Capturing Spatiotemporal and Subgrid Variability in Global Land Surface Albedo Parameterization

Akarsh Ralhan

A scholarly paper in partial fulfillment of the requirements for the degree of

Master of Science

November 2025

Department of Atmospheric and Oceanic Science, University of Maryland

College Park, Maryland

Advisor: Dr. Xin-Zhong Liang

Table of Contents

Abstract

Accurate surface albedo parameterization is critical for modeling Earth's energy balance, yet many schemes rely on static look-up tables or semi-empirical formulations that fail to capture spatiotemporal variations and complex radiative interactions. This study develops a physics-informed machine learning parameterization using 19 years (2003–2021) of MODIS BRDF data to predict direct and diffuse albedo in visible and near-infrared bands across major land cover categories (LCCs). The framework leverages ten biogeophysical predictors, including solar geometry, vegetation state, soil moisture, soil texture, topography, and background climate. It comprises a dynamic component resolving physical spatiotemporal patterns and a static correction, the Surface Albedo Localization Factor (SALF), which accounts for subgrid heterogeneity, together explaining most observed variability. The parameterization shows strong agreement with MODIS albedo (overall $R^2 = 0.81$, MAPE = 0.08), with performance ranging from $R^2 = 0.72$ in diffuse visible to $R^2 = 0.89$ in diffuse near-infrared, and generally higher accuracy in the near-infrared part. It performs well across diverse LCCs, including grasslands, shrublands, croplands, and challenging barren regions where empirical methods underperform. SALF improves accuracy across all albedo parts (average R^2 increase of 0.11 and MAPE reduction of 0.025). Feature importance analysis highlights solar zenith angle and leaf area index as dominant dynamic drivers, with soil texture and topography shaping static variability. Counterfactual experiments confirm biophysically consistent albedo responses, enhancing interpretability and model trust. This framework offers a physically grounded alternative to empirical schemes and holds strong potential for integration into Earth system models to improve surface energy exchange representation.

Acknowledgements

This work was supported by the United States Department of Agriculture - National Institute of Food and Agriculture grant number 20206801231674, awarded to the University of Maryland for the creation of a predictive decision support Dashboard for Agricultural Water use and Nutrient management (DAWN). Computing and data storage resources were provided by the Computational and Information Systems Laboratory (CISL) at the NSF National Center for Atmospheric Research (NCAR). We acknowledge the use of OpenAI's ChatGPT for assistance in developing and refining the code used in analysis and visualization.

List of Tables

List of Figures

Chapter 1. Introduction

Land surface albedo is the fraction of incoming solar radiation reflected by the Earth's surface. It critically regulates the surface energy balance by determining the amount of solar energy available for absorption. This control over energy partitioning affects both sensible and latent heat fluxes, which in turn influence surface temperature and soil moisture dynamics. These changes directly modify soil and plant canopy microclimates, with cascading effects on ecosystem processes including photosynthesis, evapotranspiration, and carbon cycling (Bala et al., 2007; Chapin et al., 2005; Kim et al., 2020; S. Wang & Davidson, 2007; Zhu & Zeng, 2015). At larger scales, albedo-mediated processes govern local and regional energy and water budgets while, through atmospheric feedback and teleconnections, influencing broader circulation patterns, hydrological cycles, and weather systems (Betts, 2000; Shuai et al., 2020; X. Zhang et al., 2022). The parameter's climate sensitivity is particularly notable; even small albedo errors (0.02-0.05) can generate significant biases in energy and hydrological budgets, especially at regional scales (Z. Wang et al., 2018; J. Yang et al., 2022). This sensitivity underscores albedo's essential role as a key parameter in both global and regional weather, climate, and Earth system models.

Surface albedo is governed by a complex interplay of solar geometry, soil, vegetation, and topographic factors, each introducing distinct yet interconnected influences. Solar geometry, including parameters such as solar zenith and azimuth angles (F. Yang et al., 2008; Z. Zheng et al., 2017), interacts with surface characteristics through complex radiative transfer processes. At the surface, soil characteristics exert substantial control over reflectance, with key determinants including water content (K. Wang et al.,

2005), mineral composition (Baumgardner et al., 1986), surface roughness (Matthias et al., 2000), particle size distribution (Sadeghi et al., 2018), and soil color (Post et al., 2000). Vegetation adds another layer of complexity, with factors such as leaf area index (Alibakhshi et al., 2020), canopy architecture (Nath & Ni-Meister, 2021), plant functional types (Leonardi et al., 2015), and the physiological and phenological states of plants (Halim et al., 2019) all contributing to dynamic changes in surface reflectance. Spectrally, green vegetation absorbs strongly in the visible range (0.4–0.7 μm) for photosynthesis, while reflecting and transmitting in the near-infrared (0.7–4.0 μm) because of high scattering by leaf tissues (Scott et al., 1968), necessitating models to distinguish between direct (Dir) and diffuse (Dif) albedos in visible (Vis) versus near-infrared (Nir) bands (Liang, et al., 2005b). Finally, topographic features, such as slope, aspect, and elevation (Hao et al., 2018; Ryu et al., 2008), and landscape heterogeneity, affect spatial variability in solar radiation interception and reflection.

Given the diverse and interdependent controls on surface albedo, from atmospheric conditions and topography to vegetation structure and soil properties, accurately resolving its dynamics in large-scale numerical models remains a major challenge. Many of these factors operate at spatial and temporal scales far finer than model resolutions, involving complex and scale-dependent interactions (e.g., pointwise vs. regional, diurnal vs. seasonal) that are not explicitly represented. As a result, land surface models (LSMs) rely on simplified albedo parameterizations, often through static lookup tables, empirically fitted equations, or limited treatment of key variables. These approaches fail to capture the full complexity of surface–radiation interactions and introduce considerable uncertainty into climate predictions (D. Xiao et al., 2011; Z. Zheng et al., 2015). Table S1 summarizes

surface albedo representations in major LSMs worldwide. For example, CLM5 and HTESSEL assign fixed soil color classes or background maps, offering limited sensitivity to soil moisture or solar zenith angle, while Noah-MP applies basic wetness adjustments but neglects angular dependence in bare soil reflectance. Vegetation albedo in models like ORCHIDEE and JSBACH is derived from static LAI or land types, limiting the ability to represent seasonal, structural, or physiological changes. Even models that use two-stream radiative transfer (e.g., JULES, CLM5) simplify canopy geometry and parameterization, resulting in biases under variable illumination or varying vegetation density. Consequently, improving albedo parameterizations is essential for reducing land–atmosphere coupling errors and enhancing the reliability of weather and climate predictions across spatial and temporal scales (Liang, et al., 2005b; Rotenberg & Yakir, 2010; Wei et al., 2016).

Numerous studies have developed albedo parameterization schemes for various LSMs by conducting near-surface observation experiments. These efforts have focused on collecting radiation and albedo data under diverse surface conditions across different regions (Guan et al., 2009; Kala et al., 2014; Sugathan et al., 2014; H. Wang et al., 2024; G. Zheng et al., 2018; Z. Zheng et al., 2015). While these studies offer valuable insights into albedo dynamics under specific conditions and enhance model performance, their findings are often not transferable or adequately tested across diverse regions and climatic zones. The spatial heterogeneity and temporal variability of surface properties pose a significant challenge in developing globally applicable parameterization schemes solely based on local measurements. To address these limitations, satellite data have become indispensable for albedo parameterization at regional and global scales, providing consistent, high-resolution, and spatially extensive datasets. In particular, the Moderate

Resolution Imaging Spectroradiometer (MODIS) measurements facilitate the accurate retrieval of direct and diffuse albedo for visible and near-infrared using a semi-empirical kernel-driven Bidirectional Reflectance Distribution Function (BRDF) model with multi-date, multi-spectral, cloud-free, atmosphere-corrected surface reflectance (Lucht et al., 2000). They are crucial to capture surface reflectance across a wide range of land types and climates, track seasonal and interannual variability, and support the development of scalable, globally applicable parameterization schemes. This capability effectively overcomes the constraints of localized observations (Jin et al., 2011; Liang, et al., 2005b; Rechid et al., 2009). Therefore, satellite remote sensing remains the only practical means of routinely generating regional or global albedo datasets.

Liang, et al., (2005b) successfully demonstrated the use of satellite data for detailed albedo parameterization. However, that study focused only on the CONUS domain and was limited to three years of available MODIS data. Building on their foundation, with the availability of more comprehensive data, this study develops a globally applicable parameterization for snow-free land surface albedo using a physics-informed machine learning (ML) approach. This framework improves albedo representation in numerical models by leveraging 19 years of MODIS BRDF data (2003–2021) and incorporating additional key variables previously unconsidered. It provides separate parameterization models for different land types and albedo parts, distinguishing direct and diffuse radiation across visible and near-infrared parts. By integrating physical principles with ML, the framework effectively captures non-linear relationships between albedo and key drivers such as solar zenith angle, soil moisture, texture, topography, and vegetation. This results

in a more robust and accurate albedo representation that is generalized for global applications.

Chapter 2. Physical Processes Governing Land Surface Albedo

Capturing the full complexity of the physical interactions that influence land surface albedo dynamics remains a significant challenge. To address this, we focus on ten key parameters chosen based on three criteria: (1) their well-established importance in controlling albedo dynamics, as supported by multiple studies; (2) their global data availability at high spatial and temporal resolution; and (3) their inclusion as output variables in most land, weather, and climate models. This selection strategy ensures that the new parameterization is both physically robust and practically applicable across different modeling systems. For each parameter, we systematically examine its dynamic role, the underlying physical processes it represents, and its relative contribution to albedo variability. Together, these analyses provide a solid foundation for incorporating these parameters into the proposed parameterization scheme.

