQI-related features showed stronger positive regions in some lag-window combinations, while pollen features showed mixed, taxon-specific, and county-specific patterns.

Temporal and spatial correlation analyses further showed that some predictors may track shared year-to-year patterns, whereas others separate counties based on geography. Predictor correlation matrices demonstrated substantial collinearity among AQI summaries and pollutants, as well as correlation or inverse correlation among some pollen taxa and meteorological variables (Figure B.2).

The dense exposure-profile heatmap reinforced that county environmental profiles differed across pollutant and AQI domains after lag-window processing (Figure B.3). This pattern helps explain why feature rankings changed across sensitivity analyses: correlated predictors can compete for importance in tree-based models.

4.3.5 Random forest models identified both air-quality and pollen features as important predictors

Average random forest feature importance across repeated runs provided the most stable pooled summary of model-informative predictors. For incidence, Pinaceae had the largest average contribution by a wide margin, followed by Ambrosia, Max AQI, PM_{2.5}, and NO₂. Gramineae/Poaceae and the AQI 90th percentile also appeared among the leading predictors, indicating that pollen features contributed to the model alongside the main air-quality variables (Table 4.3).

The pooled incidence predictions averaged across repeated runs reported mean absolute error 1.58, mean squared error 5.38, and R-squared 0.92 (Table 4.2). For mortality, Pinaceae ranked first, followed by Quercus, the AQI 90th percentile, PM2.5, and Ambrosia. PM2.5 has the strongest external evidence linking particulate air pollution with lung cancer, although it ranked below the leading pollen variables in the mortality model. Mortality performance was weaker than incidence performance, with mean absolute error 1.99, mean squared error 6.62, and R-squared 0.87 (Table 4.2).

County-specific random forest and XGBoost feature rankings showed substantial geographic heterogeneity and should be interpreted as sensitivity analyses rather than stable county-specific causal rankings (Figure B.4).

4.3.6 SHAP analyses supported context-dependent pollen contributions

SHAP analyses showed that pollen variables contributed to model predictions, but their direction and magnitude depended on model specification, endpoint, and county subgroup. In incidence subgroup 1, Median AQI and PM2.5 were dominant, whereas NO2 and the AQI 90th percentile were dominant in subgroup 2. In pooled incidence, Pinaceae, PM2.5, and NO2 were the leading variables.

For mortality, PM2.5, Median AQI, PM10, and the AQI 90th percentile were dominant in subgroup 1, whereas Ambrosia, Pinaceae, Quercus, and grasses/Poaceae were dominant in subgroup 2. In pooled mortality, Pinaceae, Quercus, and the AQI 90th percentile were leading variables. These results indicate that predictor ranking and direction differed substantially between county subgroups and that pollen did not have a single uniform effect direction (Table 4.3).

County-grouped validation of the reduced feature set showed poor performance. When the model retained only pollen variables, Max AQI, and PM2.5, incidence performance was near zero (MAE 6.29, MSE 61.43, R-squared 0.05) and mortality performance was negative (MAE 6.27, MSE 59.93, R-squared -0.19) (Table 4.2).

Held-out-county permutation importance revealed pronounced geographic heterogeneity in model reliance on the six consensus predictors (Figure B.5). PM2.5 and the AQI 90th percentile dominated many county-specific profiles, while ozone was important in selected counties, particularly for mortality. Pollen contributions were generally smaller and less consistent, although *Quercus* showed a substantial mortality contribution in Chatham County. Pinaceae and Ambrosia frequently had near-zero or negative held-out importance, indicating that their pooled predictive contributions did not transfer uniformly across counties. Negative values were interpreted as predictor redundancy or model instability rather than protective effects.

Predictive performance was poor for incidence in every held-out county, with all county-specific median R-squared values below zero. Mortality transfer was more variable: median R-squared was positive for Bexar, Chatham, Denton, Madison, and Sarasota counties but negative for the remaining six counties. These results reinforce the distinction between strong pooled model fit and limited geographic transfer.

4.3.7 Humidity changed pollen rankings but provided limited evidence for strong interactions

Adding a humidity index changed the explanatory structure of both incidence and mortality models. In the incidence model, Pinaceae remained the leading SHAP feature, followed by PM2.5, NO2, the AQI 90th percentile, and grasses/Poaceae; the humidity index itself was not among the ten leading variables. In the mortality model, Pinaceae and *Quercus* were the leading SHAP features, followed by the AQI 90th percentile, ozone, and PM2.5; the humidity index ranked below these variables but remained model-informative.

Thus, adding humidity redistributed feature rankings without eliminating the prominence of Pinaceae, supporting meteorological context dependence rather than simple attenuation of the pollen signal (Table 4.3). Friedman H-statistics remained small. Maximum pairwise H-statistics were 0.032 for incidence and 0.050 for mortality, while maximum

one-vs-all H-statistics were 0.033 for incidence and 0.042 for mortality. Thus, humidity was model-informative, but the model did not appear to rely on strong two-way interaction structures.

4.3.8 Absolute humidity analyses indicated sensitivity to pollen-meteorology context

Absolute humidity analyses provided a second meteorological sensitivity framework. For each outcome, models were evaluated with no engineered interaction terms, pollen $\times$ AH product terms, and pollen $\times$ AH plus pollen/AH ratio terms. SHAP rankings shifted across these specifications for both incidence and mortality, indicating that humidity representation affected model explanation. Engineered pollen $\times$ AH and pollen/AH terms entered the leading SHAP rankings in several models, showing that model explanations were sensitive to how humidity context was represented.

SHAP interaction and dependence plots for Max AQI and Pinaceae demonstrated nonlinear and context-dependent model behavior without establishing causal interactions. Because product and ratio terms were calculated from already selected pollen and AH lag-window variables, these results should be interpreted as model-behavior diagnostics rather than independent exposure definitions.

4.3.9 Supplementary analyses reinforced heterogeneity and confounding concerns

Additional supplementary analyses reinforced heterogeneity and confounding concerns. Alternative outcome-trend plots, temporal correlation displays, and the dense exposure-profile heatmap showed that county structure, year-to-year variation, and environmental collinearity all influenced model behavior.

Long-range projections of incidence and mortality through 2100 were presented as exploratory supplementary analyses because they require assumptions about smoking,

screening, treatment, population aging, migration, climate, and air-quality regulation that were not modeled explicitly (Figure B.6). Sex- and race-stratified outcome plots demonstrated demographic heterogeneity and highlighted potential confounding and equity considerations (Figure B.7). t-SNE and UMAP analyses showed clustering of counties in high-dimensional feature space, supporting the interpretation that county structure influences model behavior and grouped-validation difficulty (Figure B.8).

4.4 Discussion

4.4.1 Principal findings

This ecological machine-learning analysis found that pollen variables were repeatedly informative for predicting county-level lung cancer incidence and mortality. PM2.5 and AQI-related variables remained important and have the strongest external epidemiologic support, but pollen taxa - especially Pinaceae, Ambrosia, Quercus, and grasses/Poaceae - added recurring environmental signals across descriptive maps, lag-window analyses, random forest importance, SHAP summaries, humidity-adjusted sensitivity analyses, and held-out-county permutation-importance diagnostics.

Pooled random forest models showed stronger performance for incidence than for mortality. This pattern is consistent with the idea that mortality depends on additional clinical, demographic, and health-care factors not captured by the environmental feature set. However, county-grouped and leave-one-county-out evaluations showed limited and heterogeneous geographic transfer. The overall interpretation is therefore exploratory, ecological, and hypothesis-generating.

4.4.2 Interpretation of pollen-associated signal

The pollen signal likely reflects several overlapping processes. First, pollen is a direct aeroallergen exposure that can affect the respiratory tract. Second, pollen marks vegetation and land-cover patterns, which differ substantially across counties. Third, pollen is highly seasonal and may therefore capture timing of climate, humidity, and pollutant patterns. Fourth, pollen can co-occur with air pollution, and pollutant exposure may modify respiratory responses to allergens [59, 32]. Fifth, pollen exposure can contribute to airway inflammation and oxidative stress, biological processes that are relevant to respiratory disease and, more broadly, to cancer biology [81, 15, 29, 33, 46, 70, 28].

These pathways are plausible, but they are not separable in the present ecological dataset. A county with high *Quercus* or *Ambrosia* values may also differ in humidity, PM2.5, ozone, urbanization, smoking prevalence, occupational exposures, health-care access, or population structure. Consequently, pollen should be described as a model-informative environmental signature rather than a proven independent risk factor (Table 4.3).

4.4.3 Interpretation by pollen taxa

Ambrosia appeared repeatedly in incidence feature rankings and humidity-adjusted incidence SHAP summaries. Because *Ambrosia* is a highly seasonal pollen group and is often clinically relevant for allergic respiratory symptoms, its repeated importance may reflect seasonal aeroallergen burden, late-season vegetation patterns, or interactions with air-quality conditions.

Quercus appeared in both incidence and mortality explanations, including mortality SHAP summaries and 3D humidity visualizations. As a tree-pollen marker, *Quercus* may capture spring pollen season, regional vegetation structure, climate, and county geography.

Pinaceae was the leading pooled feature for both outcomes and remained highly ranked after humidity adjustment. Its persistence across specifications supports a recurring environmental signal, although its interpretation remains uncertain because Pinaceae can reflect vegetation, seasonality, humidity, and county context rather than a single independent exposure.

Gramineae/Poaceae also appeared in model explanations, especially in mortality-related rankings. Grass pollen may mark warm-season aeroallergen exposure and vegetation patterns. Its variable importance across models and counties supports a broader pollen signal rather than a singular dominant predictor.

4.4.4 Pollen, pollution, and humidity as a combined exposome

Pollen, air pollution, and humidity were jointly informative. PM_{2.5} and AQI summaries remained important, and their external lung cancer evidence is substantially stronger than that for pollen. The model therefore should not be framed as replacing air pollution with pollen. Instead, pollen appears to add environmental structure to a model already anchored by air-quality predictors (Table 4.3).

Humidity was informative mainly as a context-shaping variable. Humidity and engineered pollen-humidity terms redistributed feature rankings without eliminating the prominence of Pinaceae, while the small interaction diagnostics limited support for strong two-way pollen-humidity interactions. Future models should therefore evaluate mixture and interaction frameworks cautiously rather than treating pollen, pollution, and humidity as isolated exposures.

4.4.5 Why incidence prediction was stronger than mortality prediction

The incidence model performed better than the mortality model (Table 4.2). This is epidemiologically reasonable because incidence reflects disease occurrence, while mortality reflects a longer pathway involving disease occurrence, diagnosis timing, stage distribution, treatment access, treatment quality, comorbidity, survival, and competing causes of death.

The lower mortality performance likely reflects unmeasured clinical, demographic, and health-system determinants. Sex- and race-stratified supplementary figures show outcome heterogeneity that is not captured by pollen and air-quality features alone.

4.4.6 County-grouped and leave-one-county-out validation

Pooled random forest models achieved strong within-dataset performance, especially for incidence, whereas county-grouped validation showed limited geographic transfer. Incidence performance was near zero, and mortality performance was negative under the reduced feature-set grouped-validation design (Table 4.2).

The decline in grouped-validation performance indicates that the models likely learned county baselines, repeated county-year structure, or county-specific environmental regimes that did not transfer reliably to held-out county groups. This result is scientifically important because it prevents overinterpretation of the pooled model and emphasizes that the current analysis is best viewed as ecological and hypothesis-generating.

The leave-one-county-out analysis provided a more geographically granular assessment of transferability (Figure B.5). Incidence models produced negative median R-squared values for every held-out county, whereas mortality models achieved positive median R-squared in only five of 11 counties. The predictors that increased held-out error also differed substantially among counties. These findings indicate that both predictive performance and predictor importance were strongly dependent on county context. From

a risk-aware modeling perspective, the poor and heterogeneous transfer result is not a failure to hide; it is a warning label for decision use. Any future monitoring or intervention design based on these models should account for the risk that pollen-associated signals may not generalize beyond the counties used for model development.

4.4.7 Strengths

- The study integrates pollen, air pollution, AQI, humidity, space, and time rather than evaluating one environmental domain in isolation (Table 4.1).
- Daily-derived active-sum pollen variables provide a taxon-specific exposure basis before lag-window construction.
- Lag-window feature engineering acknowledges that annual cancer outcomes may relate to historical exposure windows rather than same-year exposure alone.
- Multiple interpretation methods were used, including feature importance, held-out-county permutation importance, SHAP, correlation matrices, interaction statistics, and 3D visualizations (Figure B.5).
- County-grouped and leave-one-county-out evaluations directly assessed geographic transferability and identified substantial county-dependent limitations in performance and predictor importance (Figure 4.6; Figure B.5; Table 4.2).

4.4.8 Limitations

- Ecological design: County-level associations cannot be translated into individual-level exposure or risk.
- Small sample size: The analysis included only 11 counties and seven annual outcome years, increasing the risk of overfitting and unstable feature rankings.

- Missing individual risk factors: Smoking history, radon exposure, occupational carcinogens, secondhand smoke, family history, and prior lung disease were not incorporated at the individual level.
- Incomplete demographic and health-system covariates: Age distribution, sex, race and ethnicity, socioeconomic status, screening, stage, treatment access, insurance, and survival determinants were not fully integrated into the predictive models.
- Pollen measurement and preprocessing uncertainty: Pollen exposure depends on monitoring coverage, prediction quality, spatial assignment, taxon-specific activity thresholds, annual aggregation, and lag-window selection.
- Lag-window selection: Candidate windows were selected using correlations with the complete seven-year outcome series before cross-validation, which may inflate pooled performance and feature-ranking stability.
- Completed pollen values: Required Collin County Gramineae/Poaceae annual values were reconstructed using year-matched means from the other 10 counties. This outcome-blind completion avoided model-stage imputation but introduced additional exposure uncertainty.
- Collinearity: Pollen, AQI, pollutants, humidity, seasonality, and geography are correlated, so single-feature importance should not be interpreted causally.
- Model-interpretation limits: SHAP, permutation importance, feature importance, t-SNE, UMAP, and H-statistics explain model behavior but do not prove biological mechanisms.
- Generalizability: Both county-grouped and leave-one-county-out evaluations showed limited geographic transfer. All incidence leave-one-county-out median R-squared values were negative, and mortality performance was positive in only

five of 11 held-out counties. The models should therefore not be used for external prediction, risk assessment, or policy decisions without substantially larger geographic validation studies (Figure 4.6; Figure B.5; Table 4.2).

4.4.9 Future directions

Future research should expand the geographic and temporal scope of the dataset. More counties and longer time series would allow stronger validation, better separation of county effects from exposure effects, and more reliable estimation of nonlinear relationships. External validation in counties not used during model development is essential.

Future models should integrate major lung cancer determinants, especially smoking prevalence, radon exposure, occupational exposures, age structure, socioeconomic conditions, lung cancer screening uptake, stage at diagnosis, treatment access, and health-care quality. These additions are particularly important for mortality models.

Pollen exposure assignment should be refined using denser monitoring networks, species-specific pollen calendars, land-cover data, vegetation indices, climate variables, remote sensing, and spatial interpolation. Mechanistic studies should examine whether pollen-pollution mixtures influence chronic airway inflammation, oxidative stress, epithelial injury, or immune responses in ways relevant to lung carcinogenesis.

Long-range projection analyses should remain a future-work concept until the model has stronger external validation and richer covariates.

4.5 Summary

This ecological machine-learning analysis supports pollen as a meaningful environmental signature in county-level lung cancer prediction. Ambrosia, Quercus, Pinaceae,

and grasses/Poaceae repeatedly contributed to model explanations, especially when evaluated alongside PM_{2.5}, AQI summaries, humidity, county structure, and temporal lag windows (Table 4.3).

Overall, pollen-associated environmental regimes were associated with spatial and temporal variation in county-level lung cancer burden. These regimes likely integrate aeroallergen exposure, vegetation, seasonality, humidity, geography, and air pollution. The ecological design and weak county-grouped and leave-one-county-out transfer preclude causal interpretation and external predictive use, but the findings support larger, better adjusted, and externally validated studies of pollen, inflammation, air pollution, humidity, and lung cancer.

Together with Chapter 3, this chapter shows that interpretable prediction can identify environmental signals while also revealing the risks of overinterpretation. Chapter 5 turns this lesson into a formal decision-learning problem by asking how a learning system can optimize policies while controlling outcome volatility.

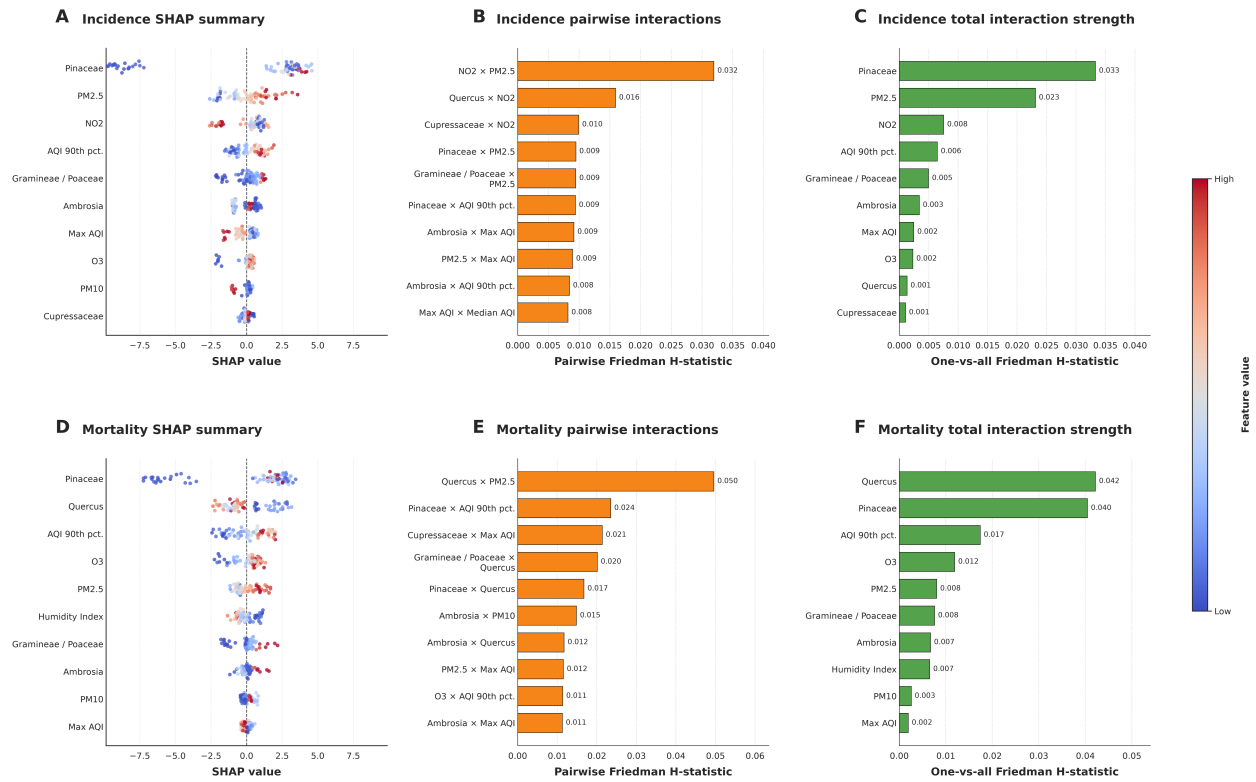


Figure 4.7: Humidity-index SHAP and interaction analysis.

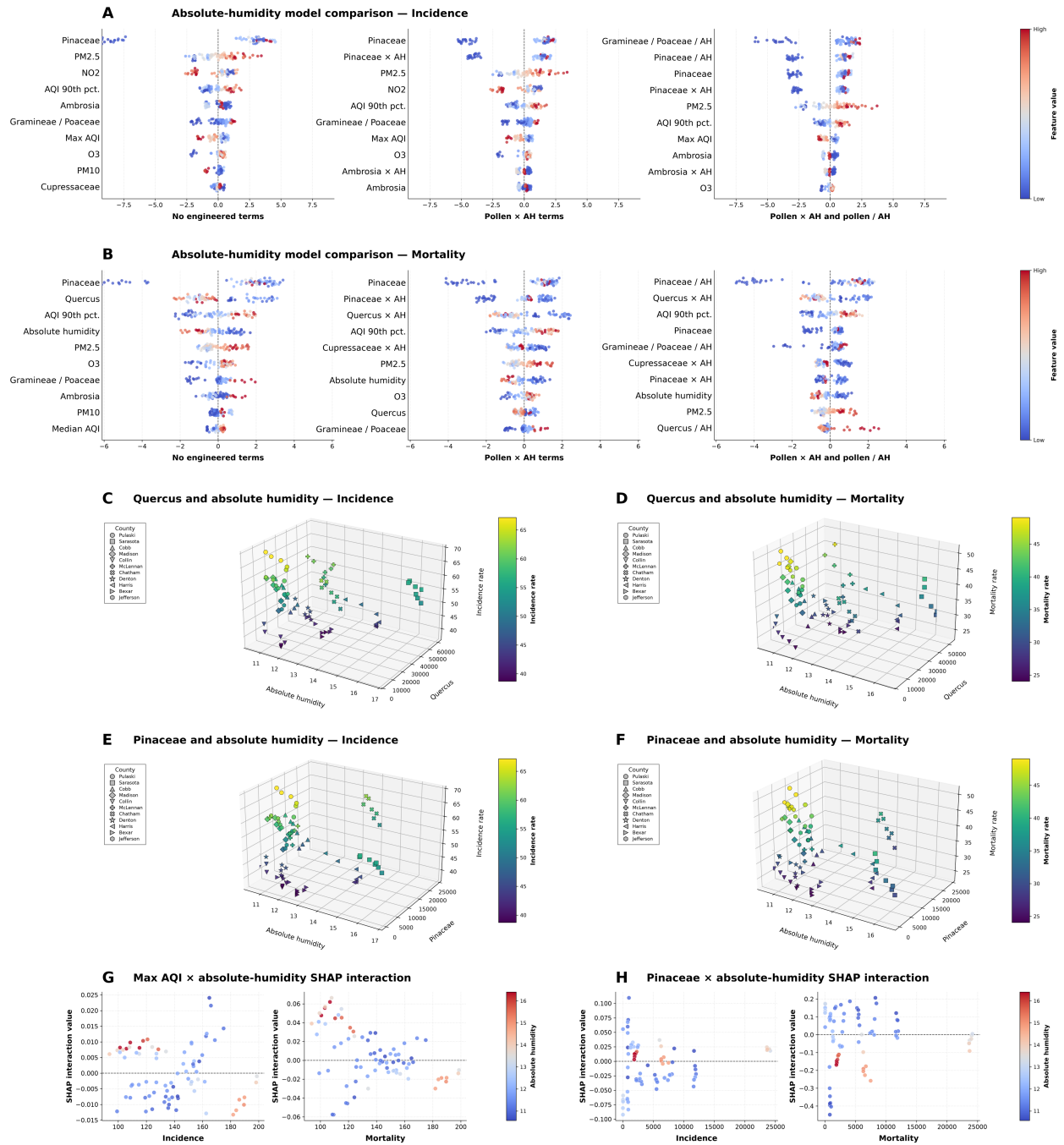


Figure 4.8: Absolute-humidity sensitivity, 3D pollen-humidity patterns, and SHAP interaction diagnostics.

Chapter 5

TOPS: Transition-based Volatility-controlled Policy Search and its Global Convergence

5.1 Background and problem formulation

Many learning problems require more than high average performance. A forest-management policy, a medical decision rule, or a robotic controller can look attractive on average but still be unacceptable if it sometimes produces very poor outcomes. In such settings, the decision maker cares about both expected reward and the stability of that reward. A method that avoids high volatility is especially valuable when failures are costly, when data are limited, or when long experimental trials are unsafe.

This idea connects the TOPS framework to the applied chapters of this dissertation. The forest-suitability chapter produces maps that help identify low-risk and high-risk adaptation contexts, such as seedling-supported suitability, adult-only suitability, future loss, and extrapolation concern. The pollen-lung cancer chapter shows that a model can perform well in pooled data but fail to transfer across county groups, creating risk if the model is used without validation. Both examples motivate learning systems that make uncertainty and variability explicit rather than optimizing average prediction or reward alone.

Reinforcement learning (RL) studies sequential decision-making, where an agent chooses actions, observes rewards, and updates its policy over time [96]. Standard RL is often risk-neutral: the goal is to maximize expected cumulative reward, even if some trajectories produce high volatility or hazardous outcomes [77, 107, 112]. Risk-neutral learning can be insufficient in applications such as financial decision-making, healthcare, robotics, and environmental management, where rare but severe outcomes may be unacceptable [58, 83, 68].

A second practical challenge is the dependence on long trajectories. Many risk-aware RL algorithms estimate risk from long uninterrupted rollouts, but long rollouts can be expensive or unsafe in safety-critical domains. Long trajectories also increase the chance of encountering hazardous states during learning [42, 48]. Learning from short trajectories or transition segments can reduce this burden and make risk-aware learning more practical in settings where complete long-run experiments are difficult to collect.

A third challenge is theoretical reliability. Policy-gradient methods are widely used, but global convergence guarantees are difficult, especially with nonlinear function approximation and risk-aware objectives. Prior work has made progress for risk-neutral policy gradients in tabular, linear, and over-parameterized neural-network settings [1, 74, 109, 62, 124]. Extending such guarantees to risk-aware objectives is harder because the objective depends not only on expected reward but also on variability.

This chapter asks the following question: *Can a risk-aware policy-gradient algorithm control reward volatility, learn from short trajectory segments, and still have a global convergence guarantee?* To answer this question, I propose **T**ransition-based **v**olatility-controlled **P**olicy **S**earch (TOPS), a risk-averse reinforcement-learning framework that uses reward volatility as its risk measure [14]. TOPS has two main contributions. First, it does not require learning from long uninterrupted trajectories; it can use short-term trajectories, segments of longer rollouts, or combinations of both. Second, it provides global convergence guarantees for neural natural policy gradient and proximal policy optimization under the assumptions stated later in the chapter.

The proof structure follows three steps. First, the mean-volatility objective is replaced by a lower-bound surrogate that can be optimized through an augmented Markov decision process. Second, actor-critic updates are analyzed for the augmented problem under over-parameterized neural-network approximation. Third, the actor and critic errors are combined to bound the optimality gap for TOPS with neural PPO and neural NPG. This

overview is provided before the technical details so that the later assumptions and theorems can be read as parts of one argument rather than as isolated proof components.

5.2 Technical background

This section introduces the technical building blocks used in the TOPS framework: reinforcement learning, policy gradients, and over-parameterized neural networks. Detailed notation is provided in Appendix C.1.

Reinforcement Learning We consider the infinite-horizon discounted Markov Decision Process (MDP) $(\mathcal{S}, \mathcal{A}, \mathcal{P}, r(s, a), \gamma)$ with state space $\mathcal{S}$, action space $\mathcal{A}$, the transition kernel $\mathcal{P} : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$, the reward function $r(s, a) : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$, the initial state $S_0 \in \mathcal{S}$ and its distribution $\mu_0 : \mathcal{S} \rightarrow [0, 1]$, and the discounted factor $\gamma \in (0, 1)$. At time step t , given a state s_t , an action a_t is taken according to policy $\pi(a_t|s_t) : \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$, generating a reward $r_t := r(s_t, a_t)$ and the next state s_{t+1} based on $p(s_{t+1}|s_t, a_t)$ the reward function is assumed to be deterministic and bounded — a constant $r_{\max} > 0$ exists such that $r_{\max} = \sup_{(s,a) \in \mathcal{S} \times \mathcal{A}} |r(s, a)|$. A change in states upon an action $(s, a, r(s, a), s')$ is termed a *transition*, where the state s' is the successive state of the state s . With a little bit abuse of notation, $r(s, a)$ is denoted as $r_{s,a}$, and $r(s_t, a_t)$ is denoted as r_t in the rest of this chapter. A *trajectory* of length T is a consecutive sequence of transitions $\{(s_t, a_t, r_t, s'_t)\}_{t=0}^{T-1}$ over a set of contiguous timestamps, where $\forall t, s'_t = s_{t+1}$. Therefore, the trajectory is also equivalently denoted by $\{(s_t, a_t, r_t, s_{t+1})\}_{t=0}^{T-1}$. To evaluate the performance of policy π , we introduce state value function $V_\pi(s) := (1 - \gamma)\mathbb{E}_{a \sim \pi(a|s)} [\sum_{t=0}^{\infty} \gamma^t r_t | S_0 = s, a_t \sim \pi(a|s_t), s_{t+1} \sim \mathcal{P}(s|s_t, a_t)]$ and state-action value function $Q_\pi(s, a) := (1 - \gamma)\mathbb{E}_{a \sim \pi(a|s)} [\sum_{t=0}^{\infty} \gamma^t r_t | s_0 = s, a_0 = a, a_t \sim \pi(a|s_t), s_{t+1} \sim \mathcal{P}(s|s_t, a_t)]$. Bounded reward implies $|V_\pi(s)| \leq r_{\max}$ and $|Q_\pi(s, a)| \leq r_{\max} \forall \pi$. Additionally, the advantage function $A_\pi : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ of policy π is defined as $A_\pi(s, a) := Q_\pi(s, a) - V_\pi(s)$.

The normalized state and state action occupancy measure of policy π is denoted by $\nu_\pi(s)$ and $\sigma_\pi(s, a) := \pi(a|s)\nu_\pi(s)$, respectively. Therefore, $\nu_\pi(s) := (1 - \gamma) \sum_{t=0}^{\infty} \gamma^t Pr(s_t = s | \mu_0, \pi, \mathcal{P})$ and $\sigma_\pi(s, a) := (1 - \gamma) \sum_{t=0}^{\infty} \gamma^t Pr(s_t = s, a_t = a | \mu_0, \pi, \mathcal{P})$, where $Pr(s_t = s | \mu_0, \pi, \mathcal{P})$ is the probability of $s_t = s$ given $\mu_0, \pi, \mathcal{P}$.

Finally, the return is defined as $G := \sum_{t=0}^{\infty} \gamma^t r_t$.

Policy Gradient Methods. In the following, we discuss two policy gradient methods, where the policy π_θ is parameterized by the parameter θ . For natural policy gradient (NPG, [52]), we first define the Fisher information matrix,

$$F(\theta) := \mathbb{E}_{(s,a) \sim \sigma_{\pi_\theta}} [\nabla_\theta \log(\pi_\theta) (\nabla_\theta \log(\pi_\theta))^\top] \quad (5.1)$$

The update of parameter θ then takes the form,

$$\theta_{k+1} = \theta_k + \eta_{\text{NPG}} (F(\theta_k))^{-1} \nabla J_\theta(\pi_{\theta_k}) \quad (5.2)$$

where $(F(\theta_{k-1}))^{-1}$ is the inverse of the Fisher information matrix $F(\theta)$ in Eq. (5.1), ∇J_θ is the objective gradient and η_{NPG} the learning rate.

In proximal policy optimization (PPO, [92]), at the k -th iteration the update of policy parameter θ takes the following form, where β is the penalty hyper-parameter:

$$\arg \max_{\theta} \mathbb{E}_{(s,a) \sim \sigma_k} [\pi_\theta A_{\theta_k / \pi_{\theta_k}} - \beta \text{KL}(\pi_\theta \| \pi_{\theta_k})]. \quad (5.3)$$

Policy Network with over-parameterized Neural Networks Policy π with the two-layer over-parameterized neural network is defined as: for $\forall (s, a) \in S \times A$,

$$f((s, a); \theta, b) := \frac{1}{\sqrt{m}} \sum_{v=1}^m b_v \text{ReLU}((s, a)^\top [\theta]_v). \quad (5.4)$$

Here (s, a) is the input and m is the width of the network. $\theta = ([\theta]_1^\top, \dots, [\theta]_m^\top)^\top \in \mathbb{R}^{m \times d}$ is the input weight matrix in the first layer of the neural network. $b = (b_1, \dots, b_m)^\top \in \mathbb{R}^{m \times 1}$ are the output weights in the second layer. We present a block diagram of a over-parameterized neural network with Figure C.1 in the Appendix C.2. At the start of training, the parameters θ, b are initialized by

$$\theta = \Theta_{\text{init}} \in \mathbb{R}^{m \times d} ([\Theta_{\text{init}}]_v \sim \mathcal{N}(0, I_d/d), \quad (5.5)$$

and $b_v \sim \text{Unif}(\{-1, 1\})$, $\forall v \in [m]$, respectively, where $\mathcal{N}$ denotes Gaussian distribution and Unif denotes uniform distribution. $f((s, a); \theta, b)$ can be simplified to $f((s, a); \theta)$ by updating only $[W]_v$ during training, and fixing b as its initialization [4]. We also restrict the possible value of θ within the space denoted by $\mathcal{D} = \{\xi \in \mathbb{R}^{md} : \|\xi - \Theta_{\text{init}}\|_2 \leq \Upsilon, \Upsilon > 0\}$. Therefore the policy is defined $\pi_\theta(a|s)$ in the following form, where τ is the temperature parameter.

$$\pi_\theta(a|s) := \frac{\exp(\tau f((s, a); \theta))}{\sum_{a' \in A} \exp(\tau f((s, a'); \theta))}. \quad (5.6)$$

Furthermore, the feature mapping of a two-layer neural network $f((s, a); \theta)$ is defined as,

$\phi_\theta := ([\phi_\theta]_1^\top, \dots, [\phi_\theta]_m^\top)^\top$, where $[\phi_\theta]_v^\top = \frac{b_v}{\sqrt{m}} \text{ReLU}((s, a)^\top [\theta]_v)$, $\forall v \in [m]$. By Eq. (5.4), it holds that $f((s, a); \theta) = \phi(s, a)^\top \theta$ and $\nabla_\theta f((s, a), \theta) = \phi(s, a)$ [109]. Furthermore, we assume that there exists a constant $M > 0$ such that,

$$\mathbb{E}_{(s,a) \sim \text{init}} \left[\sup_{(s,a) \in S \times A} |\phi((s, a)^\top \Theta_{\text{init}})|^2 \right] \leq M^2. \quad (5.7)$$

Mean-Variance & Mean-Volatility RL In a variance-constraint problem with the variance of the total reward, the objective can be formulated as,

$$\max_{\pi} J(\pi), \quad \text{subject to } \mathbb{V}(G) \leq Y \quad (5.8)$$

where $J(\pi) := \mathbb{E}_{(s,a) \sim \sigma_{\pi}}[G] = \frac{1}{1-\gamma} \mathbb{E}_{(s,a) \sim \sigma_{\pi}}[r_{s,a}]$ is the expected return, $\mathbb{V}(\cdot)$ is the variance of a random variable and $Y > 0$ is the upper bound for this variance. The constrained formulation in Eq. (5.8) is NP-hard [94], and in reality, the relaxed formulation $J_{\lambda}^G(\pi)$ defined in Eq. (5.9) is often solved instead [34, 56, 116] as follows, where λ is called variance-controlling parameter.

$$J_{\lambda}^G(\pi) := \mathbb{E}[G] - \lambda \mathbb{V}(G) = \mathbb{E}[G] - \lambda \mathbb{E}[G^2] + \lambda (\mathbb{E}[G])^2 \quad (5.9)$$

Meanwhile, [14] proposed a reward-volatility risk measure. *Volatility* is defined as the variance of per-step reward — per-step reward R is a discrete random variable with a probability mass function of $p(R = x) = \sum_{s,a} \sigma_{\pi}(s, a) \mathbb{1}\{r_{s,a} = x\}$, where $\mathbb{1}\{\cdot\}$ is the indicator function. It is easy to see that $\mathbb{E}[R] = (1-\gamma)J(\pi)$ [122]. $\mathbb{V}(R)$ is the variance of R . Likewise, $J_{\lambda}(\pi)$ is proposed as a counterpart of Eq. (5.9) in the sequel, which is defined with respect to R .

$$J_{\lambda}(\pi) := \mathbb{E}[R] - \lambda \mathbb{V}(R) = \mathbb{E}[R] - \lambda \mathbb{E}[R^2] + \lambda (\mathbb{E}[R])^2 \quad (5.10)$$

We first present the following lemma based on Lemma 1 in [14] to show that $J_{\lambda}(\pi)$ is a reasonable counterpart to $J_{\lambda}^G(\pi)$.

Lemma 1 *Given $\lambda \geq 0$, $\frac{1}{(1-\gamma)} J_{\frac{\lambda}{(1-\gamma)}}(\pi)$ is a lower-bound of $J_{\lambda}^G(\pi)$, i.e.,*

$$\frac{1}{(1-\gamma)} J_{\frac{\lambda}{(1-\gamma)}}(\pi) \leq J_{\lambda}^G(\pi).$$

A detailed proof is provided in Appendix C.4.5. Given Lemma 1, maximizing $J_{\lambda}^G(\pi)$ can be reduced to maximizing its lower bound $\frac{1}{(1-\gamma)} J_{\frac{\lambda}{(1-\gamma)}}(\pi)$. There are several advantages of

optimizing $J_\lambda(\pi)$. Compared to optimizing $\mathbb{V}(G)$, optimizing $\mathbb{V}(R)$ is computationally easier [122]. [14] argue that $\mathbb{V}(R)$ is better at capturing short-term risk and leads to smoother trajectories that avoid possible “shocks” caused by long-horizon trajectories [14].

5.3 Algorithm Formulation

In this section, we present our risk-averse policy-search algorithm. In particular, we use **(i)** reward volatility to construct the mean-volatility objective, which circumvents the long-horizon reward issue and avoids large variance, and **(ii)** over-parameterized neural network [20] as the neural network architecture of the actor and the critic to facilitate global convergence analysis.

5.3.1 Augmented MDP

As [14] shows, reward volatility has advantages over mean-variance methods, including a smoother trajectory and much-reduced variance. Therefore, in this chapter, we choose volatility as the risk measurement. Compared with the conventional mean-variance objective, the mean-volatility objective function enables the agent to learn from transitions instead of trajectories and greatly reduces the chance of getting into hazardous states due to consecutive long-horizon explorations. This can greatly help improve safety. Note that because of *compositional* expectations $(\mathbb{E}[R])^2$, double-sampling is needed, which is a heavy burden for sampling. For more details on double-sampling and the more general compositional expectations, please refer to [110, 63].

To avoid double-sampling, we resort to augmented MDP. First, note that it holds $(\mathbb{E}[R])^2 = \max_{y \in \mathbb{R}} (2\mathbb{E}[R]y - y^2)$. Then the optimization objective transforms into:

$$\max_{\pi, y} J_\lambda^y(\pi) := \mathbb{E}_{(s,a) \sim \sigma_\pi} (r_{s,a} - \lambda r_{s,a}^2 + 2\lambda r_{s,a}y) - \lambda y^2 \quad (5.11)$$

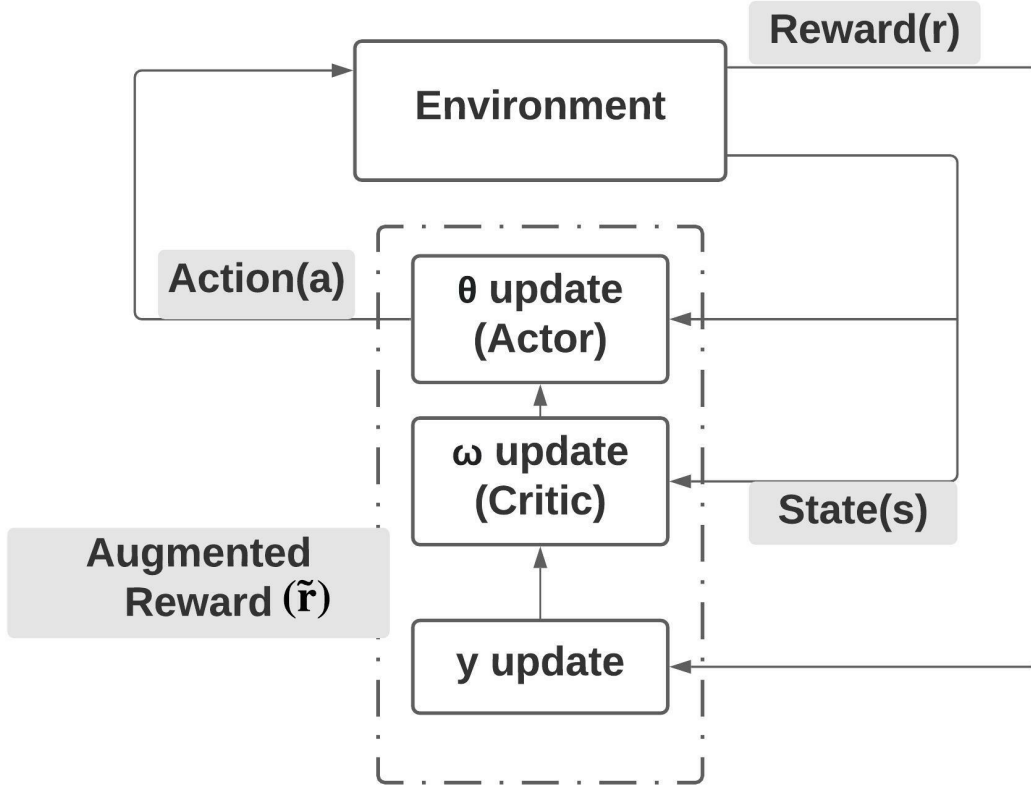


Figure 5.1: A simple block diagram of TOPS

We now introduce the augmented MDP, with the augmented reward defined as follows:

$$\tilde{r}_{s,a} := r_{s,a} - \lambda r_{s,a}^2 + 2\lambda r_{s,a} y \quad (5.12)$$

We refer to this new MDP as the *augmented MDP* $\tilde{M} = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, \tilde{r}(s, a), \gamma\}$, and denote corresponding terms by the $\tilde{\cdot}$ sign — for example, the associated state value function and state-action value function are $\tilde{V}_\pi(s)$ and $\tilde{Q}_\pi(s, a)$. $\tilde{r}(s, a)$ is denoted by $\tilde{r}_{s,a}$ for notation simplicity in the remainder of this chapter. We solve Eq. (5.11) by maximizing y and π of the augmented MDP in two steps iteratively.

5.3.2 Proposed Algorithms

We present TOPS in Algorithm 1. A block diagram of TOPS is illustrated in Figure 5.1, where there are three sets of parameters to update, i.e., θ for the actor, ω for the critic, and the auxiliary variable y . Note that the mean-volatility framework allows incorporating any off-the-shelf policy optimization methods as pointed out by [121]. Since most global optimality analysis literature is based on NPG and PPO [19, 109, 62, 124], we also use NPG and PPO as inner policy search algorithms for a fair comparison.

y update: Since Eq. (5.12) is quadratic in y , to update y in each iteration, we have $y_k = (1 - \gamma)J(\pi_k)$. However, we do not have direct access to the exact value of $J(\pi_k)$. As an alternative, we estimate this value with a sample average in the k -th iteration,

$$\hat{y}_k := \frac{1}{T} \sum_{t=1}^T r_t, \quad (5.13)$$

as an estimator of y_k , where T is the sample (batch) size.

θ update: We update π_θ with an actor-critic scheme, particularly NPG and PPO. NPG and PPO are the two most widely used policy gradient methods. According to empirical studies [113], PPO usually achieves state-of-the-art performance among on-policy policy gradient methods, and NPG has the advantage of easy hyperparameter tuning compared with PPO. On the other hand, NPG and PPO’s global convergences under the risk-neutral setting have been studied intensively and can show good results [1, 109]. Therefore, PPO and PNG are used as the inner actor algorithm of the TOPS framework to make fair comparisons with existing approaches. Additionally, over-parameterized neural networks are widely used in proving global convergence of gradient-based methods under the risk-neutral setting, which show impressive results [62, 109, 41]. The capability of a gradient-based neural network method to reach the global optimum in an over-parameterization

setting is explained in theory [20]. Therefore, we parameterize the policy in this chapter with a two-layer over-parameterized neural network. We first introduce the actor part of the two methods, respectively.

θ Update for Neural NPG Per the update rule for NPG in Eq. (5.2), we need to estimate the natural policy gradient $(F(\theta_k))^{-1} \nabla_{\theta} J(\pi_{\theta})$. However $F(\theta_k)$ is difficult to invert due to its high-dimensionality. Instead the gradient is estimated by solving $\min_{\xi \in \mathcal{D}} \|\hat{F}(\theta_k)\xi - \tau_k \hat{\nabla}_{\theta} J(\pi_{\theta_k})\|_2$, where

$$\hat{\nabla}_{\theta} J(\pi_{\theta_k}) := \frac{\tau_k}{T} \sum_{t=1}^T Q_{\omega_k}(s_t, a_t) (\phi_{\theta_k}(s_t, a_t) - \mathbb{E}_{a \sim \pi_{\theta_k}} [\phi_{\theta_k}(s_t, a'_t)]) \quad (5.14)$$

$$- \mathbb{E}_{a \sim \pi_{\theta_k}} [\phi_{\theta_k}(s_t, a'_t)]), \quad (5.15)$$

$$\begin{aligned} \hat{F}(\theta_k) := & \frac{\tau_k^2}{T} \sum_{t=1}^T \left((\phi_{\theta_k}(s_t, a_t) - \mathbb{E}_{a \sim \pi_{\theta_k}} [\phi_{\theta_k}(s_t, a'_t)]) \right. \\ & \left. (\phi_{\theta_k}(s_t, a_t) - \mathbb{E}_{a \sim \pi_{\theta_k}} [\phi_{\theta_k}(s_t, a'_t)])^{\top} \right), \end{aligned} \quad (5.16)$$

are unbiased estimations of $\nabla_{\theta} J(\pi_{\theta})$ and $F(\theta_k)$ respectively, with the help of feature mapping, θ is, therefore, updated as,

$$\tau_{k+1} = \tau_k + \eta_{\text{NPG}}, \quad (5.17)$$

$$\theta_{k+1} = (\tau_k \theta_k + \eta_{\text{NPG}} \hat{\nabla}_{\theta} J(\pi_{\theta_k})) / \tau_{k+1} \quad (5.18)$$

$$\eta_{\text{NPG}} \arg \min_{\xi \in \mathcal{D}} \|\hat{F}(\theta_k)\xi - \tau_k \hat{\nabla}_{\theta} J(\pi_{\theta_k})\|_2) / \tau_{k+1} \quad (5.19)$$

θ Update for Neural PPO Given the update rule of PPO in Eq. (5.3), the PPO's objective function $L(\theta)$ can be rewritten as $L(\theta) := \mathbb{E}_{s \sim \nu_{\pi_k}} [\mathbb{E}_{a \sim \pi_{\theta}} [Q_{\pi_k}] - \beta \text{KL}(\pi_{\theta} \parallel \pi_{\theta_k})]$. With energy-based policy $\pi \propto \exp\{\tau^{-1} f\}$, the solution to the subproblem $\hat{\pi}_{k+1} = \arg \max_{\pi} L(\theta)$ can be

obtained by solving the following [62]:

$$\theta_{k+1} = \arg \min_{\theta \in \mathcal{D}} \mathbb{E}_{(s,a) \sim \sigma} \left[\left(f_{\theta} - \tau_{k+1} (\beta^{-1} Q_{\pi_k} + \tau_k^{-1} f_{\theta_k}) \right)^2 \right] \quad (5.20)$$

The stochastic gradient method can be used to solve Eq.(5.20).

ω update To estimate the state-action function value of the augmented MDP $\tilde{Q}_{\pi}$, a critic network parameterized by ω is constructed, denoted as $\tilde{Q}_{\omega}$. Note that the critic uses the same two-layer neural network architecture as the actor defined in Eq. (5.4), which indicates that the policy network π_{θ} 's parameter θ and critic network $\tilde{Q}_{\pi}$'s parameter ω are of identical dimensions, i.e., $\theta \in \mathbb{R}^d, \omega \in \mathbb{R}^d$. The critic network is parameterized with a different set of parameters $\omega = ([\omega]_1^{\top}, \dots, [\omega]_m^{\top})^{\top} \in \mathbb{R}^{md}$, denoted by $f((s, a); \omega)$. For simplicity, we then learn $\tilde{Q}_{\omega}$ by applying the semi-gradient TD method. Other approaches, such as the Gradient TD (GTD) algorithm family [97], can also be applied. For each iteration t of the TD update,

$$\omega_{t+1} = \omega_t - \eta_{\text{TD}} \left(\tilde{Q}_{\omega_t}(s, a) \right. \quad (5.21)$$

$$\left. - (1 - \gamma) \tilde{r}_{s,a} - \gamma \tilde{Q}_{\omega_t}(s', a') \right) \nabla_{\omega} Q_{\omega_t}(s, a), \quad (5.22)$$

where η_{TD} is the learning rate for TD update.

Algorithm 1: TOPS: Transition-based Volatility-controlled Policy Search

Input: number of iterations K ; learning rates for natural policy gradient (resp. PPO) and neural TD, η_{NPG} (resp. η_{PPO}); and temperature parameters $\{\tau_k\}_{k=1}^K$.

Initialization: Initialize the policy network $f((s, a); \theta, b)$ as defined in Eq. (5.5). Set $\tau_1 = 1$. Initialize the Q-network with (b, ω_1) similarly.

For $k = 1, \dots, K$:

Sample a batch of transitions $\{s_t, a_t, r_t, s'_t\}_{t=1}^T$ following the current policy.

Set $y = \frac{1}{T} \sum_{t=1}^T r_t$.

For $t = 1, \dots, T$:

Set $\tilde{r}_t = r_t - \lambda r_t^2 + 2\lambda r_t y$ and sample $a'_t \sim \pi(a \mid s'_t)$.

Q-value update: update ω_k according to Eq. (5.21).

If the NPG update is selected, update θ_k according to Eq. (5.18).

Else if the PPO update is selected, update θ_k according to Eq. (5.20).

Output: π_{θ_K} .

5.4 Theoretical Analysis

Although NPG and PPO's global convergences under the risk-neutral setting show prominent result [1, 109], the techniques used by these methods only apply to the primal constrained-MDP case and remain challenging to apply to the analysis of the primal-dual case as in our augmented MDP, where the dual variable y is critical. For example, Lemma (5.2) of [62], a critical step in the error-bound analysis of risk-neutral PPO, cannot be applied to our primal-dual risk-averse case. In this section, we establish the global convergence rate of TOPS with both NPG and PPO. The argument is organized as follows. The assumptions first control variance, distribution shift, and approximation error for the actor-critic updates. The major theorems then state high-probability optimality-gap bounds for neural PPO and neural NPG. The detailed derivations are placed in Appendix C.4; the main text emphasizes what each result implies for short-trajectory, volatility-controlled policy learning.

5.4.1 Assumptions

We first impose regularity condition assumptions, which are common in the literature on TD analysis with a neural network approximation [19, 109, 62, 124].

Assumption 1 (*Variance upper bound*) [109] Let $\mathcal{D} = \{\alpha \in \mathbb{R}^{md} : \|\alpha - \Theta_{\text{init}}\|_2 \leq \Upsilon\}$. For all $k \in [K]$, We assume that for all $k \in [K]$, there exists an absolute constant $\sigma_\xi > 0$ such that,

$$\mathbb{E}[\|\xi_k(\delta_k)\|_2^2] \leq \tau_k^4 \sigma_\xi^2 / T, \quad \mathbb{E}[\|\xi_k(\omega_k)\|_2^2] \leq \tau_k^4 \sigma_\xi^2 / T. \quad (5.23)$$

where $\delta_k = \text{argmin}_{\delta \in \mathcal{D}} \|\hat{F}(\theta_k)\delta - \tau_k \hat{\nabla}_\theta J(\pi_{\theta_k})\|_2$ and $\xi_k(\delta) = \hat{F}(\theta_k)\delta - \tau_k \hat{\nabla}_\theta \tilde{J}(\pi_{\theta_k}) - \mathbb{E}[\hat{F}(\theta_k)\delta - \tau_k \hat{\nabla}_\theta \tilde{J}(\pi_{\theta_k})]$. The expectation is taken over σ given θ_k and ω_k .

Note that δ_k and ω_k have the same dimension due to the compatible neural network setting. $\xi_k(\delta)$ can be generalized to both δ_k and ω_k . We then introduce a regularity condition assumption on visitation measures and stationary distributions in the sequel, respectively.

Assumption 2 (*Upper bounded concentrability coefficient*) [109] ν^* and σ^* are denoted as the state and state-action visitation measures corresponding to the global optimum π^* . For all $k \in [K]$, we define the following terms:

$$\begin{aligned} \varphi_k &= \left\{ \mathbb{E}_{(s,a) \sim \sigma_{\pi_k}} \left[\left(\frac{d\sigma^*}{d(\sigma_{\pi_k})} \right)^2 \right] \right\}^{1/2}, \\ \psi_k &= \left\{ \mathbb{E}_{s \sim \nu_{\pi_k}} \left[\left(\frac{d\nu^*}{d(\nu_{\pi_k})} \right)^2 \right] \right\}^{1/2}, \\ \varphi'_k &= \left\{ \mathbb{E}_{(s,a) \sim \sigma'_{\pi_k}} \left[\left(\frac{d\sigma^*}{d(\sigma'_{\pi_k})} \right)^2 \right] \right\}^{1/2}, \\ \psi'_k &= \left\{ \mathbb{E}_{s \sim \nu'_{\pi_k}} \left[\left(\frac{d\nu^*}{d(\nu'_{\pi_k})} \right)^2 \right] \right\}^{1/2}, \end{aligned} \quad (5.24)$$

We assume that $\varphi_k, \psi_k, \varphi'_k, \psi'_k$ are uniformly upper bounded by an absolute constant $c_0 > 0$.

σ'_{π_k} and ν'_{π_k} are state-action and state stationary distribution. $\varphi_k, \psi_k, \varphi'_k, \psi'_k$ are the concentrability coefficients, which reflects how much the starting state and state-action distribution diverge from the state and state-action distribution under the optimal policy [79]. Assumption 2 impose a upper bound on such divergence. This regularity condition is commonly used in the literature. [80, 7, 38, 119, 109]. More over, we define $\varphi_k^* = \mathbb{E}_{(s,a) \sim \sigma_\pi} \left[\left(\frac{d\pi^*}{d\pi_0} - \frac{d\pi_{\theta_k}}{d\pi_0} \right)^2 \right]^{1/2}$, $\psi_k^* = \mathbb{E}_{(s,a) \sim \sigma_\pi} \left[\left(\frac{d\sigma_{\pi^*}}{d\sigma_\pi} - \frac{d\nu_{\pi^*}}{d\nu_\pi} \right)^2 \right]^{1/2}$.

5.4.2 Major Theoretical results

In the following, we present the major theoretical results, i.e., the global optimality and convergence rate of TOPS with neural PPO. We define the optimality gap $\min_{k \in [K]} (J_\lambda^{y^*}(\pi^*) - J_\lambda^{\hat{y}_k}(\pi_k))$, where π^*, y^* are respectively defined as $\pi^* := \arg \max_\pi J_\lambda(\pi)$, $\hat{y}^* := (1 - \gamma)J(\pi^*)$, and y^k is defined in Eq. (5.13). $J_\lambda^{y^*}(\pi^*)$ (resp. $J_\lambda^{\hat{y}_k}(\pi_k)$) represents the risk-averse objective under π^* (resp. π_k , i.e., the policy at the k -th iteration).

Theorem 5.1 (*Global Optimality and Rate of Convergence on neural PPO*) We set the learning rate of PPO $\eta_{\text{PPO}} = \min\{(1 - \gamma)/3(1 + \gamma)^2, 1/\sqrt{K_{\text{TD}}}\}$, the learning rate of TD update $\eta_{\text{TD}} = \min\{(1 - \gamma)/3(1 + \gamma)^2, 1/\sqrt{K_{\text{TD}}}\}$ where K_{TD} is the total iteration of TD

update, and $\beta_0 := \beta/\sqrt{K}$. Under Assumptions 3-4, we have, with a probability of $1 - \delta$,

$$\begin{aligned}
& \min_{k \in [K]} (J_\lambda^{y^*}(\pi^*) - J_\lambda^{\hat{y}^k}(\pi_k)) \\
& \leq \frac{\beta_0^2 \log |\mathcal{A}| + U + \beta_0^2 \sum_{k=1}^K (\varepsilon_k + \varepsilon'_k)}{(1 - \gamma)\beta_0\sqrt{K}} \\
& \quad + \frac{4\lambda c_3 r_{\max}(1 - \gamma)}{\sqrt{K}} \\
& \text{where } \varepsilon_k'' = \mathcal{O}(\Upsilon^3 m^{-1/2} \log(1/\delta) + \Upsilon^{5/2} m^{-1/4} \cdot \\
& \quad \sqrt{\log(1/\delta)} + \Upsilon \cdot r_{\max}^2 m^{-1/4} + \Upsilon^2 K_{\text{TD}}^{-1/2} + \Upsilon), \\
& \varepsilon_k = \tau_{k+1}^{-1} \varepsilon_k'' \varphi_{k+1}^* + \beta^{-1} \varepsilon_k'' \psi_k^*, \\
& \varepsilon'_k = |\mathcal{A}| \tau_{k+1}^{-2} \epsilon_{k+1}^2, \\
& U = 2\mathbb{E}_{s \sim \nu_{\pi^*}} [\max_{a \in \mathcal{A}} (\tilde{Q}_{\omega_0})^2] + 2\Upsilon^2
\end{aligned} \tag{5.25}$$

Similarly, we present TOPS global optimality and convergence rate with neural NPG.

Theorem 5.2 (*Global Optimality and Rate of Convergence for neural NPG*) We set the learning rate of NPG $\eta_{\text{NPG}} = 1/\sqrt{K}$, the learning rate of TD update $\eta_{\text{TD}} = \min\{(1 - \gamma)/3(1 + \gamma)^2, 1/\sqrt{K_{\text{TD}}}\}$ where K_{TD} is the total iteration of TD update, and the temperature

parameters $\tau_k = (k - 1)\eta_{\text{NPG}}$. Under Assumptions 3-4, with a probability of $1 - \delta$, we have

$$\begin{aligned}
& \min_{k \in [K]} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \\
& \leq \frac{1}{(1 - \gamma)\sqrt{K}} \left(\log |\mathcal{A}| + 9\Upsilon^2 + M^2 + 4c_3M(1 - \gamma)^2\lambda \right) \\
& \quad + \frac{1}{K} \sum_{k=1}^K (\epsilon_k) \\
& \text{where } \epsilon_k = \sqrt{8c_0}\Upsilon^{1/2}\sigma_{\xi}^{1/2}T^{-1/4} \\
& \quad + \mathcal{O}((\tau_{k+1}K^{1/2} + 1)\Upsilon^{3/2}m^{-1/4} + \Upsilon^{5/4}m^{-1/8}) \\
& \quad + c_0\mathcal{O}(\Upsilon^3m^{-1/2}\log(1/\delta) + \Upsilon^{5/2}m^{-1/4}\sqrt{\log(1/\delta)}) \\
& \quad + \Upsilon r_{\max}^2m^{-1/4} + \Upsilon^2K_{\text{TD}}^{-1/2} + \Upsilon
\end{aligned} \tag{5.26}$$

Remark Theorem 5.1 and 5.2 bound the optimality gap

$$\min_{k \in [K]} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k))$$

at the order stated in the theorem assumptions, with the dominant convergence rate summarized as $\mathcal{O}(1/\sqrt{K})$. This gap measures how close the best policy found during the first K updates is to the globally optimal risk-aware policy.

From Theorem 5.1 and 5.2, we can conclude that our risk-averse algorithm TOPS with PPO and NPG version of the actor both converge to the global optimal policy at a $\mathcal{O}(1/\sqrt{K})$ rate. We provide a detailed proof in Appendix C.4.

5.5 Experiments

In this section, we aim to empirically examine the performance of our algorithm on Mujoco robot manipulation benchmark tasks from OpenAI gym [17], as [122] does. The Mujoco benchmark is a set of challenging robot control tasks in simulated environments

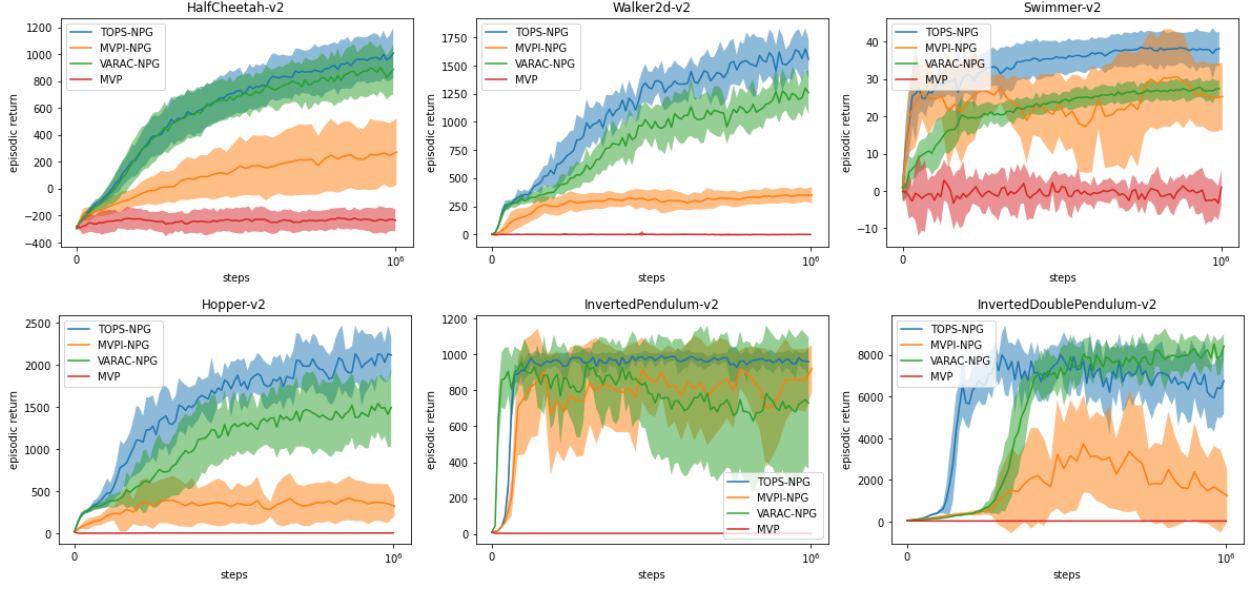


Figure 5.2: Training progress of TOPS-NPG and baseline algorithms.

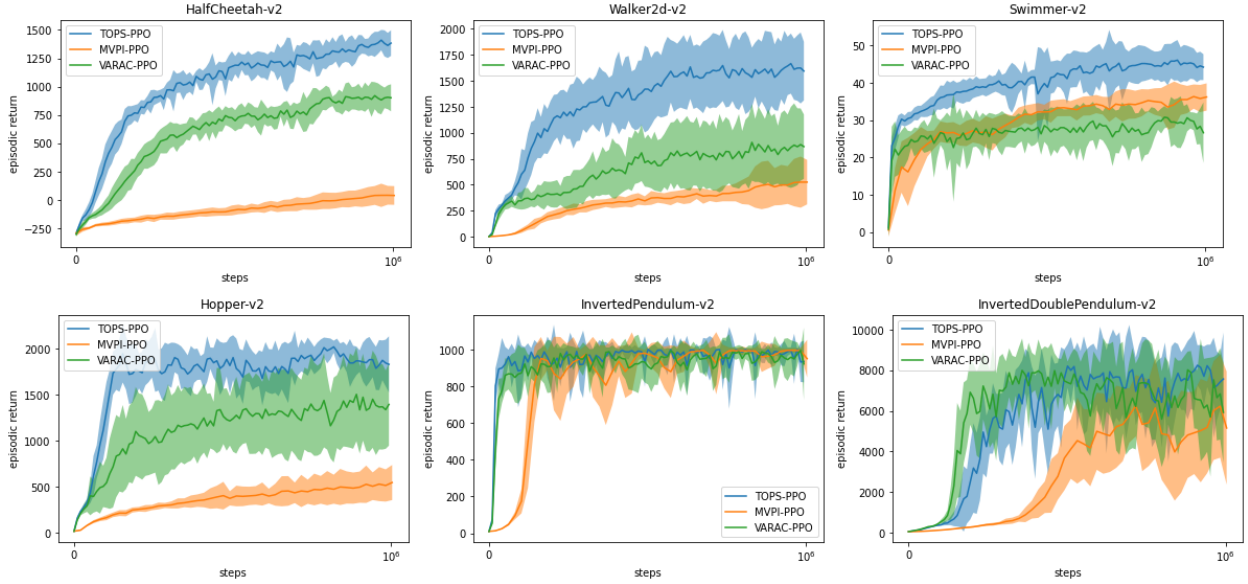


Figure 5.3: Training progress of TOPS-PPO and baseline algorithms.

designed for controller optimization in reinforcement learning [100]. In this domain, the simulated robots are expected to achieve consistent performances while avoiding failures that lead to dangerous results.

Experiment setup. We conduct our experiments in an online learning setting and include several recent risk-averse RL methods as baselines: the mean-variance policy

optimization (MVP) [116], mean-variance policy iteration (MVPI) [122], and variance-constrained actor-critic (VARAC) [124]. We set $\lambda = 1$ and run each algorithm for 10^6 steps and evaluate the algorithm every 10^4 steps for 20 episodes. All curves are averaged over 10 independent runs and use shaded areas to indicate standard errors. The experiment’s details are provided in Appendix C.3. All experiments’ parameters are tuned through rigorous grid search.

We report the learning curves of TOPS with NPG and PPO, respectively, in Figure 5.2&5.3. For MVP, since the algorithm uses coordinate gradient at each step, it does not have a PPO or NPG version, and therefore we only report its learning curve in Figure 5.2. For MVPI, we use its on-policy version for the experiment. As it works with any off-the-shelf policy search method, we implement NPG and PPO with two-layer over-parameterized neural networks as its policy search component.

There are several interesting findings of our experiments. *First*, the results show that TOPS outperforms other baselines in most of the testbeds with respect to initial learning speed, the variance during the learning process, and the steady-state mean. In particular, TOPS outperforms other methods with a large margin with respect to the mean and variance of the learning curve on `Walker2d-v2`, `Hopper-v2` and `InvertedPendulum-v2`, as seen from Figure 5.2 and Figure 5.3. *Second*, the partial order of performance level tends to be consistent across different tasks, regardless of the base method (NPG or PPO). The order of performance level (from the best to the worst) in most subfigures of Figure 5.2 is TOPS-NPG, VARAC-NPG, and MVPI-NPG. Similar results are observed in the majority of the subfigures in Figure 5.3, where the order is TOPS-PPO, VARAC-PPO, and MVPI-PPO. Compared to other methods, MVP performs poorly in all tested domains, which indicates it may not suit the tasks. Note that on domains such as `Walker2d-v2` and `Hopper-v2`, MVP’s curves show zero variance with zero mean. This is due to the fact that these domains have a sparse reward nature (i.e., the reward is 0 for most states), and zero mean and zero reward indicate that MVP simply learns nothing useful.

Overall, these results demonstrate that our TOPS algorithm can achieve state-of-the-art risk-averse performance on the challenging robot simulator testbeds.

5.6 Discussion

5.6.1 Method comparison

In risk-averse RL, variance is a more popular risk measure among its many peers [94, 69, 34, 57, 116], with the related approaches usually referred to as mean-variance RL. Variance stands out due to its advantage in interpretability and computation [71, 61]. Most mean-variance RL methods consider the variance of the total reward [34, 57, 116]. In contrast, [14] and [122] propose a reward-volatility risk measure using the variance of a per-step reward. They show that the reward-volatility method is better at capturing the short-term risk and easier to compute.

In particular, we distinguish our method with [116], [14], and [122]. [116] utilize the variance of the cumulative rewards. However, this method’s theoretical analysis is limited to the sample complexity of episodic average-reward MDP, and the choice of solvers is restricted. Both [14] and [122] use the variance of the per-step reward, which introduces a policy-dependent-reward issue. [14] solve this directly, but it is much more difficult than normal MDP due to the lack of tools, and this approach also requires double-sampling. [122] avoid these issues by proposing an augmented MDP in a flexible framework that can apply any off-the-shelf policy evaluation and control method and name it MVPI. Our algorithm is inspired by [122], but our approach differs from MVPI. At any iteration, MVPI needs to keep updating π until it maximizes the objective function defined in Eq. (5.11) with a fixed y , while TOPS is only required to update π once. This key difference enables TOPS to train faster, as shown in Section 5.5. We present an algorithm comparison between TOPS and MVPI in Algorithm 2 of the Appendix C.2. Furthermore, we can also provide a theoretical analysis of global convergence.

Second, we compare our proof technique with those of [109, 62, 124, 118]. These papers share a similar method in the first part of the analysis, but the second parts are different because each works on a different policy gradient method. Our algorithm adopts the methods of [109] and [62] for neural NPG and neural PPO, respectively, and develops a method for the Q-function value based on the two. Compared with [109] and [62], our results of convergence are in probability rather than expectation because we utilize different techniques when we characterize certain error bounds. Specifically, we develop a high probability bound for the critic part while using an expectation bound in the actor part. The closest work to ours is [124], which utilizes the variance of the cumulative rewards and therefore needs to learn from consecutive trajectories instead of non-consecutive transitions. A corresponding disadvantage is that it requires two critics to represent value functions associated with the original reward and the squared reward. On the contrary, TOPS only needs one due to the deployment of the augmented MDP. Additionally, while they present theoretical proofs, they do not provide empirical results. Moreover, in their proof, Eq.(4.15) does not hold, which leads to an invalid core conclusion regarding their Eq.(4.17), an essential part of the proof. Therefore, the validity of the theoretical analysis starting from their Eq.(4.15) remains unclear. The details are mentioned in Appendix C.4.2. The most recent work along this research line is [118]. Unlike other works, they solve safe reinforcement learning problems in the primal space, termed CRPO. In addition, the theoretical analysis of CRPO is restricted to a simplified version of NPG. In contrast, we present theoretical proof for neural NPG and PPO as the policy search algorithms. Other related work includes [125], [117], and [13]. A more detailed discussion is provided in Appendix C.5.

5.6.2 Summary

This chapter aims to answer the following question: *can risk-averse algorithms have a global convergence guarantee and learn from short trajectories?* Theoretical analysis for

both neural NPG and neural PPO with two-layer over-parameterized neural networks are presented to show that TOPS can find the global optimality at an $\mathcal{O}(1/\sqrt{K})$ converge rate. We also demonstrate the empirical success of TOPS in Mujoco robot simulation domains. The final chapter returns to the applied forest and environmental-health settings and outlines how risk-aware policy learning could be extended from this theoretical framework to practical decision support.

Chapter 6

Future Work

The three studies in this dissertation show that machine learning can support environmental and biological decision-making, but they also reveal several directions where the work can be extended. The next stage is to move from interpretable prediction toward validated, risk-aware, and stakeholder-ready decision systems. Five directions are especially important.

6.1 Apply risk-averse policy learning to environmental decision support

A natural extension is to connect the risk-averse learning framework in Chapter 5 with the applied environmental problems in Chapters 3 and 4. The forest-suitability chapter produces maps of adult-only suitability, seedling-supported suitability, future gain, future loss, non-analog uncertainty, and aggregate mill-proximity consensus. These outputs could be converted into sequential decision problems in which actions include monitoring, planting, assisted regeneration, prescribed burning, seed-source selection, or species diversification. A risk-averse policy framework could then choose actions that balance expected ecological benefit with uncertainty, cost, and the risk of poor regeneration outcomes.

The same idea could be explored in environmental health. County-level exposure prediction could be linked to policy choices such as monitoring-site placement, pollen and air-quality warning systems, or targeted data collection. In both cases, the key step is to define states, actions, transitions, rewards, costs, and risk measures in collaboration with domain experts. The TOPS framework provides a theoretical basis for decision-making when high-variance outcomes are undesirable, but future work should test whether this framework improves real environmental planning decisions.

6.2 Expand the pollen-lung cancer analysis with richer exposure and health data

The pollen and lung cancer chapter is exploratory because it uses county-level ecological data and lacks several major individual-level and contextual covariates. Future work should expand the analysis to more counties, longer time periods, and higher-resolution exposure data. Additional predictors should include smoking prevalence, radon exposure, occupational risk, socioeconomic indicators, race and sex composition, age structure, screening access, stage at diagnosis, treatment availability, health-care utilization, and urban-rural structure.

The exposure side should also be expanded. Pollen could be modeled using higher-density monitoring networks, satellite vegetation products, land-cover variables, flowering phenology, climate reanalysis, wildfire smoke, mold spores, and chemical air-pollution mixtures. These richer data would allow more advanced mixture models, spatiotemporal models, and causal-inference designs. A major goal should be to separate pollen as a direct aeroallergen exposure from pollen as a marker of vegetation, seasonality, humidity, and geography. Individual-level or small-area data would be especially valuable for avoiding ecological fallacy and for testing whether pollen-associated signals remain after controlling for established lung cancer determinants.

6.3 Strengthen forest suitability models with demographic, fire-regime, and genetic validation

The forest-suitability chapter treats seedling presence as an early regeneration signal, but seedling presence does not guarantee recruitment to sapling or canopy stages. Future work should link occurrence-based suitability maps with demographic data on seedling survival, sapling recruitment, growth, mortality, and stand development. Field validation in adult-only, seedling-supported, future-loss, and future-gain areas would help determine whether modeled classes correspond to practical regeneration outcomes.

Fire information should also be improved. Mapped wildfire intersection should be supplemented with prescribed fire records, burn severity, fire-return interval, burn season, time since fire, fuel structure, stand age, thinning history, and post-fire management. These variables would allow the analysis to distinguish wildfire-intersection effects from broader fire-regime effects. Finally, genetic and seed-source information could be incorporated to evaluate whether future suitability depends on local adaptation, assisted gene flow, or provenance-specific climate tolerance. This would move the framework from realized occurrence modeling toward a stronger biological basis for climate-adaptation planning.

6.4 Improve validation, uncertainty quantification, and causal interpretation

All three studies show that model validation determines how strongly results can be interpreted. Future forest models should include spatial block cross-validation, external validation, calibration diagnostics, threshold sensitivity analysis, and uncertainty propagation from occurrence records, climate inputs, background sampling, algorithm choice, and future climate scenarios. Non-analog climate and area-of-applicability screening should be expanded from a diagnostic map into a formal uncertainty layer used during decision-making.

For environmental health, validation should move beyond pooled prediction. County-grouped, state-grouped, time-forward, and external-region validation would clarify whether pollen and pollution signals transfer across geography and time. Causal interpretation should be strengthened through directed acyclic graphs, negative-control analyses, sensitivity to unmeasured confounding, hierarchical spatiotemporal models, and, where possible, quasi-experimental designs. For risk-averse reinforcement learning, future work should test robustness under model misspecification, partial observability, off-policy data, constrained actions, and multi-objective risk criteria.

6.5 Develop accessible, reproducible, and stakeholder-facing decision tools

A final direction is to turn the analyses into transparent decision-support tools. For forest adaptation, this could include interactive maps showing adult suitability, seedling suitability, future gain and loss, non-analog uncertainty, and aggregate infrastructure-proximity summaries. Users should be able to view uncertainty layers, compare scenarios, and identify locations where management decisions are robust or fragile. For environmental health, dashboards could support exploratory pollen, air-quality, humidity, and outcome visualization while clearly communicating that county-level results are hypothesis-generating rather than causal.

Reproducibility should be built into these tools. Future work should release documented code, model objects, processed data layers where allowed, data dictionaries, random seeds, model cards, and figure-generation scripts. Accessibility should also be considered from the beginning, including readable figures, alt text, structured tables, colorblind-safe palettes, and clear explanation of uncertainty. The long-term goal is to make interpretable and risk-aware machine learning useful not only for publication, but also for scientists, land managers, public-health researchers, and decision-makers who must act under uncertainty.

6.6 Concluding synthesis: from interpretation to risk-aware action

The long-term goal of this dissertation is to connect environmental prediction with risk-aware decision-making. Chapters 3 and 4 show that interpretable machine learning can reveal environmental signals, but they also show why those signals must be constrained by validation and uncertainty. Adult-only forest suitability is useful but risky if it is interpreted without regeneration evidence. Pooled pollen-lung cancer prediction is informative but risky if it is interpreted without county-grouped validation. Chapter 5 provides a mathematical framework for the next step: learning policies that explicitly account for

outcome volatility and that can use short trajectories when long experiments are unsafe or impractical.

Future dissertation-scale work should therefore move from static interpretation toward dynamic decision support. Forest adaptation decisions could be modeled as sequential policies that select monitoring, planting, species diversification, seed-source choice, or fire-management actions under uncertain climate and regeneration outcomes. Environmental-health applications could use risk-aware learning to prioritize monitoring, data collection, and warning systems under uncertainty about geographic transfer. In both domains, TOPS provides a starting point for decision rules that do not simply maximize expected benefit, but also account for the risk of poor outcomes.

Bibliography

- [1] A. Agarwal, S. Kakade, J. Lee, and G. Mahajan. On the theory of policy gradient methods: Optimality, approximation, and distribution shift. *Journal of Machine Learning Research*, 22(98):1–76, 2021.
- [2] AirNow and U.S. Environmental Protection Agency. Air Quality Index (AQI) Basics.
- [3] Z. Allen-Zhu, Y. Li, and Y. Liang. Learning and generalization in overparameterized neural networks, going beyond two layers. *Advances in Neural Information Processing Systems*, 32, 2019.
- [4] Z. Allen-Zhu, Y. Li, and Z. Song. A convergence theory for deep learning via overparameterization, 2019.
- [5] America’s Longleaf. Range-wide conservation plan for longleaf pine, 2025–2040, 2023. America’s Longleaf Restoration Initiative/Longleaf Partnership Council.
- [6] America’s Longleaf. Range-wide conservation plan for longleaf pine, 2025–2040, 2025. America’s Longleaf Restoration Initiative/Longleaf Partnership Council.
- [7] A. Antos, C. Szepesvári, and R. Munos. Fitted q-iteration in continuous action-space mdps. *Advances in Neural Information Processing Systems*, 20, 2007.
- [8] S. Arora, S. Du, W. Hu, Z. Li, and R. Wang. Fine-grained analysis of optimization and generalization for overparameterized two-layer neural networks. In *International Conference on Machine Learning*, pages 322–332. PMLR, 2019.
- [9] J. B. Baker and O. G. Langdon. Pinus taeda l. loblolly pine. In R. M. Burns and B. H. Honkala, editors, *Silvics of North America. Volume 1. Conifers. Agriculture Handbook 654*. U.S. Department of Agriculture, Forest Service, 1990.
- [10] M. Barbet-Massin, F. Jiguet, C. H. Albert, and W. Thuiller. Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3:327–338, 2012.
- [11] N. Barve, V. Barve, A. Jimenez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberon, and F. Villalobos. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222:1810–1819, 2011.
- [12] W. A. Bechtold and P. L. Patterson. The enhanced forest inventory and analysis program: National sampling design and estimation procedures. Technical Report General Technical Report SRS-80, U.S. Department of Agriculture, Forest Service, Southern Research Station, 2005.

- [13] J. Bhandari and D. Russo. Global optimality guarantees for policy gradient methods. *arXiv preprint arXiv:1906.01786*, 2019.
- [14] L. Bisi, L. Sabbioni, E. Vittori, M. Papini, and M. Restelli. Risk-averse trust region optimization for reward-volatility reduction. In *Proceedings of the Twenty-Ninth IJCAI*, pages 4583–4589, 7 2020. Special Track on AI in FinTech.
- [15] I. Boldogh, A. Bacsi, B. K. Choudhury, N. Dharajiya, R. Alam, T. K. Hazra, et al. Ros generated by pollen nadph oxidase provide a signal that augments antigen-induced allergic airway inflammation. *Journal of Clinical Investigation*, 115(8):2169–2179, 2005.
- [16] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.
- [17] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.
- [18] B. J. Butler and D. N. Wear. Forest ownership dynamics of southern forests. In D. N. Wear and J. G. Greis, editors, *The Southern Forest Futures Project: Technical Report*, pages 103–121. U.S. Department of Agriculture, Forest Service, Southern Research Station, 2013.
- [19] Q. Cai, Z. Yang, J. Lee, and Z. Wang. Neural temporal-difference and q-learning provably converge to global optima. *arXiv preprint arXiv:1905.10027*, 2019.
- [20] Y. Cao and Q. Gu. Generalization error bounds of gradient descent for learning over-parameterized deep relu networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3349–3356, 2020.
- [21] S. Cen, C. Cheng, Y. Chen, Y. Wei, and Y. Chi. Fast global convergence of natural policy gradient methods with entropy regularization. *Operations Research*, 2021.
- [22] Centers for Disease Control and Prevention. Allergens and Pollen. Climate and Health.
- [23] Centers for Disease Control and Prevention. Lung Cancer Risk Factors.
- [24] Centers for Disease Control and Prevention. Pollen and Your Health. Climate and Health.
- [25] C. Chen, K. Naka, and S. M. Hossain. Is shortleaf pine regeneration declining under the current forest fire regime? *Global Ecology and Conservation*, 63:e03897, 2025.
- [26] T. Chen and C. Guestrin. Xgboost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.
- [27] G. S. Collins, K. G. M. Moons, P. Dhiman, R. D. Riley, A. L. Beam, B. Van Calster, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385:e078378, 2024.