Building on the methodology of Liang, et al., (2005b), we develop distinct albedo models for each LCC, recognizing that broad-scale vegetation categories exhibit unique albedo dynamics. The physical processes discussed in this section represent general interactions governing the surface, and they vary substantially across LCC, seasons, and biomes. While these principles establish a conceptual foundation for understanding albedo dynamics, they do not constitute a universal parameterization framework, as specific relationships may differ or become negligible under varying environmental and surface conditions.

Among these governing factors, solar geometry emerges as a primary driver of albedo variability. Numerous studies highlight the dominant role of solar zenith angle

(SZA) in driving diurnal albedo variations for bare soils, often surpassing soil moisture in influence across arid and semi-arid regions with low soil moisture (Liang, et al., 2005b; H. Liu, Wang, et al., 2008; Roxy et al., 2010; F. Yang et al., 2008; Z. Zheng et al., 2017). For vegetated surfaces (e.g., croplands, grasslands, forests), SZA further modulates interactions with plant morphology and leaf structure, governing light interception and scattering dynamics (H. Liu, Tu, et al., 2008; H. Wang et al., 2021). The solar azimuth angle (SAA) also plays a critical role in albedo modulation over complex terrain, where combined SZA and SAA effects dictate shadow orientation, sub-grid solar radiation patterns, and surface-terrain interactions. This topographic sensitivity is prominent in mountainous regions, where shadowing alters solar radiation distribution and albedo.

Beyond solar geometry, surface soil moisture content (SMC) is a critical driver of surface albedo. Empirical studies consistently demonstrate an exponential decrease in albedo with increasing SMC, as wet soils exhibit lower reflectance due to enhanced light absorption by water-filled pores and reduced surface scattering (Guan et al., 2009; Z. Li et al., 2019; H. Liu, Wang, et al., 2008; K. Wang et al., 2005). This relationship is more evident in arid and semi-arid regions, where low baseline soil moisture amplifies the albedo sensitivity to even minor SMC changes (Lobell & Asner, 2002; Roxy et al., 2010). SMC varies substantially in both space and time, and its relationships with albedo can manifest as a linear decrease, exponential decline, or constant value beyond a saturation threshold, depending on the local SMC range.

Horizontal extent and vertical density can characterize vegetation canopy (Liang, et al., 2005a), quantified respectively by the fractional vegetation cover (FVEG) and the Leaf Area Index (LAI). LAI, defined as half the total green leaf area per unit of horizontal

ground surface area (Fang et al., 2019), represents the vertical structure of the canopy. In this study, FVEG is used to weigh the relative contributions of bare and vegetated surfaces, while LAI plays a vital role in modulating canopy albedo. This influence stems from vegetation's ability to alter canopy structure and radiation interception dynamics (Loew et al., 2014; W. Yang et al., 2006). The relationship between LAI and albedo is both inverse and nonlinear in the visible spectrum: albedo typically decreases with increasing LAI due to greater absorption of visible light by chlorophyll. Still, it asymptotically approaches a lower bound at higher LAI values as canopy saturation limits further absorption. In contrast, in the Nir part, albedo generally increases with LAI due to enhanced scattering by leaf cell structures (Tian et al., 2000). This relationship varies by vegetation type and season, with the strongest coupling observed during growing seasons when vegetation is most photosynthetically active (Y. Li et al., 2016; L. Wang et al., 2022). In sparse vegetation or early growth stages, albedo is highly sensitive to LAI changes due to the large contrast between vegetation and background soil reflectance (Rechid et al., 2009). However, as LAI exceeds 3, additional leaf layers have diminishing effects on albedo due to canopy saturation (Hollinger et al., 2010).

While SZA, SAA, SMC, and LAI represent time-dependent drivers of albedo, static factors such as geographic location and soil properties also play a critical role. Latitude and longitude are frequently used as proxies to capture geographic influences on albedo, including regional climate patterns and solar geometry variations (Alibakhshi et al., 2020; Hovi et al., 2019; Lukeš et al., 2014; Post et al., 2000). However, knowledge learned from specific geographic coordinates of limited data grids cannot be generalized globally. Thus, we deliberately excluded grid coordinates to avoid confounding the physical relationships

driven by local surface characteristics. Soil properties, including color (Galvdo, 1997; Post et al., 2000), surface roughness (Cierniewski et al., 2015; Matthias et al., 2000), and composition (e.g., silt, clay, sand content), further modulate albedo through their impact on light scattering and absorption. Additionally, subsurface attributes such as soil horizons, bedrock mineralogy, and geochemical characteristics (e.g., pH, salinity, and cation exchange capacity) indirectly influence albedo by altering soil moisture retention and organic matter distribution. To systematically account for geographic variations without explicitly specifying latitude and longitude, we incorporate soil texture (STX) as a categorical variable, classifying soils based on their physical composition (see Data section for dataset details).

Topography critically modulates albedo dynamics through elevation-driven land cover transitions and terrain-induced radiative effects (Hao et al., 2019). Higher elevations at low latitudes often shift vegetation from forests to grasslands, shrublands, or snowpack, increasing surface albedo (Friedl et al., 2002). Topographic variables (such as elevation (DEM), slope (SLP), and aspect (ASP)) shape microclimates, land cover, and ecological patterns (Bueno de Mesquita et al., 2018; Scherrer & Körner, 2011). However, the relationship between albedo, elevation, and regional climate remains understudied (Hu & Boos, 2017). Rugged terrain, covering ~24% of Earth's land surface, introduces anisotropic reflectance because of slope, aspect, and elevation, creating relief shadows and observation masking (Wen et al., 2022). These effects often stem from subgrid-scale heterogeneity that cannot be explicitly resolved in large-scale models. In this study, such small-scale influences are indirectly accounted for using bulk topographic parameters.

In forested regions with minimal soil exposure and stable year-round vegetation cover, meteorological variables such as near-surface relative humidity (R2M) and temperature (T2M) strongly influenced albedo dynamics by modulating leaf optical properties and canopy water content (H. Wang et al., 2024; X. Z. Zhang, 2012; Zhao et al., 2014). These variables also enable differentiation of microclimates and biotic heterogeneity within broad LCCs, for example, distinguishing albedo patterns between temperate and tropical evergreen forests, where divergent humidity and thermal regimes alter canopy structure and spectral signatures.

Following Liang, et al., (2005b), the albedo parameterization developed here integrates both dynamic and static drivers of surface reflectivity. Dynamic factors include solar geometry (SZA, SAA), soil moisture (SMC), and vegetation (LAI), which influence albedo variability from diurnal to interannual scales. Static factors such as soil texture (STX), meteorological (R2M, T2M), and topographic features (DEM, SLP, ASP) establish the baseline surface reflectance. We included the meteorological variables to capture microclimatic effects on canopy and soil optical properties. The next section outlines the data sources, resolution, and preprocessing methods used to incorporate these variables into a unified albedo modeling framework.

Chapter 3. Data

MODIS MCD12C1 Version 6.1, a dataset providing annual global land cover classifications at a 0.05° spatial resolution (~5.6 km), derived from the 500-meter MCD12Q1 product through spatial aggregation (Friedl et al., 2019; Sulla-Menashe et al., 2019). The dataset incorporates multiple classification systems (such as UMD, LCCS, and BCG) with the International Geosphere-Biosphere Program (IGBP) scheme serving as the primary system (Loveland et al., 2000). The IGBP system categorizes landscapes into 17 distinct classes, including forests, barren land, savannas, shrublands, grasslands, croplands, and urban areas, thereby capturing the unique dynamic signatures associated with various land types. Moreover, the dataset offers global coverage along with the percentage of land grid occupied by each class, a critical feature for our methodological approach (discussed later). We selected the MCD12C1 product over alternative land cover maps because of its compatible spatiotemporal resolution with the other datasets used in this study and its demonstrated overall accuracy of 73.6% (Sulla-Menashe et al., 2019). We omitted IGBP categories 0 (water) and 15 (snow and ice) from the analysis in this study to focus solely on the snow-free land surface albedo.

The MODIS MCD43C1 BRDF/Albedo product (v061) provides essential surface albedo data, including directional hemispherical (black-sky) and bi-hemispherical (white-sky) reflectance across two broad spectral bands: visible (0.3–0.7 μm) and near-infrared (0.7–5.0 μm) (Schaaf & Wang, 2021). We derive these parameters daily using MODIS observations. The albedo values are computed through spectral-to-broadband conversions

and can be represented using polynomial functions (Lucht et al., 2000). The BRDF model is mathematically expressed as:

$$\boldsymbol{\alpha_{s,b,\lambda}} = \sum_{k=1}^{3} \mathbf{f_{k,\lambda}}\left(\mathbf{g_{0k}^{bs}} + \mathbf{g_{1k}^{bs}\theta^2} + \mathbf{g_{2k}^{bs}\theta^3}\right)$$
$$\boldsymbol{\alpha_{s,d,\lambda}} = \sum_{k=1}^{3} \mathbf{f_{k,\lambda}} \mathbf{g_k^{ws}} \quad (1)$$

where α denotes satellite-derived albedo; subscript *b* and superscript *bs* denote black-sky (direct beam) radiation, while subscript *d* and superscript *ws* represent the white-sky (diffuse) radiation; λ is the spectral band; and θ is the zenith angle (in radian). g_{jk}^{bs} and g_k^{ws} are the fitting coefficients from Table I of (Lucht et al., 2000). This model effectively captures surface anisotropy and enables accurate estimation of albedo under varying illumination conditions.

A dedicated sub-module of the CWRF regional climate model (X.-Z. Liang et al., 2012) computed the SZA and SAA, accounting for the Earth's position and orientation to the sun. These angles fulfill two key roles: first, they facilitate the calculation of albedo from the bidirectional reflectance distribution function (BRDF); second, they serve as input features for the physics-informed ML parameterization.

For SMC, we employed the ERA5-Land reanalysis dataset. This product offers an improved quality at finer spatial resolution of 0.1° (~9 km) compared to the driving ERA5 product (~31 km) through direct assimilation of comprehensive surface air temperature and precipitation observations (Hersbach et al., 2020; Muñoz-Sabater et al., 2021). We selected ERA5-Land SMC because of its good accuracy (with correlations exceeding 0.8 against ground measurements), its superior performance relative to ERA-Interim (M. Li et al., 2020; Pelosi et al., 2020), and its ability to capture diurnal variability. Specifically, we

utilized the volumetric soil water layer 1 data (0–7 cm, measured in m^3/m^3), acknowledging that topsoil conditions directly influence albedo while vertical diffusion processes connect surface moisture to deeper layers (Idso et al., 1975).

For LAI, we utilized the GLASS dataset (Z. Xiao et al., 2016), derived from MODIS and AVHRR reflectance data. This dataset offers global coverage at 0.05° spatial resolution, with superior data quality and temporal consistency compared to alternatives like MOD15 and GEOV1 products (S. Liang et al., 2021). Additionally, GLASS LAI exhibits enhanced spatial integrity, particularly in tropical regions (J. Li & Xiao, 2020), and its performance has been validated against high-resolution LAI maps, yielding an RMSE of 0.79 and an R^2 of 0.81 (Z. Xiao et al., 2016).

We used the Digital Soil Map of the World (DSMW) (Sanchez et al., 2003) to acquire the soil texture (STX) information. This comprehensive dataset, derived from the FAO-UNESCO Soil Map of the World at a 1:5,000,000 scale, categorizes soils into 117 distinct types organized into 26 major soil groupings and six additional land units, yielding a total of 123 classes as defined by the DOMSOI attribute, which indicates the dominant soil or land unit in each mapping area.

To account for topographic influences on albedo, we incorporated elevation, slope cosine, and the standard deviation of the sub-pixel aspect (which captures sub-grid anisotropy) as input features. These variables were sourced from the standardized global multivariate product provided by EarthEnv (Amatulli et al., 2018).

To represent climate background effects on albedo, we incorporated long-term (2000-2020) daily climatology of near-surface air temperature (T2M) and relative humidity (R2M) from the ERA5-Land dataset as static features. Table 1 presents a summary of the

final spatial and temporal resolutions, along with the modifications made compared to the original resolution.

Variable	Data Source	Original Spatial Resolution	Original Temporal Resolution	Final Spatial Resolution	Final Temporal Resolution
Land Cover (IGBP classes)	MODIS MCD12Q1 v6.1	0.05°	Annual	0.05°	Annual
Surface Albedo (Black-sky & White-sky)	MODIS MCD43C1 v061	0.05°	Daily	0.05°	Daily
Solar Zenith Angle (SZA), Solar Azimuth Angle (SAA)	CWRF regional climate model (X.-Z. Liang et al., 2012)	0.05°	6 Hourly	0.05°	6 Hourly
Soil Moisture Content (SMC)	ERA5-Land Reanalysis	0.1°	6 Hourly	0.05°	6 Hourly
Leaf Area Index (LAI)	GLASS (MODIS/AVHRR)	0.05°	8-day	0.05°	Daily
Soil Texture (STX)	Digital Soil Map of the World (DSMW/FAO-UNESCO)	1:5M scale	Static	0.05°	Static
Topography (DEM, SLP, ASP)	Global Multivariate Product (Amatulli et al. 2018)	0.05°	Static	0.05°	Static
Near-Surface Air Temperature (T2M) and Relative Humidity (R2M)	ERA5-Land Reanalysis	0.1°	Hourly	0.05°	Daily climatology

Table 1: Overview of key variables, their respective data sources, and spatio-temporal preprocessing steps applied to align datasets for parameterization development.

Chapter 4. Study Framework and Methodology

The parameterization framework integrates two components: (1) Dynamic — an ML-based model trained on time-and-space varying (SZA, SAA, LAI, and SMC) and space-only varying (DEM, SLP, ASP, STX, T2M, and R2M) predictors to resolve physical albedo dynamics; and (2) Static — the surface albedo localization factor (SALF), a space-explicit but time-invariant correction to empirically approximate subgrid scale effects and unobserved local features. The ML model quantifies explicit albedo dependences on physical predictors that vary with space and/or time, while SALF provides geographic corrections to minimize residuals between modeled and observed albedo. This hybrid approach combines physics-informed ML prediction with empirical correction to enhance global albedo representation.

4.1 Physics-Informed ML Dynamic Parameterization

As we have already seen, albedo is governed by multiple factors, each of which exhibits complex, nonlinear interactions with strong spatial and temporal variability. Traditional parameterizations often simplify these interactions, relying on fixed empirical relationships or look-up tables that may fail to capture the full range of observed albedo behavior under different environmental conditions. To address these limitations, we employ a physics-informed ML framework, which explicitly integrates knowledge of physical processes into the model design. In this context, “physics-informed” means that our selection of input variables is guided by established physical principles and prior research identifying key drivers of albedo variability, as mentioned in Section 2. By choosing predictors grounded in process understanding, we ensure that the model does not merely capture statistical

correlations but learns relationships that reflect the underlying physics of land-atmosphere radiative exchanges.

To model these complex, nonlinear relationships, we employ XGBoost (eXtreme Gradient Boosting), a supervised ML algorithm particularly well-suited for learning complex, nonlinear relationships and interactions among variables (Chen & Guestrin, 2016). XGBoost builds an ensemble of decision trees in sequence, with each tree trained to correct the residual errors of its predecessors, allowing the model to capture intricate dependencies and threshold effects often seen in albedo dynamics, such as abrupt changes during snow melt or vegetation transitions. Compared to simpler models, XGBoost offers significant advantages for this study, including the ability to model nonlinear responses without requiring predefined functional forms, high computational efficiency suitable for large, high-resolution datasets, and built-in regularization techniques such as L1 and L2 penalties that help prevent overfitting and enhance generalization across diverse LCCs. Due to these advantages, we avoided more complex, computationally expensive, and slower neural network methods. Importantly, XGBoost also provides feature importance metrics, which enable us to assess the relative influence of each physical variable on albedo variability and ensure consistency with physical understanding. We also conducted a counterfactual analysis, which systematically identifies both local and global sensitivity to uncover the relationships the model has learned between predictors and outputs. This approach enhances our understanding of the nature of dynamical relationships and improves the model's explainability.

4.2 Localized Static Bias Correction

Despite capturing key drivers of albedo dynamics, the framework inherently omits factors such as soil roughness, plant species type, canopy height, and sub-story vegetation, etc. These elements are challenging to observe via satellite, frequently occur at sub-grid scales, and are difficult to predict in numerical models. Inspired by Liang, et al., (2005b), SALF addresses these gaps through regression-derived spatial corrections that minimize discrepancies between ML-predicted and observed albedo, offering a direct and realistic means to capture the geographic influence of surface characteristics not represented by the primary dynamic model. Notably, SALF varies with geographic location and the albedo part (i.e., direct beam versus diffuse radiation). While ML effectively models the dynamic albedo variations, SALF serves as a localized statistical correction that accounts for unresolved static effects specific to local surface characteristics.

4.3 Machine Learning Approach

The parameterization framework begins with data preprocessing and model training (Figure 1, Step 1). Direct and diffuse albedo for Vis and Nir bands were separately calculated at 6-hour intervals (00Z, 06Z, 12Z, 18Z) using MODIS BRDF kernel parameters (Equation 1) to capture diurnal and interannual variability. While the temporal sampling is limited to 6-hour intervals, the extensive spatial coverage ensures that the complete range of SZA is captured. High-confidence BRDF retrievals (QA = 0) are retained, and snow-covered grids are excluded using MODIS snow fraction data to ensure data quality. The resulting dataset serves as ground truth for supervised ML training.