- [28] F. Colotta, P. Allavena, A. Sica, C. Garlanda, and A. Mantovani. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. *Carcinogenesis*, 30(7):1073–1081, 2009.
- [29] A. Csillag, I. Boldogh, K. Pazmandi, Z. Magyarics, P. Gogolak, S. Sur, et al. Pollen-induced oxidative stress influences both innate and adaptive immune responses via altering dendritic cell functions. *Journal of Immunology*, 184(5):2377–2385, 2010.
- [30] I. Csiszár and J. Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
- [31] W. Dabney, Z. Kurth-Nelson, N. Uchida, C. Starkweather, D. Hassabis, R. Munos, and M. Botvinick. A distributional code for value in dopamine-based reinforcement learning. *Nature*, 577(7792):671–675, 2020.
- [32] G. D’Amato, L. Cecchi, M. D’Amato, and G. Liccardi. Climate change and air pollution: effects on pollen allergy and other allergic respiratory diseases. *Allergo Journal International*, 23(1):17–23, 2014.
- [33] N. Dharajiya, B. K. Choudhury, A. Bacsi, I. Boldogh, R. Alam, and S. Sur. Role of pollen nad(p)h oxidase in allergic inflammation. *Current Opinion in Allergy and Clinical Immunology*, 8(1):57–62, 2008.
- [34] D. Di Castro, A. Tamar, and S. Mannor. Policy gradients with variance related risk criteria. *arXiv preprint arXiv:1206.6404*, 2012.
- [35] C. F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carre, J. R. Garcia Marquez, B. Gruber, B. Lafourcade, P. J. Leita, T. Munkemuller, C. McClean, P. E. Osborne, B. Reineking, B. Schroder, A. K. Skidmore, D. Zurell, and S. Lautenbach. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36:27–46, 2013.
- [36] S. Du, X. Zhai, B. Póczos, and A. Singh. Gradient descent provably optimizes over-parameterized neural networks. *arXiv preprint arXiv:1810.02054*, 2018.
- [37] J. Elith and J. R. Leathwick. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40:677–697, 2009.
- [38] A. Farahmand, M. Ghavamzadeh, C. Szepesvári, and S. Mannor. Regularized policy iteration with nonparametric function spaces. *The Journal of Machine Learning Research*, 17(1):4809–4874, 2016.
- [39] L. Feng, X. Tian, Y. A. El-Kassaby, J. Qiu, Z. Feng, J. Sun, G. Wang, and T. Wang. Predicting suitable habitats of *Melia azedarach* L. in china using data mining. *Scientific Reports*, 12:12617, 2022.
- [40] J. H. Friedman and B. E. Popescu. Predictive learning via rule ensembles. *The Annals of Applied Statistics*, 2(3):916–954, 2008.

- [41] Z. Fu, Z. Yang, and Z. Wang. Single-timescale actor-critic provably finds globally optimal policy. *arXiv preprint arXiv:2008.00483*, 2020.
- [42] J. Garcia and F. Fernández. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(1):1437–1480, 2015.
- [43] H. Gu, X. Guo, X. Wei, and R. Xu. Mean-field multi-agent reinforcement learning: A decentralized network approach. *arXiv preprint arXiv:2108.02731*, 2021.
- [44] A. Guisan and W. Thuiller. Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8:993–1009, 2005.
- [45] G. B. Hamra, N. Guha, A. Cohen, F. Laden, O. Raaschou-Nielsen, J. M. Samet, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environmental Health Perspectives*, 122(9):906–911, 2014.
- [46] D. Hanahan. Hallmarks of cancer: new dimensions. *Cancer Discovery*, 12(1):31–46, 2022.
- [47] B. B. Hanberry, H. S. He, and B. J. Palik. Pseudoabsence generation strategies for species distribution models. *PLoS ONE*, 7(8):e44486, 2012.
- [48] A. Hans, D. Schneegaß, A. Schäfer, and S. Udluft. Safe exploration for reinforcement learning. In *ESANN*, pages 143–148. Citeseer, 2008.
- [49] T. J. Hawbaker, M. K. Vanderhoof, G. L. Schmidt, Y.-J. Beal, J. J. Picotte, J. D. Takacs, J. T. Falgout, and J. L. Dwyer. The landsat burned area algorithm and products for the conterminous united states. *Remote Sensing of Environment*, 244:111801, 2020.
- [50] International Agency for Research on Cancer. Outdoor air pollution a leading environmental cause of cancer deaths. Press Release No. 221, 2013.
- [51] L. R. Iverson, M. P. Peters, A. M. Prasad, and S. N. Matthews. Analysis of climate change impacts on tree species of the eastern US: Results of DISTRIB-II modeling. *Forests*, 10:302, 2019.
- [52] S. Kakade. A natural policy gradient. In *Advances in neural information processing systems*, NIPS’01, page 15311538, Cambridge, MA, USA, 2001. MIT Press.
- [53] S. Kakade and J. Langford. Approximately optimal approximate reinforcement learning. In *International Conference on Machine Learning*, International Conference on Machine Learning ’02, page 267274, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.
- [54] T. Konstantopoulos, Z. Zerakidze, and G. Sokhadze. Radon–nikodým theorem. In *International Encyclopedia of Statistical Science*, pages 1161–1164. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

- [55] M. Kubo, R. Banno, H. Manabe, and M. Minoji. Implicit regularization in over-parameterized neural networks. *arXiv preprint arXiv:1903.01997*, 2019.
- [56] P. L.A. and M. Ghavamzadeh. Actor-critic algorithms for risk-sensitive mdps. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.
- [57] P. La and M. Ghavamzadeh. Actor-critic algorithms for risk-sensitive mdps. *Advances in Neural Information Processing Systems*, 26, 2013.
- [58] T. Lai, H. Xing, and Z. Chen. Meanvariance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics*, 5(2A), Jun 2011.
- [59] H. C. Y. Lam, D. Jarvis, and E. Fuertes. Interactive effects of allergens and air pollution on respiratory health: a systematic review. *Science of the Total Environment*, 757:143924, 2021.
- [60] E. R. Lawson. Pinus echinata mill. shortleaf pine. In R. M. Burns and B. H. Honkala, editors, *Silvics of North America. Volume 1. Conifers. Agriculture Handbook 654*. U.S. Department of Agriculture, Forest Service, 1990.
- [61] D. Li and W. Ng. Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical finance*, 10(3):387–406, 2000.
- [62] B. Liu, Q. Cai, Z. Yang, and Z. Wang. Neural trust region/proximal policy optimization attains globally optimal policy. *Advances in Neural Information Processing Systems*, 32, 2019.
- [63] B. Liu, J. Liu, M. Ghavamzadeh, S. Mahadevan, and M. Petrik. Finite-sample analysis of proximal gradient td algorithms. In *Proceedings of the Conference on Uncertainty in AI (UAI)*, pages 504–513, 2015.
- [64] J. M. Lobo, A. Jimenez-Valverde, and R. Real. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17:145–151, 2008.
- [65] R. E. Lohrey and S. V. Kossuth. Pinus elliottii engelm. slash pine. In R. M. Burns and B. H. Honkala, editors, *Silvics of North America. Volume 1. Conifers. Agriculture Handbook 654*. U.S. Department of Agriculture, Forest Service, 1990.
- [66] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
- [67] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
- [68] A. Majumdar and M. Pavone. How should a robot assess risk? towards an axiomatic theory of risk in robotics. In *Robotics Research*, pages 75–84. Springer, 2020.

- [69] S. Mannor and J. Tsitsiklis. Mean-variance optimization in markov decision processes. *arXiv preprint arXiv:1104.5601*, 2011.
- [70] A. Mantovani, P. Allavena, A. Sica, and F. Balkwill. Cancer-related inflammation. *Nature*, 454:436–444, 2008.
- [71] H. Markowitz and P. Todd. *Mean-variance analysis in portfolio choice and capital markets*, volume 66. John Wiley & Sons, 2000.
- [72] L. McInnes, J. Healy, and J. Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv:1802.03426*, 2018.
- [73] S. McNulty, P. Caldwell, T. W. Doyle, K. Johnsen, Y. Liu, J. Mohan, J. Prestemon, and G. Sun. Forests and climate change in the southeast USA. In K. Ingram, K. Dow, L. Carter, and J. Anderson, editors, *Climate of the Southeast United States: Variability, Change, Impacts, and Vulnerability*, pages 165–189. Island Press, 2013.
- [74] J. Mei, C. Xiao, C. Szepesvari, and D. Schuurmans. On the global convergence rates of softmax policy gradient methods. In *International Conference on Machine Learning*, pages 6820–6829. PMLR, 2020.
- [75] H. Meyer and E. Pebesma. Predicting into unknown space? estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12:1620–1633, 2021.
- [76] R. J. Mitchell, Y. Liu, J. J. O’Brien, K. J. Elliott, G. Starr, C. F. Miniati, and J. K. Hiers. Future climate and fire interactions in the southeastern region of the united states. *Forest Ecology and Management*, 327:316–326, 2014.
- [77] V. Mnih, K. Kavukcuoglu, D. Silver, A. Rusu, J. Veness, M. Bellemare, A. Graves, M. Riedmiller, A. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, Feb. 2015.
- [78] C. Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 3 edition, 2025.
- [79] R. Munos. Performance bounds in l_p -norm for approximate value iteration. *SIAM journal on control and optimization*, 46(2):541–561, 2007.
- [80] R. Munos and C. Szepesvári. Finite-time bounds for fitted value iteration. *Journal of Machine Learning Research*, 9(5), 2008.
- [81] N. J. Nassikas, H. Luttmann-Gibson, S. L. Rifas-Shiman, E. Oken, D. R. Gold, and M. B. Rice. Acute exposure to pollen and airway inflammation in adolescents. *Pediatric Pulmonology*, 59:1313–1320, 2024.

- [82] National Cancer Institute. SEER Cancer Stat Facts: Lung and Bronchus Cancer. Surveillance, Epidemiology, and End Results Program.
- [83] D. Parker. Managing risk in healthcare: understanding your safety culture using the manchester patient safety framework (mapsaf). *Journal of nursing management*, 17(2):218–222, 2009.
- [84] M. P. Peters, A. M. Prasad, S. N. Matthews, and L. R. Iverson. Climate change tree atlas, version 4, 2020. U.S. Department of Agriculture, Forest Service, Northern Research Station and Northern Institute of Applied Climate Science. DOI: 10.2737/Climate-Change-Tree-Atlas-v4.
- [85] S. J. Phillips, R. P. Anderson, and R. E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190:231–259, 2006.
- [86] K. M. Potter, C. M. Oswalt, and J. M. Guldin. Range-wide assessment of recent longleaf pine (*Pinus palustris* mill.) area and regeneration trends. *Forests*, 15:1255, 2024.
- [87] T. Qiu, S. Sharma, C. W. Woodall, and J. S. Clark. Niche shifts from trees to fecundity to recruitment that determine species response to climate change. *Frontiers in Ecology and Evolution*, 9:719141, 2021.
- [88] A. Rahimi and B. Recht. Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning. *Advances in Neural Information Processing Systems*, 21, 2008.
- [89] D. R. Roberts, V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schroder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F. Dormann. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40:913–929, 2017.
- [90] S. Satpathi, H. Gupta, S. Liang, and R. Srikant. The role of regularization in over-parameterized neural networks. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 4683–4688. IEEE, 2020.
- [91] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz. Trust region policy optimization. In *International Conference on Machine Learning*, pages 1889–1897. PMLR, 2015.
- [92] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- [93] L. Shani, Y. Efroni, and S. Mannor. Adaptive trust region policy optimization: Global convergence and faster rates for regularized mdps. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5668–5675, 2020.
- [94] M. Sobel. The variance of discounted markov decision processes. *Journal of Applied Probability*, 19(4):794–802, 1982.

- [95] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8:25, 2007.
- [96] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA, 2018.
- [97] R. Sutton, H. Maei, D. Precup, S. Bhatnagar, D. Silver, C. Szepesvári, and E. Wiewiora. Fast gradient-descent methods for temporal-difference learning with linear function approximation. In *International Conference on Machine Learning*, pages 993–1000, 2009.
- [98] Tall Timbers Research. Southeast firemap data readme, 2024. Updated May 22, 2024. Tall Timbers Research, Tallahassee, Florida.
- [99] C. Teske, M. K. Vanderhoof, T. J. Hawbaker, J. Noble, and J. K. Hiers. Using the landsat burned area products to derive fire history relevant for fire management and conservation in the state of florida, southeastern usa. *Fire*, 4(2):26, 2021.
- [100] E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems*, pages 5026–5033. IEEE, 2012.
- [101] M. C. Turner, Z. J. Andersen, A. Baccarelli, W. R. Diver, S. M. Gapstur, C. A. Pope, et al. Outdoor air pollution and cancer: an overview of the current evidence and public health recommendations. *CA: A Cancer Journal for Clinicians*, 70(6):460–479, 2020.
- [102] U.S. Department of Agriculture Forest Service. Forest inventory and analysis, n.d. U.S. Department of Agriculture, Forest Service.
- [103] U.S. Department of Agriculture Forest Service. Forest inventory and analysis, n.d. U.S. Department of Agriculture, Forest Service.
- [104] U.S. Environmental Protection Agency. Particulate Matter (PM) Basics.
- [105] USDA Forest Service. Forest inventory and analysis national core field guide, 2023. U.S. Department of Agriculture, Forest Service.
- [106] L. J. P. van der Maaten and G. E. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
- [107] O. Vinyals, I. Babuschkin, W. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. Choi, R. Powell, T. Ewalds, P. Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.
- [108] E. von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, and J. P. Vandenbroucke. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*, 370:1453–1457, 2007.

- [109] L. Wang, Q. Cai, Z. Yang, and Z. Wang. Neural policy gradient methods: Global optimality and rates of convergence, 2019.
- [110] M. Wang, E. X. Fang, and H. Liu. Stochastic compositional gradient descent: algorithms for minimizing compositions of expected-value functions. *Mathematical Programming*, 161(1-2):419–449, 2017.
- [111] T. Wang, A. Hamann, D. L. Spittlehouse, and C. Carroll. Locally downscaled and spatially customizable climate data for historical and future periods for north america. *PLoS ONE*, 11(6):e0156720, 2016.
- [112] W. Wang, J. Li, and X. He. Deep reinforcement learning for nlp. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 19–21, 2018.
- [113] J. Weng, A. Duburcq, K. You, and H. Chen. Mujoco benchmark, 2020.
- [114] J. L. Willis, M. M. Bataineh, D. C. Clabo, E. D. Dickens, E. Galeano, S. B. Jack, J. S. Kush, C. D. Nelson, A. D. Polinko, K. M. Quigley, M. A. Sayer, J. M. Varner, and W. C. Zipperer. Longleaf pine (*Pinus palustris*). In S. McNulty and M. A. J. Callaham, editors, *Silvics of North America*. U.S. Department of Agriculture, Forest Service, 2026.
- [115] World Health Organization. Climate change, air quality, and pollen allergies.
- [116] T. Xie, B. Liu, Y. Xu, M. Ghavamzadeh, Y. Chow, D. Lyu, and D. Yoon. A block coordinate ascent algorithm for mean-variance optimization. In *Advances in Neural Information Processing Systems*, 2018.
- [117] P. Xu, J. Chen, D. Zou, and Q. Gu. Global convergence of langevin dynamics based algorithms for nonconvex optimization. In *Advances in Neural Information Processing Systems*, 2018.
- [118] T. Xu, Y. Liang, and G. Lan. Crpo: A new approach for safe reinforcement learning with convergence guarantee. In *International Conference on Machine Learning*, pages 11480–11491. PMLR, 2021.
- [119] L. Yang and M. Wang. Reinforcement learning in feature space: Matrix bandit, kernels, and regret bound. In *International Conference on Machine Learning*, pages 10746–10756. PMLR, 2020.
- [120] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.
- [121] S. Zhang, B. Liu, and S. Whiteson. Mean-variance policy iteration for risk-averse reinforcement learning. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

- [122] S. Zhang, B. Liu, and W. Whiteson. Mean-variance policy iteration for risk-averse reinforcement learning, 2020.
- [123] Y. Zhao and T. Wang. Predicting the global fundamental climate niche of lodgepole pine for climate change adaptation. *Frontiers in Forests and Global Change*, 6:1084797, 2023.
- [124] H. Zhong, E. Fang, Z. Yang, and Z. Wang. Risk-sensitive deep rl: Variance-constrained actor-critic provably finds globally optimal policy, 2020.
- [125] D. Zou, Y. Cao, D. Zhou, and Q. Gu. Gradient descent optimizes over-parameterized deep relu networks. *Machine Learning*, 109(3):467–492, 2020.

Appendix A

Supplementary Material for Forest Suitability Modeling

This appendix contains the supplementary figures and supplementary table associated with Chapter 3.

A.1 Supplementary plots, tables, and figure notes

Supplementary Methods S3. Data provenance and preprocessing

Occurrence-data preparation. Occurrence records were obtained from the U.S. Department of Agriculture Forest Service Forest Inventory and Analysis database (FIADB) and organized into adult-tree and seedling data streams. Adult-tree presences were restricted to live trees with a diameter at breast height of at least 12.7 cm, and sampled locations where the focal species was not recorded formed the candidate absence pool. Repeated observations at the same geographic location were consolidated, and the resulting records were aligned to the common analytical coordinate domain; the latest available measurement year was retained for temporal linkage with mapped fire history. Seedling observations came from a separate FIA seedling data stream and were processed independently from adult-tree observations. Locations lacking valid coordinates or corresponding climate information were excluded during downstream preprocessing, and separate balanced datasets were prepared for adult-tree, seedling, and wildfire-stratified analyses.

Climate-data preparation. Historical and future climate information was obtained from downscaled ClimateNA-ERA products prepared for supplied geographic locations and projection grids (Wang et al., 2016). The source climate outputs contained monthly, seasonal, and annual fields. Monthly and seasonal fields were excluded during preprocessing, while

annual bioclimatic variables were retained as the candidate predictor set used for missing-value checks, standardization, VIF screening, model fitting, and interpretation. Historical location-level climate values were aligned to the occurrence domain for the four 30-year normal periods, whereas mapped historical and future predictions were evaluated on the 0.1° analytical grid. Future projections used five comparable climate products: one 8-GCM ensemble and four individual GCMs. These future climate layers were applied to adult-tree Random Forest models calibrated under the 1961–1990 historical baseline. All five products were processed under the same SSPs, future periods, and species-specific projection workflow; suitability averaging occurred after prediction.

Fire-history-data preparation. Mapped burned-area and fire-history information was obtained from the Southeast FireMap, a regional Landsat-based product developed by Tall Timbers Research (Tall Timbers Research, 2024). The product uses 30 m Landsat burned-area data to identify annual burn detections from 1994 through 2023 and includes summary fire-history metrics. Annual burn probabilities were thresholded at 90%, spatial smoothing was applied, and detections smaller than 11 connected pixels (approximately 2.45 acres) were removed (Hawbaker et al., 2020; Teske et al., 2021). The source product is designed to capture burned areas associated with both prescribed fire and wildfire activity, although fire type was not distinguished in the analytical classification used here. A point representation of the mapped fire-history polygons was used for occurrence linkage. For the primary classification, occurrence records measured during 1994–2023 were matched to mapped fire history within approximately 100 m, and only fire detections occurring in or before the year each record was measured were considered. Records with one or more qualifying detections were classified as wildfire-intersected; records with no qualifying match were classified as non-wildfire-intersected. Spatial sensitivity was evaluated at 250, 500, and 1,000 m using the same temporal rule.

Table A.1: Summary of data origin, temporal coverage, spatial basis, preprocessing, and analytical use for the occurrence, climate, and fire-history data streams.

Data stream	General source and coverage	Spatial basis	Preprocessing summary	Analytical use
Adult-tree occurrence	USDA Forest Service Forest Inventory and Analysis database (FIADB) tree records; source measurements extend from the late 1960s through 2023 across the broader southeastern and eastern U.S. sampling domain.	Plot-linked geographic coordinates used within the common analytical domain.	Adult presences were restricted to live trees at least 12.7 cm in diameter at breast height; candidate absences came from sampled locations where the focal species was not recorded. Repeated observations were consolidated, and measurement year was retained for temporal fire-history linkage.	Adult-tree SDMs, adult–seedling comparisons, model interpretation, and future projections.
Seedling occurrence	USDA Forest Service Forest Inventory and Analysis database (FIADB) seedling records; measurement years 1995–2023 across the focal species datasets.	Plot-linked seedling locations aligned to the common analytical coordinate domain.	Seedling records were processed separately from adult-tree observations, with measurement year retained for temporal fire-history linkage.	Seedling SDMs, regeneration comparisons, and wildfire-stratified analyses.
Historical climate	Downscaled ClimateNA-ERA climate products for 1961–1990, 1971–2000, 1981–2010, and 1991–2020 normal periods.	Location-associated climate values for occurrence modeling and a 0.1° analytical grid for mapping.	Monthly and seasonal fields were removed; annual bioclimatic variables were retained, standardized, and screened by VIF.	Historical calibration, baseline sensitivity, model interpretation, and spatial prediction.
Future climate	ClimateNA-derived CMIP6 products for SSP126, SSP245, and SSP370 and the 2011–2040, 2041–2070, and 2071–2100 periods.	Future climate layers evaluated on the 0.1° analytical grid.	One 8-GCM ensemble and four individual GCMs were processed comparably; suitability surfaces were generated first and then averaged.	Future suitability, historical–future change, and extrapolation screening.
Fire history	Tall Timbers Research Southeast FireMap; 30 m Landsat-based annual burn detections and fire-history metrics for 1994–2023.	Mapped fire-history polygons represented as points for occurrence linkage; primary matching distance 100 m, with 250–1,000 m sensitivity distances.	Annual burn detections were linked to occurrence locations and restricted to fires detected in or before the year each occurrence was measured. The source product captures burned areas associated with prescribed fire and wildfire activity, but fire type was not distinguished in the binary analytical classification.	Wildfire stratification, spatial-distance sensitivity, and wildfire-conditioned PCA and Random Forest comparisons.

Note. This table summarizes the provenance and processing information needed to interpret the analytical workflow; local file names and computer-specific processing paths are omitted. FIA = Forest Inventory and Analysis; FIADB = Forest Inventory and Analysis Database; GCM = general circulation model; SSP = Shared Socioeconomic Pathway; VIF = variance inflation factor.

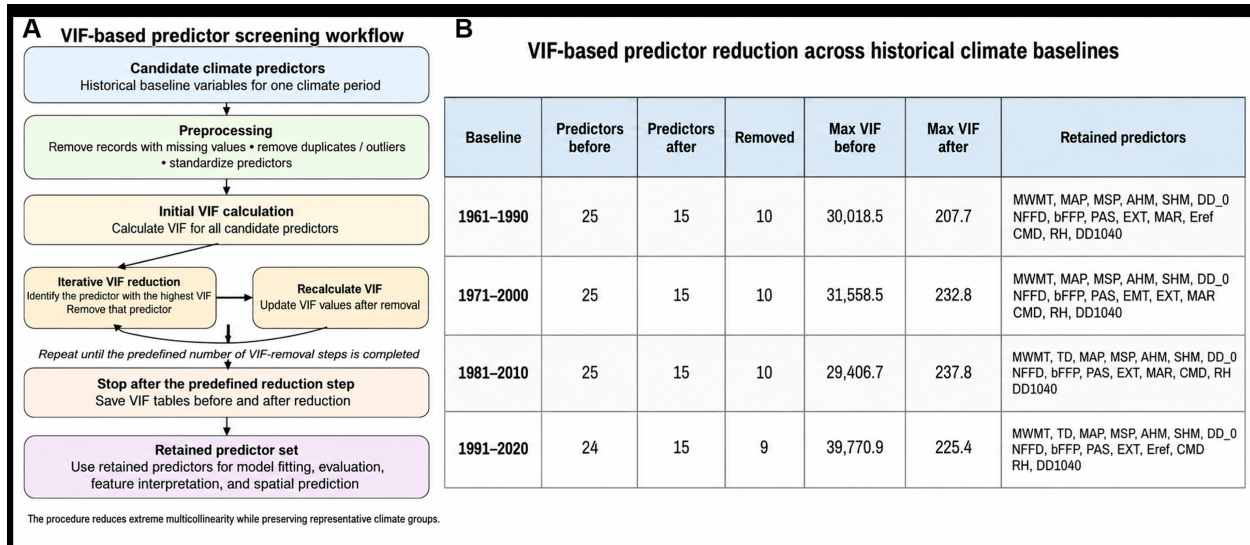


Figure A.1: VIF-based climate-predictor screening workflow and retained predictor sets across historical climate baselines.

Detailed description for Figure A.1. VIF-based climate-predictor screening workflow and retained predictor sets across historical climate baselines. Panel A shows the workflow used to reduce multicollinearity among candidate climate predictors before species distribution modeling. Candidate predictors were assembled separately for each historical baseline window, preprocessed by removing records with missing values, duplicates, and outliers, and standardized before VIF calculation. Variance inflation factors were then calculated for all candidate predictors, and the predictor with the highest VIF was removed iteratively. VIF values were recalculated after each removal, and the procedure was repeated until the predefined reduction step was completed. The retained predictor set was then used for model fitting, model evaluation, feature interpretation, and spatial prediction. Panel B summarizes VIF-based predictor reduction for four 30-year historical climate baselines: 1961-1990, 1971-2000, 1981-2010, and 1991-2020. The table reports the number of candidate predictors before and after reduction, the number removed, maximum VIF before and after reduction, and the retained predictors for each baseline. The procedure retained 15 predictors for each baseline, reducing the predictor set from 25 to 15 variables for 1961-1990, 1971-2000, and 1981-2010, and from 24 to 15 variables for 1991-2020, where MAR was unavailable. Retained predictors represented major climate-filter groups used throughout the analysis, including thermal energy, cold limitation, moisture availability, evaporative demand or water deficit, heat-moisture balance, humidity, and seasonality. Because retained predictor sets differed slightly among baselines and several retained variables represented related climate gradients, subsequent interpretation emphasized broader climate-filter groups rather than treating individual predictors as independent causal drivers.

Table A.2: Model performance metrics for seven species distribution modeling algorithms under the 1991-2020 climate baseline.

Species	Method	Accuracy	ROC AUC	Kappa	TSS
Loblolly pine	SVM	0.858	0.912	0.716	0.718
Loblolly pine	GLM	0.854	0.910	0.709	0.710
Loblolly pine	GBM	0.866	0.923	0.732	0.733
Loblolly pine	RF	0.872	0.933	0.744	0.750
Loblolly pine	NBM	0.831	0.896	0.662	0.676
Loblolly pine	XGBoost	0.876	0.935	0.753	0.754
Loblolly pine	MaxEnt	0.853	0.910	0.706	0.708
Shortleaf pine	SVM	0.822	0.880	0.643	0.664
Shortleaf pine	GLM	0.818	0.876	0.637	0.638
Shortleaf pine	GBM	0.844	0.902	0.687	0.689
Shortleaf pine	RF	0.860	0.918	0.720	0.726
Shortleaf pine	NBM	0.754	0.838	0.508	0.511
Shortleaf pine	XGBoost	0.863	0.916	0.726	0.728
Shortleaf pine	MaxEnt	0.816	0.874	0.632	0.633
Slash pine	SVM	0.899	0.950	0.798	0.802
Slash pine	GLM	0.897	0.952	0.795	0.798
Slash pine	GBM	0.910	0.963	0.820	0.823
Slash pine	RF	0.914	0.966	0.829	0.832
Slash pine	NBM	0.885	0.943	0.771	0.784
Slash pine	XGBoost	0.916	0.966	0.831	0.833
Slash pine	MaxEnt	0.897	0.951	0.794	0.796
Longleaf pine	SVM	0.859	0.915	0.718	0.733
Longleaf pine	GLM	0.870	0.916	0.741	0.744
Longleaf pine	GBM	0.885	0.931	0.770	0.773
Longleaf pine	RF	0.889	0.941	0.778	0.787
Longleaf pine	NBM	0.841	0.906	0.683	0.705
Longleaf pine	XGBoost	0.892	0.938	0.785	0.788
Longleaf pine	MaxEnt	0.871	0.916	0.741	0.744

Note. Models were fitted separately for each southern yellow pine species. ROC AUC is the receiver operating characteristic area under the curve; Kappa is Cohen's Kappa; TSS is the true skill statistic. SVM = support vector machine; GLM = generalized linear model; GBM = gradient boosting machine; RF = Random Forest; NBM = naive Bayes model; XGBoost = extreme gradient boosting; MaxEnt = maximum entropy.

ROC curves across four southern pine species (1991–2020)

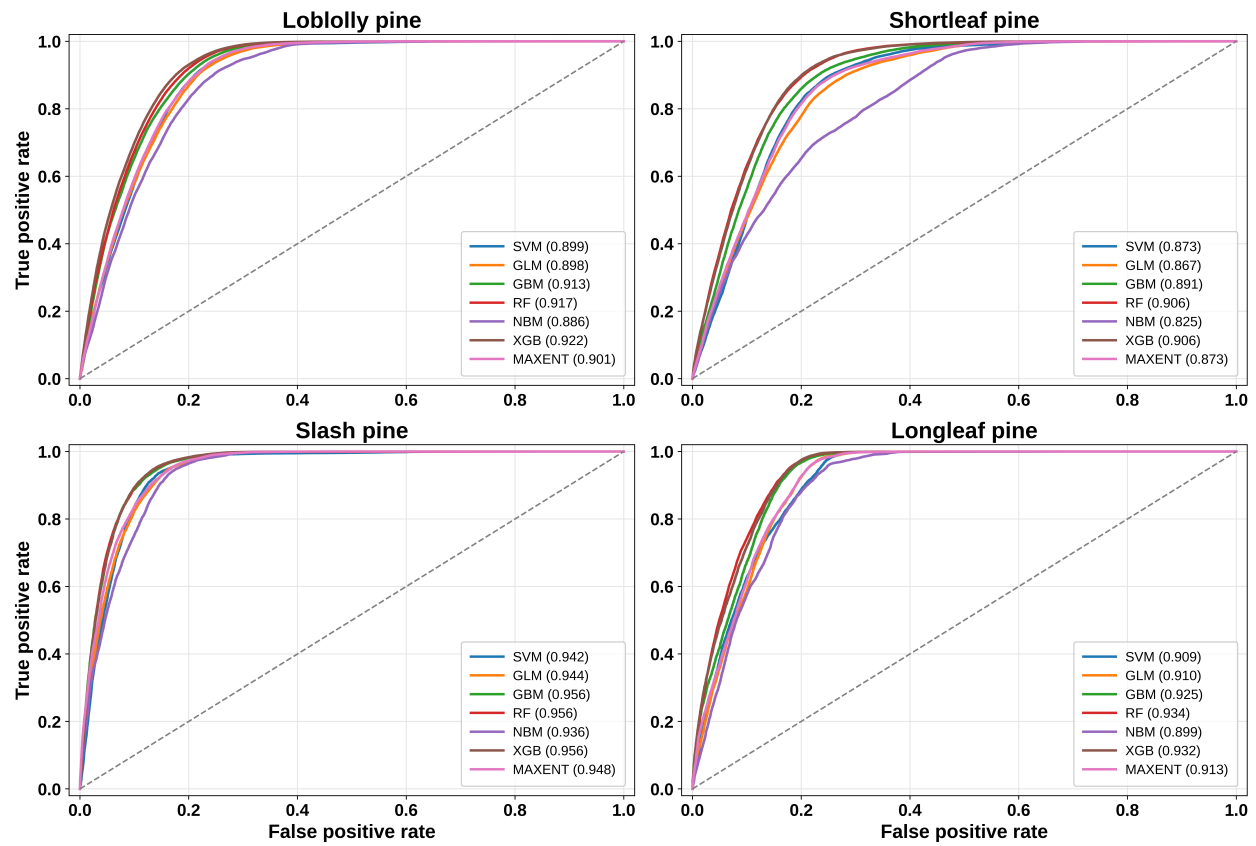


Figure A.2: Receiver operating characteristic curves for seven species distribution modeling algorithms under the 1991-2020 climate baseline.

Detailed description for Figure A.2. Receiver operating characteristic curves for seven species distribution modeling algorithms under the 1991-2020 climate baseline. ROC curves show discrimination performance for seven species distribution modeling algorithms fitted separately for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*). Algorithms include support vector machine (SVM), generalized linear model (GLM), gradient boosting machine (GBM), Random Forest (RF), naive Bayes model (NBM), XGBoost (XGB), and MaxEnt. The x-axis shows the false positive rate, and the y-axis shows the true positive rate across classification thresholds. The dashed diagonal line represents no-discrimination performance, and curves farther above this line indicate stronger discrimination between presence and background or pseudoabsence data. Values in parentheses in each legend indicate area under the ROC curve (AUC). The 1991-2020 baseline was used as the most contemporary historical climate window for model evaluation and interpretation.

Random Forest MDI feature importance across 10 runs (1991–2020)

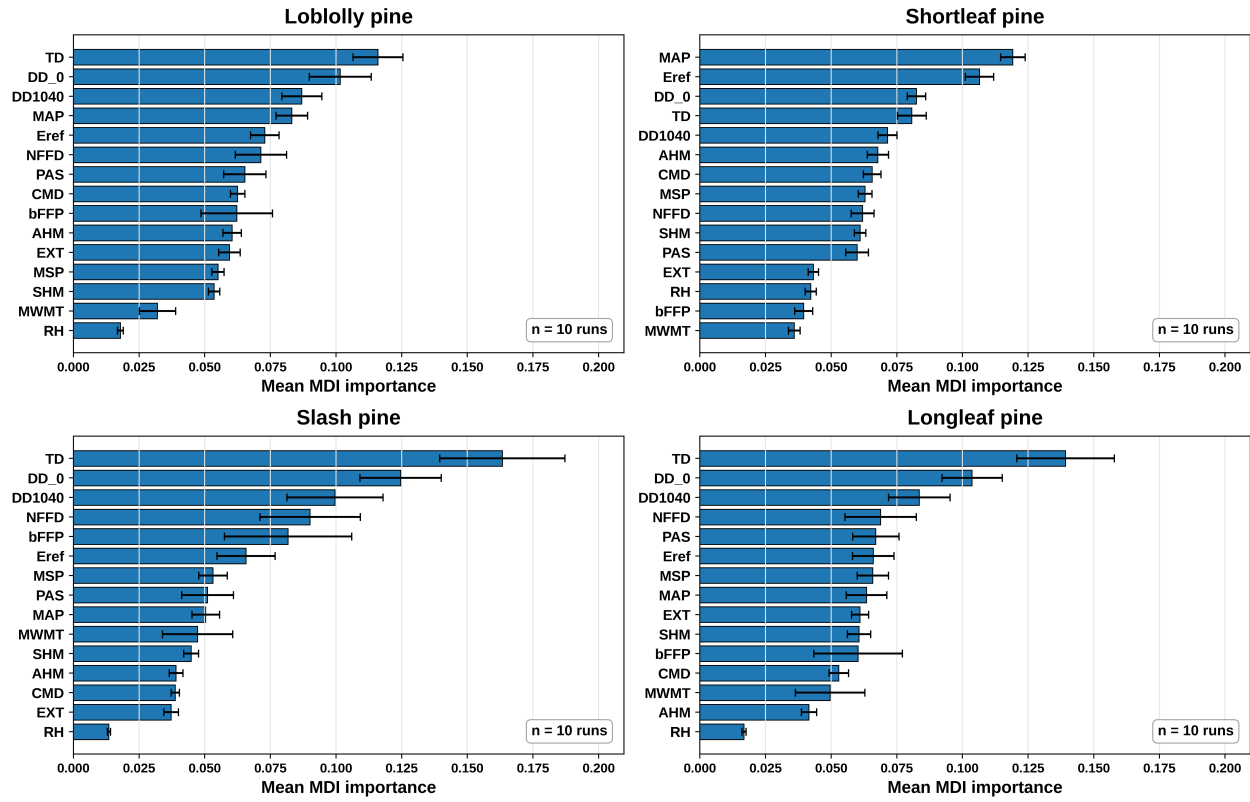


Figure A.3: Random Forest MDI feature-importance rankings across 10 model runs under the 1991-2020 climate baseline.

Detailed description for Figure A.3. Random Forest MDI feature-importance rankings across 10 model runs under the 1991-2020 climate baseline. Mean decrease in impurity (MDI) feature-importance values are shown for Random Forest species distribution models fitted separately for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*). Bars represent mean MDI importance across 10 model runs, and horizontal error bars indicate variation among runs. Predictors are ranked within each species from highest to lowest mean importance. Across species, the strongest predictors differed: temperature difference/continentality (TD), cold limitation (DD_0), degree-day accumulation (DD1040), and growing-season variables were highly ranked for loblolly, slash, and longleaf pine, whereas shortleaf pine showed stronger importance of moisture and water-balance variables, especially mean annual precipitation (MAP) and reference evaporation (Eref). These rankings support the interpretation that southern pine suitability is structured by species-specific combinations of moisture availability, thermal energy, cold limitation, evaporative demand, and seasonality rather than by a single universal climate driver. MDI values are interpreted as model-based indicators of predictive contribution and should not be treated as causal effects of individual climate variables, especially because correlated predictors can influence feature-importance rankings.

Random Forest SHAP feature importance across 10 runs (1991–2020)

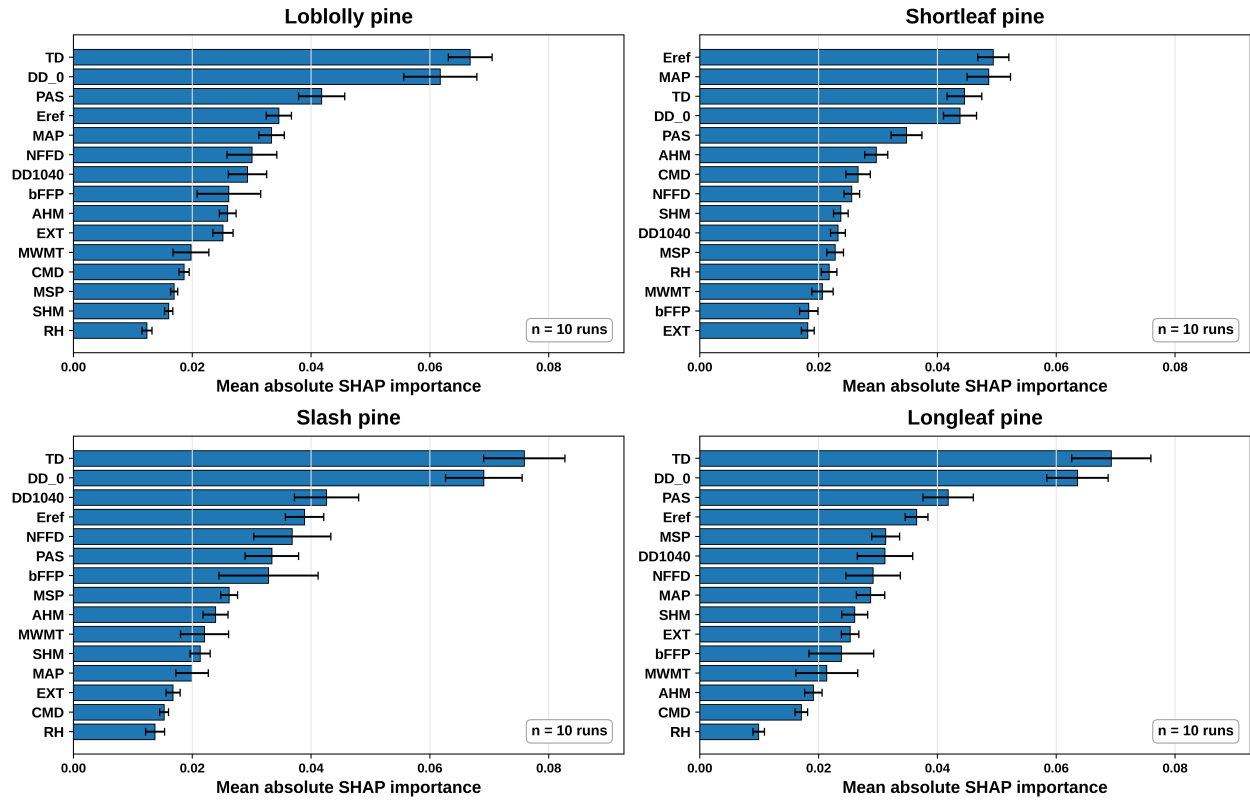
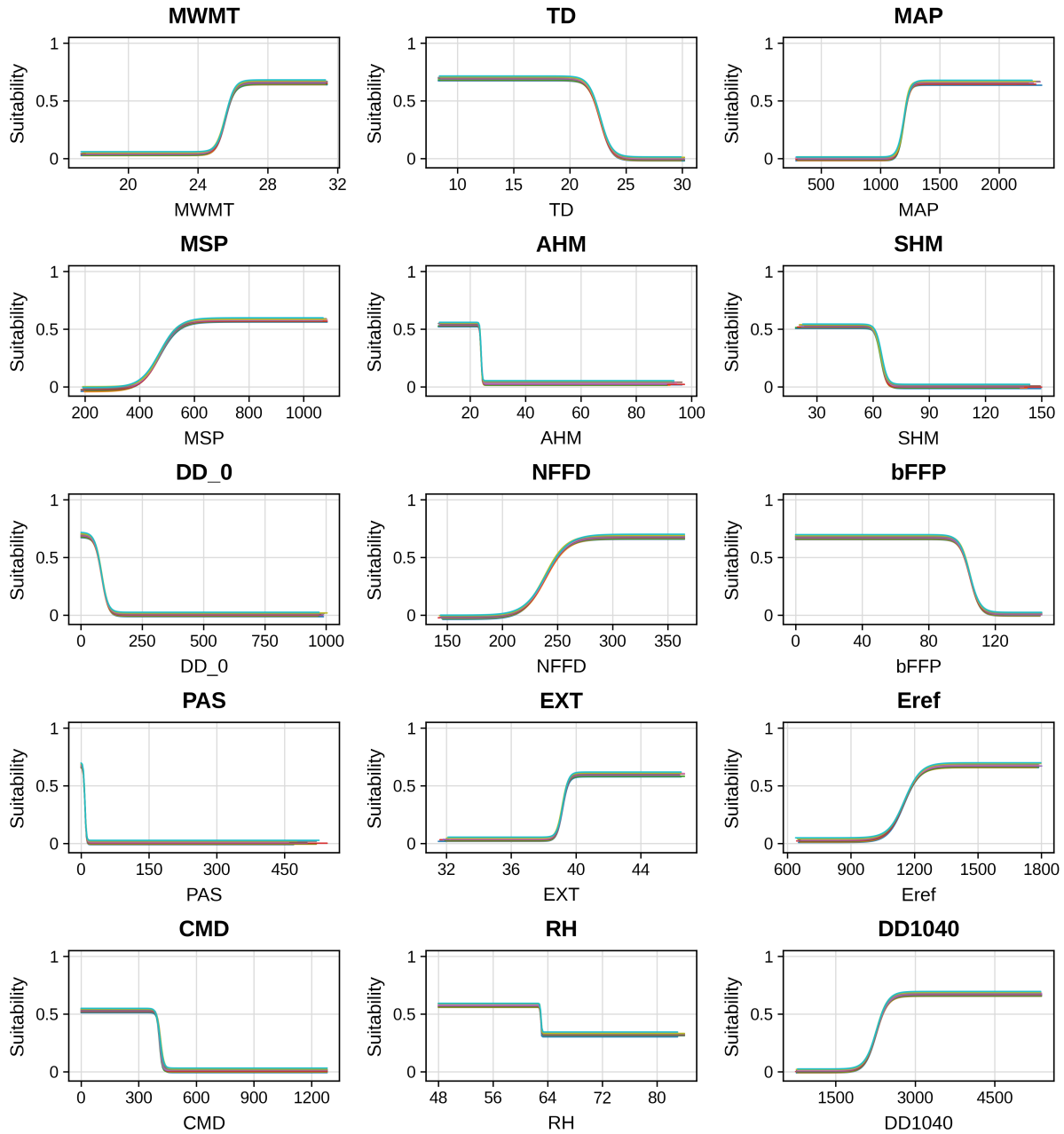


Figure A.4: Random Forest SHAP feature-importance rankings across 10 model runs under the 1991-2020 climate baseline.