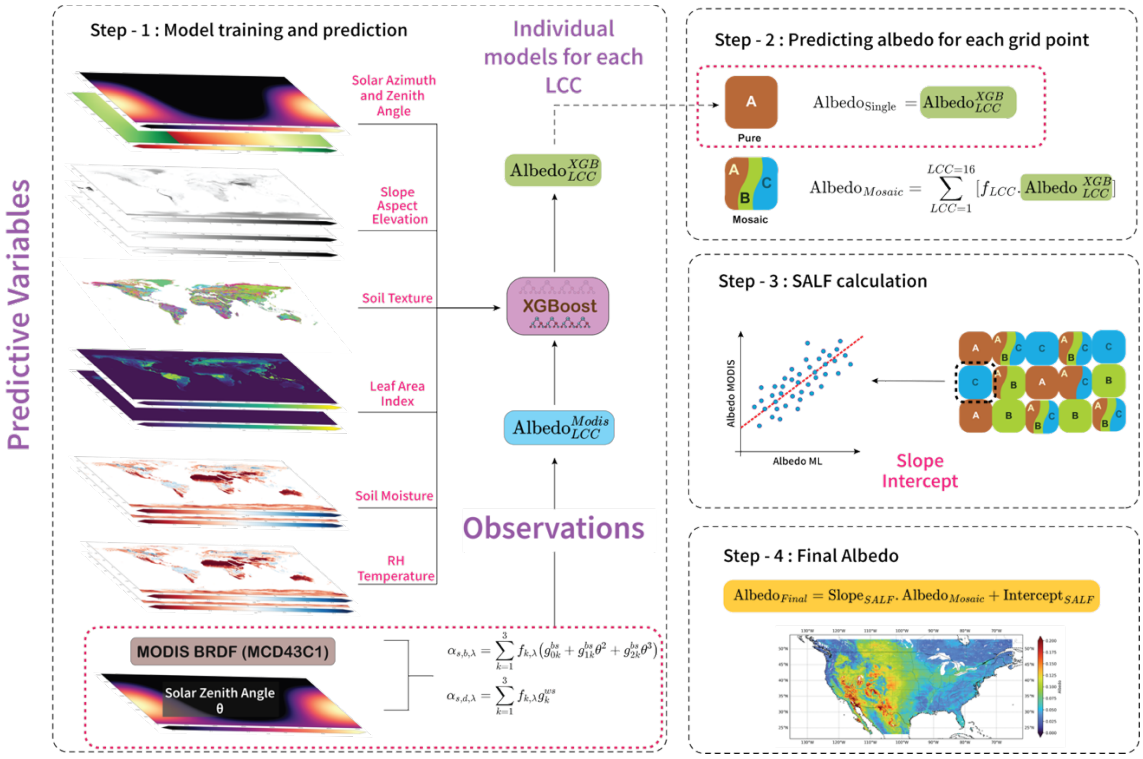

Figure 1. Schematic overview of the parameterization development methodology. The framework consists of four main steps: (1) Model training and prediction using XGBoost based on multiple predictive features for Pure grids in each LCC type; (2) Mosaic grid albedo estimation using weighted average of individual LCC models; (3) SALF calculation for each grid point by determining slope and intercept; (4) final albedo estimation.

Our model inputs consist of ten predictors or input features that are all geographically distributed, divided into time-varying variables (SZA, SAA, LAI, SMC) and static variables (STX, R2M, T2M, DEM, ASP, and SLP). SZA, SAA, and SMC are provided at a 6-hour temporal resolution, while LAI, originally available at an 8-day resolution, is linearly interpolated to a daily scale assuming gradual variation. The time-varying dataset spans 19 years (2003–2021), ensuring robust temporal coverage. Notably, the cosine of SZA is used as an input feature, consistent with its widespread use in existing albedo parameterization schemes. For STX, categorical vector data are assigned unique

numerical values and converted into a raster format. To ensure consistency across the dataset, all variables are regridded to a 0.05° spatial resolution when not already at this scale.

Due to the albedo's strong temporal persistence, random train-test splitting risks data leakage by allowing highly correlated samples to appear in both sets. While our model architecture is not explicitly temporal, we maintain temporal integrity through year-based splitting: training on 13 years, validating on 2006 and 2014 (with 5-fold cross-validation for hyperparameter tuning), and testing on four independent years (2004, 2008, 2012, 2016). This approach prevents inflation of performance metrics that would occur if temporally adjacent samples were distributed across training and test sets.

As previously mentioned, separate models are developed for each LCC. However, because an individual grid may contain varying fractions of multiple LCCs, we classify grids into two categories: Pure (100% coverage of an individual LCC) and Mosaic (mixed LCC). Model training is performed exclusively on Pure grids to minimize the complexities associated with mixed land cover and ensure robust learning for each LCC. In total, we train 60 distinct models, 15 for each LCC across four albedo parts (Dir_Vis, Dir_Nir, Dif_Vis, Dif_Nir). For a Mosaic grid, the final albedo is estimated by summing up the predicted values from individual LCC models, weighted by their fractional coverage within the grid (Figure 1, Step 2).

Finally, the SALF is computed for each grid by fitting a linear regression between the model predictions and the observed albedo (Figure 1, Step 3). Consistent with the dynamic model development, we applied the same temporal partitioning (13 training years, 2006/2014 for validation, and 2004/2008/2012/2016 for testing) to obtain SALF. The

regression process produces four global SALF datasets, one for each albedo part. The corresponding SALF corrections are then applied to all future model predictions, yielding the final albedo estimates (Figure 1, Step 4).

4.4 Model Evaluation Metrics

To assess the performance of our ML models of albedo parameterization, we employ two primary evaluation metrics: the coefficient of determination (R^2) and the mean absolute percentage error (MAPE).

R^2 measures how well a model explains the variation in the dependent variable (e.g., albedo). It ranges from 0, indicating no explanatory power, to 1, indicating the model perfectly accounts for all variability in the data. R^2 is computed using:

$$\mathbf{R^2} = \left(\mathbf{1} - \frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{y_i} - \hat{\mathbf{y}}_\mathbf{i})^2}{\sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{y_i} - \bar{\mathbf{y}})^2}\right) \tag{2}$$

where n is the number of observations, y_i is the observed value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the observed values.

MAPE measures forecasting accuracy, offering an intuitive and scale-independent way to compare errors across datasets and models. It indicates the average prediction error relative to the observed values and is defined as:

$$\mathbf{MAPE} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \left| \frac{\mathbf{y_i} - \hat{\mathbf{y}}_\mathbf{i}}{\mathbf{y_i}} \right| \tag{3}$$

Chapter 5. Results

5.1 Comparison Between Model and Observations

Figure 2 compares the model's performance (R^2 and MAPE) for direct (Dir) and diffuse (Dif) albedo in visible (Vis) and near-infrared (Nir) bands against MODIS observations across LCCs. These results are based exclusively on the testing dataset comprising Pure grids (the same type used for training) and are presented for the dynamic parameterization component, that is, before applying the static SALF correction. To aid comparison, similar LCCs have been grouped and their results discussed together below.

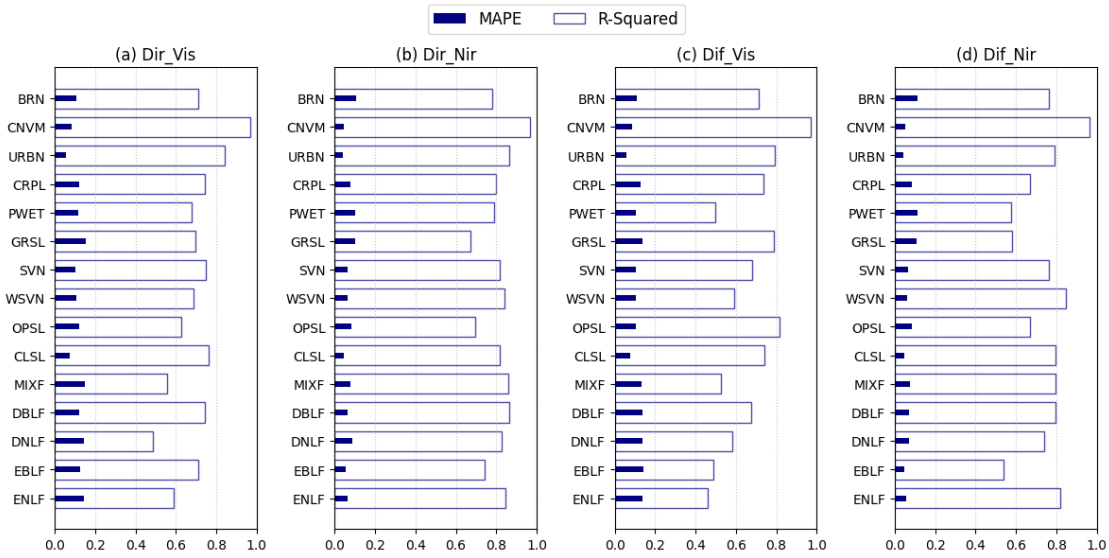

Figure 2. The figure compares model performance for different LCCs, showing MAPE (navy solid bars) and R^2 (open bars). The four panels represent different albedo parts: (a) Dir_Vis; (b) Dir_Nir; (c) Dif_Vis; and (d) Dif_Nir.

Forests

Forests in this study fall into five classes: evergreen needleleaf (ENLF), evergreen broadleaf (EBLF), deciduous needleleaf (DNLF), deciduous broadleaf (DBLF), and mixed forests (MIXF). Model skill in the visible range is moderate. For Dir_Vis, R^2 spans from

0.49 in DNLF to 0.74 in DBLF, averaging 0.61 across forest types. Dif_Vis shows a slight drop, with R^2 between 0.46 and 0.68, while MAPE stays consistent at 0.12–0.15. In contrast, the near-infrared parts perform markedly better: Dir_Nir achieves a mean R^2 of 0.82 and Dif_Nir 0.76, with MAPE generally below 0.08.

These differences reflect both the amount and the quality of training data, as well as the natural variability of forest albedo. ENLF occupies only 1.6% of the Earth's surface, with less than 5% of its grids classified as Pure. DNLF and DBLF are even rarer, covering just 0.2% (1% Pure) and 1.3% (9% Pure) of the globe, respectively (Figure S1). EBLF is more widespread, at about 5% of the surface and 60% Pure grids, yet its location in tropical and subtropical regions means frequent cloud cover, which limits high-quality BRDF retrievals. Snow cover in boreal and temperate needleleaf forests further reduces usable observations.

Variability in the observations themselves also shapes performance (Figure S2). Direct albedo fluctuates more than diffuse because of diurnal solar geometry, and Nir albedo varies more than Vis. Evergreen forests, with their year-round leaf cover, show little interannual change, whereas deciduous forests—especially DBLF—display pronounced seasonal swings in direct albedo. On top of these patterns, structural differences in canopy geometry, species mix, forest density, leaf optical properties, and understory vegetation introduce additional complexity. Many of these features remain unmeasured, leaving parts of forest behavior beyond the reach of current dynamic models.

Shrublands and Savannahs

Shrublands, dominated by woody shrubs with sparse grasses, are prevalent in arid and semi-arid regions, while savannahs—grasses interspersed with scattered trees or shrubs—

occupy tropical and subtropical zones with distinct wet/dry seasons. Together, these LCCs cover ~25% of the Earth's surface (Figure S1). Closed Shrublands (CLSL, 0.2% coverage; 10% Pure) and Open Shrublands (OPSL, 8.5% coverage; 35% Pure) exhibit strong model performance. CLSL achieves R^2 of ~ 0.80 across all albedo parts, while OPSL shows R^2 ranging from 0.63 (Dir_Vis) to 0.82 (Dif_Vis). Across all albedo parts, MAPE remains low, typically under 0.09. Savannahs, woody (WSVN) and non-woody (SVN) show minimal performance differences, with superior accuracy in the Nir parts ($R^2 \approx 0.77$, MAPE ≈ 0.063) compared to the Vis parts ($R^2 \approx 0.61$, MAPE ≈ 0.12), owing to Nir's sensitivity to sparse woody cover and reduced soil reflectance interference.

Grasslands (GRSL)

GRSL, the largest vegetated LCC, covers 16.3% of the Earth's surface (40% Pure) and exhibits high spatiotemporal heterogeneity in albedo dynamics. GRSL shows significant spatiotemporal heterogeneity in its albedo dynamics. This variability is reflected in higher standard deviations, particularly in the diffuse albedo, driven by seasonal fluctuations in soil exposure and vegetation density. Despite these challenges, the model achieves a robust R^2 of ~0.70 across albedo parts, with MAPE values of 0.12 (Vis) and 0.09 (Nir).

Permanent Wetlands (PWET)

PWET, characterized by persistent water-vegetation mixtures, covers 0.7% of the Earth's surface, with only 5% of grids classified as Pure. The spectral signal in these areas is challenging due to the partial inundation and heterogeneous water-vegetation mosaics, causing overlap and mixing in reflectance characteristics. The model achieves moderately high accuracy, particularly in the Nir part ($R^2 \approx 0.82$), leveraging Nir's sensitivity to vegetation-water interactions. Vis performance is lower ($R^2 \approx 0.66$), due to spectral overlap

between water and leaf reflectance. MAPE remains consistent across parts, averaging ~0.077, indicating consistent model skill despite the complex spectral mixing and relatively sparse training data in these wetland areas.

Cropland (CRPL), Urban (URBN), and Cropland-Vegetation Mosaics (CNVM)

CRPL, the largest human-modified land type, covers 6% of the Earth's surface (30% Pure) and exhibits the highest albedo variability across all categories due to pronounced seasonal agricultural cycles (sowing, growth, harvest). Models achieve robust performance ($R^2 \approx 0.70$), with MAPE of 0.12 (Vis) and 0.08 (Nir). CNVM (0.5% coverage) and URBN (0.3%) achieve exceptional accuracy ($R^2 > 0.92$, MAPE < 0.05), surpassing all other categories. These two categories exhibit moderate albedo variability, but their relatively uniform surface composition at the spatial resolution of 0.05 degrees contributes to consistent model performance. Cropland mosaics represent a mixture of cropland and natural vegetation, while urban areas consist primarily of concrete, asphalt, and scattered greenery. In particular, urban areas benefit from spectral similarity across cities globally at this resolution, which effectively masks finer-scale heterogeneity within the grid cells.

Barren Land (BRN)

BRN regions, characterized by <10% vegetative cover year-round, exhibit significant albedo variability due to diverse surface compositions, ranging from bright dunes in North Africa and Arabia to darker volcanic terrains in the Americas and Eurasia. Covering ~10% of Earth's surface, these arid landscapes (sand flats, salt lakes, rocky outcrops) pose persistent challenges for albedo modeling, as global land-atmosphere models struggle to replicate observed spatial dynamics (Oleson et al., 2003; Pinty et al., 2000; Roesch et al., 2002). Despite this complexity, the framework achieves relatively robust performance,

with an average $R^2 \approx 0.71$ (peaking at 0.78 for Dir_Nir) and MAPE ≈ 0.11. This robustness is partly attributable to high-quality observational data: ~70% of BRN grids are Pure, and cloud-/snow-free conditions prevalent in arid regions enhance retrieval accuracy. Importantly, the inclusion of soil texture (STX) as a predictive variable significantly improved the model's ability to resolve spatial heterogeneity across these geologically diverse surfaces.

Overall Model Performance Summary

The parameterization achieves a global average R^2 of 0.73 and MAPE of 0.093 across all LCCs and albedo parts. Its performance varies substantially between spectral domains, driven by differences in surface structure, composition, temporal variability (e.g., standard deviation), and statistical factors such as data availability. These variations highlight that land-type-specific models can offer improved predictive power and interpretability by better capturing the unique biophysical characteristics of each ecosystem. The model performs significantly better in Nir parts (average R^2: 0.77, MAPE: 0.072) than in Vis parts (average R^2: 0.68, MAPE: 0.11).

While models were trained exclusively on Pure grids, Mosaic-type grids (with multiple mixed LCC) dominate globally. As outlined in the Methods section, the Mosaic approach estimates albedo as a weighted average of predictions from individual Pure LCC models. For ease of comparison, the dominant LCC is assigned to each Mosaic grid (e.g., a grid is labeled GRSL if grasslands are dominant amongst all others).

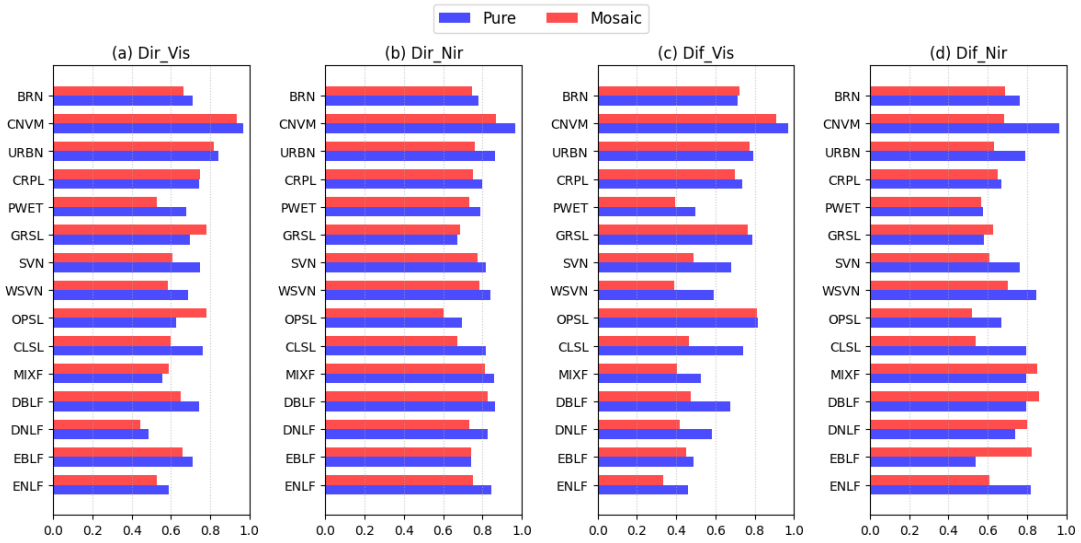

Figure 3. Comparison of R^2 for each LCC type under different grid types, Pure (blue bars) and Mosaic (red bars). To simplify interpretation, we assigned each Mosaic grid a dominant LCC (for example, a grid dominated by grasslands is labeled as GRSL). The four panels represent different albedo parts: (a) Dir_Vis; (b) Dir_Nir; (c) Dif_Vis; and (d) Dif_Nir.

Despite the training being restricted to Pure grids, the Mosaic approach achieved robust predictive performance across all (dominant) LCCs (Figure 3). Averaged over all four albedo parts, the reduction in R^2 values compared to Pure category predictions ranged from 0.025 to 0.075 points, indicating that the loss in accuracy due to added heterogeneity is modest and expected. This performance decline is primarily due to the compositional complexity of Mosaic grids and the uneven distribution of Pure pixels across classes. In some LCCs, such as CNVM, URBN, and DNLF, fewer than 10% of the global pixels are Pure, meaning that the model had limited training exposure to the full heterogeneity of these classes. As a result, these categories may suffer from underfitting to unseen static features or overfitting to skewed samples. Importantly, these discrepancies are later corrected by the SALF correction step, which is incorporated for this exact reason.

Nevertheless, most LCCs often retained high R^2 scores even under the Mosaic configuration, suggesting a good degree of generalization and robustness in the parameterization. Interestingly, in certain classes, especially forest LCCs, the Mosaic approach sometimes improved performance, plausibly by cancelling LCC-specific biases, because these grids contain other LCCs that have better performance. This weighted averaging strategy enables stronger predictions from one LCC model to compensate for errors in another. The methodology is also beneficial in transitional zones like GRSL-SVN and GRSL-CRPL interfaces, where the LCC dataset is more prone to classification errors.