Detailed description for Figure A.4. Random Forest SHAP feature-importance rankings across 10 model runs under the 1991-2020 climate baseline. Mean absolute SHAP values are shown for Random Forest species distribution models fitted separately for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliotii*), and longleaf pine (*P. palustris*). Bars represent the mean absolute SHAP value across 10 model runs, and horizontal error bars show among-run variation. Predictors are ranked within each species by mean SHAP importance, with larger values indicating greater average contribution to model predictions. Under the 1991-2020 baseline, loblolly, slash, and longleaf pine were most strongly associated with temperature seasonality and cold-limitation gradients, especially TD and DD_0, while shortleaf pine showed the strongest importance of moisture and water-balance variables, particularly MAP and Eref. These species-specific SHAP rankings support the interpretation that southern pine suitability is structured by different combinations of thermal energy, cold limitation, moisture availability, evaporative demand, and seasonality rather than by a single universal climate predictor. Because SHAP values are summarized here as absolute magnitudes, they indicate the strength of each predictor's contribution but not whether higher predictor values increase or decrease suitability. SHAP results are interpreted as model-based predictive associations within realized climate space, not as causal effects of individual climate variables.

RF response curves across 10 runs: Loblolly pine (1991–2020)

All retained predictors

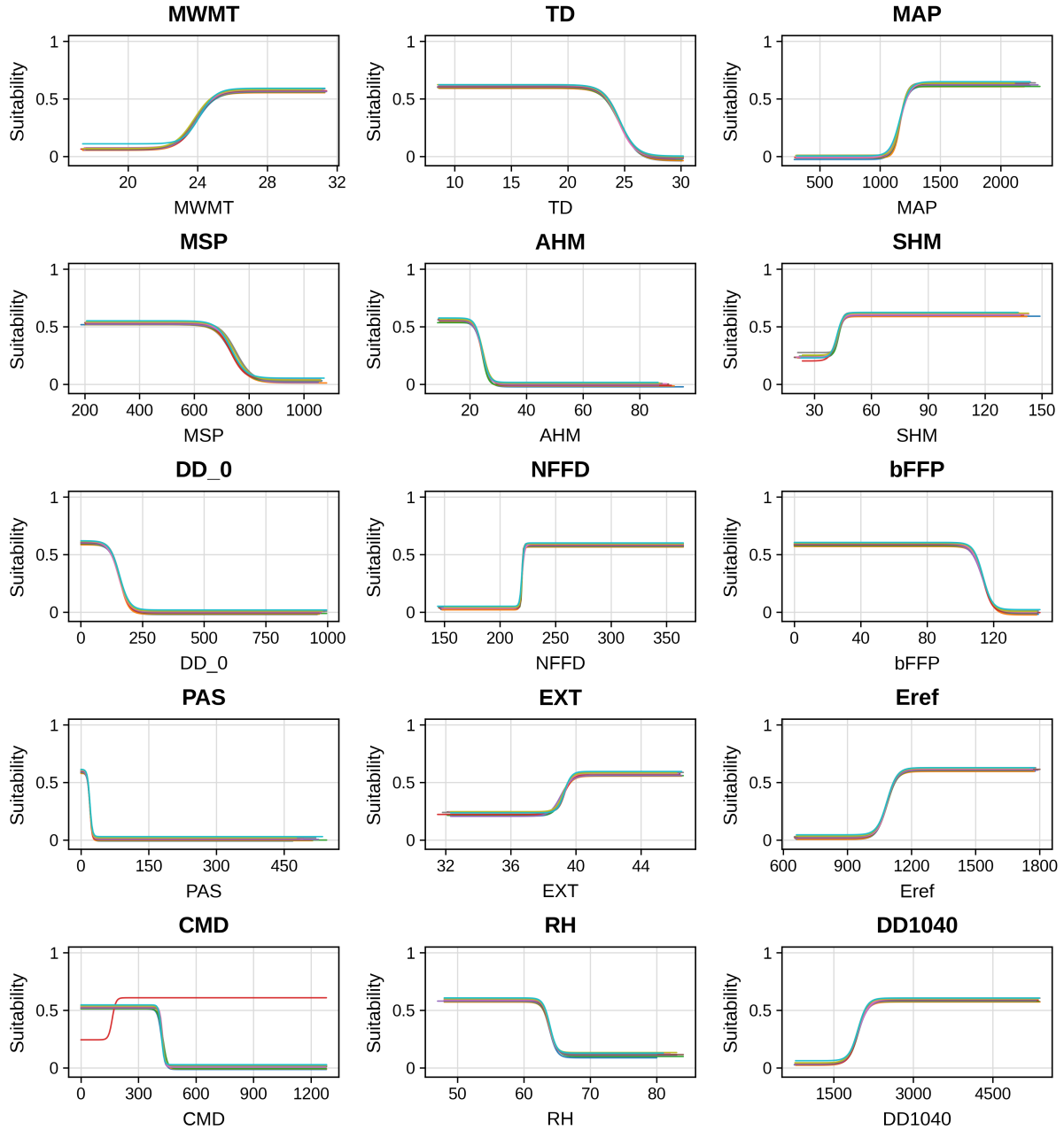


Each line represents one RF run (n = 10 runs).

Figure A.5: Random Forest response curves for retained climate predictors across 10 model runs under the 1991–2020 climate baseline for loblolly pine.

RF response curves across 10 runs: Shortleaf pine (1991–2020)

All retained predictors

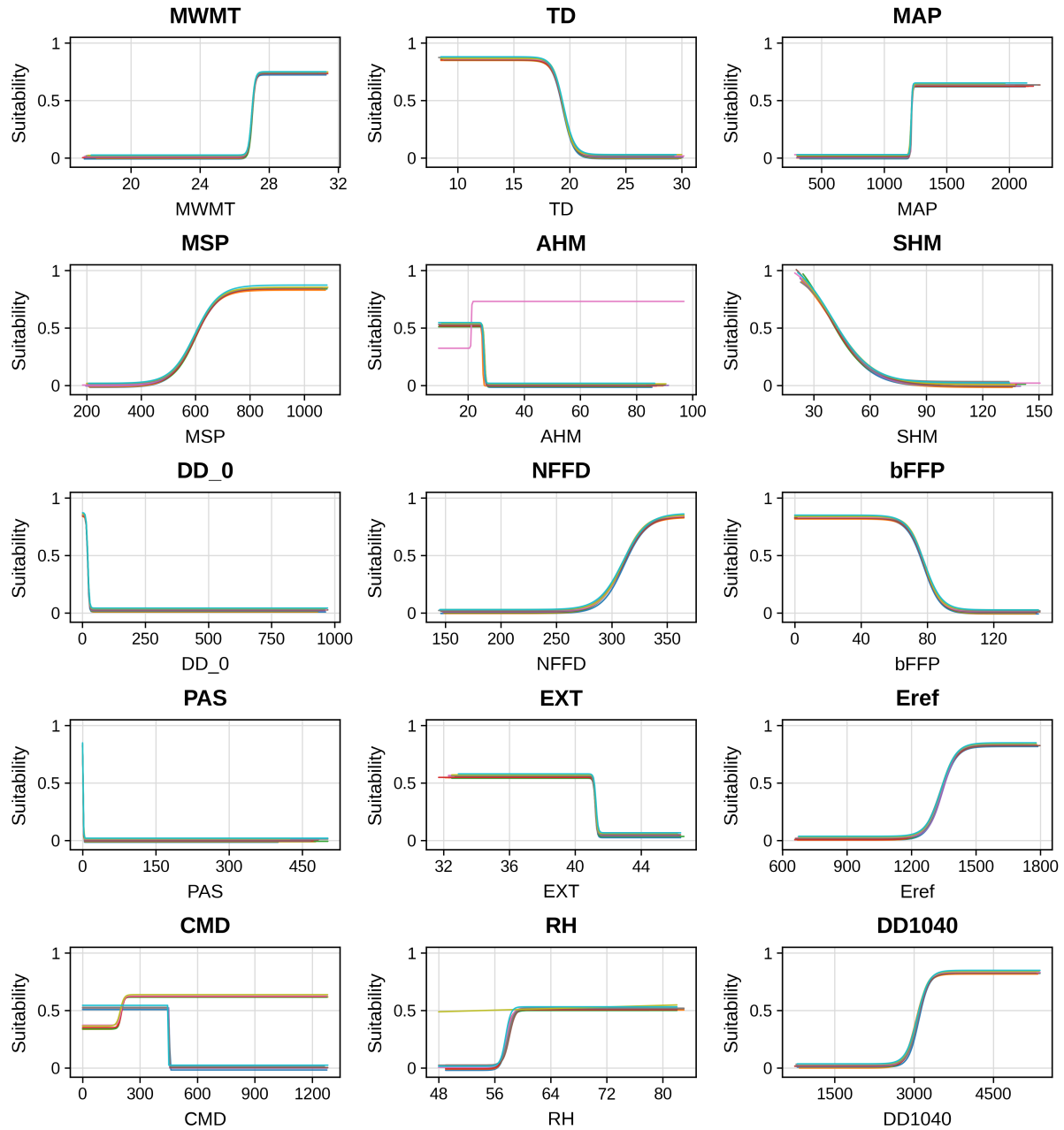


Each line represents one RF run (n = 10 runs).

Figure A.6: Random Forest response curves for retained climate predictors across 10 model runs under the 1991–2020 climate baseline for shortleaf pine.

RF response curves across 10 runs: Slash pine (1991–2020)

All retained predictors

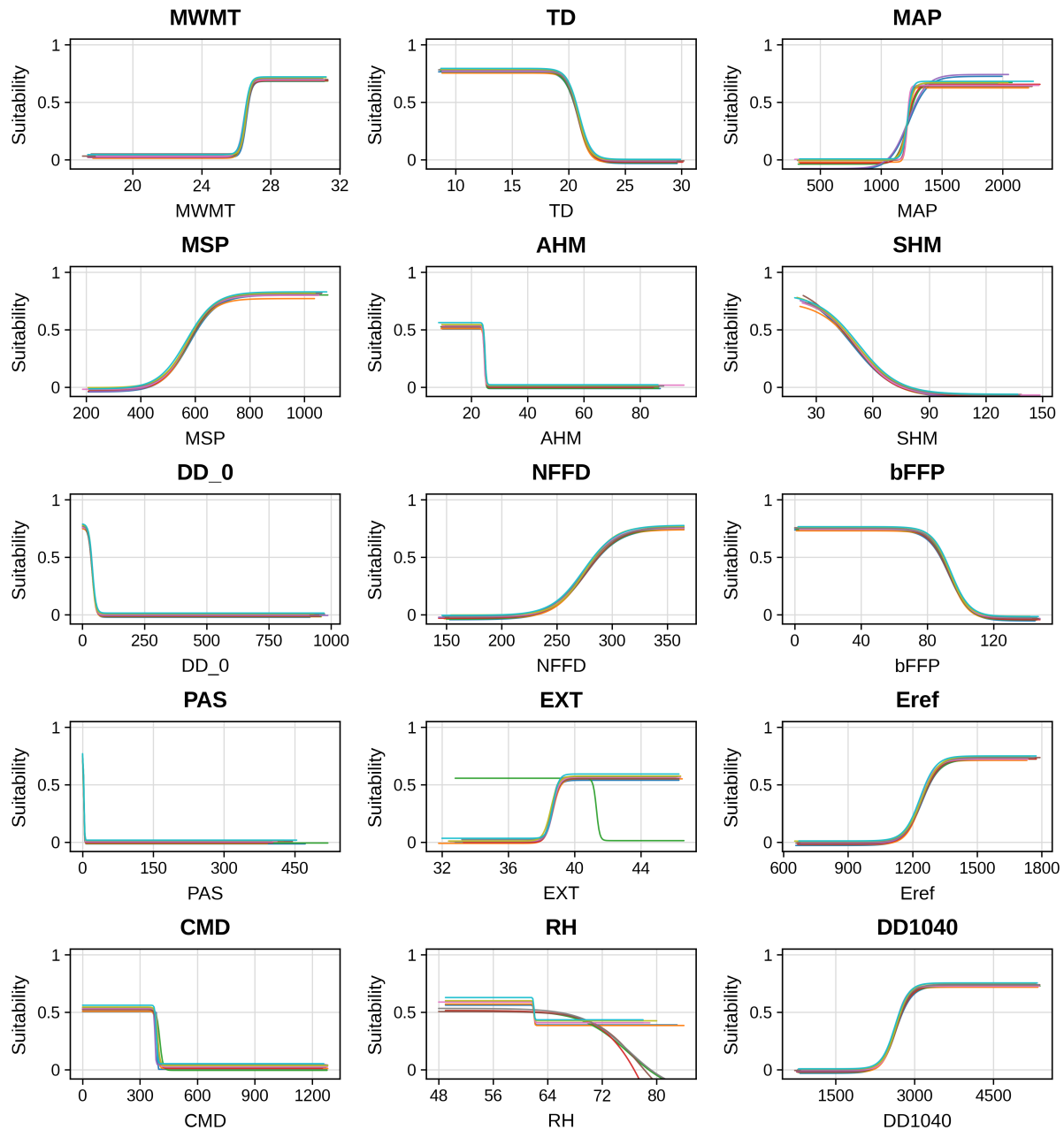


Each line represents one RF run (n = 10 runs).

Figure A.7: Random Forest response curves for retained climate predictors across 10 model runs under the 1991–2020 climate baseline for slash pine.

RF response curves across 10 runs: Longleaf pine (1991–2020)

All retained predictors



Each line represents one RF run (n = 10 runs).

Figure A.8: Random Forest response curves for retained climate predictors across 10 model runs under the 1991–2020 climate baseline for longleaf pine.

Detailed description for Figures A.5–A.8. Random Forest response curves for all retained climate predictors across 10 model runs under the 1991-2020 climate baseline. Response curves show modeled climatic suitability for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*). Each colored line represents one Random Forest run, with 10 runs shown for each predictor. The y-axis shows predicted suitability scaled from 0 to 1, and the x-axis shows the original value of each climate predictor. Predictors include variables representing thermal energy, temperature seasonality, precipitation, heat-moisture balance, cold limitation, growing-season length, precipitation as snow, evaporative demand, climatic moisture deficit, humidity, and degree-day accumulation. The close overlap among curves indicates that response shapes were generally stable across repeated Random Forest runs.

Average Feature Importances (MDI) of Loblolly Pine for 1991–2020

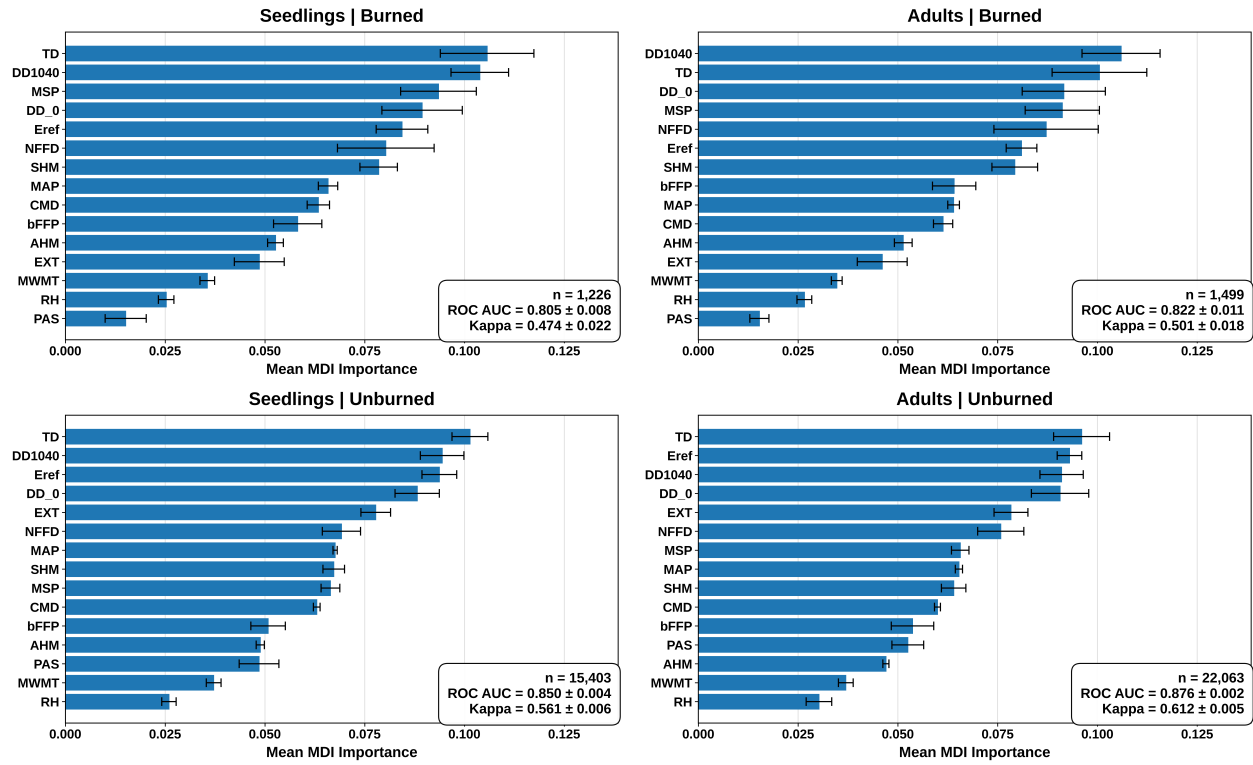


Figure A.9: Wildfire-stratified Random Forest MDI feature-importance rankings by life stage and wildfire-intersection status under the 1991–2020 climate baseline for loblolly pine.

Average Feature Importances (MDI) of Shortleaf Pine for 1991–2020

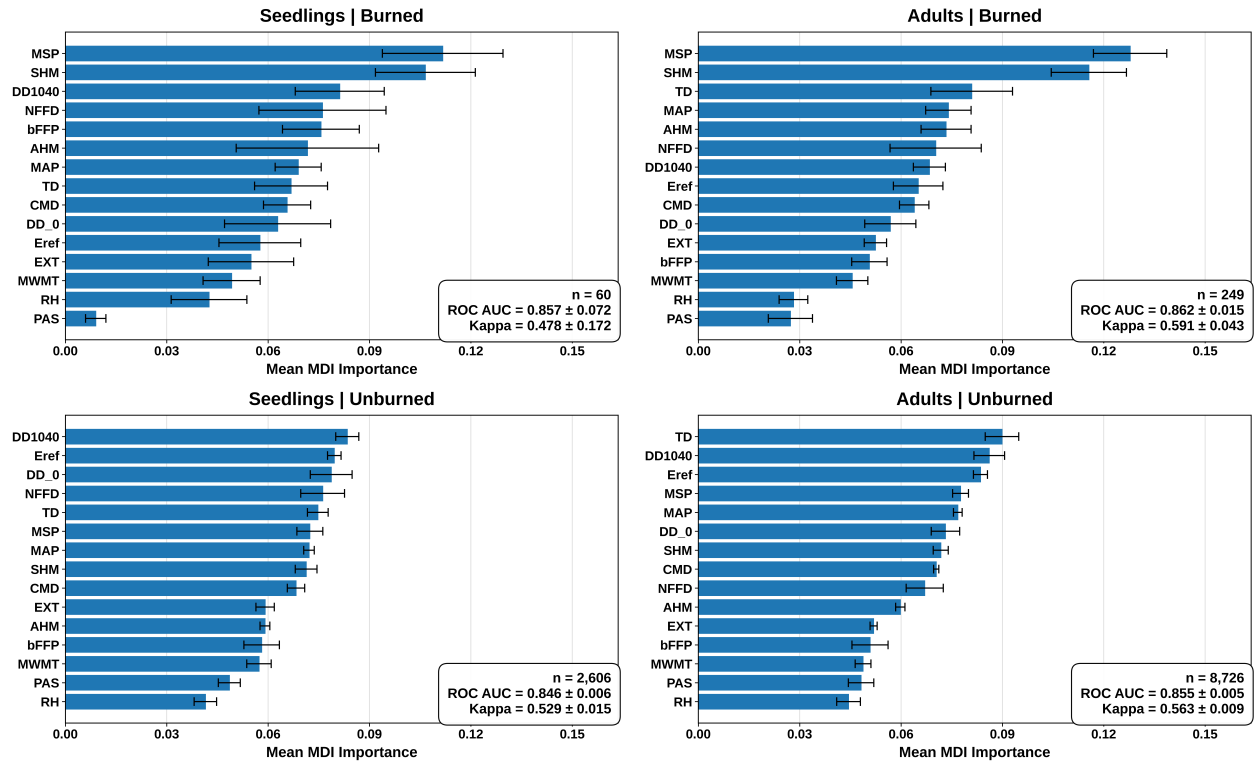


Figure A.10: Wildfire-stratified Random Forest MDI feature-importance rankings by life stage and wildfire-intersection status under the 1991–2020 climate baseline for shortleaf pine.

Average Feature Importances (MDI) of Slash Pine for 1991–2020

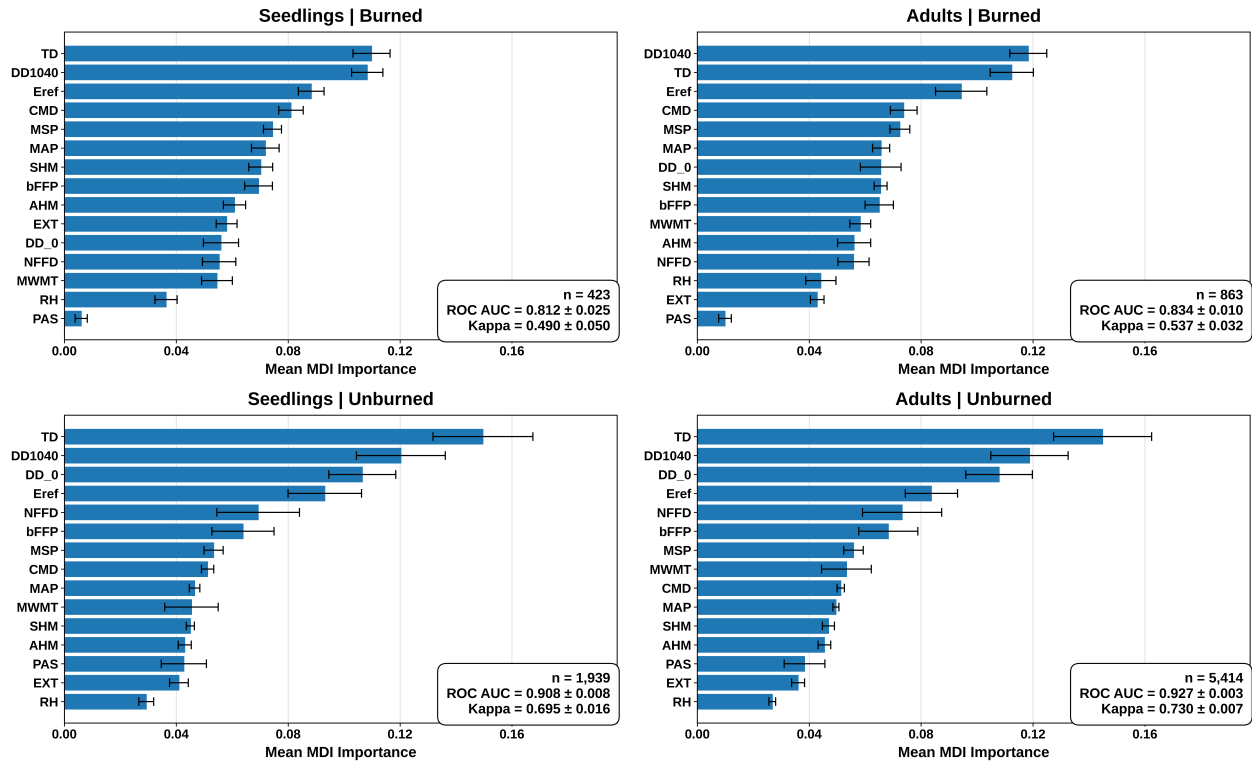


Figure A.11: Wildfire-stratified Random Forest MDI feature-importance rankings by life stage and wildfire-intersection status under the 1991–2020 climate baseline for slash pine.

Average Feature Importances (MDI) of Longleaf Pine for 1991–2020

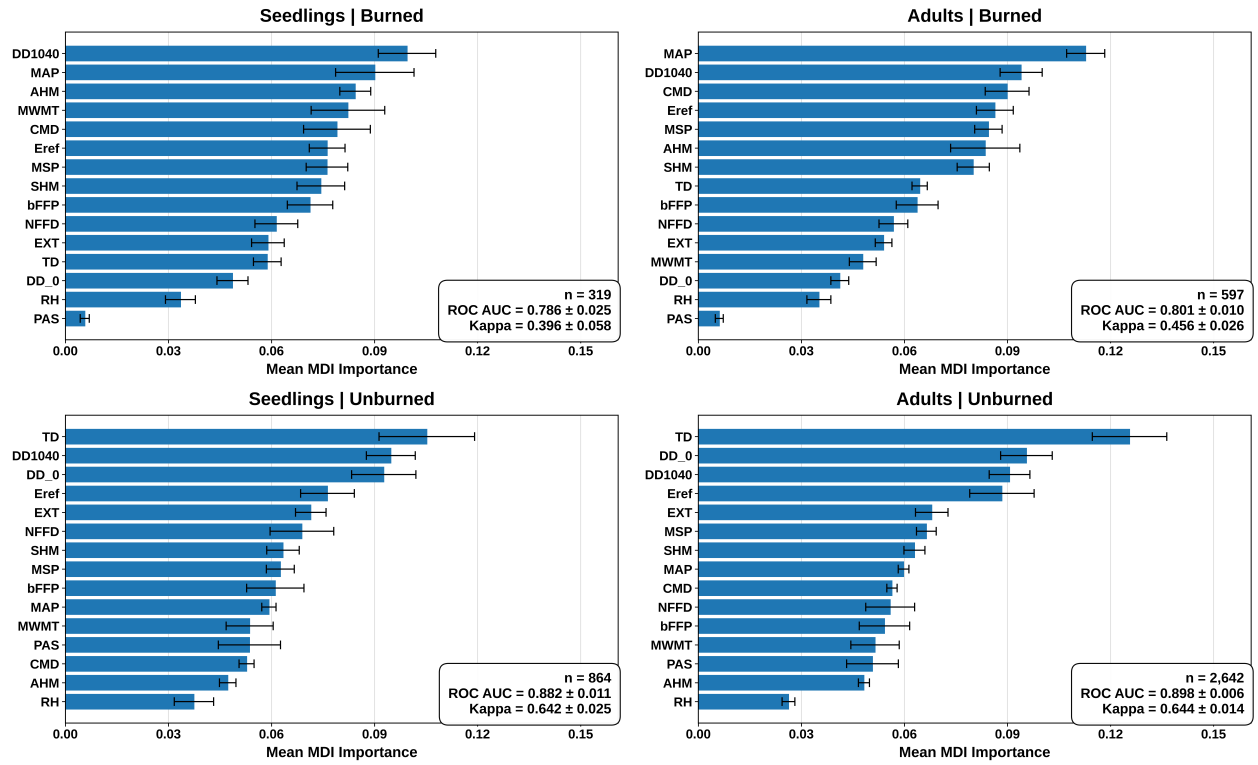


Figure A.12: Wildfire-stratified Random Forest MDI feature-importance rankings by life stage and wildfire-intersection status under the 1991–2020 climate baseline for longleaf pine.

Detailed description for Figures A.9–A.12. Wildfire-stratified Random Forest MDI feature-importance rankings by species, life stage, and wildfire-intersection status under the 1991-2020 climate baseline. Mean decrease in impurity (MDI) feature-importance values are shown for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*). Within each species, Random Forest models were stratified into seedling wildfire-intersected, adult-tree wildfire-intersected, seedling non-wildfire-intersected, and adult-tree non-wildfire-intersected groups. Bars show mean MDI importance for retained climate predictors, and horizontal error bars indicate among-run variation. Predictors are ranked within each panel from highest to lowest mean importance. Inset boxes report sample size, model accuracy, and Cohen's Kappa for each stratified model. Across species, MDI rankings showed that wildfire-intersected seedling and adult-tree models often emphasized moisture, seasonal precipitation, evaporative demand, and heat-moisture variables, including MSP, MAP, Eref, SHM, AHM, and CMD. Species-level patterns differed: loblolly pine emphasized MSP, Eref, SHM, MAP, and CMD in wildfire-intersected groups; shortleaf pine showed strong importance of MSP, SHM, MAP, AHM, and CMD in wildfire-intersected groups; slash pine wildfire-intersected groups emphasized Eref, CMD, MAP, MSP, SHM, EMT, and DD_0; and longleaf pine wildfire-intersected groups emphasized MAP, CMD, MSP, Eref, SHM, and AHM. Non-wildfire-intersected groups generally placed greater importance on Eref, DD_0, EMT, NFFD, MSP, CMD, and related thermal or frost-timing predictors. These patterns indicate that the relative importance of moisture, thermal, cold-limitation, and growing-season filters differed among species, life stages, and wildfire-intersection groups. MDI results are interpreted as disturbance-conditioned realized climate associations rather than direct causal effects of wildfire or individual climate predictors.

Average SHAP of Loblolly Pine for 1991–2020

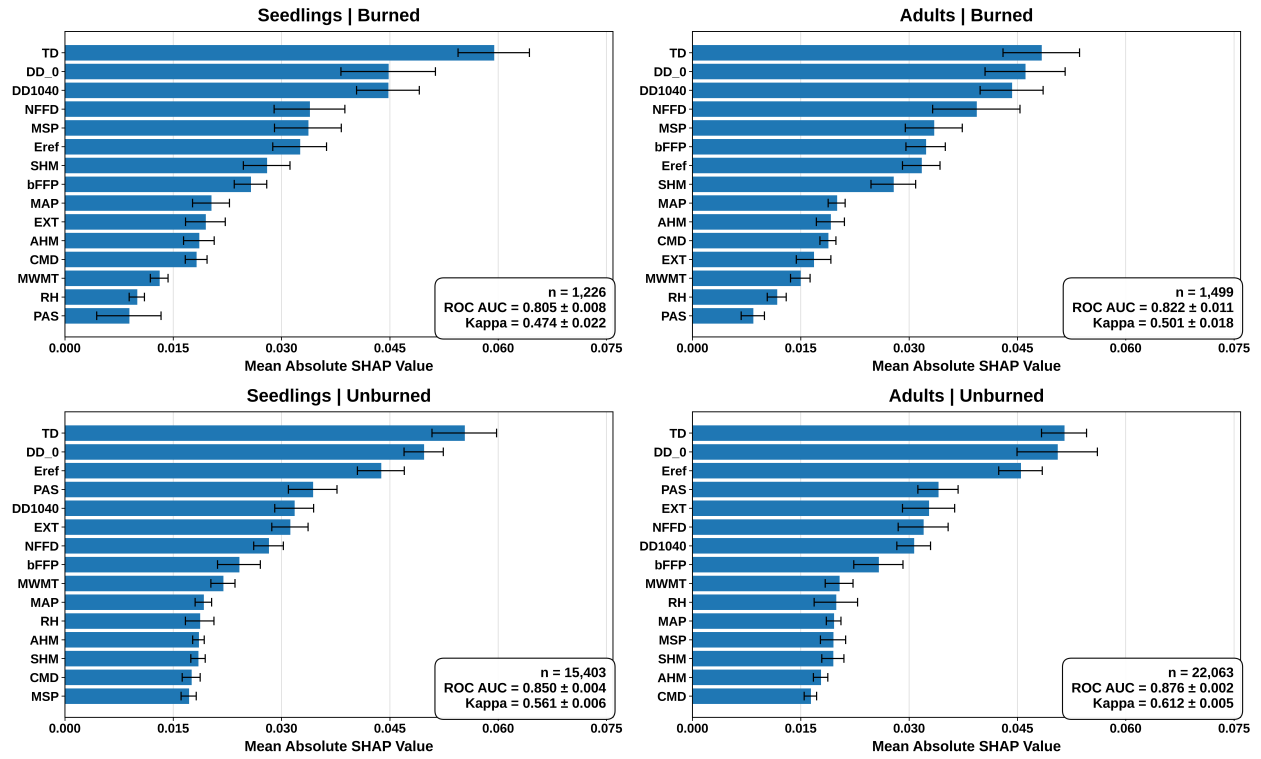


Figure A.13: SHAP predictor importance by species, life stage, and wildfire-intersection status under the 1991–2020 climate baseline for loblolly pine.

Average SHAP of Shortleaf Pine for 1991–2020

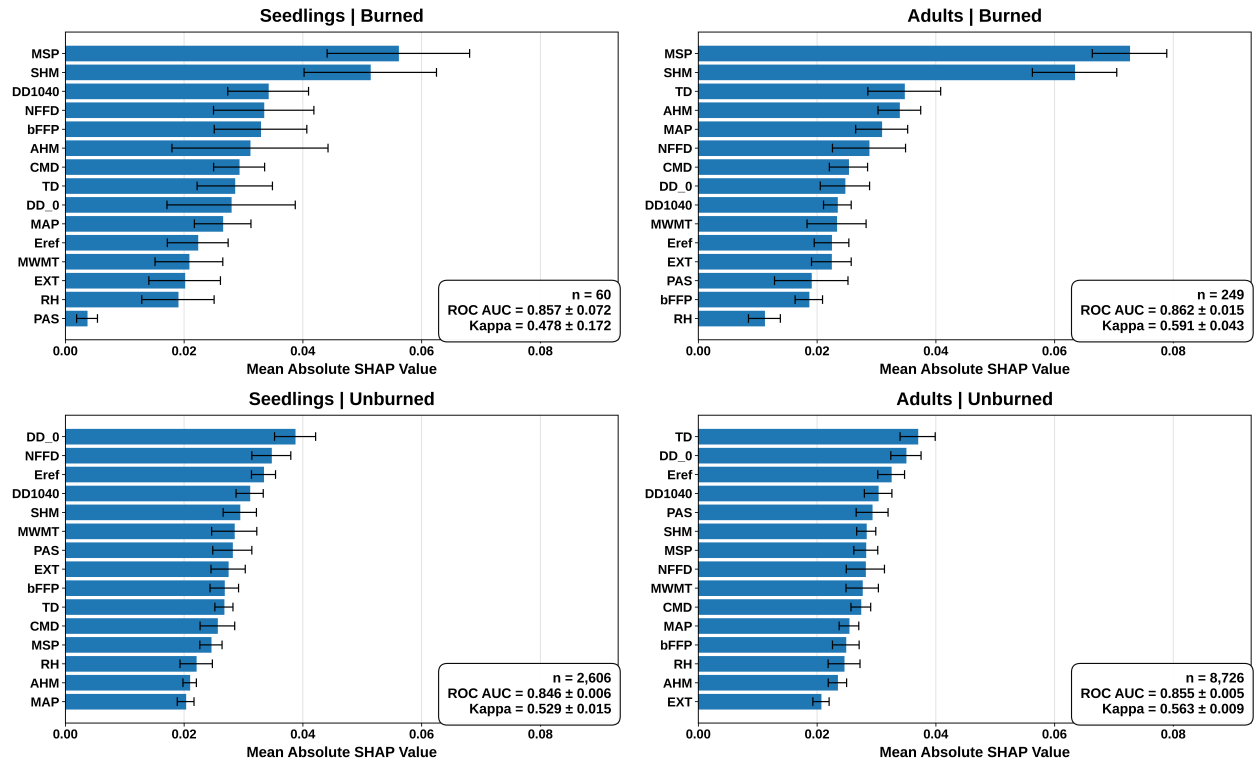


Figure A.14: SHAP predictor importance by species, life stage, and wildfire-intersection status under the 1991–2020 climate baseline for shortleaf pine.

Average SHAP of Slash Pine for 1991–2020

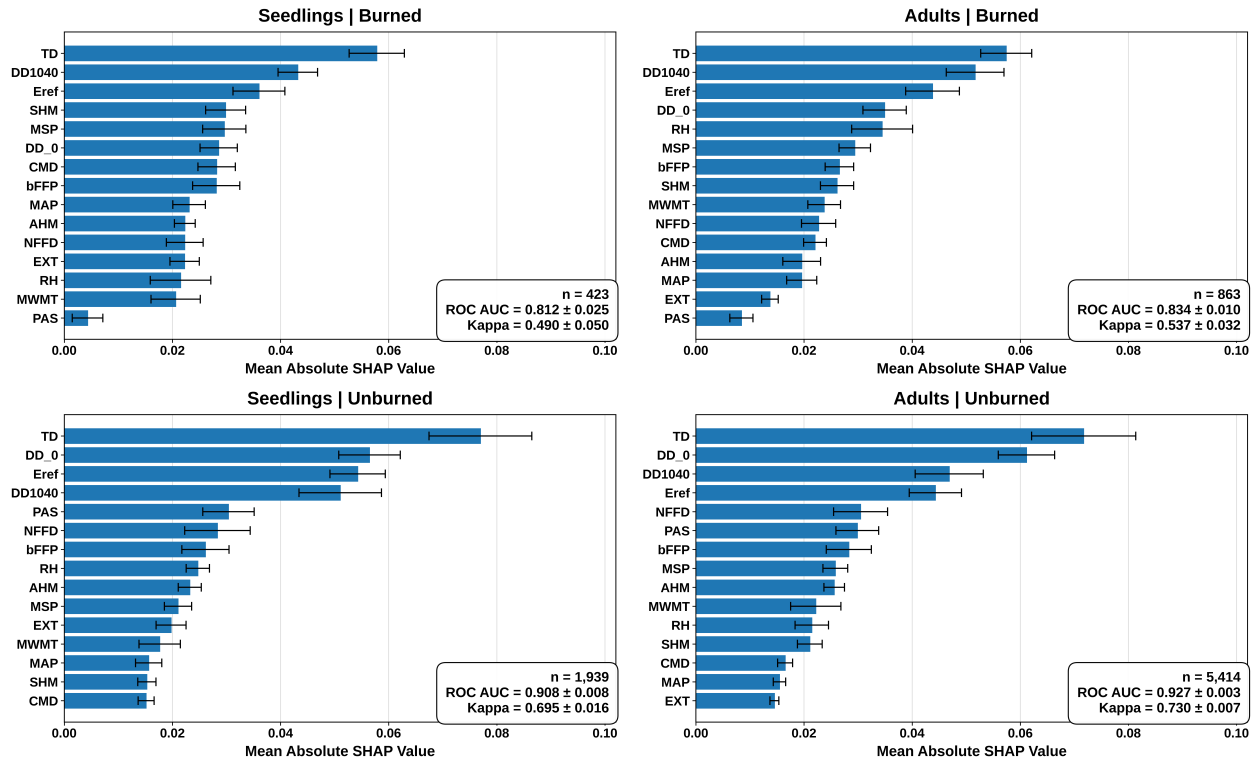


Figure A.15: SHAP predictor importance by species, life stage, and wildfire-intersection status under the 1991–2020 climate baseline for slash pine.