Overall, the Mosaic method performed reliably even prior to SALF correction, demonstrating the strength of LCC-specific parameterization and its ability to generalize to heterogeneous landscapes. That the model, trained on a small fraction of the total grid space, delivers consistent results across globally diverse surfaces is a strong validation of the modeling framework's robustness.

5.2 Performance Enhancement by Localized Correction

The inclusion of SALF led to consistent improvements in model skill across all albedo parts and LCCs. These gains are driven by enhanced spatial representation of surface characteristics in the climatological albedo fields. All performance metrics discussed in this section are computed over the full set of grid cells, including both Pure and Mosaic types. Figure 4 and Supplemental Figure S6 illustrate the model performance metrics, R^2 and MAPE, respectively, before and after the implementation of the SALF component. For the actual magnitude difference between R^2 and MAPE, refer to Figures S3 and S4, respectively.

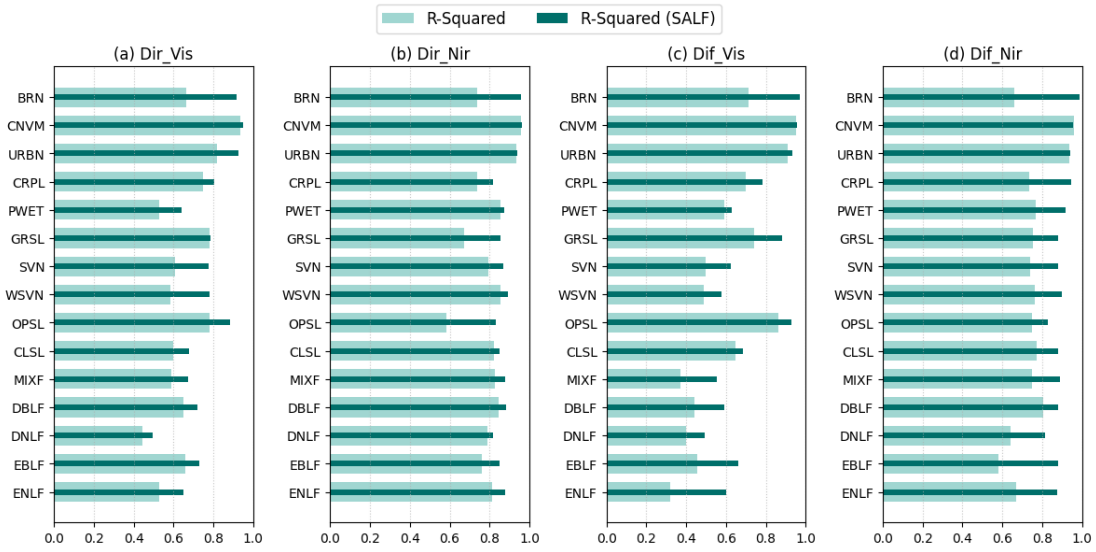

Figure 4. Comparison of changes in R^2 with the inclusion of SALF (solid dark green bars) for each LCC. The four panels represent different albedo parts: (a) Dir_Vis; (b) Dir_Nir; (c) Dif_Vis; and (d) Dif_Nir. Metrics are computed over all available grids, including both Pure and Mosaic grid types.

Forest classes exhibited notable improvements, particularly in the diffuse albedo. The change in R^2 varied from 0.03 (DNLF – Dir_NIR) to 0.3 (EBLF – Dif_Nir), with the final R^2 exceeding 0.85 in the Nir albedo parts (Figure 4). Shrublands, both CLSL and OPSL, showed gains in R^2 ranging from 0.05 to 0.1 points, except for the Dir_Nir part in OPSL, which had one of the highest gains of ~0.25 points (Figure S3). Overall, CLSL and OPSL achieved a MAPE reduction of 0.036 points (Figure S4). Both savannas (SVN and WSVN) showed similar performance, with final R^2 values ranging from 0.6 (Dif_Vis) to 0.89 (Dif_Nir). Grasslands (GRSL) exhibited a negligible change in R^2 in the Dir_Vis part but showed a slight increase of ~0.15 in all other albedo parts, resulting in final values of ~0.8 across all parts. Croplands (CRPL) consistently demonstrated gains across all albedo parts, with final R^2 values ranging from 0.78 (Dif_Vis) to 0.92 (Dif_Nir). High-performing

categories such as URBN and CNVM maintained strong performance, with R^2 values exceeding 0.90 after the SALF. SALF demonstrated significant improvements in barren regions (BRN), which have historically posed challenges for parameterization schemes. Following the implementation of SALF, these areas consistently achieved R^2 values exceeding 0.90 across all albedo parts. These regions also exhibited the highest average increase in R^2 (Figure S3) and a large reduction in MAPE (Figure S4). This performance demonstrates SALF's enhanced capability to capture substantial sub-pixel heterogeneity and spatial variability in albedo across deserts and semideserts—aspects that conventional approaches relying solely on static soil color or texture classifications fail to adequately represent.

Averaged globally, R^2 increases by 0.11 (from 0.70 to 0.81), with the largest gain in Dif_Nir (+0.14, from 0.75 to 0.89) and the smallest in Dir_Nir (+0.08, from 0.80 to 0.88). Among visible bands, Dif_Vis improves by +0.12 (0.60 → 0.72) and Dir_Vis by +0.10 (0.66 → 0.76). NIR parts consistently achieve higher final R^2 than VIS parts, with Dir_Nir (0.88) and Dif_Nir (0.90) outperforming Dir_Vis (0.76) and Dif_Vis (0.72). MAPE reductions mirror these trends, averaging –0.024 globally, with the largest decreases in Dir_Nir (–0.29) and Dir_Vis (–0.29), followed by Dif_Vis (–0.23) and Dif_Nir (–0.17). Since SALF preserves temporal structure, these improvements are attributable solely to its enhanced spatial parameterization of albedo climatology.

5.3 Feature Importance

XGBoost quantifies feature importance using three key metrics: Gain (accuracy improvement per split), Cover (number of affected samples), and Frequency (split usage rate). These metrics help identify the most influential features driving model predictions.

By developing separate models for each LCC, we uncovered unique combinations of key factors influencing albedo predictions (Figure 5). This approach not only enhances the interpretability of our ML models but also provides deeper insights into albedo dynamics and their interactions with surface characteristics when combined with counterfactual analysis (detailed in the next subsection).

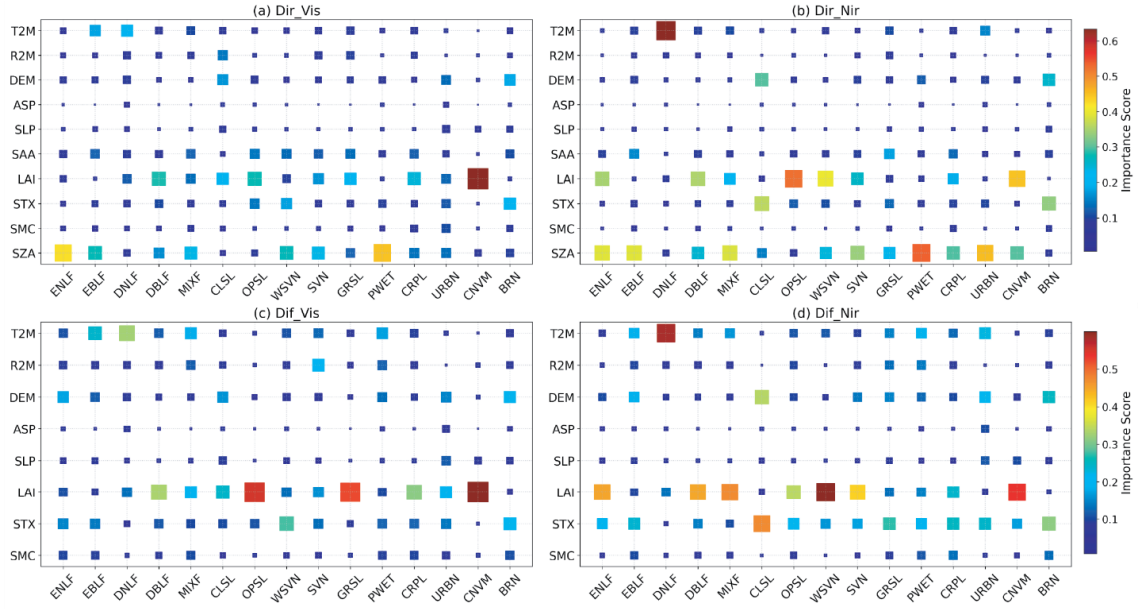

Figure 5. The heatmap illustrates the relative importance scores of different features across various LCC types for each albedo part: (a) Dir_Vis; (b) Dir_Nir; (c) Dif_Vis; and (d) Dif_Nir. Square markers represent the significance of each feature, with size indicating importance and color intensity representing the magnitude of importance scores.

In evergreen forests (ENLF, EBLF), SZA and SAA strongly govern direct albedo (Dir_Vis/Nir), highlighting the fundamental control of solar geometry on canopy radiative transfer (Hollinger et al., 2010; Lukeš et al., 2013). Interestingly, despite their evergreen nature, these forests exhibit significant LAI variability - particularly pronounced in ENLF (0.5-5) compared to EBLF (2.0-5.0 m^2/m^2). This structural contrast leads to divergent albedo responses: in EBLF, where LAI is already high, additional foliage contributes little

to albedo modulation. In contrast, in ENLF, LAI emerges as a key predictor of Dir_Nir albedo, but not Dir_Vis, consistent with the findings of Alibakhshi et al., (2020). This decoupling reflects wavelength-specific photon-canopy interactions: chlorophyll and associated pigments drive strong absorption in the Vis, rendering Vis albedo largely insensitive to further changes in foliage density. Conversely, the Nir part, governed by leaf internal structure and inter-leaf scattering, is more responsive to canopy thickness and foliage arrangement (Gates et al., 1965). As a result, during the growing season when LAI variation is maximal, Nir albedo displays heightened sensitivity to structural dynamics, whereas Vis albedo remains relatively invariant due to photosynthetic saturation. (Hollinger et al., 2010) also observed that ENLF albedos lack clear seasonal patterns (e.g., spring increases or autumn declines), with mid-year albedo minima driven by summer solar elevation changes, consistent with the strong influence of the SZA identified in this study.

LAI plays a more critical role for DBLF in the Vis parts, as evergreen forests tend to maintain stable canopy conditions throughout the year (Z. Liu et al., 2015). In DBLF, which are widespread across temperate regions such as the eastern United States, Europe, and East Asia, LAI exhibits pronounced seasonal fluctuations due to leaf emergence and senescence. These seasonal dynamics make LAI a primary driver of albedo variability in these regions. In contrast, DNLF shows a stronger sensitivity to near-surface temperature, especially in the Nir part. These forests are typically located in boreal regions, where annual temperature cycles are more pronounced. Prior studies, (e.g. Alibakhshi et al., 2020; Leonardi et al., 2015) have observed that albedo in DNLF regions fluctuates strongly with temperature, while LAI tends to have lower variability and importance. Additionally, SZA emerged as an influential factor in both DBLF and MIXF, emphasizing the role of solar

geometry in modulating albedo in these forest types. This effect likely arises from the horizontally extensive and structurally complex nature of broadleaf canopies, which contrasts with the vertically layered and more uniform architecture of needleleaf forests (Rago et al., 2021).

CLSL are defined by dense woody vegetation (>60% canopy cover) with overlapping foliage, while OPSL have sparse cover (10–60%) and exposed soil gaps. Both consist of shrubs under 2 meters tall. CLSL covers only 0.4% of Earth's land surface, whereas OPSL is widespread (~16%). For CLSL, STX, R2M, and DEM emerged as major static predictors, while LAI captures the dynamic variability. Elevation gradients further shape shrubland distribution: in low latitudes, land cover transitions from forests to grasslands/shrublands and eventually to barren/snow-covered terrain at higher elevations (Friedl et al., 2002). This pattern aligns with albedo trends observed over the Tibetan Plateau and South Asia, where elevation-driven microclimates alter surface reflectance (Hu & Boos, 2017). Static factors like topography and soil properties thus provide critical insights into regional albedo variability and ecosystem-climate interactions.

Open shrublands (OPSL) exhibit albedo behavior like grasslands (GRSL), both characterized by herbaceous vegetation with minimal tree/shrub cover (<10%). These LCCs share significant geographical overlap and frequently co-occur in Mosaic grids, making them the most common pairing among all possible LCC combinations (Figure S5). Seasonal LAI changes drive nearly identical albedo patterns in these categories, as observed in East Africa, where LAI and albedo show a strong correlation of 0.8 (Abera et al., 2019). Annual LAI variability in OPSL and GRSL, linked to vegetation growth cycles, makes LAI the dominant predictor of albedo. Also, at the 0.05° resolution (~5 km), these

surfaces appear spatially homogeneous, and their frequent geographic overlap (e.g., transitional arid-to-grassland zones) further amplifies similarities in radiative responses.

Savannahs (SVN) and woody savannahs (WSVN) differ in canopy structure and vegetation density. SVNs are characterized by 10–30% tree cover, with trees exceeding 2 meters in height and a grassy understory. These systems transition to grasslands if woody plants diminish or to shrublands/forests if grasses disappear. WSVN, in contrast, features denser tree cover (30–60%) but retains a similar grassy layer. Phenological patterns vary regionally: SVN and WSVN in Australia and South America are predominantly evergreen, while African savannahs are largely deciduous, reflecting adaptations to seasonal rainfall (Hill et al., 2011). SZA and LAI dominate direct albedo predictions, while LAI and STX jointly drive diffuse albedo.

Permanent wetlands (PWET) feature importance varies across spectral parts: SZA and LAI dominate direct albedo (Dir_Vis/Nir), while STX and elevation drive Dif_Vis albedo. LAI's stronger influence in Nir parts arises from higher water absorption in these wavelengths compared to visible light. In wetlands (PWET), where water is abundant, this spectral sensitivity better distinguishes vegetation-water variations, improving LAI's ability to predict albedo.

SZA and LAI dominate direct albedo predictions in croplands (CRPL), where diverse crops and seasonal cycles pose challenges. LAI tracks crop growth phases, while SZA captures reflectance changes from phenology, bare soil, or post-harvest residues. Notably, CRPLs and GRSLs exhibit highly similar feature importance patterns. After the GRSL-OPSL pair, the GRSL-CRPL combination is the most frequent in mosaic pixels (Figure S5), underscoring their ecological and structural parallels—technically, cultivated

crops are grasses in terms of phenology and spatial uniformity. For Dif_Vis albedo, LAI remains the primary predictor. In the Nir part, LAI, soil STX, and SMC collectively contribute. SMC influence was limited, likely because downscaled reanalysis-derived SMC does not accurately capture variations from human-managed irrigation systems in croplands. Yet, CRPL showed SMC's highest relative contribution among all LCCs. CRPL albedo is also influenced by other factors, particularly crop type, which directly impacts dynamics. Unfortunately, global crop type data is unavailable, but we've explored crop-specific albedo dynamics over the CONUS region, which will be discussed in subsequent papers.

For Cropland-Vegetation Mosaics (CNVM), LAI alone governs albedo across all parts, reflecting its control over vegetation density and structure. In urban (URBN) areas, many features got relatively similar scores with no characteristic pattern. Static factors collectively have a greater contribution than dynamic factors, except in Dir_Nir, where SZA dominated.

For BRN, static features like STX and topography (DEM, aspect, slope) were stronger predictors than dynamic features. This pattern aligns with our understanding of the complexities posed by barren areas in radiation and land interaction studies. As previously noted, these regions exhibit highly dynamic spatiotemporal albedo behavior, making it incorrect to classify them under a Pure category for radiation-based analyses. STX and topographic features aid in parametrizing and reducing variability in capturing diverse albedo signatures, aligning with methods validated in arid regions by (Tsvetsinskaya et al., 2006).

5.4 Counterfactual Analysis

Counterfactual analysis is a model interrogation technique that evaluates how predictions respond to controlled changes in individual input variables while keeping all other features constant. It systematically probes both local and global sensitivity to reveal the relationships the model has learned between predictors and outputs. In this study, we perturbed three key dynamic predictors (SZA, LAI, and SMC) and examined how these controlled changes influenced predicted direct albedo (Dir_Vis and Dir_Nir). This enabled us to assess whether model responses align with established biophysical principles. While feature importance scores from XGBoost offer global insights into which predictors most strongly influence outputs, they do not reveal the directionality or plausibility of these effects. Counterfactual analysis complements feature importance by explicitly testing whether changes in these critical variables, for example, whether increasing LAI produces the expected decrease in Dir_Vis albedo due to canopy absorption, thereby enhancing both the interpretability and physical credibility of the model. The following discusses the implications of these counterfactual experiments and highlights the insights they provide into model behavior across different LCCs.

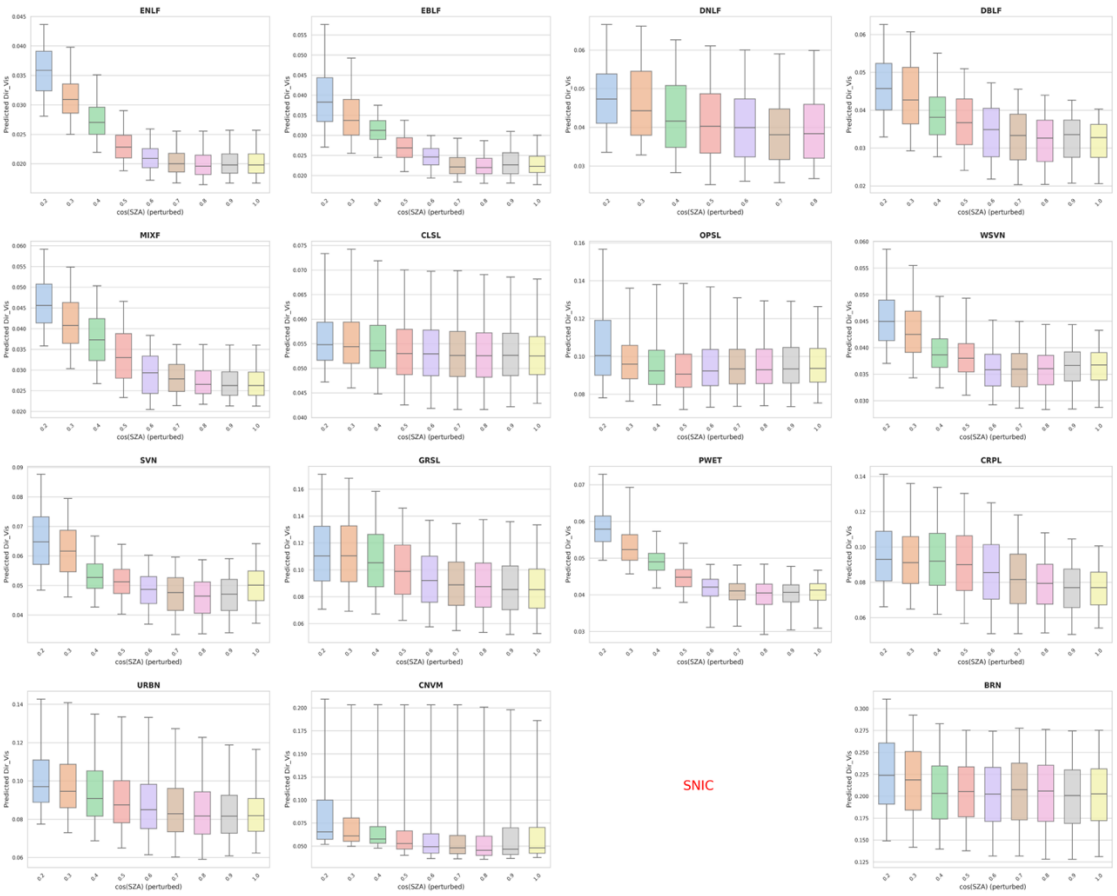

Figure 6. Counterfactual analysis of predicted Dir_Vis albedo to key biophysical parameters across 15 LCCs. Each panel shows the response of the XGBoost model predictions to perturbations in cos(SZA) for Dir_Vis albedo, with boxplots representing the distribution of predicted values.