Average SHAP of Longleaf Pine for 1991–2020

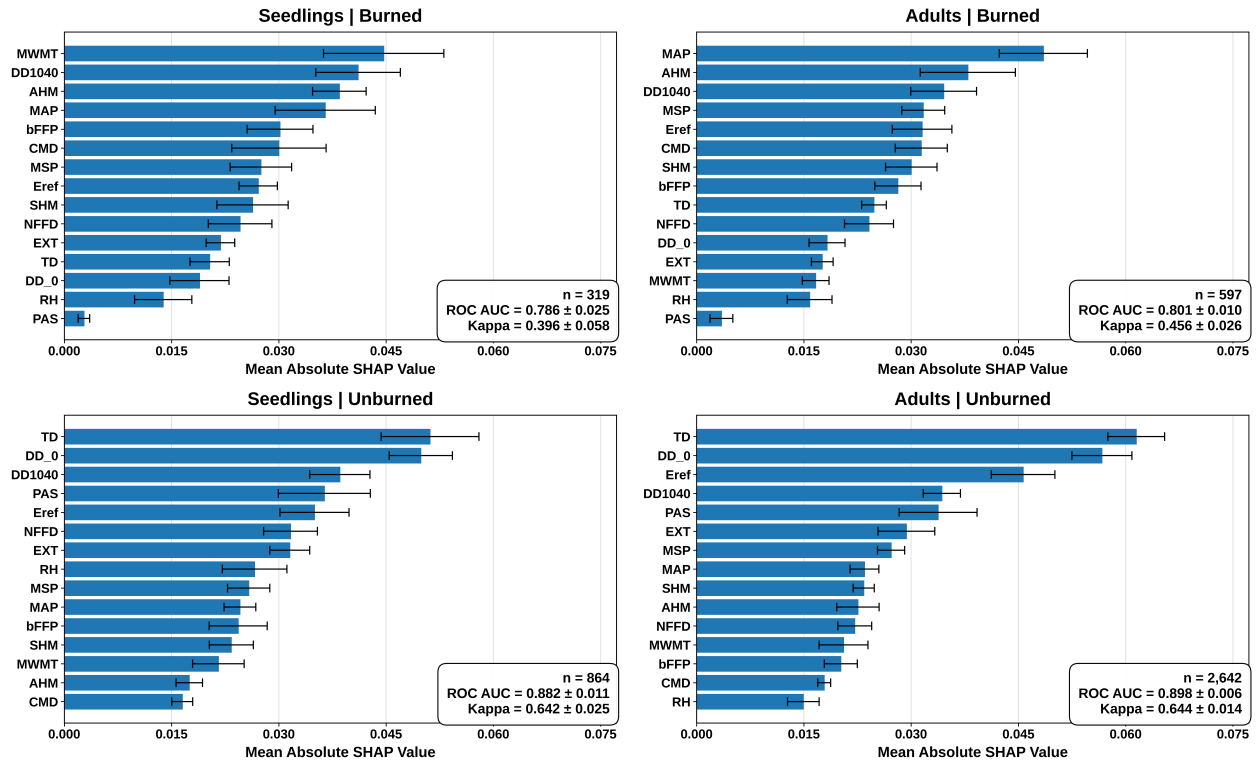


Figure A.16: SHAP predictor importance by species, life stage, and wildfire-intersection status under the 1991–2020 climate baseline for longleaf pine.

Detailed description for Figures A.13–A.16. SHAP predictor importance by species, life stage, and wildfire-intersection status under the 1991-2020 climate baseline. Horizontal bars show average Random Forest SHAP feature importance for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*). Within each species, models were stratified into seedling wildfire-intersected, adult-tree wildfire-intersected, seedling non-wildfire-intersected, and adult-tree non-wildfire-intersected groups. Error bars show among-run variation, and inset boxes report sample size, accuracy, and Kappa. Across species, cold limitation, evaporative demand, seasonal precipitation, frost-free-period variables, heat-moisture predictors, and moisture-availability variables were important, but rankings differed among species, life stages, and wildfire-intersection groups. Loblolly pine showed strong contributions from DD_0, MSP, Eref, NFFD, EMT, and PAS across several strata. Shortleaf pine wildfire-intersected groups emphasized moisture and heat-moisture variables such as MSP, SHM, MAP, AHM, and CMD, whereas non-wildfire-intersected groups also emphasized Eref, DD_0, EMT/EXT, PAS, and RH. Slash pine showed high importance of DD_0 and Eref, with additional contributions from MSP, CMD, SHM, NFFD, PAS, and frost-free-period variables. Longleaf pine wildfire-intersected groups emphasized MAP, AHM, CMD, MSP, and SHM, whereas non-wildfire-intersected groups showed stronger contributions from DD_0, Eref, EMT, PAS, NFFD, and related thermal or frost-timing predictors. Results are interpreted as disturbance-conditioned realized climate associations rather than direct causal effects of wildfire or individual climate predictors.

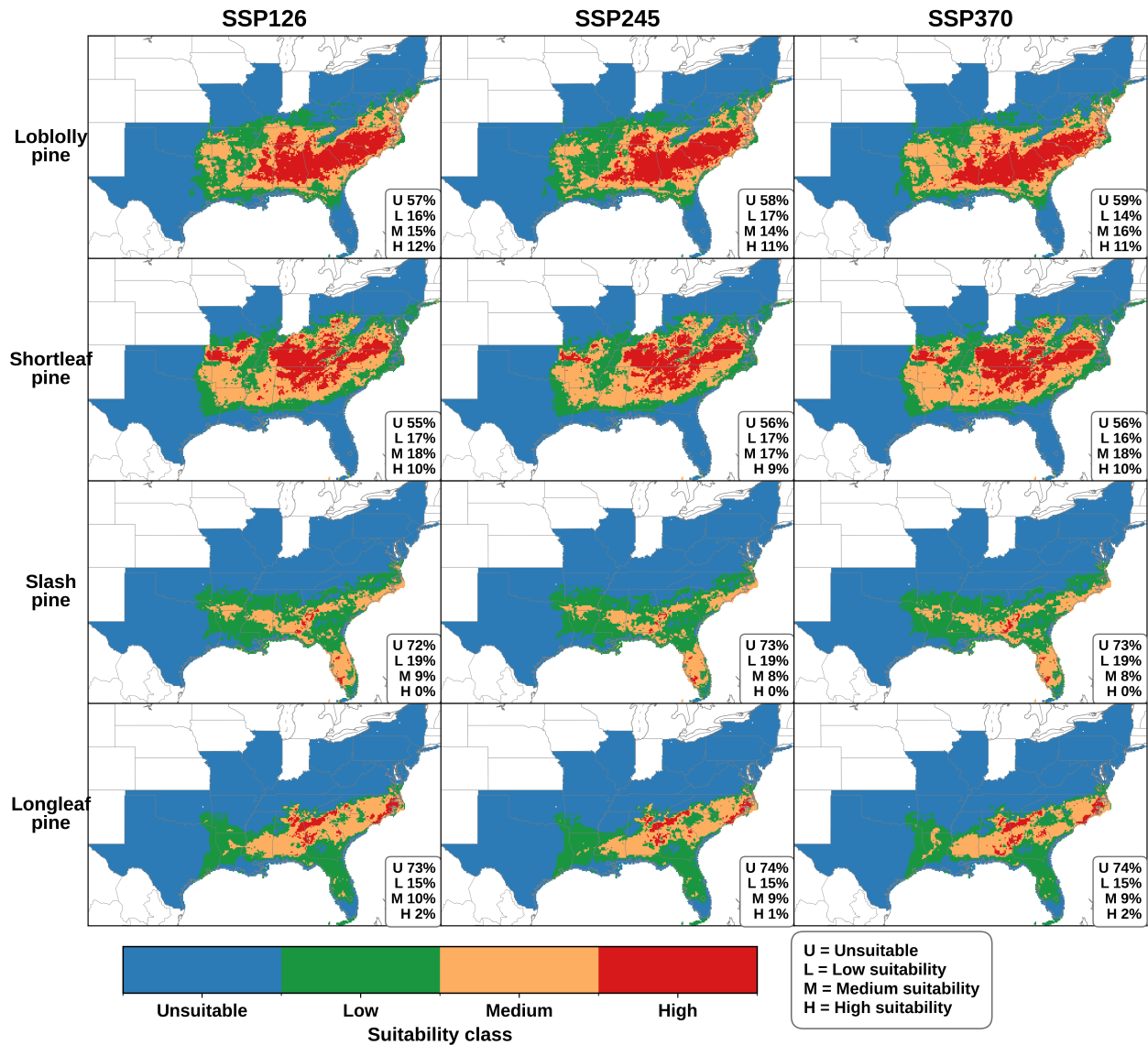


Figure A.17: Early-century future climatic suitability projections for southern yellow pines using the 1961-1990 baseline model and an eight-GCM climate ensemble.

Detailed description for Figure A.17. Early-century future climatic suitability projections for southern yellow pines using the 1961-1990 baseline model and an eight-GCM climate ensemble. Maps show projected climatic suitability for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*) for the 2011-2040 future period. Species are arranged by row, and emissions scenarios are arranged by column: SSP126, SSP245, and SSP370. Species distribution models calibrated under the 1961-1990 historical climate baseline were projected to future climate conditions summarized from eight general circulation models. Suitability predictions were classified into four categories: unsuitable, low, medium, and high suitability. Inset values within each map report the percentage of mapped cells in each class. Across the early-century projections, species identity produced stronger spatial differences than SSP scenario: loblolly and shortleaf pine retained broad areas of medium-to-high modeled suitability across the interior Southeast, whereas slash and longleaf pine showed more restricted high-suitability patterns concentrated in the warmer Gulf and Atlantic Coastal Plain portions of the study domain. These maps provide a scenario-level visualization of projected realized climatic suitability and should be interpreted as climate-associated suitability patterns rather than direct forecasts of establishment success, timber supply, dispersal, planting feasibility, or physiological tolerance.

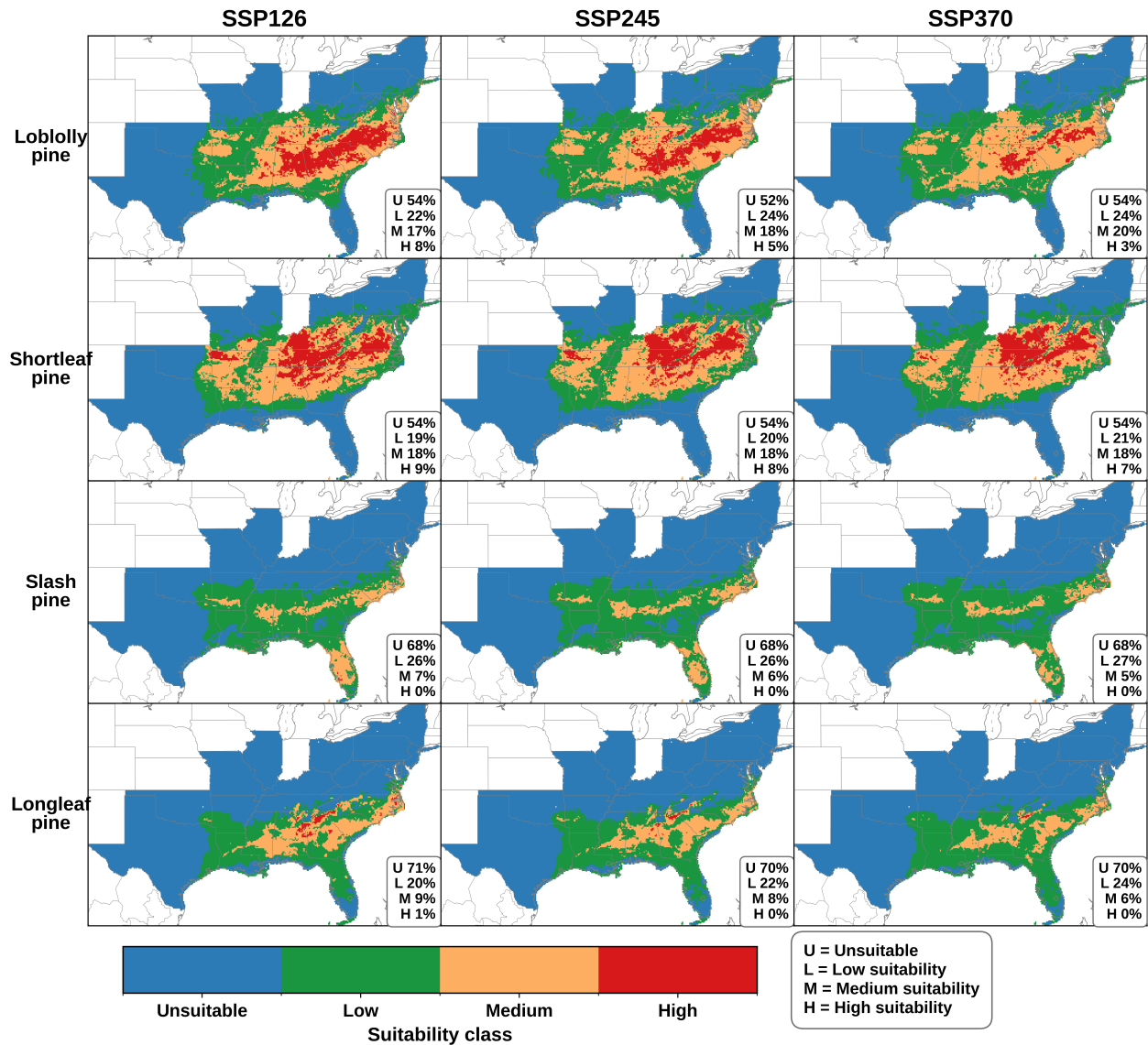


Figure A.18: Mid-century future climatic suitability projections for southern yellow pines using the 1961-1990 baseline model and eight-GCM climate projections.

Detailed description for Figure A.18. Mid-century future climatic suitability projections for southern yellow pines using the 1961-1990 baseline model and eight-GCM climate projections. Maps show projected climatic suitability for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*) for the 2041-2070 future period. Species are arranged by row, and emissions scenarios are arranged by column: SSP126, SSP245, and SSP370. Species distribution models calibrated under the 1961-1990 historical climate baseline were projected to mid-century climate conditions summarized across eight general circulation models. Predicted suitability was classified into four categories: unsuitable, low, medium, and high. Inset values within each map report the percentage of mapped cells in each class. Across mid-century projections, loblolly and shortleaf pine retained broad areas of medium-to-high modeled suitability across the interior Southeast, although high-suitability area declined under higher-emissions scenarios. Slash pine and longleaf pine showed more restricted medium-suitability areas, mainly concentrated in the Gulf and Atlantic Coastal Plain, with little to no high-suitability area by 2041-2070. These maps provide scenario-specific visualizations of projected realized climatic suitability and should be interpreted as climate-associated suitability patterns rather than direct forecasts of establishment success, regeneration, migration, timber supply, planting feasibility, or physiological tolerance.

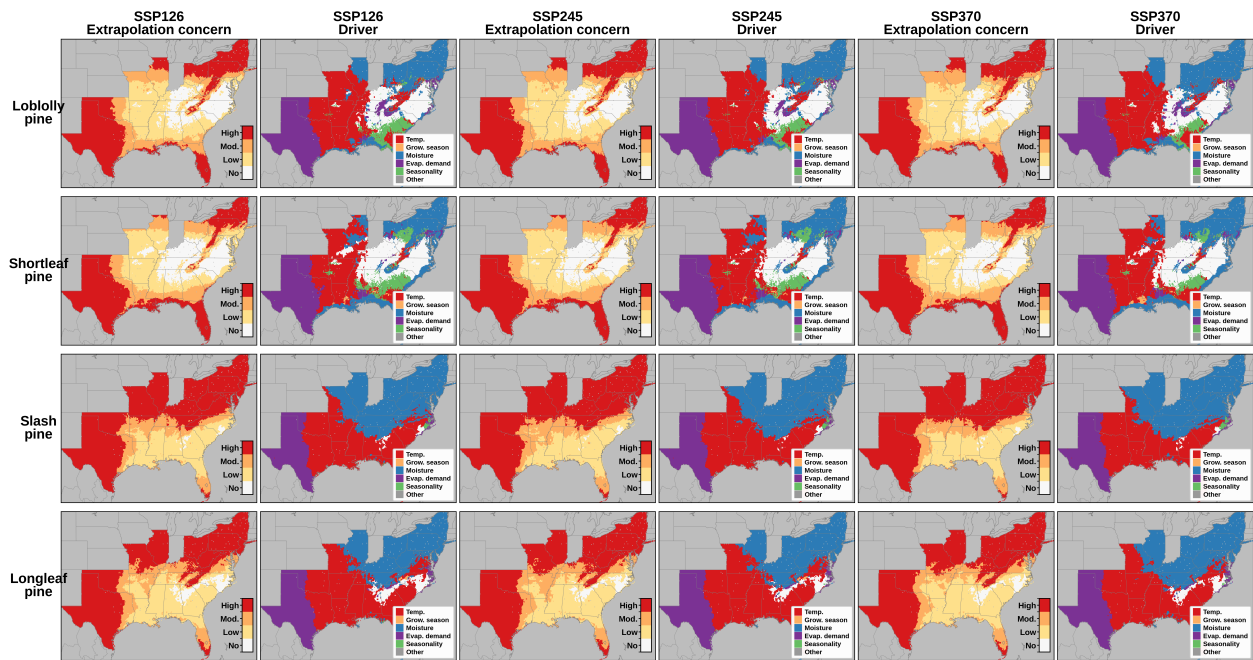


Figure A.19: Non-analog climate uncertainty for early-century future projections using the 1961-1990 baseline model and eight-GCM climate projections.

Detailed description for Figure A.19. Non-analog climate uncertainty for early-century future projections using the 1961-1990 baseline model and eight-GCM climate projections. Maps show non-analog climate uncertainty and the dominant climate-driver group responsible for extrapolation concern when 1961-1990 models are projected to 2011-2040 future climate under SSP126, SSP245, and SSP370. For each species-scenario combination, the uncertainty panel classifies future grid cells by extrapolation concern, and the driver panel identifies the climate-filter group most responsible for the non-analog signal. These outputs identify areas where future suitability predictions should be interpreted with greater caution because projected climate conditions are near or outside the historical calibration domain. Non-analog uncertainty is not a suitability class; rather, it provides a complementary screening layer for evaluating where future SDM projections are more extrapolative and less directly supported by historical training data.

A.2 Detailed captions for Chapter 3 main figures

The following detailed captions correspond to the Chapter 3 main figures. Detailed notes for Appendix A supplementary figures are placed immediately after the corresponding supplementary figures in Section A.1.

Figure 3.1. Spatial distributions of adult-tree and seedling occurrence records for four southern yellow pine species in the southeastern United States. Adult-tree occurrence records are shown in the left column and seedling occurrence records are shown in the right column. Panels show distributions for loblolly pine (*Pinus taeda*; A, E), slash pine (*P. elliottii*; B, F), shortleaf pine (*P. echinata*; C, G), and longleaf pine (*P. palustris*; D, H). Each point represents a retained occurrence record used in the analysis, with multiple records within the same spatial unit collapsed to reduce pseudoreplication. Sample sizes are shown above each map. Adult-tree records represent established populations and legacy persistence, whereas seedling records provide a more contemporary signal of regeneration and early establishment. Across species, adult and seedling records show broad spatial overlap but differ in local density and spatial continuity, supporting their treatment as separate ecological data streams in the climate-regeneration modeling framework. **Fig-**

ure 3.2. Adult-seedling climate-space separation by wildfire-intersection status for four southern yellow pine species during the 1991-2020 climate baseline. Occurrence records for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*) were projected into a shared principal component analysis climate space using retained climate predictors from the 1991-2020 baseline. Panels compare records without mapped wildfire intersection and with mapped wildfire intersection for each species. Points show occurrence records in climate space, with colors indicating the plotted occurrence categories. Black arrows connect the adult-tree centroid to the seedling

centroid, representing the direction and magnitude of adult-seedling climate-space displacement. Boxed annotations report adult-tree and seedling sample sizes, centroid distance, 95% confidence interval, and associated p value for each species-fire-status group. PC1 and PC2 explain 65.3% and 19.1% of total climate-space variation, respectively; the inset biplot shows the correlation of selected retained climate variables with the shared PCA axes. This visualization evaluates whether seedling recruitment occupies the same realized climate space as established trees and whether wildfire-intersected landscapes modify adult-seedling climate associations. Because wildfire intersection does not describe the full fire regime, including prescribed fire, fire frequency, severity, seasonality, or time since fire, these patterns should be interpreted as disturbance-conditioned climate associations rather than direct causal effects of fire. **Figure 3.3.** Adult versus seedling modeled climate suitability for four southern yellow pine species under the 1991-2020 climate baseline. Maps compare gridded modeled climatic suitability for adult-tree and seedling occurrence data streams for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*) at 0.1° spatial resolution. Rows represent species, and columns show adult suitability, seedling suitability, and binary adult-seedling agreement or disagreement. Continuous suitability maps are scaled from 0 to 1, with higher values indicating greater modeled realized climatic suitability. Adult suitability represents the climate signal associated with established populations and legacy persistence, whereas seedling suitability represents the climate signal associated with recent regeneration and early establishment. Agreement maps classify grid cells as suitable for both adult trees and seedlings, adult-only suitable, seedling-only suitable, or unsuitable for both groups after applying the selected suitability threshold. Areas suitable for both adult trees and seedlings indicate seedling-supported suitability and may represent stronger candidates for persistence and regeneration-supported resilience. Adult-only suitable areas indicate places where established trees may persist but modeled seedling suitability is

weaker, which may flag areas needing regeneration diagnosis rather than proving regeneration failure. Seedling-only suitable areas indicate locations where modeled regeneration suitability occurs outside the strongest adult-suitability footprint. These maps support the interpretation that adult persistence and seedling recruitment are related but not interchangeable, but all management interpretation requires independent survival, sapling recruitment, and site-level validation. **Figure 3.4.** Average future climatic suitability and suitability change for four southern yellow pine species relative to the 1961-1990 baseline model. Top panels show average projected future climatic suitability for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*) across the future climate projections represented in the analysis. Suitability values were derived by applying species distribution models calibrated under the 1961-1990 historical baseline to future climate layers and summarizing projected suitability across the selected GCM, SSP, and future-period combinations. Projection classes indicate unsuitable, low, medium, and high modeled suitability. Bottom panels show binary suitability change relative to the 1961-1990 baseline model after applying the selected suitability threshold. Stable suitable areas are grid cells classified as suitable under both baseline and future climate; loss areas are cells suitable under the baseline but unsuitable under future climate; gain areas are cells unsuitable under the baseline but suitable under future climate; and unsuitable-both areas are unsuitable under both baseline and future climate. These maps translate continuous SDM projections into management-relevant spatial categories for identifying potential persistence zones, projected suitability losses, and possible future-gain areas for southern pine adaptation planning. Because occurrence-based SDMs are interpreted as realized climatic suitability, these projections should be read as modeled climate associations rather than direct estimates of physiological tolerance, dispersal, planting feasibility, or regeneration success.

Figure 3.5. Sample size and wildfire stratification of adult-tree and seedling occurrence records for four southern yellow pine species during the 1991-2020 baseline period.

Occurrence records for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliotii*), and longleaf pine (*P. palustris*) were separated by life stage and wildfire-intersection status. Adult-tree records represent established populations and legacy persistence, whereas seedling records represent recent regeneration and early establishment. Records were classified as wildfire-intersected or non-wildfire-intersected by overlaying occurrence locations with mapped wildfire layers. Panel A shows the absolute number of retained presence records in each species-life-stage-fire-status group on a log-scaled y-axis. Panel B shows the percentage composition of the same four groups within each species. Colors distinguish adult wildfire-intersected, adult non-wildfire-intersected, seedling wildfire-intersected, and seedling non-wildfire-intersected records. Across species, adult non-wildfire-intersected records formed the largest group, whereas seedling wildfire-intersected records were consistently the smallest or among the smallest groups. This sample imbalance should be considered when interpreting wildfire-conditioned models, especially for seedling records. Because mapped wildfire intersection does not capture prescribed fire, fire frequency, severity, seasonality, time since fire, or complete fire history, wildfire-stratified results are interpreted as disturbance-associated patterns rather than direct causal effects of fire. **Figure 3.6.** Multi-algorithm

comparison of species distribution model performance across southern yellow pine species and historical climate baselines. Violin plots show validation ROC AUC values for seven species distribution modeling algorithms - support vector machine (SVM), generalized linear model (GLM), gradient boosting machine (GBM), Random Forest (RF), naive Bayes model (NBM), extreme gradient boosting (XGB), and MaxEnt - calibrated separately for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliotii*), and longleaf pine (*P. palustris*). Columns represent species, and rows represent the four 30-year historical climate baselines used to evaluate calibration-window sensitivity: 1961-1990, 1971-2000, 1981-2010, and 1991-2020. Higher ROC AUC values indicate stronger discrimination between presence and background or pseudoabsence data.

Across species and baselines, XGBoost and Random Forest showed the most consistently high performance, GBM also performed strongly for several species, SVM, GLM, and MaxEnt showed intermediate performance, and naive Bayes generally produced the lowest ROC AUC values. Slash pine showed the strongest overall climatic separability, whereas shortleaf pine showed lower but still useful discrimination, consistent with its broader and more heterogeneous realized climate space. The limited variation among baselines indicates that top-level discrimination was relatively stable across historical climate windows, although baseline choice affected climate-variable interpretation more strongly than ROC AUC performance. **Figure 3.7.** Relative importance of climate-filter groups in Random Forest species distribution models for four southern yellow pine species under the 1991-2020 climate baseline. Heatmaps summarize the relative contribution of grouped climate predictors to Random Forest models for loblolly pine (*Pinus taeda*), shortleaf pine (*P. echinata*), slash pine (*P. elliottii*), and longleaf pine (*P. palustris*). The upper panel shows SHAP-based climate-filter importance, and the lower panel shows mean decrease in impurity (MDI) importance. Individual retained predictors were grouped into eight biologically interpretable climate filters: thermal energy, cold limitation, temperature extremes, growing-season/frost timing, moisture availability, evaporative demand or water deficit, heat-moisture balance, and seasonality or continentality. Values represent relative group importance within each species and metric, with rows summing approximately to 1 after rounding; darker colors indicate greater relative importance. Across both SHAP and MDI summaries, moisture availability was the dominant climate-filter group for all four species and was especially important for shortleaf pine. Thermal energy, cold limitation, growing-season/frost timing, evaporative demand, and seasonality also contributed substantially, whereas temperature extremes had consistently low relative importance. These results support the interpretation that southern pine suitability is structured by multiple climate dimensions rather than by a single universal predictor. Because SHAP and MDI are model-interpretation metrics, their values should be interpreted as predictive

associations within the realized climate space, not as causal effects of individual climate variables. **Figure 3.8.** Random Forest response curves for selected climate predictors in southern yellow pine species distribution models under the 1991-2020 climate baseline. Panels show species-specific marginal response curves from Random Forest models for shortleaf pine (*Pinus echinata*), slash pine (*P. elliottii*), longleaf pine (*P. palustris*), and loblolly pine (*P. taeda*). The selected predictors represent major climate-filter groups identified in this chapter, including moisture availability, evaporative demand or water deficit, temperature seasonality, cold limitation, thermal energy, and growing-season length. The y-axis shows predicted climatic suitability, scaled from 0 to 1, and the x-axis shows the original values of each climate predictor. Black curves represent the fitted Random Forest response for each predictor, while the rug marks and gray density shading along the bottom of each panel show the distribution of predictor values in the model data. Scores shown in the upper-left corner of each panel indicate the model-derived importance of the plotted predictor. Shortleaf pine shows strong responses to moisture and water-deficit variables, particularly MAP and Eref, whereas slash, longleaf, and loblolly pine show pronounced threshold responses to temperature seasonality, cold limitation, thermal-energy, and growing-season predictors such as TD, DD_0, DD1040, and NFFD. These response curves support the interpretation that southern pine suitability is structured by species-specific combinations of moisture, thermal, and seasonal climate filters rather than by a single universal climate driver. Because Random Forest response curves describe predictive associations within the realized climate space, they should not be interpreted as direct causal or physiological tolerance thresholds. MAP, mean annual precipitation; Eref, reference evaporation or evaporative demand; TD, temperature difference/continentality; DD_0, degree-days below 0 degrees C; DD1040, thermal degree-day accumulation; NFFD, number of frost-free days; PAS, precipitation as snow. **Figure 3.9.** Relative importance of wildfire-stratified climate filters in Random Forest models for southern yellow pines. Heatmaps compare SHAP-based and MDI-based relative importance of

grouped climate predictors for loblolly, shortleaf, slash, and longleaf pine under the 1991-2020 baseline. For each species, Random Forest models were stratified into adult-tree wildfire-intersected, adult-tree non-wildfire-intersected, seedling wildfire-intersected, and seedling non-wildfire-intersected groups. Columns show climate-filter groups, and cell values show relative group importance; darker colors indicate stronger predictive contribution. Wildfire-stratified importance patterns differed most clearly in seedling models, indicating that fire-associated recruitment records can emphasize different combinations of moisture, heat-moisture, frost-timing, and thermal filters than non-wildfire-intersected records. Results are interpreted as disturbance-conditioned realized climate associations rather than causal fire-regime effects.

Appendix B
Supplementary Material for Pollen and Lung Cancer Modeling

This appendix contains the supplementary figures associated with Chapter 4.

B.1 Supplementary plots and figure notes

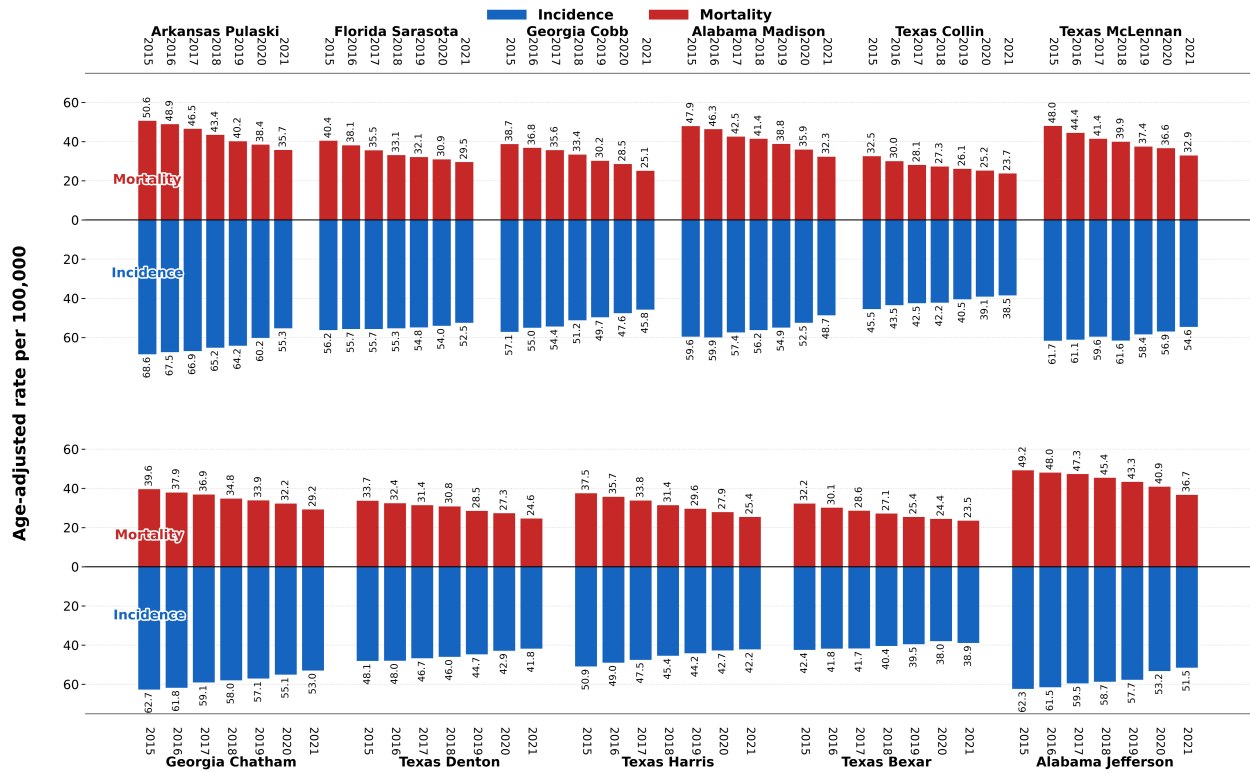


Figure B.1: Alternative outcome trend displays.

Detailed description for Figure B.1. Alternative outcome trend displays. Panels A and B show bar-based summaries of incidence and mortality by county and year, reinforcing the county baseline differences and general temporal decline shown in Figure 4.1.

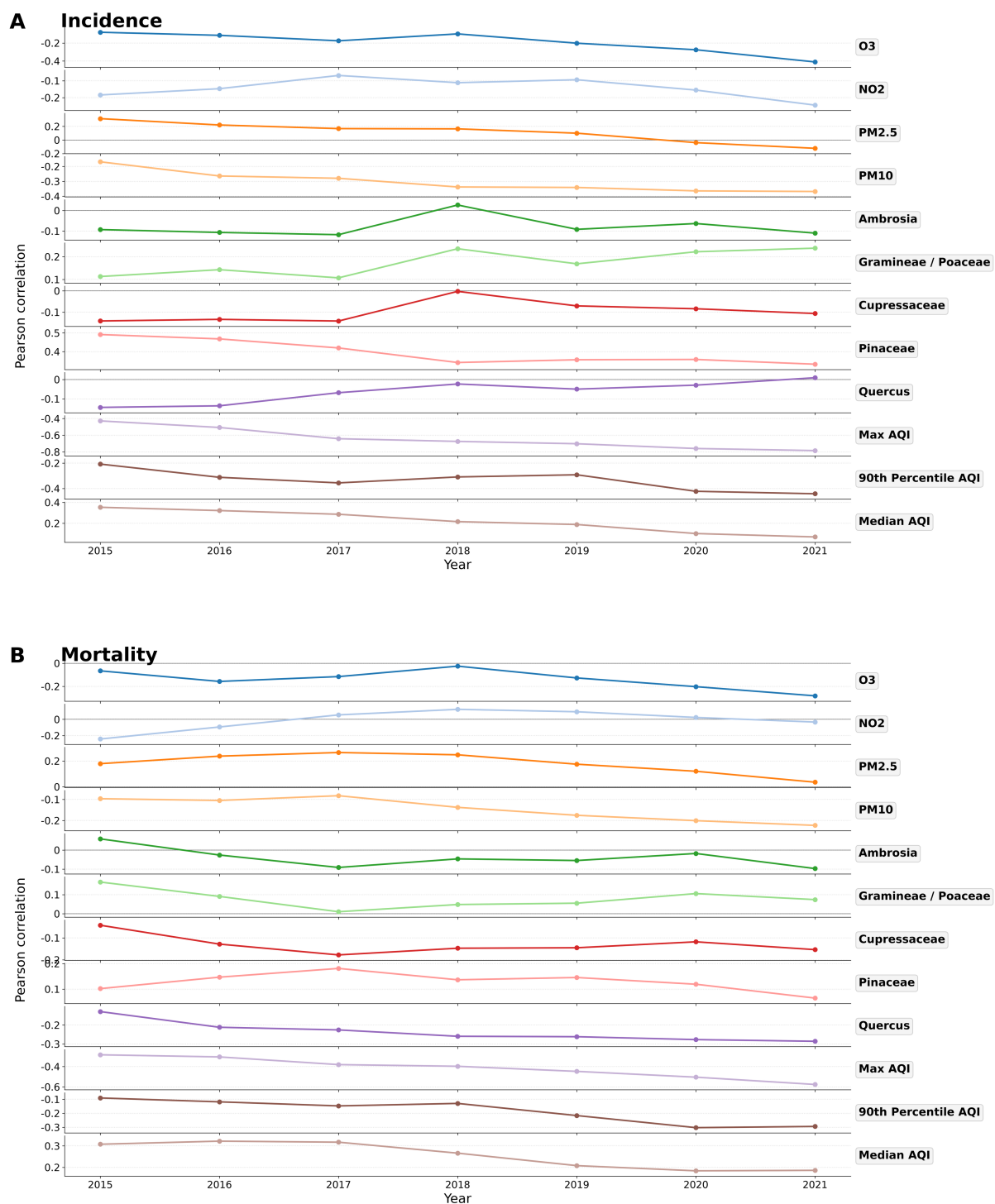


Figure B.2: Temporal correlations by outcome.

Detailed description for Figure B.2. Temporal correlations by outcome. Panels A and B show annual Pearson correlations between features and incidence or mortality. Relationships vary by year and endpoint, supporting the use of nonlinear predictive models and cautioning against simple static correlations.

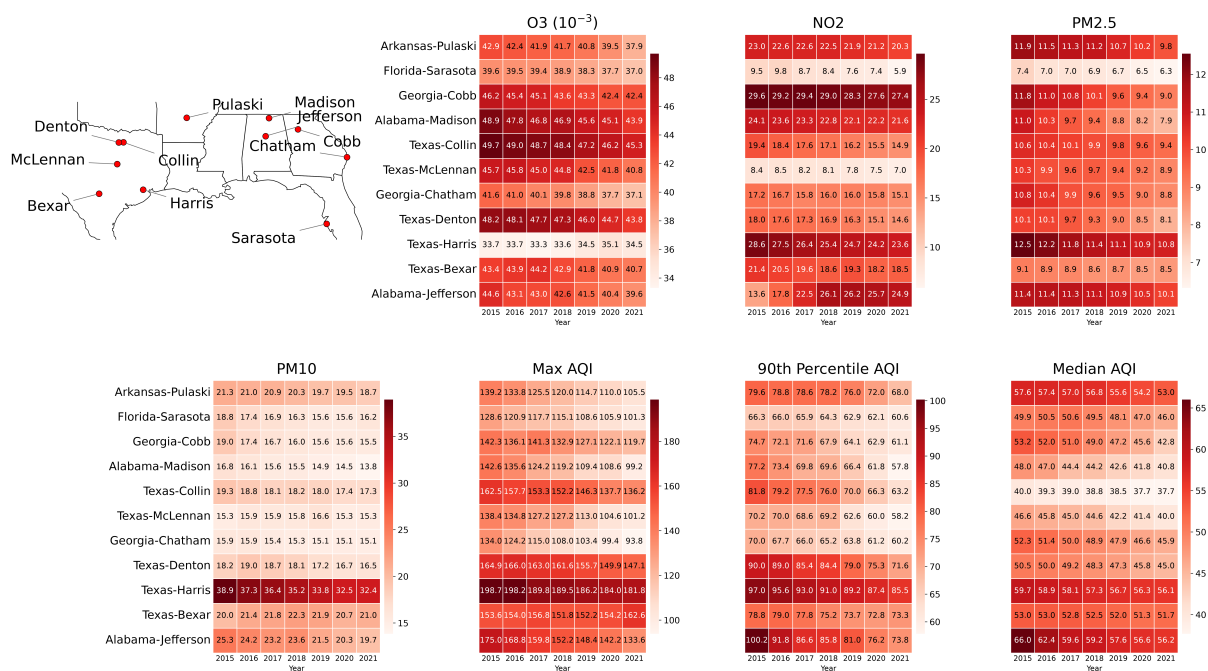


Figure B.3: Diagnostic environmental exposure heatmap.

Detailed description for Figure B.3. Diagnostic environmental exposure heatmap. This dense diagnostic view illustrates county-level environmental exposure profiles after lag-window construction. It is retained as supplementary material because it illustrates heterogeneity but is too dense for a primary result.

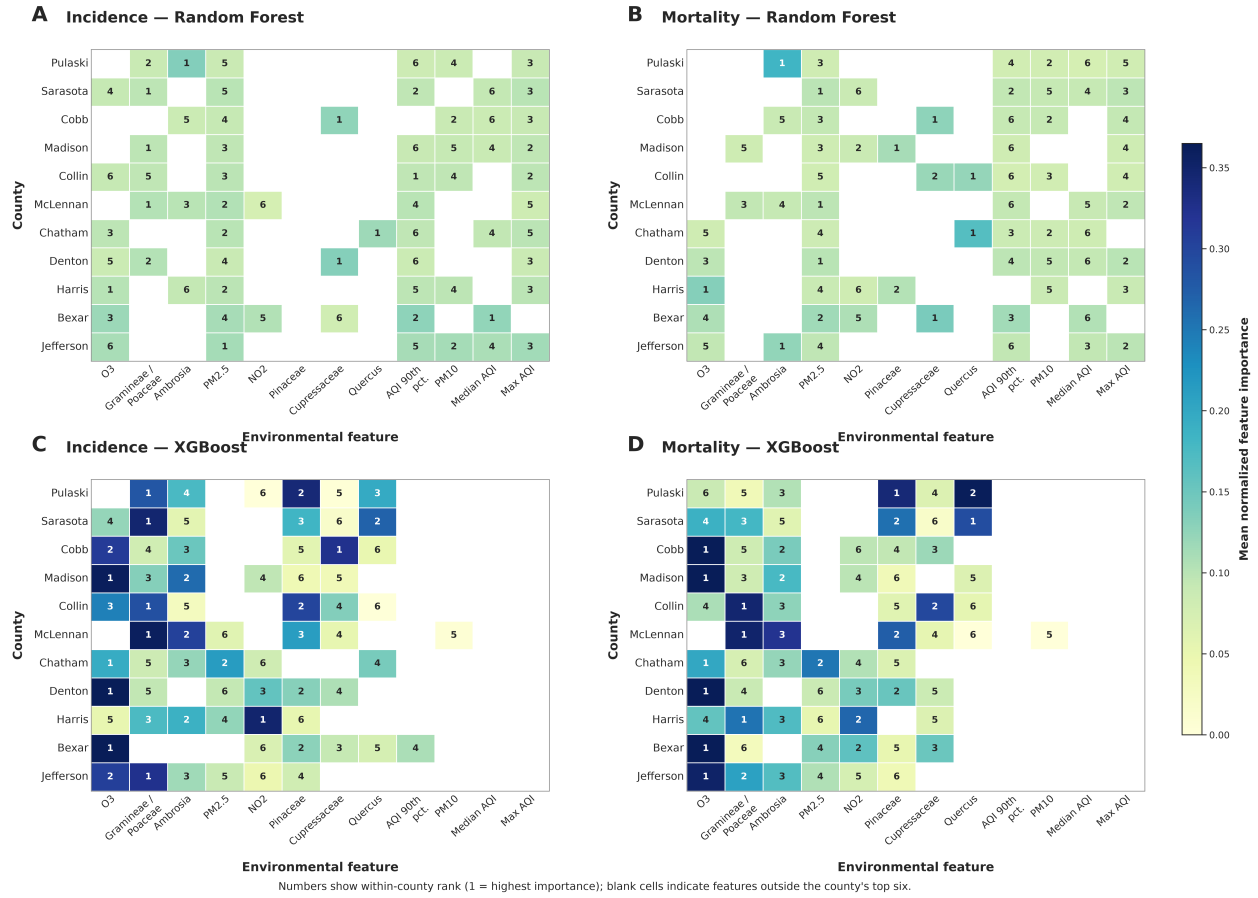


Figure B.4: County-specific tree-based feature rankings.

Detailed description for Figure B.4. County-specific tree-based feature rankings. Panels A and B show county-specific random forest and XGBoost feature importance. The heterogeneity across counties supports pooled and subgroup analyses but should not be interpreted as stable county-specific causal rankings.

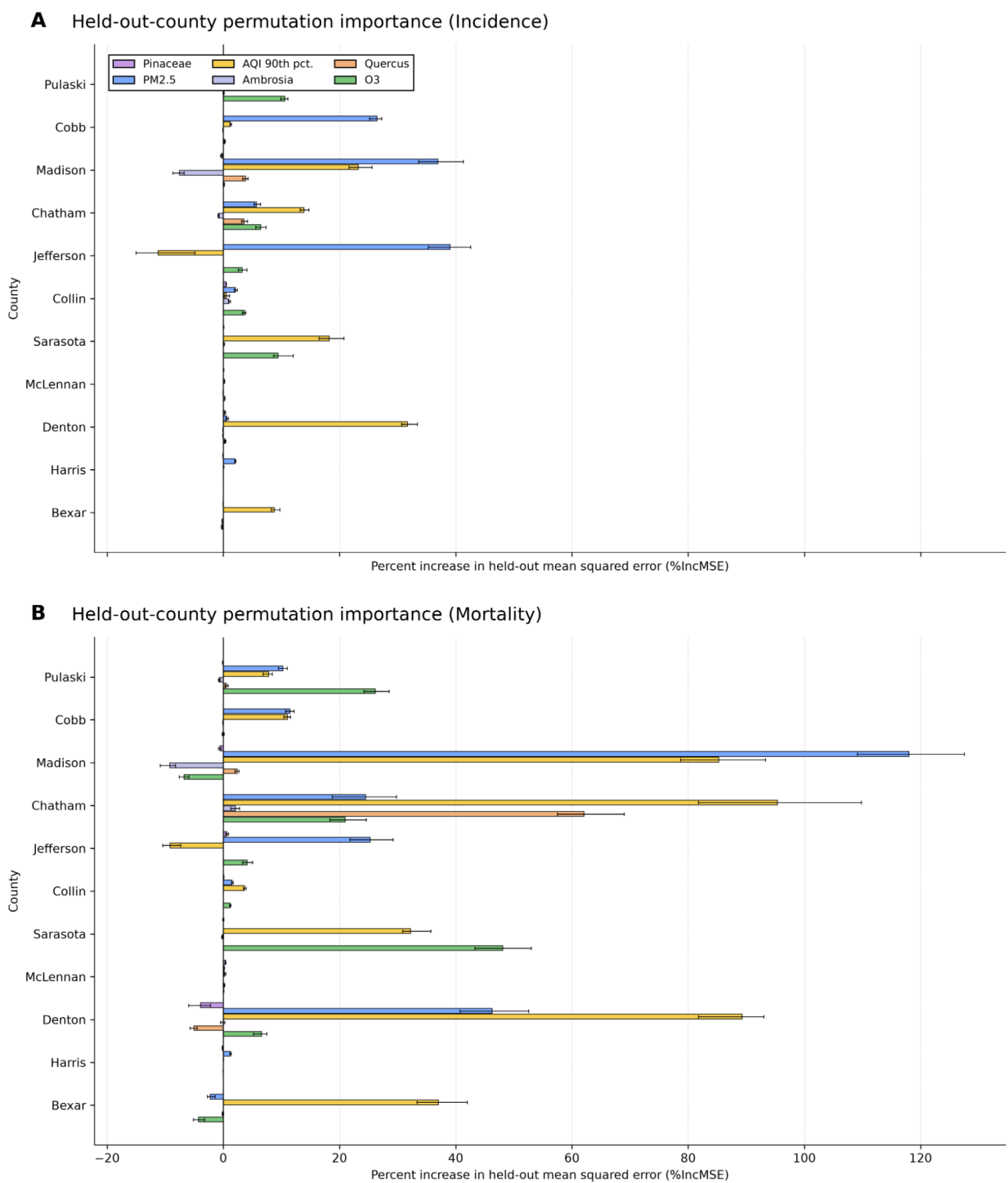


Figure B.5: Held-out-county permutation importance.

Detailed description for Figure B.5. Held-out-county permutation importance. Panels A and B show leave-one-county-out permutation importance for incidence and mortality, respectively, using six consensus predictors: Pinaceae, PM2.5, the AQI 90th percentile, Ambrosia, Quercus, and ozone. For each county, random forest models were trained on the remaining 10 counties and evaluated on the seven annual observations from the held-out county. Bars show the median percent increase in held-out mean squared error across 50 repeated model fits, and horizontal error bars show the interquartile range. Positive values indicate that permuting a feature worsened held-out prediction, whereas near-zero or negative values indicate limited transferable importance, redundancy, or instability rather than protective effects.

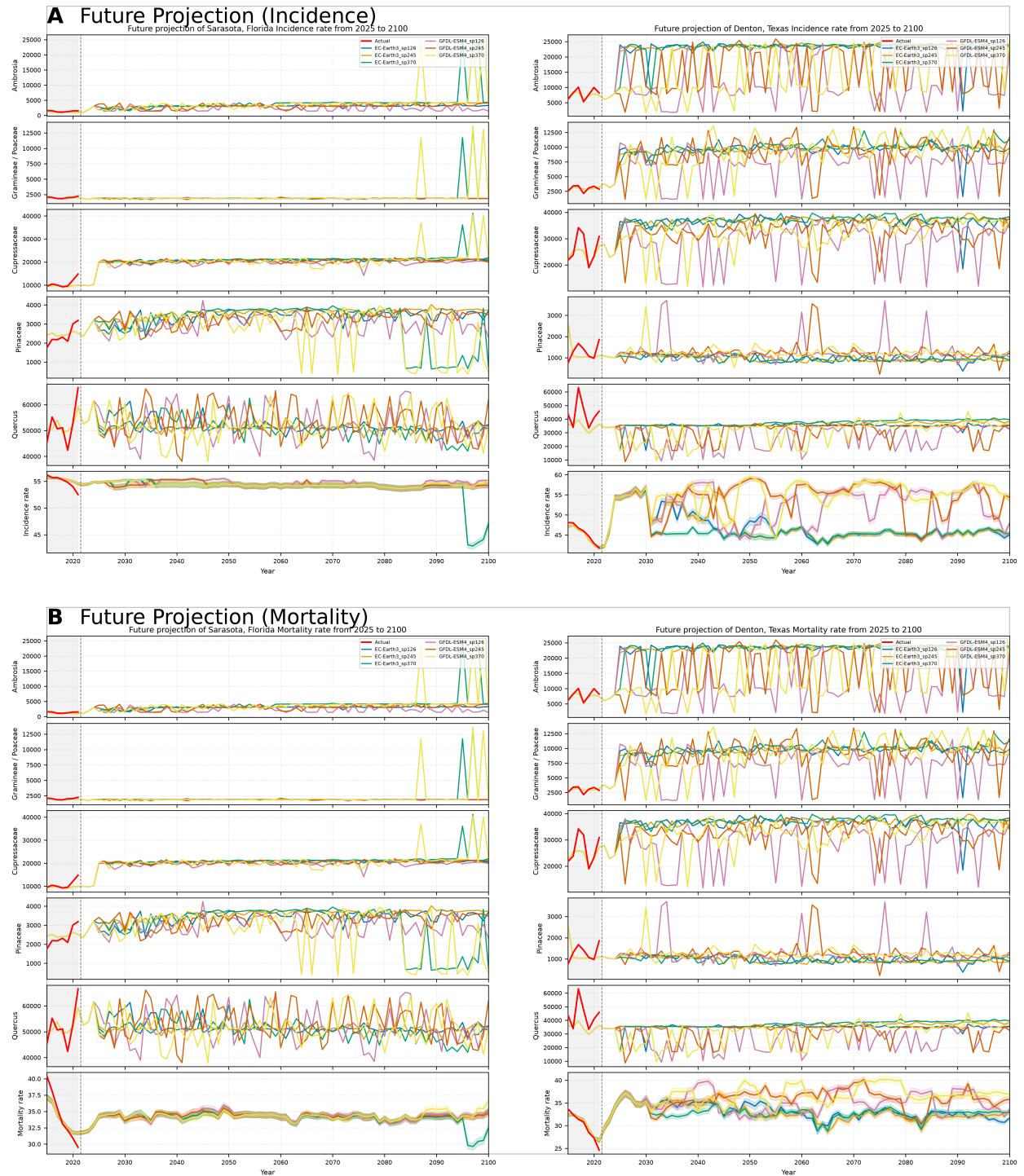


Figure B.6: Exploratory future-projection concept.

Detailed description for Figure B.6. Exploratory future-projection concept. Panels A and B show exploratory future-projection concepts for incidence and mortality. These outputs should not be presented as validated forecasts and are retained only to illustrate a possible future scenario-modeling extension.

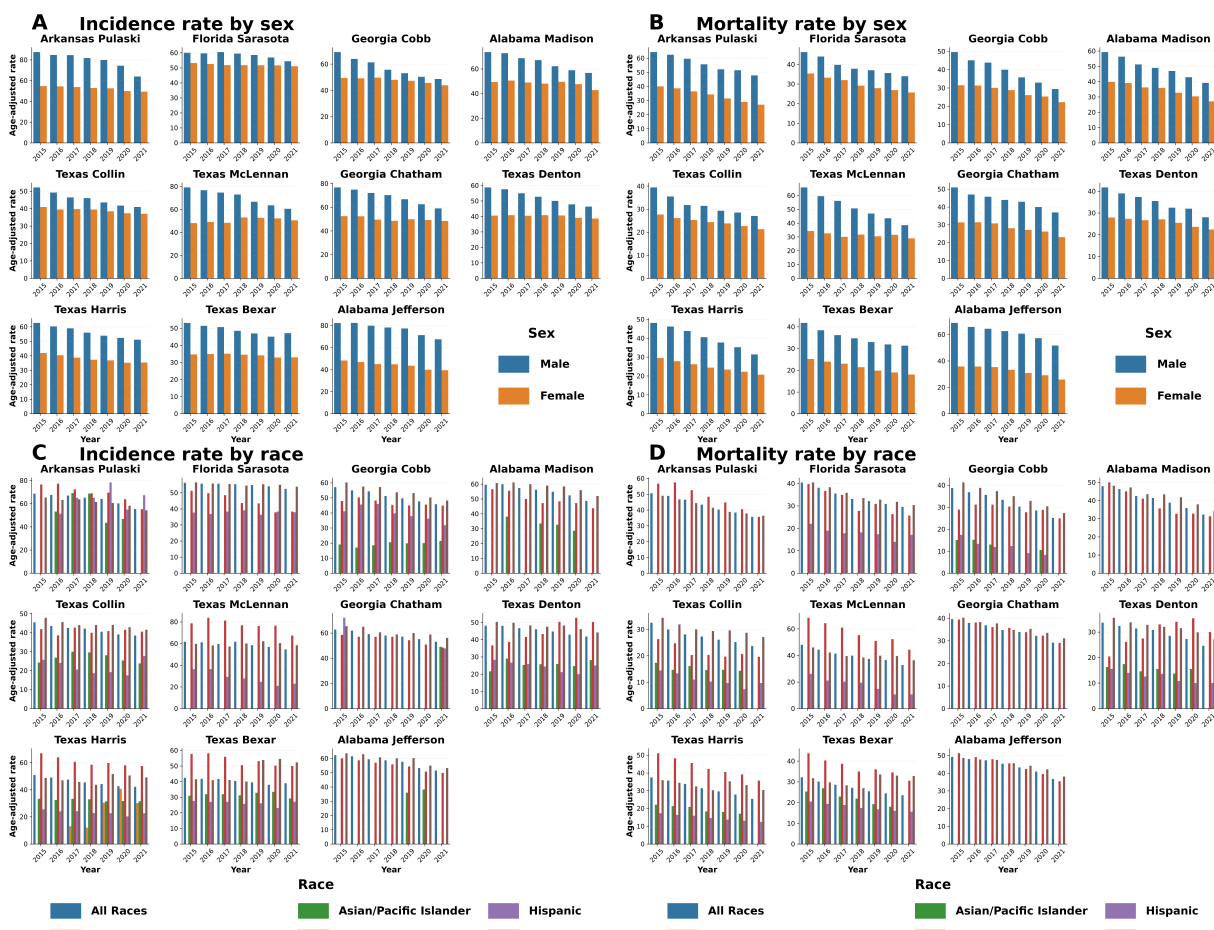
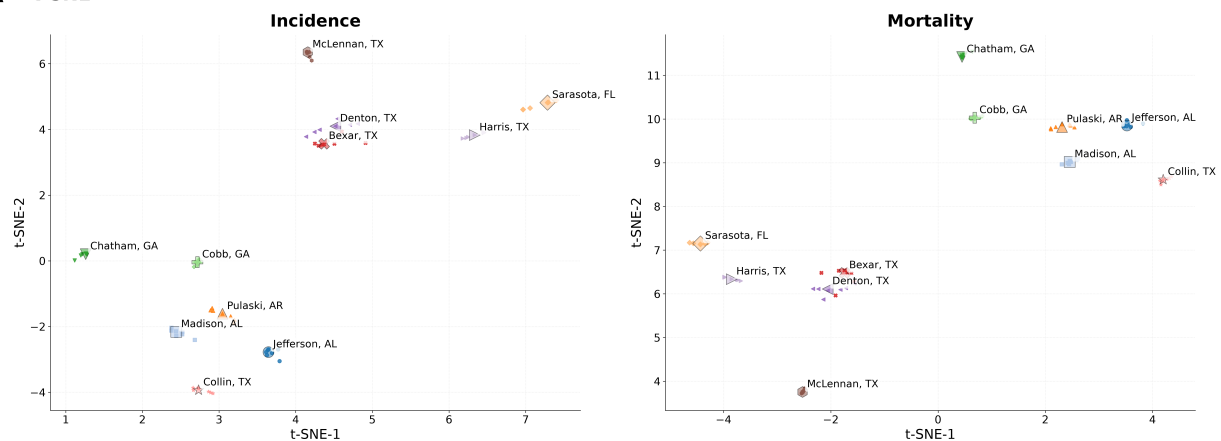


Figure B.7: Sex- and race-stratified lung cancer outcome patterns.

Detailed description for Figure B.7. Sex- and race-stratified lung cancer outcome patterns. Panels A-D show incidence and mortality stratified by sex and race. These figures support the discussion that demographic structure, equity, and health-system variables should be incorporated into future models.

A t-SNE



B UMAP

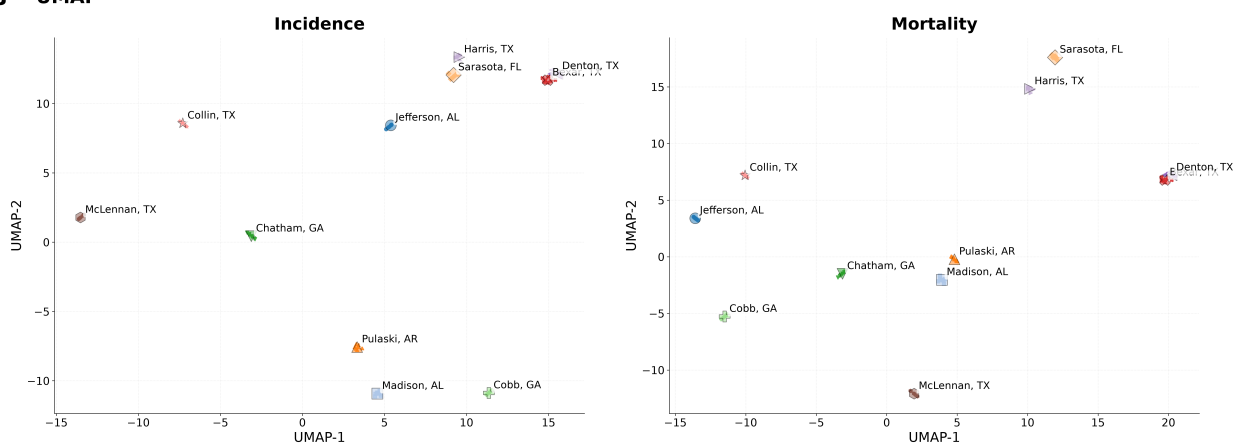


Figure B.8: t-SNE and UMAP clustering of county profiles.

Detailed description for Figure B.8. t-SNE and UMAP clustering of county profiles. Panels A and B show dimensionality-reduction visualizations of county environmental and outcome profiles, supporting the interpretation that county context partly drives pooled prediction and grouped-validation difficulty.

B.2 Detailed captions for Chapter 4 main figures

The following detailed captions correspond to the Chapter 4 main figures. Detailed notes for Appendix B supplementary figures are placed immediately after the corresponding supplementary figures in Section B.1.

Figure 4.1. Spatial and temporal distribution of county-level lung cancer incidence and mortality. Panel A displays average lung cancer incidence and mortality patterns across the selected counties and surrounding region. Panel B displays annual incidence and mortality heatmaps from 2015 to 2021. These descriptive ecological summaries establish spatial heterogeneity, high burden in Pulaski County, lower rates in several Texas counties, and a general decline over time.

Figure 4.2. Pollen station levels overlaid on average lung cancer incidence and mortality. Panel A uses incidence as the background outcome; Panel B uses mortality. Marker sizes represent relative pollen levels for Ambrosia, Gramineae/Poaceae, Cupressaceae, Pinaceae, and Quercus. The figure supports inclusion of pollen as a central environmental signal, but it is descriptive and hypothesis-generating.

Figure 4.3. Weekly pollen reconstruction in Pulaski and Sarasota counties. Panels A and B compare observed weekly pollen levels, shown as blue bars, with predicted values, shown as red lines and points, for Ambrosia, Gramineae/Poaceae, Cupressaceae, Pinaceae, and Quercus in Pulaski County, Arkansas, and Sarasota County, Florida, respectively. The multi-year series are included as descriptive illustrations of taxon-specific seasonal timing and magnitude; weekly values were not used as predictors in the lung cancer models.

Figure 4.4. Lag-window correlations, predictor collinearity, and temporal/spatial correlation. Panel A shows window-lag Pearson correlation surfaces for pollen, pollutant, and AQI predictors in the Pulaski incidence analysis. Panel B shows triangular predictor-correlation

matrices for the incidence and mortality model inputs. Panels C and D show endpoint-specific temporal correlations across 2015–2021 and spatial correlations across the selected counties for incidence and mortality, respectively. Together, the panels demonstrate that pollen, pollutants, and AQI summaries have distinct lag patterns but also substantial temporal, spatial, and cross-predictor dependence.

Figure 4.5. Pooled random forest feature importance and predictive performance. Panels A and B show mean decrease in impurity (MDI) feature importance for incidence and mortality, respectively, averaged across repeated random forest runs; horizontal error bars show among-run variability. Panels C and D compare averaged predicted and observed county-year incidence and mortality rates, respectively. Point colors identify counties, dashed lines indicate perfect agreement, and inset statistics report mean absolute error, mean squared error, and R-squared. These pooled results describe fit to the available county-year structure and should be interpreted alongside the county-grouped validation in Figure 4.6.

Figure 4.6. SHAP interpretation and county-grouped validation. Panels A–C show incidence SHAP beeswarm plots for subgroup 1, subgroup 2, and the pooled model after removing Max AQI, respectively; Panel D shows observed-versus-predicted incidence under reduced-feature county-grouped validation. Panels E–G show the corresponding mortality SHAP plots, and Panel H shows county-grouped mortality predictions. In the SHAP panels, features are ordered by mean absolute SHAP magnitude, horizontal position indicates contribution to the model prediction, and color denotes low-to-high feature values. The grouped prediction panels use county colors, a dashed 1:1 line, and performance metrics. Poor grouped performance indicates limited transfer to held-out county groups and supports cautious interpretation.

Figure 4.7. Humidity-index SHAP and interaction analysis. Panel A shows the incidence SHAP summary, Panel B shows the strongest pairwise Friedman H-statistics for incidence,

and Panel C shows incidence one-vs-all interaction strength. Panels D–F present the corresponding mortality analyses. In the SHAP panels, point color represents low-to-high feature values. The generally small H-statistics indicate limited evidence for strong interaction structures, although adding the humidity index changes the model explanation and feature-ranking pattern.

Figure 4.8. Absolute-humidity sensitivity, 3D pollen-humidity patterns, and SHAP interaction diagnostics. Panels A and B compare incidence and mortality SHAP rankings across models with no engineered terms, pollen-by-absolute-humidity product terms, and product-plus-ratio terms. Panels C and D show three-dimensional relationships among *Quercus*, absolute humidity, and incidence or mortality, respectively; Panels E and F show the corresponding *Pinaceae* relationships. Marker symbols distinguish counties, and point colors represent observed outcome rates. Panels G and H show Max AQI-by-absolute-humidity and *Pinaceae*-by-absolute-humidity SHAP interaction values for incidence and mortality, with point color indicating absolute humidity. These panels are model-behavior diagnostics and should not be interpreted as causal effects.

Appendix C

Supplementary Material for TOPS Framework

C.1 Notation Systems

- $(\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma)$ with state space $\mathcal{S}$, action space $\mathcal{A}$, the transition kernel $\mathcal{P}$, the reward function r , the initial state S_0 and its distribution μ_0 , and the discounted factor γ .
- $r_{\max} > 0$ is a constant as the upper bound of the reward.
- State value function $V_\pi(s)$ and state-action value function $Q_\pi(s, a)$.
- The normalized state and state action occupancy measure of policy π is denoted by $\nu_\pi(s)$ and $\sigma_\pi(s, a)$
- T is the length of a *trajectory*.
- The return is defined as G . $J(\pi)$ is the expectation of G .
- Policy π_θ is parameterized by the parameter θ .
- τ is the temperature parameter in the softmax parameterization of the policy.
- $F(\theta)$ is the Fisher information matrix.
- $\eta_T D$ is the learning rate of TD update. Similarly, $\eta_N PG$ is the learning rate of NPG update. $\eta_P PO$ is the learning rate of PPO update.
- β is the penalty factor of KL difference in PPO update.
- $f((s, a); \theta)$ is the two-layer over-parameterized neural network, with m as its width.
- ϕ_θ is the feature mapping of the neural network.

- $\mathcal{D}$ is the parameter space for θ , with Υ as its radius.
- $M > 0$ is a constant as the initialization upper bound on θ .
- $J_\lambda^G(\pi)$ is the mean-variance objective function.
- $J_\lambda(\pi)$ is the reward-volatility objective function, with λ as the penalty factor.
- $J_\lambda^y(\pi)$ is the transformed reward-volatility objective function, with y as the auxiliary variable.
- $\tilde{r}$ is the reward for the augmented MDP. Similarly, $\tilde{V}_\pi(s)$ and $\tilde{Q}_\pi(s, a)$ are state value function and state-action value function of the augmented MDP, respectively. $\tilde{J}(\pi)$ is the risk-neutral objective of the augmented MDP.
- $\hat{y}_k$ is an estimator of y at k -th iteration.
- ω is the parameter of critic network.
- $\delta_k = \operatorname{argmin}_{\delta \in \mathcal{D}} \|\hat{F}(\theta_k)\delta - \tau_k \hat{\nabla}_\theta J(\pi_{\theta_k})\|_2$.
- $\xi_k(\delta) = \hat{F}(\theta_k)\delta - \tau_k \hat{\nabla}_\theta \tilde{J}(\pi_{\theta_k}) - \mathbb{E}[\hat{F}(\theta_k)\delta - \tau_k \hat{\nabla}_\theta \tilde{J}(\pi_{\theta_k})]$.
- σ_ξ is a constant associated with the upper bound of the gradient variance.
- $\varphi_k, \psi_k, \varphi'_k, \psi'_k$ are the concentrability coefficients, upper bounded by a constant $c_0 > 0$.
- $\varphi_k^* = \mathbb{E}_{(s,a) \sim \sigma_\pi} \left[\left(\frac{d\pi^*}{d\pi_0} - \frac{d\pi_{\theta_k}}{d\pi_0} \right)^2 \right]^{1/2}$.
- $\psi_k^* = \mathbb{E}_{(s,a) \sim \sigma_\pi} \left[\left(\frac{d\sigma_{\pi^*}}{d\sigma_\pi} - \frac{d\nu_{\pi^*}}{d\nu_\pi} \right)^2 \right]^{1/2}$.
- K is the total number of iterations. Similarly, K_{TD} is the total number of TD iterations.
- $c_3 > 0$ is a constant as to quantify the difference in risk-neutral objective between optimal policy and any policy.

C.2 Algorithm Details

We provide a comparison between MVPI and TOPS.

Algorithm 2: A comparison between TOPS and MVPI

For $k = 1, \dots, K$:

Step 1: $y_k := (1 - \gamma)J(\pi_k)$.

Step 2: $\tilde{J}(\pi_{\theta_k}) := \mathbb{E}_{(s,a) \sim \sigma_{\pi_{\theta_k}}} (r_{s,a} - \lambda r_{s,a}^2 + 2\lambda r_{s,a} y_k)$.

If MVPI: set $\theta_k := \arg \max_{\theta} \tilde{J}(\pi_{\theta_k})$. This is achieved by the iterations from Line 8 to Line 11 in Algorithm 3.

Else if TOPS:

If the NPG update is selected, update θ_k according to Eq. (5.18).

Else if the PPO update is selected, update θ_k according to Eq. (5.20).

Output: π_{θ_K} .

Note that neither NPG nor PPO solve $\theta_k := \arg \max_{\theta} (\tilde{J}(\pi_{\theta_k}))$ directly, but instead solve an approximation optimization problem at each iteration. We provide pseudo-code for the implementation of MVPI and VARAC in Algorithm 3 and 4.

Algorithm 3: MVPI with over-parameterized networks

Input: number of iterations K ; learning rates for natural policy gradient (resp. PPO) TD, η_{NPG} (resp. η_{PPO}); and temperature parameters $\{\tau_k\}_{k=1}^K$.

Initialization: Initialize the policy network $f((s, a); \theta, b)$ as defined in Eq. (5.5). Set $\tau_1 = 1$. Initialize the Q-network with (b, ω_1) similarly.

For $k = 1, \dots, K$:

Sample a batch of transitions $\{s_t, a_t, r_t, s'_t\}_{t=1}^T$ following the current policy.

Set $y = \frac{1}{T} \sum_{t=1}^T r_t$.

For $t = 1, \dots, T$:

Set $\tilde{r}_t = r_t - \lambda r_t^2 + 2\lambda r_t y$ and sample $a'_t \sim \pi(a \mid s'_t)$.

Repeat until convergence:

Q-value update: update ω_k according to Eq. (5.21).

If the NPG update is selected, update θ_k according to Eq. (5.18).

Else if the PPO update is selected, update θ_k according to Eq. (5.20).

Output: π_{θ_K} .

Algorithm 4: VARAC

Input: number of iterations K ; learning rates for natural policy gradient (resp. PPO) TD, η_{NPG} (resp. η_{PPO}); and temperature parameters $\{\tau_k\}_{k=1}^K$.

Initialization: Initialize the policy network $f((s, a); \theta, b)$ as defined in Eq. (5.5). Set $\tau_1 = 1$. Initialize the Q-network with (b, ω_1) similarly.

For $k = 1, \dots, K$:

Sample a batch of transitions $\{s_t, a_t, r_t, s'_t\}_{t=1}^T$ following the current policy.

Set $y = \frac{1}{T} \sum_{t=1}^T r_t$.

Q-value update: update both networks' ω_k according to Eq. (5.21).

Output Q_k and W_k .

Update θ_k with NPG or PPO.

Output: π_{θ_K} .

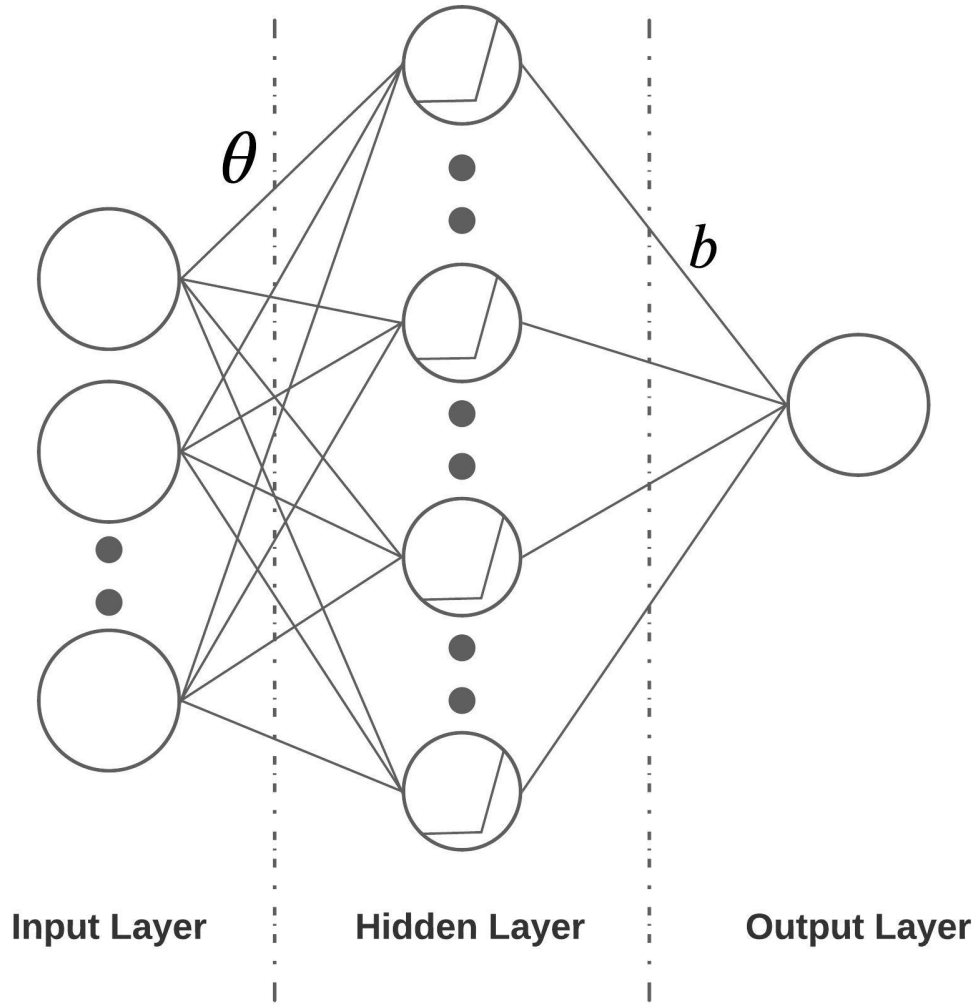


Figure C.1: A block diagram of over-parameterized neural network

C.3 Experimental Details

Note that although the mean-volatility method can be adapted to off-policy methods [122], in this chapter, for the ease of the theoretical analysis, our proposed method is an on-policy actor-critic algorithm.

C.3.1 Testbeds

We use six Mujoco tasks from Open AI gym [17] as testbeds. They are HalfCheetah-v2, Hopper-V2, Swimmer-V2, Walker2d-V2, InvertedPendulum-v2, and InvertedDoublePendulum-v2.

C.3.2 Hyper-parameter Settings

In the experiment we set $\lambda = 1$. We then tune learning rate for different algorithms. For MVP, we use the same setting as [122]. For MVPI, TOPS and VARAC with neural NPG, we tune the learning rate of the actor network from $\{0.1, 1 \times 10^{-2}, 1 \times 10^{-3}, 7 \times 10^{-4}\}$ and the learning rate of the critic network from $\{1 \times 10^{-2}, 1 \times 10^{-3}, 7 \times 10^{-4}\}$. For MVPI, TOPS and VARAC with neural PPO, we tune the learning rate of the actor network from $\{3 \times 10^{-3}, 3 \times 10^{-4}, 3 \times 10^{-5}\}$ and the learning rate of the critic network from $\{1 \times 10^{-2}, 1 \times 10^{-3}, 1 \times 10^{-4}\}$.

C.3.3 Computing infrastructure

We conducted our experiments on a GPU GTX 970 and GPU GTX 1080Ti.

C.4 Theoretical Analysis Details

In this section, we discuss the theoretical analysis in detail. We first present the overview in section C.4.1. Then we provide additional assumptions in section C.4.2. In the rest of the section, we present all the supporting lemmas and the proof for Theorem 5.1 and 5.2.

C.4.1 Overview

We provide Figure C.2 to illustrate the structure of the theoretical analysis. First, under Assumption 3 and 4, as well as Lemma 13. we can obtain Lemma 14, 15 and 16.

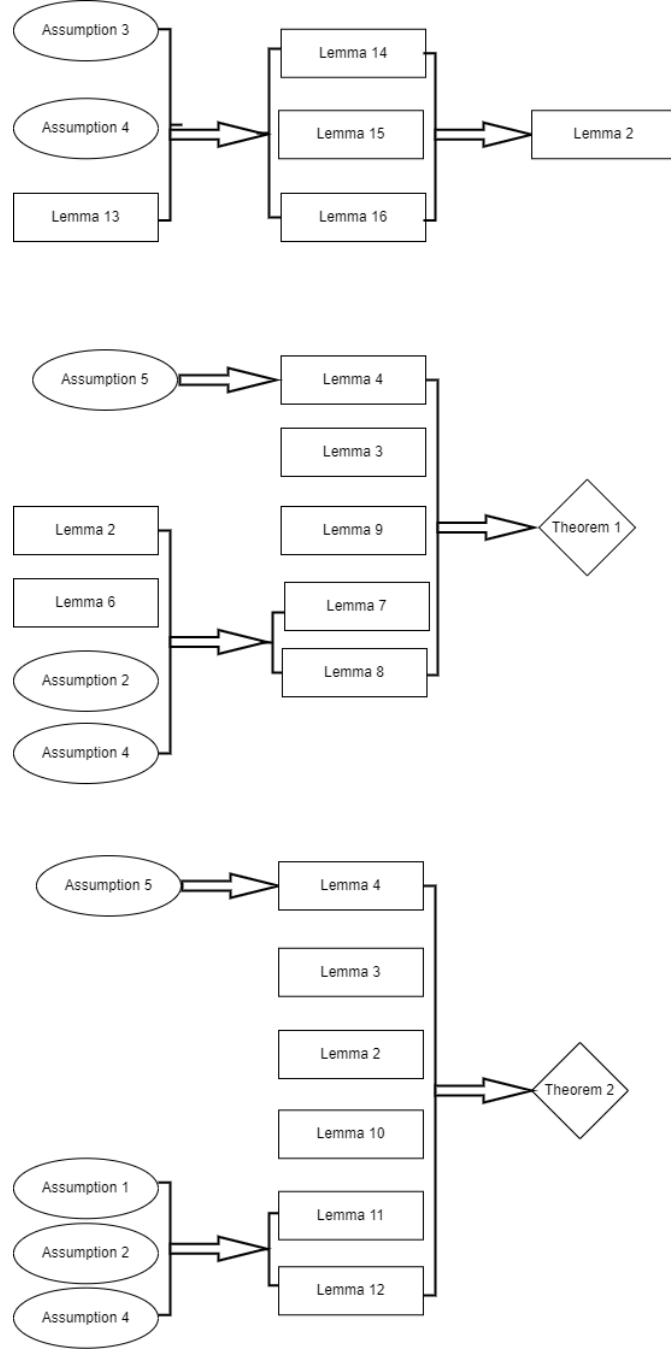


Figure C.2: A flow chart of the theoretical analysis

These are the building blocks of lemma 2, which is a shared component in the analysis of both NPG and PPO. The shared components also include Lemma 3, as well as Lemma 4 obtained under Assumption 5. For PPO analysis, under Assumption 2 and 4, we obtain Lemma 7 and 8 from Lemma 2 and 6, Then combined with Lemma 3, 4 and 9, we obtain

Theorem 5.1, the major result of PPO analysis. Likely for NPG analysis, we first obtain Lemma 11 and 12 under Assumption 1, 2 and 4. Then together with Lemma 2, 3, 4 and 10, we obtain Theorem 5.2, the major result of NPG analysis.

C.4.2 Additional Assumptions

Assumption 3 (*Action-value function class*) We define

$$\mathcal{F}_{\Upsilon, \infty} := \left\{ f(s, a; \theta) = f_0(s, a) + \int \mathbb{1}\{\theta^\top(s, a) > 0\}(s, a)^\top \iota(\theta) d\mu(w) : \|\iota(\theta)\|_\infty \leq \Upsilon/\sqrt{d} \right\} \quad (\text{C.1})$$

Where $\mu : \mathbb{R}^d \rightarrow [0, 1]$ is a probability density function of $\mathcal{N}(0, I_d/d)$. $f_0(s, a)$ is the two-layer neural network corresponding to the initial parameter Θ_{init} , and $\iota : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a weighted function. We assume that $\tilde{Q}_\pi \in \mathcal{F}_{\Upsilon, \infty}$ for all π .

Assumption 4 (*Regularity of stationary distribution*) For any policy π , any unit vector $x \in \mathbb{R}^d$ with $\|x\|_2 = 1$, and any $u > 0$, we assume that there exists a constant $c > 0$ such that

$$\mathbb{E}_{(s,a) \sim \sigma_\pi} [\mathbb{1}\{|x^\top(s, a)| \leq u\}] \leq cu.$$

Assumption 3 is a mild regularity condition on Q_π , as $\mathcal{F}_{\Upsilon, \infty}$ is a sufficiently rich function class and approximates a subset of the reproducing kernel Hilbert space (RKHS) [88]. Similar assumptions are widely imposed [80, 7, 38, 119, 109]. Assumption 4 is a regularity condition on the transition kernel $\mathcal{P}$. Such regularity holds so long as σ_π has an upper bound density, satisfying most Markov chains.

In [124] Lemma 4.15, they make a mistake in the proof. They accidentally flip a sign in $y^* - \bar{y}$ when transitioning from the first equation in the proof to Eq.(4.15). This invalidates the conclusion in Eq.(4.17), an essential part of the proof. We tackle this issue by proposing the next assumption.

Assumption 5 (*Convergence Rate of $J(\pi)$*) *We assume π^* (the optimal policy to the risk-averse objective function $J_\lambda(\pi)$) converges to the risk-neutral objective $J(\pi)$ for both NPG and PPO with the over-parameterized neural network to be $\mathcal{O}(1/\sqrt{k})$. Specifically, there exists a constant $c_3 > 0$ such that,*

$$J(\pi^*) - J(\pi_k) \leq \frac{c_3}{\sqrt{k}} \quad (\text{C.2})$$

It was proved [109, 62] that the optimal policy w.r.t the risk-neutral objective $J(\pi)$ obtained by NPG and PPO method with the over-parameterized two-layer neural network converges to the globally optimal policy at a rate of $\mathcal{O}(1/\sqrt{K})$, where K is the number of iteration. Since our method uses similar settings, we assume the convergence rates of risk-neutral objective $J(\pi)$ in our method follow their results.

In the following subsections, we study TOPS’s convergence of global optimality and provide a proof sketch.

C.4.3 Proof of Theorem 5.1

We first present the analysis of policy evaluation error, which is induced by TD update in Line 8 of Algorithm 1. We characterize the policy evaluation error in the following lemma:

Lemma 2 (Policy Evaluation Error) *We set learning rate of TD $\eta_{TD} = \min\{(1 - \gamma)/3(1 + \gamma)^2, 1/\sqrt{K_{TD}}\}$. Under Assumption 3 and 4, it holds that, with probability of $1 - \delta$,*

$$\begin{aligned} & \|\tilde{Q}_{\omega_k} - \tilde{Q}_{\pi_k}\|_{\nu_{\pi_k}}^2 \\ &= \mathcal{O}(\Upsilon^3 m^{-1/2} \log(1/\delta) + \Upsilon^{5/2} m^{-1/4} \sqrt{\log(1/\delta)} \\ & \quad + \Upsilon r_{\max}^2 m^{-1/4} + \Upsilon^2 K_{TD}^{-1/2} + \Upsilon), \end{aligned} \quad (\text{C.3})$$

where $\tilde{Q}_{\pi_k}$ is the Q-value function of the augmented MDP, and $\tilde{Q}_{\omega_k}$ is its estimator at the k -th iteration. We provide the proof and its supporting lemmas in Appendix C.4.6. In the following, we establish the error induced by the policy update. Eq. (5.11) can be re-expressed as

$$J_{\lambda}^y(\pi) = \sum_{s,a} \sigma_{\pi}(r_{s,a} - \lambda r_{s,a}^2 + 2\lambda r_{s,a} y_{k+1}) - \lambda y_{k+1}^2 \quad (\text{C.4})$$

It can be shown that $\forall \pi, \max_y J_{\lambda}^y(\pi) = J_{\lambda}(\pi)$ [116, 122]. We denote the optimal policy to the augmented MDP associated with y^* by $\pi^*(y^*)$. By definition, it is obvious that π^* and $\pi^*(y^*)$ are equivalent. For simplicity, we will use the unified term π^* in the rest of the chapter. We present Lemma 3 and 4.

Lemma 3 (Policy's Performance Difference) *For mean-volatility objective w.r.t. auxiliary variable y as $J_{\lambda}^y(\pi)$ defined in Eq. (C.4). For any policy π and π' , we have the following,*

$$J_{\lambda}^y(\pi') - J_{\lambda}^y(\pi) = (1 - \gamma)^{-1} \mathbb{E}_{s \sim \nu_{\pi'}} [\mathbb{E}_{a \sim \pi'} [\tilde{Q}_{\pi,y}]] \quad (\text{C.5})$$

$$- \mathbb{E}_{a \sim \pi} [\tilde{Q}_{\pi,y}]], \quad (\text{C.6})$$

where $\tilde{Q}_{\pi,y}$ is the state-action value function of the augmented MDP, and its rewards are associated with y .

Proof 1 When y is fixed,

$$\begin{aligned} & J_\lambda^y(\pi') - J_\lambda^y(\pi) \\ &= \sum_{s,a} \sigma_{\pi'} \tilde{r}_{s,a} - \sum_{s,a} \sigma_\pi \tilde{r}_{s,a} = \tilde{J}(\pi') - \tilde{J}(\pi) \end{aligned} \quad (\text{C.7})$$

$$(\text{C.8})$$

We then follow Lemma 6.1 in [53]:

$$\tilde{J}(\pi') - \tilde{J}(\pi) = (1 - \gamma)^{-1} \mathbb{E}_{(s,a) \sim \sigma_{\pi'}} [\tilde{A}_\pi] \quad (\text{C.9})$$

where $\tilde{A}_\pi = \tilde{Q}_\pi - \tilde{V}_\pi$ is the advantage function of policy π . Meanwhile,

$$\mathbb{E}_{a \sim \pi'} [\tilde{A}_\pi] = \mathbb{E}_{a \sim \pi'} [\tilde{Q}_\pi] - \tilde{V}_\pi = \mathbb{E}_{a \sim \pi'} [\tilde{Q}_\pi] - \mathbb{E}_{a \sim \pi} [\tilde{Q}_\pi] \quad (\text{C.10})$$

From Eq. (C.8), Eq. (C.9) and Eq. (C.10), we complete the proof.

Lemma 3 is inspired by [53] and adopted by most work on global convergence [1, 62, 118]. Next, we derive an upper bound for the error of the critic update in Line 5 of Algorithm 1:

Lemma 4 (*y* Update Error) We characterize the error induced by the estimation of auxiliary variable y w.r.t the optimal value y^* at k -th iteration as, $J_\lambda^{y^*}(\pi^*) - J_\lambda^{\hat{y}^k}(\pi^*) = \frac{2c_3 r_{\max}(1-\gamma)\lambda}{\sqrt{k}}$, where $r_{\max}$ is the bound of the original reward, and c_3 is a constant error term.

Proof 2 We start from the subproblem objective defined in Eq. (C.4) with y^* and $\hat{y}_k$:

$$\begin{aligned}
& J_\lambda^{y^*}(\pi^*) - J_\lambda^{\hat{y}_k}(\pi^*) \\
&= \left(\sum_{s,a} \sigma_{\pi^*}(r_{s,a} - \lambda r_{s,a}^2 + 2\lambda r_{s,a} y^*) - \lambda y^{*2} \right) - \\
&\quad \left(\sum_{s,a} \sigma_{\pi^*}(r_{s,a} - \lambda r_{s,a}^2 + 2\lambda r_{s,a} \hat{y}_k) - \lambda \hat{y}_k^2 \right) \\
&= 2\lambda \left(\sum_{s,a} \sigma_{\pi^*} r_{s,a} \right) (y^* - \hat{y}_k) - \lambda (y^{*2} - \hat{y}_k^2) \\
&= \lambda \langle y^* - \hat{y}_k, 2(1 - \gamma)J(\pi^*) - y^* - \hat{y}_k \rangle \tag{C.11}
\end{aligned}$$

$$= (1 - \gamma) \lambda \langle y^* - \hat{y}_k, J(\pi^*) - \hat{J}(\pi_k) \rangle \tag{C.12}$$

where we obtain the final two equalities by the definition of J_π and y . Because $r_{s,a}$ is upper-bounded by a constant $r_{\max}$, we have $|y^* - \hat{y}_k| \leq 2r_{\max}$. Under Assumption 5 we have,

$$J_\lambda^{y^*}(\pi^*) - J_\lambda^{\hat{y}_k}(\pi^*) = \frac{2c_3 r_{\max} (1 - \gamma) \lambda}{\sqrt{k}} \tag{C.13}$$

Thus we finish the proof.

From Lemma 3 and 4, we can also obtain the following Lemma.

Lemma 5 (Performance Difference on π and y) For mean-volatility objective w.r.t. auxiliary variable y as $J_\lambda^y(\pi)$ defined in Eq. (C.4). For any π, y and the optimal π^*, y^* , we have the following,

$$J_\lambda^{y^*}(\pi^*) - J_\lambda^y(\pi) = (1 - \gamma)^{-1} \mathbb{E}_{s \sim \nu_{\pi^*}} [\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi, y}]] \tag{C.14}$$

$$- \mathbb{E}_{a \sim \pi} [\tilde{Q}_{\pi, y}] + \frac{2c_3 r_{\max} (1 - \gamma) \lambda}{\sqrt{k}}. \tag{C.15}$$

where $\tilde{Q}_{\pi, y}$ is the state-action value function of the augmented MDP, and its rewards are associated with y .

Proof 3 *It is easy to see that $J_\lambda^{y*}(\pi^*) - J_\lambda^y(\pi) = J_\lambda^{y*}(\pi^*) - J_\lambda^y(\pi^*) + J_\lambda^y(\pi^*) - J_\lambda^y(\pi)$. Then replace $J_\lambda^y(\pi^*) - J_\lambda^y(\pi)$ with Lemma 3 and $J_\lambda^{y*}(\pi^*) - J_\lambda^y(\pi^*)$ with Lemma 4, we finish the proof.*

Lemma 5 quantifies the performance difference of $J_\lambda^y(\pi)$ between any pair π, y and the optimal π^*, y^* , while Lemma 3 only quantifies the performance difference of $J_\lambda^y(\pi)$ between π and π' when y is fixed.

We now study the global convergence of TOPS with neural PPO as the policy update component. First, we define the neural PPO update rule.

Lemma 6 [62]. *Let $\pi_{\theta_k} \propto \exp\{\tau_k^{-1} f_{\theta_k}\}$ be an energy-based policy. We define the update*

$$\hat{\pi}_{k+1} = \arg \max_{\pi} \mathbb{E}_{s \sim \nu_k} [\mathbb{E}_{\pi} [Q_{\omega_k}] - \beta_k KL(\pi_{\theta} \parallel \pi_{\theta_k})]$$

, where Q_{ω_k} is the estimator of the exact action-value function $Q^{\pi_{\theta_k}}$. We have

$$\hat{\pi}_{k+1} \propto \exp\{\beta_k^{-1} Q_{\omega_k} + \tau_k^{-1} f_{\theta_k}\} \quad (\text{C.16})$$

And to represent $\hat{\pi}_{k+1}$ with $\pi_{\theta_{k+1}} \propto \exp\{\tau_{k+1}^{-1} f_{\theta_{k+1}}\}$, we solve the following subproblem,

$$\begin{aligned} \theta_{k+1} = \arg \min_{\theta \in \mathbb{D}} \mathbb{E}_{(s,a) \sim \sigma_k} [& (f_{\theta}(s, a) - \tau_{k+1}(\beta_k^{-1} Q_{\omega_k}(s, a) \\ & + \tau_k^{-1} f_{\theta_k}(s, a)))^2] \end{aligned} \quad (\text{C.17})$$

We analyze the policy improvement error in Line 10 of Algorithm 1. [62] proves that the policy improvement error can be characterized similarly to the policy evaluation error as in Eq. (C.3). Recall $\tilde{Q}_{\omega_k}$ is the estimator of Q-value, f_{θ_k} the energy function for policy, and $f_{\hat{\theta}}$ its estimator. We characterize the policy improvement error as follows: Under Assumptions 3 and 4, we set the learning rate of PPO $\eta_{\text{PPO}} = \min\{(1-\gamma)/3(1+\gamma)^2 1/\sqrt{K_{\text{TD}}}\}$,

and with a probability of $1 - \delta$:

$$\begin{aligned}
& \| (f_{\hat{\theta}} - \tau_{k+1}(\beta^{-1}\tilde{Q}_{\omega_k} + \tau_k^{-1}f_{\theta_k})) \|^2 \\
& = \mathcal{O}(\Upsilon^3 m^{-1/2} \log(1/\delta) + \Upsilon^{5/2} m^{-1/4} \sqrt{\log(1/\delta)} \\
& \quad + \Upsilon r_{\max}^2 m^{-1/4} + \Upsilon^2 K_{\text{TD}}^{-1/2} + \Upsilon).
\end{aligned} \tag{C.18}$$

We quantify how the errors propagate in neural PPO [62] in the following.

Lemma 7 [62]. *(Error Propagation) We have,*

$$\begin{aligned}
& \left| \mathbb{E}_{s \sim \nu_{\pi^*}} \left[\mathbb{E}_{a \sim \pi^*} [\log(\pi_{\theta_{k+1}} / \pi_{k+1})] - \mathbb{E}_{a \sim \pi_{\theta_k}} \right. \right. \\
& \left. \left. [\log(\pi_{\theta_{k+1}} / \pi_{k+1})] \right] \right| \leq \tau_{k+1}^{-1} \varepsilon_k'' \varphi_{k+1}^* + \beta^{-1} \varepsilon_k'' \psi_k^*
\end{aligned} \tag{C.19}$$

$$\tag{C.20}$$

ε_k'' are defined in Eq.(C.3) as well as Eq.(C.18). $\varphi_k^* = \mathbb{E}_{(s,a) \sim \sigma_{\pi}} \left[\left(\frac{d\pi^*}{d\pi_0} - \frac{d\pi_{\theta_k}}{d\pi_0} \right)^2 \right]^{1/2}$, $\psi_k^* = \mathbb{E}_{(s,a) \sim \sigma_{\pi}} \left[\left(\frac{d\sigma_{\pi^*}}{d\sigma_{\pi}} - \frac{d\nu_{\pi^*}}{d\nu_{\pi}} \right)^2 \right]^{1/2} \cdot \frac{d\pi^*}{d\pi_0}, \frac{d\pi_{\theta_k}}{d\pi_0}, \frac{d\sigma_{\pi^*}}{d\sigma_{\pi}}, \frac{d\nu_{\pi^*}}{d\nu_{\pi}}$ are the Radon-Nikodym derivatives [54]. We denote RHS in Eq. (C.20) by $\varepsilon_k = \tau_{k+1}^{-1} \varepsilon_k'' \varphi_{k+1}^* + \beta^{-1} \varepsilon_k'' \psi_k^*$. Lemma 7 essentially quantifies the error from which we use the two-layer neural network to approximate the action-value function and policy instead of having access to the exact ones. Please refer to [62] for complete proofs of Lemma 6 and 7.

$$\begin{aligned}
& \left| \mathbb{E}_{s \sim \nu_{\pi^*}} \left[\mathbb{E}_{a \sim \pi^*} [\log(\pi_{\theta_{k+1}} / \pi_{k+1})] - \right. \right. \\
& \left. \left. \mathbb{E}_{a \sim \pi_{\theta_k}} [\log(\pi_{\theta_{k+1}} / \pi_{k+1})] \right] \right| \leq \tau_{k+1}^{-1} \varepsilon_k'' \varphi_{k+1}^* + \beta^{-1} \varepsilon_k'' \psi_k^*
\end{aligned} \tag{C.21}$$

We then characterize the difference between energy functions in each step [62]. Under the optimal policy π^* ,

Lemma 8 [62]. (Stepwise energy-function difference) Under the conditions of Lemma 7, we have

$$\mathbb{E}_{s \sim \nu_{\pi^*}} [\|\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}\|_{\infty}^2] \leq 2\varepsilon'_k + 2\beta_k^{-2} U, \quad (\text{C.22})$$

$$(\text{C.23})$$

where $\varepsilon'_k = |\mathcal{A}| \tau_{k+1}^{-2} \epsilon_{k+1}^2$

and $U = 2\mathbb{E}_{s \sim \nu_{\pi^*}} [\max_{a \in \mathcal{A}} (\tilde{Q}_{\omega_0})^2] + 2\Upsilon^2$.

Proof 4 By the triangle inequality, we get the following,

$$\|\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}\|_{\infty}^2 \quad (\text{C.24})$$

$$\leq 2(\|\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k} - \beta^{-1} \tilde{Q}_{\omega_k}\|_{\infty}^2 + \|\beta^{-1} \tilde{Q}_{\omega_k}\|_{\infty}^2) \quad (\text{C.25})$$

We take the expectation of both sides of Eq. (C.25) with respect to $s \sim \nu_{\pi^*}$. With the 1-Lipshitz continuity of $\tilde{Q}_{\omega_k}$ in ω and $\|\omega_k - \Theta_{\text{init}}\|_2 \leq \Upsilon$, we have,

$$\mathbb{E}_{\nu_{\pi^*}} [\|\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}\|_{\infty}^2] \quad (\text{C.26})$$

$$\leq 2(|\mathcal{A}| \tau_{k+1}^{-2} \epsilon_{k+1}^2 + \mathbb{E}_{s \sim \nu_{\pi^*}} [\max_{a \in \mathcal{A}} (\tilde{Q}_{\omega_0})^2] + \Upsilon^2) \quad (\text{C.27})$$

Thus complete the proof.

We then derive a difference term associated with π_{k+1} and π_{θ_k} , where at the k -th iteration π_{k+1} is the solution for the following subproblem,

$$\pi_{k+1} = \arg \max_{\pi} \left(\mathbb{E}_{s \sim \nu_{\pi_k}} [\mathbb{E}_{a \sim \pi} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \beta \text{KL}(\pi \| \pi_{\theta_k})] \right) \quad (\text{C.28})$$

and π_{θ_k} is the policy parameterized by the two-layered over-parameterized neural network. The following lemma establishes the one-step descent of the KL-divergence in the policy space:

Lemma 9 (*One-step difference of π*) For π_{k+1} and π_{θ_k} , we have

$$\begin{aligned}
& KL(\pi^* \parallel \pi_{\theta_k}) - KL(\pi^* \parallel \pi_{\theta_{k+1}}) \\
& \geq (\mathbb{E}_{a \sim \pi^*} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{k+1}})] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{k+1}})]) \\
& \quad + \beta^{-1} (\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\tilde{Q}_{\pi_k, \hat{y}_k}]) \\
& \quad + \frac{1}{2} \|\pi_{\theta_{k+1}} - \pi_{\theta_k}\|_1^2 + (\mathbb{E}_{a \sim \pi_{\theta_k}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}] \\
& \quad - \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}])
\end{aligned} \tag{C.29}$$

Proof 5 We start from

$$\begin{aligned}
KL(\pi^* \parallel \pi_{\theta_k}) - KL(\pi^* \parallel \pi_{\theta_{k+1}}) &= \mathbb{E}_{a \sim \pi^*} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})] \\
(\text{By definition, } KL(\pi_{\theta_{k+1}} \parallel \pi_{\theta_k}) &= \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})]) \\
&= (\mathbb{E}_{a \sim \pi^*} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})] - \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})] + \\
&\quad KL(\pi_{\theta_{k+1}} \parallel \pi_{\theta_k})
\end{aligned}$$

We then add and subtract terms,

$$\begin{aligned}
&= \mathbb{E}_{a \sim \pi^*} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})] - \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})] + KL \\
&\quad (\pi_{\theta_{k+1}} \parallel \pi_{\theta_k}) + \beta^{-1} (\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\tilde{Q}_{\pi_k, \hat{y}_k}]) \\
&\quad - \beta^{-1} (\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\tilde{Q}_{\pi_k, \hat{y}_k}]) \\
&\quad + \mathbb{E}_{a \sim \pi_{\theta_k}} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\log(\pi_{\theta_{k+1}} \pi_{\theta_k})] \tag{C.30}
\end{aligned}$$

Rearrange the terms and we get,

$$\begin{aligned}
&= (\mathbb{E}_{a \sim \pi^*} [\log(\pi_{\theta_{k+1}}) - \log(\pi_{\theta_k}) - \beta^{-1} \tilde{Q}_{\pi_k, \hat{y}_k}] \\
&\quad - \mathbb{E}_{a \sim \pi_{\theta_k}} [\log(\pi_{\theta_{k+1}}) - \log(\pi_{\theta_k}) - \beta^{-1} \tilde{Q}_{\pi_k, \hat{y}_k}]) \\
&\quad + \beta^{-1} (\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\tilde{Q}_{\pi_k, \hat{y}_k}]) + KL \\
&\quad (\pi_{\theta_{k+1}} \parallel \pi_{\theta_k}) + (\mathbb{E}_{a \sim \pi_{\theta_k}} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})] - \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} \\
&\quad [\log(\pi_{\theta_{k+1}} \pi_{\theta_k})]) \tag{C.31}
\end{aligned}$$

Recall that $\pi_{k+1} \propto \exp\{\tau_k^{-1} f_{\theta_k} + \beta^{-1} \tilde{Q}_{\pi_k}^y\}$. We define the two normalization factors associated with ideal improved policy π_{k+1} and the current parameterized policy π_{θ_k} as,

$$Z_{k+1}(s) := \sum_{a' \in \mathcal{A}} \exp\{\tau_k^{-1} f_{\theta_k}(s, a') + \beta^{-1} \tilde{Q}_{\pi_k}^y(s, a')\} \tag{C.32}$$

$$Z_{\theta_{k+1}}(s) := \sum_{a' \in \mathcal{A}} \exp\{\tau_{k+1}^{-1} f_{\theta_{k+1}}(s, a')\} \tag{C.33}$$

We then have,

$$\pi_{k+1}(a|s) = \frac{\exp\{\tau_k^{-1}f_{\theta_k}(s, a) + \beta^{-1}\tilde{Q}_{\pi_k}^y(s, a)\}}{Z_{k+1}(s)}, \quad (\text{C.34})$$

$$\pi_{\theta_{k+1}}(a|s) = \frac{\exp\{\tau_{k+1}^{-1}f_{\theta_{k+1}}(s, a)\}}{Z_{\theta_{k+1}}(s)} \quad (\text{C.35})$$

For any π, π' and k , we have,

$$\mathbb{E}_{a \sim \pi}[\log Z_{\theta_{k+1}}] - \mathbb{E}_{a \sim \pi'}[\log Z_{\theta_{k+1}}] = 0 \quad (\text{C.36})$$

$$\mathbb{E}_{a \sim \pi}[\log Z_{k+1}] - \mathbb{E}_{a \sim \pi'}[\log Z_{k+1}] = 0 \quad (\text{C.37})$$

Now we look back at a few terms on RHS from Eq.(C.31):

$$\mathbb{E}_{a \sim \pi^*}[\log(\pi_{\theta_k}) + \beta^{-1}\tilde{Q}_{\pi_k, \hat{y}_k}] \quad (\text{C.38})$$

$$- \mathbb{E}_{a \sim \pi_{\theta_k}}[\log(\pi_{\theta_k}) + \beta^{-1}\tilde{Q}_{\pi_k, \hat{y}_k}] \quad (\text{C.39})$$

$$\begin{aligned} &= (\mathbb{E}_{a \sim \pi^*}[\tau_k^{-1}f_{\theta_k} + \beta^{-1}\tilde{Q}_{\pi_k, \hat{y}_k} - \log Z_{\theta_{k+1}}] \\ &\quad - \mathbb{E}_{a \sim \pi_{\theta_k}}[\tau_k^{-1}f_{\theta_k} + \beta^{-1}\tilde{Q}_{\pi_k, \hat{y}_k} - \log Z_{\theta_{k+1}}]) \\ &= \mathbb{E}_{a \sim \pi^*} \left[\log \frac{\exp\{\tau_k^{-1}f_{\theta_k} + \beta^{-1}\tilde{Q}_{\pi_k, \hat{y}_k}\}}{Z_{k+1}} \right] \\ &\quad - \mathbb{E}_{a \sim \pi_{\theta_k}} \left[\log \frac{\exp\{\tau_k^{-1}f_{\theta_k} + \beta^{-1}\tilde{Q}_{\pi_k, \hat{y}_k}\}}{Z_{k+1}} \right] \\ &= \mathbb{E}_{a \sim \pi^*}[\log \pi_{k+1}] - \mathbb{E}_{a \sim \pi_{\theta_k}}[\log \pi_{k+1}] \end{aligned} \quad (\text{C.40})$$

For Eq. (C.40), we obtain the first equality by Eq. (C.35). Then, by swapping Eq. (C.36) with Eq. (C.37), we obtain the second equality. We achieve the concluding step with the

definition in Eq. (C.34). Following a similar logic, we have,

$$\begin{aligned}
& \mathbb{E}_{a \sim \pi_{\theta_k}} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})] - \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} [\log(\frac{\pi_{\theta_{k+1}}}{\pi_{\theta_k}})] \\
&= \mathbb{E}_{a \sim \pi_{\theta_k}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} - \log Z_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k} + \log Z_{\theta_k}] - \\
& \quad \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} - \log Z_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k} + \log Z_{\theta_k}] \\
&= \mathbb{E}_{a \sim \pi_{\theta_k}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}] - \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} -
\end{aligned} \tag{C.41}$$

$$\tau_k^{-1} f_{\theta_k}] \tag{C.42}$$

Finally, by using the Pinsker's inequality [30], we have,

$$KL(\pi_{\theta_{k+1}} \parallel \pi_{\theta_k}) \geq 1/2 \|\pi_{\theta_{k+1}} - \pi_{\theta_k}\|_1^2 \tag{C.43}$$

Plugging Eqs. (C.40), (C.42), and (C.43) into Eq. (C.31), we have

$$\begin{aligned}
& KL(\pi^* \parallel \pi_{\theta_k}) - KL(\pi^* \parallel \pi_{\theta_{k+1}}) \\
& \geq (\mathbb{E}_{a \sim \pi^*} [\log(\pi_{\theta_{k+1}}) - \log(\pi_{\theta_k})] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\log(\pi_{\theta_{k+1}}) \\
& \quad - \log(\pi_{\theta_k})]) + \beta^{-1} (\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\tilde{Q}_{\pi_k, \hat{y}_k}]) \\
& \quad + \frac{1}{2} \|\pi_{\theta_{k+1}} - \pi_{\theta_k}\|_1^2 + (\mathbb{E}_{a \sim \pi_{\theta_k}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}] \\
& \quad - \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}])
\end{aligned} \tag{C.44}$$

Rearranging the terms, we obtain Lemma 9.

Lemma 9 serves as an intermediate-term for the major result's proof. We obtain upper bounds by telescoping this term in Theorem 5.1. Now we are ready to present the proof for Theorem 5.1.

Proof 6 First we take expectation of both sides of Eq. (C.29) with respect to $s \sim \nu_{\pi^*}$ from Lemma 9 and insert Eq (C.20) to obtain,

$$\begin{aligned} & \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* || \pi_{\theta_{k+1}})] - \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* || \pi_{\theta_k})] \\ & \leq \varepsilon_k - \beta^{-1} \mathbb{E}_{s \sim \nu_{\pi^*}} [\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\tilde{Q}_{\pi_k, \hat{y}_k}]] \\ & - 1/2 \mathbb{E}_{s \sim \nu_{\pi^*}} [\|\pi_{\theta_{k+1}} - \pi_{\theta_k}\|_1^2] - \mathbb{E}_{s \sim \nu_{\pi^*}} [\mathbb{E}_{a \sim \pi_{\theta_k}} \end{aligned} \quad (C.45)$$

$$[\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}] - \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}]] \quad (C.46)$$

Then, by Lemma 3, we have,

$$\beta^{-1} \mathbb{E}_{s \sim \nu_{\pi^*}} [\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \mathbb{E}_{a \sim \pi_{\theta_k}} [\tilde{Q}_{\pi_k, \hat{y}_k}]] \quad (C.47)$$

$$= \beta^{-1} (1 - \gamma) (J_{\lambda}^{\hat{y}_k}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi)) \quad (C.48)$$

And with Hölder's inequality, we have,

$$\mathbb{E}_{s \sim \nu_{\pi^*}} [\mathbb{E}_{a \sim \pi_{\theta_k}} [\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}] - \mathbb{E}_{a \sim \pi_{\theta_{k+1}}} \quad (C.49)$$

$$[\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}]] \quad (C.50)$$

$$= \mathbb{E}_{s \sim \nu_{\pi^*}} [\langle \tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}, \pi_{\theta_k} - \pi_{\theta_{k+1}} \rangle] \quad (C.51)$$

$$\leq \mathbb{E}_{s \sim \nu_{\pi^*}} [\|\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}\|_{\infty} \|\pi_{\theta_k} - \pi_{\theta_{k+1}}\|_1] \quad (C.52)$$

Insert Eqs. (C.48) and (C.51) into Eq. (C.45), we have,

$$\begin{aligned}
& \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \parallel \pi_{\theta_{k+1}})] - \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \parallel \pi_{\theta_k})] \\
& \leq \varepsilon_k - (1 - \gamma)\beta^{-1} (J_{\lambda}^{\hat{y}_k}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi)) - 1/2 \mathbb{E}_{s \sim \nu_{\pi^*}} \\
& \quad [\|\pi_{\theta_{k+1}} - \pi_{\theta_k}\|_1^2] + \mathbb{E}_{s \sim \nu_{\pi^*}} [\|\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}\|_{\infty} \\
& \quad \|\pi_{\theta_k} - \pi_{\theta_{k+1}}\|_1] \\
& \leq \varepsilon_k - (1 - \gamma)\beta^{-1} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi) - J_{\lambda}^{y^*}(\pi^*) \\
& \quad + J_{\lambda}^{\hat{y}_k}(\pi^*)) + 1/2 \mathbb{E}_{s \sim \nu_{\pi^*}} [\|\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}\|_{\infty}^2] \\
& \leq \varepsilon_k - (1 - \gamma)\beta^{-1} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi)) \\
& \quad + (1 - \gamma)\beta^{-1} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi^*)) \\
& \quad + 1/2 \mathbb{E}_{s \sim \nu_{\pi^*}} [\|\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}\|_{\infty}^2]. \tag{C.53}
\end{aligned}$$

The second inequality holds by using the inequality $2AB - B^2 \leq A^2$, with a minor abuse of notations. Here, $A := \|\tau_{k+1}^{-1} f_{\theta_{k+1}} - \tau_k^{-1} f_{\theta_k}\|_{\infty}$ and $B := \|\pi_{\theta_k} - \pi_{\theta_{k+1}}\|_1$. Then, by plugging in Lemma 4 and Eq. (C.23) we end up with,

$$\begin{aligned}
& \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \parallel \pi_{\theta_{k+1}})] - \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \parallel \pi_{\theta_k})] \\
& \leq \varepsilon_k - (1 - \gamma)\beta^{-1} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \tag{C.54}
\end{aligned}$$

$$+ (1 - \gamma)\beta^{-1} \left(\frac{2c_3 M(1 - \gamma)\lambda}{\sqrt{k}} \right) + (\varepsilon'_k + \beta_k^{-2} U) \tag{C.55}$$

Rearrange Eq. (C.54), we have

$$\begin{aligned}
& (1 - \gamma)\beta^{-1} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \\
& \leq \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \parallel \pi_{\theta_k})] - \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \parallel \pi_{\theta_{k+1}})] \\
& \quad + \left(\frac{2c_3 M(1 - \gamma)^2 \lambda}{\beta \sqrt{k}} \right) + \varepsilon_k + \varepsilon'_k + \beta_k^{-2} U \tag{C.56}
\end{aligned}$$

And then telescoping Eq. (C.56) results in,

$$\begin{aligned}
& (1 - \gamma) \sum_{k=1}^K \beta^{-1} \min_{k \in [K]} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \\
& \leq (1 - \gamma) \sum_{k=1}^K \beta^{-1} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \\
& \leq \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \| \pi_0)] - \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \| \pi_K)] \\
& \quad + \lambda r_{\max} (1 - \gamma)^2 \sum_{k=1}^K \beta^{-1} \left(\frac{2c_3}{\sqrt{k}} \right) + U \sum_{k=1}^K \beta_k^{-2} \\
& \quad + \sum_{k=1}^K (\varepsilon_k + \varepsilon'_k) \tag{C.57}
\end{aligned}$$

We complete the final step in Eq.(C.57) by plugging in Lemma 4 and Eq. (C.20). Per the observation we make in the proof of Theorem 5.2,

1. $\mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \| \pi_0)] \leq \log \mathcal{A}$ due to the uniform initialization of policy.
2. $KL(\pi^* \| \pi_K)$ is a non-negative term.

We now have,

$$\begin{aligned}
& \min_{k \in [K]} J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k) \\
& \leq \frac{\log |\mathcal{A}| + UK\beta^{-2} + \sum_{k=1}^K (\varepsilon_k + \varepsilon'_k)}{(1 - \gamma)K\beta^{-1}} \\
& \quad + \lambda r_{\max} (1 - \gamma) \left(\frac{2c_3}{\sqrt{k}} \right) \tag{C.58}
\end{aligned}$$

Replacing β with $\beta_0 \sqrt{K}$ finishes the proof.

C.4.4 Proof of Theorem 5.2

In the following part, we focus the convergence of neural NPG. We first define the following terms under neural NPG update rule.

Lemma 10 [109] *For energy-based policy π_θ , we have policy gradient and Fisher information matrix,*

$$\nabla_\theta J(\pi_\theta) = \tau \mathbb{E}_{d_{\pi_\theta}(s,a)}[Q_{\pi_\theta}(s,a)(\phi_\theta(s,a) - \mathbb{E}_{\pi_\theta}[\phi_\theta(s,a')])] \quad (\text{C.59})$$

$$F(\theta) = \tau^2 \mathbb{E}_{d_{\pi_\theta}(s,a)}[(\phi_\theta(s,a) - \mathbb{E}_{\pi_\theta}[\phi_\theta(s,a')]) (\phi_\theta(s,a) - \mathbb{E}_{\pi_\theta}[\phi_\theta(s,a')])^\top] \quad (\text{C.60})$$

We then derive an upper bound for $J_\lambda^{y^*}(\pi^*) - J_\lambda^{y^*}(\pi_k)$ for the neural NPG method in the following lemma:

Lemma 11 *(One-step difference of π) It holds that, with probability of $1 - \delta$,*

$$\begin{aligned} & (1 - \gamma)(J_\lambda^{\hat{y}_k}(\pi^*) - J_\lambda^{\hat{y}_k}(\pi_k)) \\ & \leq \eta_{\text{NPG}}^{-1} \mathbb{E}_{s \sim \nu_{\pi^*}} [\text{KL}(\pi^* \| \pi_k) - \text{KL}(\pi^* \| \pi_{k+1})] + \\ & \quad \eta_{\text{NPG}}(9\Upsilon^2 + r_{\max}^2) + 2c_0\epsilon'_k + \eta_{\text{NPG}}^{-1}\epsilon''_k, \end{aligned} \quad (\text{C.61})$$

where (C.62)

$$\begin{aligned} \epsilon'_k &= \mathcal{O}(\Upsilon^3 m^{-1/2} \log(1/\delta) + \Upsilon^{5/2} m^{-1/4} \sqrt{\log(1/\delta)} \\ & \quad + \Upsilon r_{\max}^2 m^{-1/4} + \Upsilon^2 K_{\text{TD}}^{-1/2} + \Upsilon), \end{aligned} \quad (\text{C.63})$$

$$\begin{aligned} \epsilon''_k &= 8\eta_{\text{NPG}} \Upsilon^{1/2} c_0 \sigma_\xi^{1/2} T^{-1/4} \\ & \quad + \mathcal{O}((\tau_{k+1} + \eta_{\text{NPG}}) \Upsilon^{3/2} m^{-1/4} \\ & \quad + \eta_{\text{NPG}} \Upsilon^{5/4} m^{-1/8}), \end{aligned} \quad (\text{C.64})$$

c_0 is defined in Assumption 2 and σ_ξ is defined in Assumption 1. Meanwhile, Υ is the radius of the parameter space, m is the width of the neural network, and T is the sample batch size.

Proof 7 We start from the following,

$$KL(\pi^* || \pi_k) - KL(\pi^* || \pi_{k+1}) - KL(\pi_{k+1} || \pi_k) \quad (C.65)$$

$$= \mathbb{E}_{a \sim \pi^*} \left[\log\left(\frac{\pi_{k+1}}{\pi_k}\right) \right] - \mathbb{E}_{a \sim \pi_{k+1}} \left[\log\left(\frac{\pi_{k+1}}{\pi_k}\right) \right] \quad (C.66)$$

$$(by \text{ KL's definition}). \quad (C.67)$$

We now show the building blocks of the proof. First, we add and subtract a few terms to RHS of Eq. (C.66) then take the expectation of both sides with respect to $s \sim \nu_{\pi^*}$. Rearrange these terms, we get,

$$\mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* || \pi_k) - KL(\pi^* || \pi_{k+1}) - KL(\pi_{k+1} || \pi_k)] \quad (C.68)$$

$$= \eta_{\text{NPG}} \mathbb{E}_{s \sim \nu_{\pi^*}} [\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \mathbb{E}_{a \sim \pi_k} [\tilde{Q}_{\pi_k, \hat{y}_k}]] \quad (C.69)$$

$$+ H_k \quad (C.70)$$

where H_k is denoted by,

$$\begin{aligned} H_k := & \mathbb{E}_{s \sim \nu_{\pi^*}} \left[\mathbb{E}_{a \sim \pi^*} \left[\log\left(\frac{\pi_{k+1}}{\pi_k}\right) - \eta_{\text{NPG}} \tilde{Q}_{\omega_k} \right] \right. \\ & \left. - \mathbb{E}_{a \sim \pi_k} \left[\log\left(\frac{\pi_{k+1}}{\pi_k}\right) - \eta_{\text{NPG}} \tilde{Q}_{\omega_k} \right] \right] \end{aligned} \quad (C.71)$$

$$+ \eta_{\text{NPG}} \mathbb{E}_{s \sim \nu_{\pi^*}} [\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\omega_k} - \tilde{Q}_{\pi_k, \hat{y}_k}]] \quad (C.72)$$

$$- \mathbb{E}_{a \sim \pi_k} [\tilde{Q}_{\omega_k} - \tilde{Q}_{\pi_k, \hat{y}_k}] \quad (C.73)$$

$$+ \mathbb{E}_{s \sim \nu_{\pi^*}} [\mathbb{E}_{a \sim \pi_k} [\log\left(\frac{\pi_{k+1}}{\pi_k}\right)]] \quad (C.74)$$

$$- \mathbb{E}_{a \sim \pi_{k+1}} [\log\left(\frac{\pi_{k+1}}{\pi_k}\right)] \quad (C.75)$$

By Lemma 3, we have

$$\eta_{\text{NPG}} \mathbb{E}_{s \sim \nu_{\pi^*}} [\mathbb{E}_{a \sim \pi^*} [\tilde{Q}_{\pi_k, \hat{y}_k}] - \mathbb{E}_{a \sim \pi_k} [\tilde{Q}_{\pi_k, \hat{y}_k}]] \quad (\text{C.76})$$

$$= \eta_{\text{NPG}} (1 - \gamma) (J_{\lambda}^{\hat{y}_k}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \quad (\text{C.77})$$

Insert Eqs. (C.77) back to Eq. (C.70), we have,

$$\begin{aligned} & \eta_{\text{NPG}} (1 - \gamma) (J_{\lambda}^{\hat{y}_k}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \\ &= \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \parallel \pi_k) - KL(\pi^* \parallel \pi_{k+1}) - KL(\pi_{k+1} \parallel \pi_k)] \\ & \quad - H_k \\ &\leq \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \parallel \pi_k) - KL(\pi^* \parallel \pi_{k+1}) - KL(\pi_{k+1} \parallel \pi_k)] \\ & \quad + |H_k| \end{aligned} \quad (\text{C.78})$$

We reach the final inequality of Eq. (C.78) by algebraic manipulation. Second, we follow Lemma 5.5 of [109] and obtain an upper bound for Eq. (C.75). Specifically, with probability of $1 - \delta$,

$$\begin{aligned} & \mathbb{E}_{a \sim \text{init}} [|H_k| - \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi_{k+1} \parallel \pi_k)]] \\ &\leq \eta_{\text{NPG}}^2 (9\Upsilon^2 + r_{\max}^2) + 2\eta_{\text{NPG}} c_0 \epsilon'_k + \epsilon''_k \end{aligned} \quad (\text{C.79})$$

The expectation is taken over randomness. With these building blocks of Eqs. (C.78) and (C.79), we are now ready to reach the concluding inequality. Plugging Eqs. (C.79) back into Eq. (C.78), we end up with, with probability of $1 - \delta$,

$$\begin{aligned} & \eta_{\text{NPG}} (1 - \gamma) (J_{\lambda}^{\hat{y}_k}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \\ &\leq \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \parallel \pi_k) - KL(\pi^* \parallel \pi_{k+1})] \end{aligned} \quad (\text{C.80})$$

$$+ \eta_{\text{NPG}}^2 (9\Upsilon^2 + r_{\max}^2) + 2\eta_{\text{NPG}} c_0 \epsilon'_k + \epsilon''_k \quad (\text{C.81})$$

Dividing both sides of Eq. (C.81) by η_{NPG} completes the proof. The details are included in the Appendix.

We have the following Lemma to bound the error terms H_k defined in Eq. (C.75) of Lemma 11.

Lemma 12 [109]. *Under Assumptions 4, we have*

$$\begin{aligned} & \mathbb{E}_{a \sim \text{init}} \left[|H_k| - \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi_{k+1} \| \pi_k)] \right] \\ & \leq \eta_{\text{NPG}}^2 (9\Upsilon^2 + r_{\max}^2) + \eta_{\text{NPG}} (\varphi'_k + \psi'_k) \epsilon'_k + \epsilon''_k \end{aligned} \quad (\text{C.82})$$

$$(\text{C.83})$$

Here the expectation is taken over all the randomness. We have $\epsilon'_k := \|Q_{\omega_k} - Q_{\pi_k}\|_{\nu_{\pi_k}}^2$ and

$$\begin{aligned} \epsilon''_k &= \sqrt{2}\Upsilon^{1/2} \eta_{\text{NPG}} (\varphi_k + \psi_k) \tau_k^{-1} \left\{ \mathbb{E}_{(s,a) \sim \sigma_{\pi_{\theta_k}}} [\|\xi_k(\delta_k)\|_2^2] \right. \\ & \quad \left. + \mathbb{E}_{(s,a) \sim \sigma_{\pi_{\omega_k}}} [\|\xi_k(\omega_k)\|_2^2] \right\}^{1/2} \end{aligned} \quad (\text{C.84})$$

$$+ \mathcal{O}((\tau_{k+1} + \eta_{\text{NPG}}) \Upsilon^{3/2} m^{-1/4} + \eta_{\text{NPG}} \Upsilon^{5/4} m^{-1/8}). \quad (\text{C.85})$$

Recall $\xi_k(\omega_k)$ and $\xi_k(\delta_k)$ are defined in Assumption 1, while $\varphi_k, \psi_k, \varphi'_k$, and ψ'_k are defined in Assumption 2.

Please refer to [109] for complete proof. Finally, we are ready to show the proof for Theorem 5.2.

Proof 8 First, we combine Lemma 4 and 11 to get the following:

$$\begin{aligned} & (1 - \gamma) (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}^k}(\pi^*) + J_{\lambda}^{\hat{y}^k}(\pi^*) - J_{\lambda}^{\hat{y}^k}(\pi_k)) \\ & \leq \eta_{\text{NPG}}^{-1} \mathbb{E}_{s \sim \nu_{\pi^*}} [KL(\pi^* \| \pi_k) - KL(\pi^* \| \pi_{k+1})] \\ & \quad + \eta_{\text{NPG}} (9\Upsilon^2 + r_{\max}^2) + 2c_0 \epsilon'_k + \eta_{\text{NPG}}^{-1} \epsilon''_k \\ & \quad + \frac{2c_3 M (1 - \gamma)^2 \lambda}{\sqrt{k}} \end{aligned} \quad (\text{C.86})$$

We can then see this:

1. $\mathbb{E}_{s \sim \nu_{\pi^*}} [\text{KL}(\pi^* \parallel \pi_1)] \leq \log |\mathcal{A}|$ due to the uniform initialization of policy.

2. $\text{KL}(\pi^* \parallel \pi_{K+1})$ is a non-negative term.

And by setting $\eta_{\text{NPG}} = 1/\sqrt{K}$ and telescoping Eq. (C.86), we obtain,

$$\begin{aligned}
& (1 - \gamma) \min_{k \in [K]} (J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \\
& \leq (1 - \gamma) \frac{1}{K} \sum_{k=1}^K \mathbb{E}(J_{\lambda}^{y^*}(\pi^*) - J_{\lambda}^{\hat{y}_k}(\pi_k)) \\
& \leq \frac{1}{\sqrt{K}} (\mathbb{E}_{s \sim \nu_{\pi^*}} [\text{KL}(\pi^* \parallel \pi_1)] + 9\Upsilon^2 + r_{\max}^2) + \frac{1}{K} \sum_{k=1}^K \\
& \quad (2\sqrt{K}c_0\epsilon'_k + \eta_{\text{NPG}}^{-1}\epsilon''_k + \frac{2c_3M(1-\gamma)^2\lambda}{\sqrt{k}})
\end{aligned} \tag{C.87}$$

plug ϵ'_k and ϵ''_k defined in Lemma 11 into Eq.(C.87), and set ϵ_k as,

$$\begin{aligned}
\epsilon_k &= \sqrt{8}c_0\Upsilon^{1/2}\sigma_{\xi}^{1/2}T^{-1/4} \\
&+ \mathcal{O}((\tau_{k+1}K^{1/2} + 1)\Upsilon^{3/2}m^{-1/4} + \Upsilon^{5/4}m^{-1/8}) \\
&+ c_0\mathcal{O}(\Upsilon^3m^{-1/2}\log(1/\delta) + \Upsilon^{5/2}m^{-1/4}\sqrt{\log(1/\delta)}) \\
&+ \Upsilon r_{\max}^2m^{-1/4} + \Upsilon^2K_{\text{TD}}^{-1/2} + \Upsilon)
\end{aligned} \tag{C.88}$$

we complete the proof.

C.4.5 Proof of Lemma 1

Proof 9 First, we have $\mathbb{E}[G] = \frac{1}{1-\gamma}\mathbb{E}[R]$, i.e., the per-step reward R is an unbiased estimator of the cumulative reward G . Second, it is proved that $\mathbb{V}(G) \leq \frac{\mathbb{V}(R)}{(1-\gamma)^2}$ [14]. Given $\lambda \geq 0$,

summing up the above equality and inequality, we have

$$\frac{1}{(1-\gamma)} J_{\frac{\lambda}{(1-\gamma)}}(\pi) = \frac{1}{(1-\gamma)} \left(\mathbb{E}[R] - \frac{\lambda}{(1-\gamma)} \mathbb{V}(R) \right) \quad (\text{C.89})$$

$$\leq \mathbb{E}[G] - \lambda \mathbb{V}(G) = J_{\lambda}^G(\pi). \quad (\text{C.90})$$

It completes the proof.

C.4.6 Proof of Lemma 2

We first provide the supporting lemmas for Lemma 2. We define the local linearization of $f((s, a); \theta)$ defined in Eq. (5.4) at the initial point Θ_{init} as,

$$\hat{f}((s, a); \theta) = \frac{1}{\sqrt{m}} \sum_{v=1}^m b_v \mathbb{1}\{[\Theta_{\text{init}}]_v^\top(s, a) > 0\} [\theta]_v^\top(s, a) \quad (\text{C.91})$$

$$(\text{C.92})$$

We then define the following function spaces,

$$\mathcal{F}_{\Upsilon, m} := \left\{ \frac{1}{\sqrt{m}} \sum_{v=1}^m b_v \mathbb{1}\{[\Theta_{\text{init}}]_v^\top(s, a) > 0\} [\theta]_v^\top(s, a) : \|\theta - \Theta_{\text{init}}\|_2 \leq \Upsilon \right\}, \quad (\text{C.93})$$

and

$$\bar{\mathcal{F}}_{\Upsilon, m} := \left\{ \frac{1}{\sqrt{m}} \sum_{v=1}^m b_v \mathbb{1}\{[\Theta_{\text{init}}]_v^\top(s, a) > 0\} [\theta]_v^\top(s, a) : \|\theta - \Theta_{\text{init}}\|_\infty \leq \Upsilon / \sqrt{md} \right\}. \quad (\text{C.94})$$

$[\Theta_{\text{init}}]_r \sim \mathcal{N}(0, I_d/d)$ and $b_r \sim \text{Unif}(\{-1, 1\})$ are the initial parameters. By the definition, $\bar{\mathcal{F}}_{\Upsilon, m}$ is a subset of $\mathcal{F}_{\Upsilon, m}$. The following lemma characterizes the deviation of $\bar{\mathcal{F}}_{\Upsilon, m}$ from $\mathcal{F}_{\Upsilon, \infty}$.

Lemma 13 (*Projection Error*)[88]. *Let $f \in \mathcal{F}_{\Upsilon, \infty}$, where $\mathcal{F}_{\Upsilon, \infty}$ is defined in Assumption 3. For any $\delta > 0$, it holds with probability at least $1 - \delta$ that*

$$\|\Pi_{\bar{\mathcal{F}}_{\Upsilon, m}} f - f\|_{\varsigma} \leq \Upsilon m^{-1/2} [1 + \sqrt{2 \log(1/\delta)}] \quad (\text{C.95})$$

where ς is any distribution over $S \times A$.

Please refer to [88] for a detail proof.

Lemma 14 (*Linearization Error*) Under Assumption. 4, for all $\theta \in \mathcal{D}$, where $\mathcal{D} = \{\xi \in \mathbb{R}^{md} : \|\xi - \Theta_{\text{init}}\|_2 \leq \Upsilon\}$, it holds that,

$$\mathbb{E}_{\nu_{\pi}} \left[\left(f((s, a); \theta) - \hat{f}((s, a); \theta) \right)^2 \right] \leq \frac{4c_1 \Upsilon^3}{\sqrt{m}} \quad (\text{C.96})$$

where $c_1 = c \sqrt{\mathbb{E}_{\mathcal{N}(0, I_d/d)} [1/\|(s, a)\|_2^2]}$, and c is defined in Assumption. 4.

Proof 10 We start from the definitions in Eq. (5.4) and Eq. (C.92),

$$\begin{aligned} & \mathbb{E}_{\nu_{\pi}} \left[\left(f((s, a); \theta) - \hat{f}((s, a); \theta) \right)^2 \right] \\ &= \mathbb{E}_{\nu_{\pi}} \left[\left(\frac{1}{\sqrt{m}} \left| \sum_{v=1}^m \left((\mathbb{1}\{[\theta]_v^{\top}(s, a) > 0\} - \mathbb{1}\{[\Theta_{\text{init}}]_v^{\top}(s, a) > 0\}) b_v [\theta]_v^{\top}(s, a) \right) \right| \right)^2 \right] \\ &\leq \frac{1}{m} \mathbb{E}_{\nu_{\pi}} \left[\left(\sum_{v=1}^m \left(\left| \mathbb{1}\{[\theta]_v^{\top}(s, a) > 0\} - \mathbb{1}\{[\Theta_{\text{init}}]_v^{\top}(s, a) > 0\} \right| \left| b_v \right| \left| [\theta]_v^{\top}(s, a) \right| \right) \right)^2 \right] \end{aligned} \quad (\text{C.97})$$

The above inequality follows from $|\sum W| \leq \sum |W|$, where

$$W = (\mathbb{1}\{[\theta]_v^\top(s, a) > 0\} - \mathbb{1}\{[\Theta_{\text{init}}]_v^\top(s, a) > 0\}) \\ \times b_v[\theta]_v^\top(s, a).$$

The parameter Θ_{init} is defined in Eq. (5.5). Next, since $\mathbb{1}\{[\Theta_{\text{init}}]_v^\top(s, a) > 0\} \neq \mathbb{1}\{[\theta]_v^\top(s, a) > 0\}$, we have,

$$|[\Theta_{\text{init}}]_v^\top(s, a)| \leq |[\theta]_v^\top(s, a) - [\Theta_{\text{init}}]_v^\top(s, a)| \\ \leq \|[\theta]_v - [\Theta_{\text{init}}]_v\|_2, \quad (\text{C.98})$$

where we obtain the last inequality from the Cauchy-Schwartz inequality. We also assume that $\|(s, a)\|_2 \leq 1$ without loss of generality [62, 109]. Eq. (C.98) further implies that,

$$|\mathbb{1}\{[\theta]_v^\top(s, a) > 0\} - \mathbb{1}\{[\Theta_{\text{init}}]_v^\top(s, a) > 0\}| \\ \leq \mathbb{1}\{|[\Theta_{\text{init}}]_v^\top(s, a)| \leq \|[\theta]_v - [\Theta_{\text{init}}]_v\|_2\} \quad (\text{C.99})$$

Then plug Eq. (C.99) and the fact that $|b_v| \leq 1$ back to Eq. (C.97), we have the following,

$$\begin{aligned}
& \mathbb{E}_{\nu_\pi} \left[\left(f((s, a); \theta) - \hat{f}((s, a); \theta) \right)^2 \right] \\
& \leq \frac{1}{m} \mathbb{E}_{\nu_\pi} \left[\left(\sum_{v=1}^m \mathbb{1} \left\{ \left| [\Theta_{\text{init}}]_v^\top(s, a) \right| \leq \left\| [\theta]_v - [\Theta_{\text{init}}]_v \right\|_2 \right\} \right. \right. \\
& \quad \left. \left. \left| [\theta]_v^\top(s, a) \right| \right)^2 \right] \\
& \leq \frac{1}{m} \mathbb{E}_{\nu_\pi} \left[\left(\sum_{v=1}^m \mathbb{1} \left\{ \left| [\Theta_{\text{init}}]_v^\top(s, a) \right| \leq \left\| [\theta]_v - [\Theta_{\text{init}}]_v \right\|_2 \right\} \right. \right. \\
& \quad \left. \left(\left([\theta]_v - [\Theta_{\text{init}}]_v \right)^\top(s, a) \right) + \left| [\Theta_{\text{init}}]_v^\top(s, a) \right| \right)^2 \right] \\
& \leq \frac{1}{m} \mathbb{E}_{\nu_\pi} \left[\left(\sum_{v=1}^m \mathbb{1} \left\{ \left| [\Theta_{\text{init}}]_v^\top(s, a) \right| \leq \left\| [\theta]_v - [\Theta_{\text{init}}]_v \right\|_2 \right\} \right. \right. \\
& \quad \left. \left(\left\| [\theta]_v - [\Theta_{\text{init}}]_v \right\|_2 + \left| [\Theta_{\text{init}}]_v^\top(s, a) \right| \right)^2 \right] \\
& \leq \frac{1}{m} \mathbb{E}_{\nu_\pi} \left[\left(\sum_{v=1}^m \mathbb{1} \left\{ \left| [\Theta_{\text{init}}]_v^\top(s, a) \right| \leq \left\| [\theta]_v - [\Theta_{\text{init}}]_v \right\|_2 \right\} \right. \right. \\
& \quad \left. \left. 2 \left\| [\theta]_v - [\Theta_{\text{init}}]_v \right\|_2 \right)^2 \right] \tag{C.100}
\end{aligned}$$

We obtain the second inequality by the fact that $|A| \leq |A-B| + |B|$. Then follow the Cauchy-Schwartz inequality and $\|(s, a)\|_2 \leq 1$ we have the third equality. By inserting Eq. (C.98) we achieve the fourth inequality. We continue Eq. (C.100) by following the Cauchy-Schwartz

inequality and plugging $\|[\theta] - [\Theta_{\text{init}}]\|_2 \leq \Upsilon$,

$$\begin{aligned}
& \mathbb{E}_{\nu_\pi} \left[\left(f((s, a); \theta) - \hat{f}((s, a); \theta) \right)^2 \right] \\
& \leq \frac{4\Upsilon^2}{m} \mathbb{E}_{\nu_\pi} \left[\sum_{v=1}^m \mathbb{1}\{ |[\Theta_{\text{init}}]_v^\top(s, a)| \leq \|[\theta]_v - [\Theta_{\text{init}}]_v\|_2 \} \right] \\
& = \frac{4\Upsilon^2}{m} \sum_{v=1}^m P_{\nu_\pi} |[\Theta_{\text{init}}]_v^\top(s, a)| \leq \|[\theta]_v - [\Theta_{\text{init}}]_v\|_2 \\
& \leq \frac{4c\Upsilon^2}{m} \sum_{v=1}^m \frac{\|[\theta]_v - [\Theta_{\text{init}}]_v\|_2}{\|[\Theta_{\text{init}}]_v\|_2} \\
& \leq \frac{4c\Upsilon^2}{m} \left(\sum_{v=1}^m \|[\theta]_v - [\Theta_{\text{init}}]_v\|_2^2 \right)^{-1/2} \left(\sum_{v=1}^m \frac{1}{\|[\Theta_{\text{init}}]_v\|_2^2} \right)^{-1/2} \\
& \leq \frac{4c_1\Upsilon^3}{\sqrt{m}}
\end{aligned} \tag{C.101}$$

We obtain the second inequality by imposing Assumption 4 and the third by following the Cauchy-Schwartz inequality. Finally, we set $c_1 := c\sqrt{\mathbb{E}_{\mathcal{N}(0, I_d/d)}[1/\|(s, a)\|_2^2]}$. Thus, we complete the proof.

In the t -th iterations of TD iteration, we denote the temporal difference terms w.r.t $\hat{f}((s, a); \theta_t)$ and $f((s, a); \theta_t)$ as

$$\delta_t^0((s, a), (s, a)'; \theta_t) = \hat{f}((s, a)'; \theta_t) - \gamma \hat{f}((s, a); \theta_t) \tag{C.102}$$

$$- r_{s,a}, \tag{C.103}$$

$$\delta_t^\theta((s, a), (s, a)'; \theta_t) = f((s, a)'; \theta_t) - \gamma f((s, a); \theta_t) \tag{C.104}$$

$$- r_{s,a}. \tag{C.105}$$

For notation simplicity. in the sequel we write $\delta_t^0((s, a), (s, a)'; \theta_t)$ and $\delta_t^\theta((s, a), (s, a)'; \theta_t)$ as δ_t^0 and δ_t^θ . We further define the stochastic semi-gradient $g_t(\theta_t) := \delta_t^\theta \nabla_\theta f((s, a); \theta_t)$, its population mean $\bar{g}_t(\theta_t) := \mathbb{E}_{\nu_\pi}[g_t(\theta_t)]$. The local linearization of $\bar{g}_t(\theta_t)$ is $\hat{g}_t(\theta_t) := \mathbb{E}_{\nu_\pi}[\delta_t^0 \nabla_\theta \hat{f}((s, a); \theta_t)]$. We denote them as $g_t, \bar{g}_t, \hat{g}_t$ respectively for simplicity.

Lemma 15 *Under Assumption. 4, for all $\theta_t \in \mathcal{D}$, where $\mathcal{D} = \{\xi \in \mathbb{R}^{md} : \|\xi - \Theta_{init}\|_2 \leq \Upsilon\}$, it holds with probability of $1 - \delta$ that,*

$$\begin{aligned} & \|\bar{g}_t - \hat{g}_t\|_2 \\ &= \mathcal{O}\left(\Upsilon^{3/2}m^{-1/4}\left(1 + (m \log \frac{1}{\delta})^{-1/2}\right) + \Upsilon^{1/2}r_{\max}m^{-1/4}\right) \end{aligned} \quad (\text{C.106})$$

Proof 11 *By the definition of $\bar{g}_t$ and $\hat{g}_t$, we have*

$$\begin{aligned} & \|\bar{g}_t - \hat{g}_t\|_2^2 \\ &= \left\| \mathbb{E}_{\nu_\pi}[\delta_t^\theta \nabla_\theta f((s, a); \theta_t) - \delta_t^0 \nabla_\theta \hat{f}((s, a); \theta_t)] \right\|_2^2 \\ &= \left\| \mathbb{E}_{\nu_\pi}[(\delta_t^\theta - \delta_t^0) \nabla_\theta f((s, a); \theta_t) + \delta_t^0 (\nabla_\theta f((s, a); \theta_t) - \nabla_\theta \hat{f}((s, a); \theta_t))] \right\|_2^2 \\ &\leq 2\mathbb{E}_{\nu_\pi}[(\delta_t^\theta - \delta_t^0)^2 \|\nabla_\theta f((s, a); \theta_t)\|_2^2] + \\ &\quad 2\mathbb{E}_{\nu_\pi}[(|\delta_t^0| \|\nabla_\theta f((s, a); \theta_t) - \nabla_\theta \hat{f}((s, a); \theta_t)\|_2)^2] \end{aligned} \quad (\text{C.107})$$

$$(\text{C.108})$$

We obtain the inequality because $(A + B)^2 \leq 2A^2 + 2B^2$. We first upper bound $\mathbb{E}_{\nu_\pi}[(\delta_t^\theta - \delta_t^0)^2 \|\nabla_\theta f((s, a); \theta_t)\|_2^2]$ in Eq. (C.108). Since $\|(s, a)\|_2 \leq 1$, we have $\|\nabla_\theta f((s, a); \theta_t)\|_2 \leq 1$.

Then by definition, we have the following first inequality,

$$\begin{aligned}
& \mathbb{E}_{\nu_\pi} \left[\left(\delta_t^\theta - \delta_t^0 \right)^2 \left\| \nabla_\theta f((s, a); \theta_t) \right\|_2^2 \right] \\
& \leq \mathbb{E}_{\nu_\pi} \left[\left(f((s, a); \theta_t) - \hat{f}((s, a); \theta_t) - \gamma \left(f((s', a'); \theta_t) \right. \right. \right. \\
& \quad \left. \left. \left. - \hat{f}((s', a'); \theta_t) \right) \right)^2 \right] \\
& \leq \mathbb{E}_{\nu_\pi} \left[\left(\left| f((s, a); \theta_t) - \hat{f}((s, a); \theta_t) \right| + \left| f((s', a'); \theta_t) \right. \right. \right. \\
& \quad \left. \left. \left. - \hat{f}((s', a'); \theta_t) \right| \right)^2 \right] \tag{C.109}
\end{aligned}$$

$$\begin{aligned}
& \leq 2\mathbb{E}_{\nu_\pi} \left[\left(f((s, a); \theta_t) - \hat{f}((s, a); \theta_t) \right)^2 \right] + 2\mathbb{E}_{\nu_\pi} \\
& \quad \left[\left(f((s', a'); \theta_t) - \hat{f}((s', a'); \theta_t) \right)^2 \right] \\
& \leq 4\mathbb{E}_{\nu_\pi} \left[\left(f((s, a); \theta_t) - \hat{f}((s, a); \theta_t) \right)^2 \right] \leq \frac{16c_1\Upsilon^3}{\sqrt{m}} \tag{C.110}
\end{aligned}$$

We obtain the second inequality by $|\gamma| \leq 1$, then obtain the third inequality by the fact that $(A + B)^2 \leq 2A^2 + 2B^2$. We reach the final step by inserting Lemma 14. We then proceed to upper bound $\mathbb{E}_{\nu_\pi} \left[|\delta_t^0| \left\| \nabla_\theta f((s, a); \theta_t) - \nabla_\theta \hat{f}((s, a); \theta_t) \right\|_2 \right]$. From Hölder's inequality, we have,

$$\begin{aligned}
& \mathbb{E}_{\nu_\pi} \left[\left(|\delta_t^0| \left\| \nabla_\theta f((s, a); \theta_t) - \nabla_\theta \hat{f}((s, a); \theta_t) \right\|_2 \right)^2 \right] \\
& \leq \mathbb{E}_{\nu_\pi} \left[(\delta_t^0)^2 \right] \mathbb{E}_{\nu_\pi} \left[\left\| \nabla_\theta f((s, a); \theta_t) - \nabla_\theta \hat{f}((s, a); \theta_t) \right\|_2^2 \right] \tag{C.111}
\end{aligned}$$

We first derive an upper bound for first term in Eq.(C.111), starting from its definition,

$$\begin{aligned}
& \mathbb{E}_{\nu_\pi} [(\delta_t^0)^2] \\
&= \mathbb{E}_{\nu_\pi} \left[\left[\hat{f}((s', a'); \theta_t) - \gamma \hat{f}((s, a); \theta_t) - r_{s,a} \right]^2 \right] \\
&\leq 3\mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s', a'); \theta_t) \right)^2 \right] + 3\mathbb{E}_{\nu_\pi} \left[\left(\gamma \hat{f}((s, a); \theta_t) \right)^2 \right] \quad (\text{C.112}) \\
&\quad + 3\mathbb{E}_{\nu_\pi} \left[r_{s,a}^2 \right] \\
&\leq 6\mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s, a); \theta_t) \right)^2 \right] + 3r_{\max}^2 \\
&= 6\mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) + \hat{f}((s, a); \theta_{\pi^*}) \right. \right. \\
&\quad \left. \left. - Q_\pi + Q_\pi \right)^2 \right] + 3r_{\max}^2 \\
&\leq 18\mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right)^2 \right] + 18\mathbb{E}_{\nu_\pi} \\
&\quad \left[\left(\hat{f}((s, a); \theta_{\pi^*}) - Q_\pi \right)^2 \right] + 18\mathbb{E}_{\nu_\pi} \left[\left(Q_\pi \right)^2 \right] + 3r_{\max}^2 \\
&\leq 72\Upsilon^2 + 18\mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s, a); \theta_{\pi^*}) - Q_\pi \right)^2 \right] \\
&\quad + 21(1 - \gamma)^{-2}r_{\max}^2 \quad (\text{C.113})
\end{aligned}$$

We obtain the first and the third inequality by the fact that $(A + B + C)^2 \leq 3A^2 + 3B^2 + 3C^2$.

Recall $r_{\max}$ is the boundary for reward function r , which leads to the second inequality. We

obtain the last inequality in Eq. (C.113) following the fact that $|\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*})| \leq$

$\|\theta_t - \theta_{\pi^*}\| \leq 2\Upsilon$ and $Q_\pi \leq (1 - \gamma)^{-1}r_{\max}$. Since $\bar{\mathcal{F}}_{\Upsilon, m} \subset \mathcal{F}_{\Upsilon, m}$, by Lemma 13, we have,

$$E_{\nu_\pi} \left[\left(\hat{f}((s, a); \theta_{\pi^*}) - Q_\pi \right)^2 \right] \leq \frac{\Upsilon^2 (1 + \sqrt{2 \log(1/\delta)})^2}{m} \quad (\text{C.114})$$

$$(\text{C.115})$$

Combine Eq. (C.113) and Eq. (C.115), we have with probability of $1 - \delta$,

$$\begin{aligned} & \mathbb{E}_{\nu_\pi} [(\delta_t^0)^2] \\ & \leq 72\Upsilon^2(1 + \frac{\log(1/\delta)}{m}) + 21(1 - \gamma)^{-2}r_{\max}^2 \end{aligned} \quad (\text{C.116})$$

Lastly we have

$$\begin{aligned} & \mathbb{E}_{\nu_\pi} [\|\nabla_\theta f((s, a); \theta_t) - \nabla_\theta \hat{f}((s, a); \theta_t)\|_2^2] \\ & = \mathbb{E}_{\nu_\pi} \left[\left(\frac{1}{m} \sum_{v=1}^m (\mathbb{1}\{[\theta]_v^\top(s, a) > 0\} - \mathbb{1}\{[\Theta_{\text{init}}]_v^\top(s, a) > 0\})^2 (b_v)^2 \|(s, a)\|_2^2 \right) \right] \\ & \leq \mathbb{E}_{\nu_\pi} \left[\frac{1}{m} \sum_{v=1}^m (\mathbb{1}\{|\Theta_{\text{init}}]_v^\top(s, a)| \leq \|[\theta]_v - [\Theta_{\text{init}}]_v\|_2\}) \right] \\ & \leq \frac{c_1 \Upsilon}{\sqrt{m}} \end{aligned} \quad (\text{C.117})$$

We obtain the first inequality by following Eq. (C.99) and the fact that $|b_v| \leq 1$ and $\|(s, a)\|_2 \leq 1$. Then for the rest, we follow the similar argument in Eq. (C.101). To finish the proof, we plug Eq. (C.110), Eq. (C.116) and Eq. (C.117) back to Eq. (C.108),

$$\begin{aligned} & \|\bar{g}_t - \hat{g}_t\|_2^2 \\ & \leq 2 \left(\frac{16c_1 \Upsilon^3}{\sqrt{m}} + \left(72\Upsilon^2(1 + \frac{\log(1/\delta)}{m}) + 21(1 - \gamma)^{-2}r_{\max}^2 \right) \frac{c_1 \Upsilon}{\sqrt{m}} \right) \\ & = \frac{176c_1 \Upsilon^3}{\sqrt{m}} + \frac{144c_1 \Upsilon^3 \log(1/\delta)}{m^{3/2}} + \frac{42c_1 \Upsilon r_{\max}^2}{(1 - \gamma)^{-2} \sqrt{m}} \end{aligned} \quad (\text{C.118})$$

Then we have,

$$\begin{aligned}
& \|\bar{g}_t - \hat{g}_t\|_2 \\
& \leq \sqrt{\frac{176c_1\Upsilon^3}{\sqrt{m}} + \frac{144c_1\Upsilon^3 \log(1/\delta)}{m^{3/2}} + \frac{42c_1\Upsilon r_{\max}^2}{(1-\gamma)^{-2}\sqrt{m}}} \\
& \leq \sqrt{\frac{176c_1\Upsilon^3}{\sqrt{m}}} + \sqrt{\frac{144c_1\Upsilon^3 \log(1/\delta)}{m^{3/2}}} + \sqrt{\frac{42c_1\Upsilon r_{\max}^2}{(1-\gamma)^{-2}\sqrt{m}}} \\
& = \mathcal{O}\left(\Upsilon^{3/2}m^{-1/4}\left(1 + \left(m \log \frac{1}{\delta}\right)^{-1/2}\right) + \Upsilon^{1/2}r_{\max}m^{-1/4}\right) \tag{C.119}
\end{aligned}$$

Next, we provide the following lemma to characterize the variance of g_t .

Lemma 16 (*Variance of the Stochastic Update Vector*)[62]. *There exists a constant $\xi_g^2 = \mathcal{O}(\Upsilon^2)$ independent of t . such that for any $t \leq T$, it holds that*

$$\mathbb{E}_{\nu_\pi}[\|g_t(\theta_t) - \bar{g}_t(\theta_t)\|_2^2] \leq \xi_g^2 \tag{C.120}$$

A detailed proof can be found in [62]. Now we provide the proof for Lemma 2.

Proof 12

$$\begin{aligned}
& \|\theta_{t+1} - \theta_{\pi^*}\|_2^2 \\
& = \|\Pi_{\mathcal{D}}(\theta_t - \eta g_t(\theta_t)) - \Pi_{\mathcal{D}}(\theta_{\pi^*} - \eta \hat{g}_t(\theta_{\pi^*}))\|_2^2 \\
& \leq \|(\theta_t - \theta_{\pi^*}) - \eta(g_t(\theta_t) - \hat{g}_t(\theta_{\pi^*}))\|_2^2 \\
& = \|\theta_t - \theta_{\pi^*}\|_2^2 - 2\eta(g_t(\theta_t) - \hat{g}_t(\theta_{\pi^*}))^\top (\theta_t - \theta_{\pi^*}) \\
& \quad + \eta^2 \|g_t(\theta_t) - \hat{g}_t(\theta_{\pi^*})\|_2^2 \tag{C.121}
\end{aligned}$$

The inequality holds due to the definition of $\Pi_{\mathcal{D}}$. We first upper bound $\|g_t(\theta_t) - \hat{g}_t(\theta_{\pi^*})\|_2^2$ in Eq. (C.121),

$$\begin{aligned}
& \|g_t(\theta_t) - \hat{g}_t(\theta_{\pi^*})\|_2^2 \\
&= \|g_t(\theta_t) - \bar{g}_t(\theta_t) + \bar{g}_t(\theta_t) - \hat{g}_t(\theta_t) + \hat{g}_t(\theta_t) - \hat{g}_t(\theta_{\pi^*})\|_2^2 \\
&\leq 3 \left(\|g_t(\theta_t) - \bar{g}_t(\theta_t)\|_2^2 + \|\bar{g}_t(\theta_t) - \hat{g}_t(\theta_t)\|_2^2 + \right. \\
&\quad \left. \|\hat{g}_t(\theta_t) - \hat{g}_t(\theta_{\pi^*})\|_2^2 \right) \tag{C.122}
\end{aligned}$$

The inequality holds due to fact that $(A + B + C)^2 \leq 3A^2 + 3B^2 + 3C^2$. Two of the terms on the right hand side of Eq. (C.122) are characterized in Lemma 15 and Lemma 16. We therefore characterize the remaining term,

$$\begin{aligned}
& \|\hat{g}_t(\theta_t) - \hat{g}_t(\theta_{\pi^*})\|_2^2 \\
&= \mathbb{E}_{\nu_{\pi}} \left[(\delta_t^0(\theta_t) - \delta_t^0(\theta_{\pi^*}))^2 \|\nabla_{\theta} \hat{f}((s, a); \theta_t)\|_2^2 \right] \\
&\leq \mathbb{E}_{\nu_{\pi}} \left[\left(\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right) - \gamma \left(\hat{f}((s', a'); \theta_t) \right. \right. \right. \\
&\quad \left. \left. \left. - \hat{f}((s', a'); \theta_{\pi^*}) \right) \right)^2 \right] \\
&\leq \mathbb{E}_{\nu_{\pi}} \left[\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right)^2 \right] + 2\gamma \mathbb{E}_{\nu_{\pi}} \\
&\quad \left[\left(\hat{f}((s', a'); \theta_t) - \hat{f}((s', a'); \theta_{\pi^*}) \right) \left(\hat{f}((s, a); \theta_t) \right. \right. \\
&\quad \left. \left. - \hat{f}((s, a); \theta_{\pi^*}) \right) \right] \\
&\quad + \gamma^2 \mathbb{E}_{\nu_{\pi}} \left[\left(\hat{f}((s', a'); \theta_t) - \hat{f}((s', a'); \theta_{\pi^*}) \right)^2 \right] \tag{C.123}
\end{aligned}$$

We obtain the first inequality by the fact that $\|\nabla_{\theta} \hat{f}((s, a); \theta_t)\|_2 \leq 1$. Then we use the fact that (s, a) and (s', a') have the same marginal distribution as well as $\gamma < 1$ for the second inequality. Follow the Cauchy-Schwarz inequality and the fact that (s, a) and (s', a') have

the same marginal distribution, we have

$$\begin{aligned}
& \mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s', a'); \theta_t) - \hat{f}((s', a'); \theta_{\pi^*}) \right) \left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right) \right] \\
& \leq \mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s', a'); \theta_t) - \hat{f}((s', a'); \theta_{\pi^*}) \right) \right] \mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right) \right] \\
& = \mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s', a'); \theta_t) - \hat{f}((s', a'); \theta_{\pi^*}) \right)^2 \right] \tag{C.124}
\end{aligned}$$

We plug Eq. (C.124) back to Eq. (C.123),

$$\begin{aligned}
& \left\| \hat{g}_t(\theta_t) - \hat{g}_t(\theta_{\pi^*}) \right\|_2^2 \\
& \leq (1 + \gamma)^2 \mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right)^2 \right]. \tag{C.125}
\end{aligned}$$

Next, we upper bound $(g_t(\theta_t) - \hat{g}_t(\theta_{\pi^*}))^\top (\theta_t - \theta_{\pi^*})$. We have,

$$\begin{aligned}
& (g_t(\theta_t) - \hat{g}_t(\theta_{\pi^*}))^\top (\theta_t - \theta_{\pi^*}) \\
& = (g_t(\theta_t) - \bar{g}_t(\theta_t))^\top (\theta_t - \theta_{\pi^*}) + (\bar{g}_t(\theta_t) - \hat{g}_t(\theta_t))^\top (\theta_t - \theta_{\pi^*}) \\
& \quad + (\hat{g}_t(\theta_t) - \hat{g}_t(\theta_{\pi^*}))^\top (\theta_t - \theta_{\pi^*}) \tag{C.126}
\end{aligned}$$

One term on the right hand side of Eq. (C.126) are characterized by Lemma 16. We continue to characterize the remaining terms. First, by Hölder's inequality, we have

$$\begin{aligned}
& (\bar{g}_t(\theta_t) - \hat{g}_t(\theta_t))^\top (\theta_t - \theta_{\pi^*}) \\
& \geq - \left\| \bar{g}_t(\theta_t) - \hat{g}_t(\theta_t) \right\|_2 \left\| \theta_t - \theta_{\pi^*} \right\|_2 \\
& \geq - 2\Upsilon \left\| \bar{g}_t(\theta_t) - \hat{g}_t(\theta_t) \right\|_2 \tag{C.127}
\end{aligned}$$

We obtain the second inequality since $\|\theta_t - \theta_{\pi^*}\|_2 \leq 2\Upsilon$ by definition. For the last term,

$$\begin{aligned}
& (\hat{g}_t(\theta_t) - \hat{g}_t(\theta_{\pi^*}))^\top (\theta_t - \theta_{\pi^*}) \\
&= \mathbb{E}_{\nu_\pi} \left[\left(\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right) - \gamma \left(\hat{f}((s', a'); \theta_t) \right. \right. \right. \\
&\quad \left. \left. \left. - \hat{f}((s', a'); \theta_{\pi^*}) \right) \right) \left(\nabla_\theta \hat{f}((s, a); \theta_t) \right)^\top (\theta_t - \theta_{\pi^*}) \right] \\
&= \mathbb{E}_{\nu_\pi} \left[\left(\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right) - \gamma \left(\hat{f}((s', a'); \theta_t) \right. \right. \right. \\
&\quad \left. \left. \left. - \hat{f}((s', a'); \theta_{\pi^*}) \right) \right) \left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right) \right] \\
&\geq \mathbb{E}_{\nu_\pi} \left[\left(\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right) \right)^2 \right] - \\
&\quad \gamma \mathbb{E}_{\nu_\pi} \left[\left(\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right) \right)^2 \right] \\
&= (1 - \gamma) \mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right)^2 \right], \tag{C.128}
\end{aligned}$$

where the inequality follows from Eq. (C.124). Combine Eqs. (C.121), (C.122), (C.125), (C.126), (C.127) and (C.128), we have,

$$\begin{aligned}
& \|\theta_{t+1} - \theta_{\pi^*}\|_2^2 \\
&\leq \|\theta_t - \theta_{\pi^*}\|_2^2 - (2\eta(1 - \gamma) - 3\eta^2(1 + \gamma)^2) \\
&\quad \mathbb{E}_{\nu_\pi} \left[\left(\hat{f}((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right)^2 \right] \\
&\quad + 3\eta^2 \|\bar{g}_t - \hat{g}_t\|_2^2 + 4\eta\Upsilon \|\bar{g}_t - \hat{g}_t\|_2 + 4\Upsilon\eta |\xi_g| \\
&\quad + 3\eta^2 \xi_g^2 \tag{C.129}
\end{aligned}$$

We then bound the error terms by rearrange Eq. (C.129). First, we have, with probability of $1 - \delta$,

$$\begin{aligned}
& \mathbb{E}_{\nu_\pi} \left[\left(f((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right)^2 \right] \\
&= \mathbb{E}_{\nu_\pi} \left[\left(f((s, a); \theta_t) - \hat{f}((s, a); \theta_t) + \hat{f}((s, a); \theta_t) \right. \right. \\
&\quad \left. \left. - \hat{f}((s, a); \theta_{\pi^*}) \right)^2 \right] \\
&\leq 2 \mathbb{E}_{\nu_\pi} \left[\left(f((s, a); \theta_t) - \hat{f}((s, a); \theta_t) \right)^2 + \left(\hat{f}((s, a); \theta_t) \right. \right. \\
&\quad \left. \left. - \hat{f}((s, a); \theta_{\pi^*}) \right)^2 \right] \\
&\leq (\eta(1 - \gamma) - 1.5\eta^2(1 + \gamma)^2)^{-1} \left(\|\theta_t - \theta_{\pi^*}\|_2^2 \right. \\
&\quad \left. - \|\theta_{t+1} - \theta_{\pi^*}\|_2^2 + 4\Upsilon\eta|\xi_g| + 3\eta^2\xi_g^2 \right) + \epsilon_g
\end{aligned} \tag{C.130}$$

where

$$\begin{aligned}
\epsilon_g &= \mathcal{O}(\Upsilon^3 m^{-1/2} \log(1/\delta) + \Upsilon^{5/2} m^{-1/4} \sqrt{\log(1/\delta)}) \\
&\quad + \Upsilon r_{\max}^2 m^{-1/4})
\end{aligned} \tag{C.131}$$

We obtain the first inequality by the fact that $(A + B)^2 \leq 2A^2 + 2B^2$. Then by Eq. (C.129), Lemma 14 and Lemma 15, we reach the final inequality. By telescoping Eq. (C.130) for $t = 1$ to T , we have, with probability of $1 - \delta$,

$$\begin{aligned}
& \|f((s, a); \theta_T) - \hat{f}((s, a); \theta_{\pi^*})\|^2 \\
&\leq \frac{1}{T} \sum_{t=1}^T \mathbb{E}_{\nu_\pi} \left[\left(f((s, a); \theta_t) - \hat{f}((s, a); \theta_{\pi^*}) \right)^2 \right] \\
&\leq T^{-1} (2\eta(1 - \gamma) - 3\eta^2(1 + \gamma)^2)^{-1} (\|\theta_{\text{init}} - \theta_{\pi^*}\| + \\
&\quad 4\Upsilon T\eta|\xi_g| + 3T\eta^2\xi_g^2) + \epsilon_g
\end{aligned} \tag{C.132}$$

Set $\eta = \min\{1/\sqrt{T}, (1-\gamma)/3(1+\gamma)^2\}$, which implies that $T^{-1/2}(2\eta(1-\gamma) - 3\eta^2(1+\gamma)^2)^{-1} \leq 1/(1-\gamma)^2$, then we have, with probability of $1 - \delta$,

$$\begin{aligned}
& \|f((s, a); \theta_T) - \hat{f}((s, a); \theta_{\pi^*})\| \\
& \leq \frac{1}{(1-\gamma)^2\sqrt{T}} (\|\Theta_{\text{init}} - \theta_{\pi^*}\|_2^2 + 4\Upsilon\sqrt{T}|\xi_g| \\
& \quad + 3\xi_g^2) + \epsilon_g \\
& \leq \frac{\Upsilon^2 + 4\Upsilon\sqrt{T}|\xi_g| + 3\xi_g^2}{(1-\gamma)^2\sqrt{T}} + \epsilon_g \\
& = \mathcal{O}(\Upsilon^3 m^{-1/2} \log(1/\delta) + \Upsilon^{5/2} m^{-1/4} \sqrt{\log(1/\delta)} \\
& \quad + \Upsilon r_{\max}^2 m^{-1/4} + \Upsilon^2 T^{-1/2} + \Upsilon)
\end{aligned} \tag{C.133}$$

We obtain the second inequality by the fact that $\|\Theta_{\text{init}} - \theta_{\pi^*}\|_2 \leq \Upsilon$. Then by definition we replace $\tilde{Q}_{\omega_k}$ and $\tilde{Q}_{\pi_k}$

C.5 Additional method comparisons

C.5.1 Global Optimality of Policy Search Methods

A major challenge of existing RL research is the lack of theoretical justification, such as sample complexity analysis, mainly because the objective function of policy search in RL is often nonconvex. It is challenging to determine if a policy search approach is guaranteed to reach the global optimal. Besides, the RL architecture components are usually parameterized by neural networks in practice. Its nonlinearity and complex nature render the analysis significantly difficult [124].

The theoretical understanding of policy gradient methods is also under tentative study. Work on this topic has been done mostly in tabular and linear parametrization settings for different variants of policy gradient. For example, [21] and [93] establish a non-asymptotic

convergence guarantee for natural policy gradient (NPG, [52]) and trusted region policy optimization (TRPO, [91]), respectively. [74] show converge rate for softmax parametrization, while [1] analyze multiple variants of policy gradient. On the other side of the spectrum, [62, 109] prove the global convergence and optimality of various policy gradient algorithms with over-parameterized neural networks. Furthermore, [124] apply the global optimality analysis to variance-constrained actor-critic risk-averse control with cumulative average rewards, and proposed a corresponding variance-constrained actor-critic (VARAC) algorithm. However, the analysis procedure is complicated due to the risk constraints on cumulative rewards, and the algorithm’s experimental performance remains unverified. Therefore, it remains interesting if there can be simplified global optimality analysis with verifiable experimental studies for risk-averse policy search methods.

C.5.2 Over-parameterized Neural Networks in RL

Overparameterization, a technique of deploying more parameters than necessary, improves the performance of neural networks [120]. The learning ability and generalization of over-parameterized neural networks have been studied extensively [36, 3, 8]. Integration with over-parameterized neural networks can be found in multiple RL topics. One line of work is to prove the global optimality of RL algorithms in a non-linear approximation setting [109, 62, 124]. They use ReLU activation over-parameterized neural networks with policy gradient methods such as NPG and PPO. Our work also belongs to this category. Other works include [43], which also deploy a two-layered ReLU activation over-parameterized neural network on mean-field multi-agent reinforcement learning problem. Regularization with over-parameterized neural networks is also investigated recently [55, 90].