The relationship between direct Vis/Nir albedo and SZA can be explained by two competing radiative processes: (1) shadowing effects, which dominate at low sun angles (high SZA), and (2) multiple scattering within the canopy, which becomes increasingly important as the sun approaches zenith. The relative influence of these processes depends on canopy structure—particularly its density, heterogeneity, and gap fraction—leading to distinct albedo responses across LCCs. In closed-canopy forests (ENLF, EBLF, DBLF), where foliage is dense and vertically continuous, albedo decreases sharply at low sun

angles (SZA > 45^{o}) as elongated shadows trap photons. However, once solar elevation increases, shadows shorten, and multiple scattering between leaves stabilizes reflectance, causing albedo to plateau rather than decline further (Figure 6). In contrast, open or structurally heterogeneous canopies (e.g., GRSL, OPSL) display a more linear decrease in albedo with increasing *cos*(SZA). These landscapes are characterized by shorter vegetation, higher gap fractions, and lower vertical complexity, which allow direct solar beams to illuminate the ground surface more uniformly across a wide range of sun angles. As SZA decreases, both shadowing and volumetric scattering play relatively minor roles, leading to a smoother angular response. Finally, sparse or discontinuous canopies (WSVN, SVN) along with some other LCCs display a rebound or slight increase in albedo near *cos*(SZA) ≈ 1 (sun at zenith). This bump could be attributed to enhanced illumination of the ground through canopy gaps and reduced horizontal shadow displacement.

An intriguing aspect of albedo dynamics emerges when examining how LAI differentially affects the Dir_Vis and Dir_Nir albedo. In the Vis range, increasing LAI typically leads to a reduction in albedo due to enhanced absorption by chlorophyll-rich foliage, which supports greater photosynthetic activity (Figure 7). In contrast, in the Nir range, albedo often increases with LAI because leaves tend to reflect and transmit Nir radiation rather than absorb it. As LAI increases, multiple scattering among leaves becomes more prominent, enhancing the likelihood of Nir photons escaping the canopy and thereby raising albedo (Figure S9).

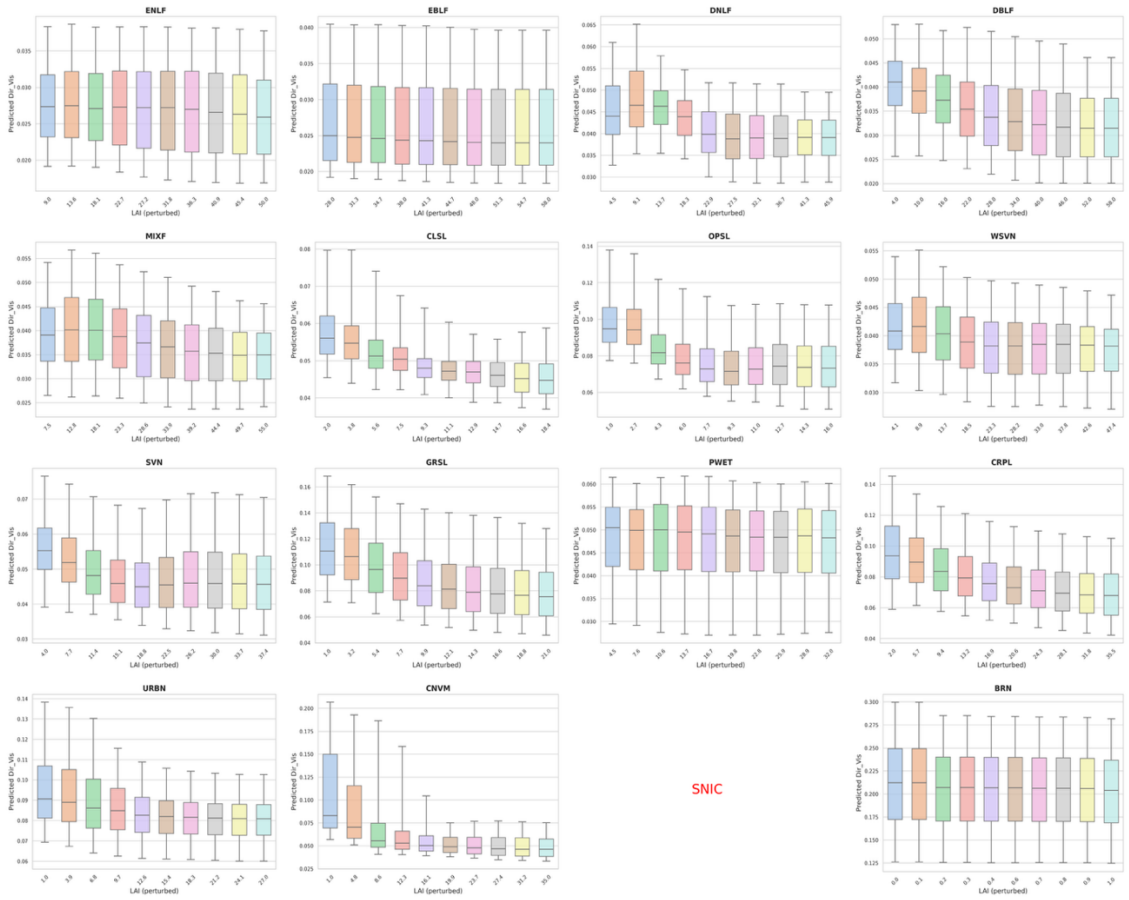

Figure 7. Same as Figure 6, but for LAI as the perturbed variable.

Mixed forests (MIXF) present a more complex response. Here, direct Nir albedo initially decreases with increasing LAI but begins to rise again beyond a certain threshold (Figure S9). This non-monotonic relationship has been documented in temperate mixed forests, including findings by Alibakhshi et al. (2020), who reported a similar dip and subsequent rise in direct Nir albedo during the peak growing season. At very low LAI values (<1.5), the forest understory and ground surface—including soil, litter, or herbaceous vegetation—are partially exposed and contribute modestly to reflectance. In this regime, shadows cast by leafless trunks further suppress albedo. As LAI increases beyond this threshold, however, the canopy becomes sufficiently developed to enter a

scattering-dominated regime, in which the combined effects of leaf reflectance and multiple scattering enhance Nir albedo.

Across most classes, the impact of soil moisture on albedo is modest—approximately an order of magnitude weaker than effects from SZA or LAI. Nonetheless, a subtle decrease in Dir_Vis/Nir albedo with increasing SMC is observable in many ecosystems, reflecting the darkening effect of wetter soils (Figure S7 and Figure S8).

Chapter 6. Discussion and Conclusions

This study develops a globally applicable, physics-informed ML parameterization of snow-free land surface albedo using 19 years (2003–2021) of MODIS BRDF observations. The framework comprises two complementary components: a dynamic component, in which XGBoost models predict direct and diffuse albedo in the visible and near-infrared bands separately for 15 LCCs using ten biogeophysical predictors; and a static component—the surface albedo localization factor (SALF)—that applies spatially explicit corrections to account for subgrid heterogeneity and unobserved local influences. The predictors capture key controls on albedo variability, including solar geometry, vegetation state, soil properties, topography, and background climate.

The parameterization shows strong agreement with MODIS albedo across all LCCs and albedo parts, with an overall R^2 of 0.81 and MAPE of 0.08. Accuracy is generally higher in the near-infrared band (up to $R^2 = 0.89$ for Dif_Nir) than in the visible band (minimum $R^2 = 0.72$ for Dif_Vis), and the model performs robustly even in challenging barren regions where empirical methods often underperform. Incorporating SALF improves the performance across all albedo parts, yielding an average R^2 gain of 0.11 and reducing MAPE by 0.024, underscoring its value in capturing localized and static heterogeneity, particularly in regions with complex surface characteristics or low temporal variability.

A notable strength is the framework's generalization: although trained solely on Pure grids (a minority of the global dataset), it performs remarkably well on Mosaic grids—which dominate global coverage—demonstrating that the model has effectively learned the underlying albedo dynamics. Feature importance analysis enhances interpretability,

revealing that solar zenith angle (SZA) and leaf area index (LAI) dominate dynamic variability, while soil texture (STX) and topography govern static variability. Each LCC exhibits a distinct predictor signature consistent with its biogeophysical characteristics, providing a basis for targeted refinement and improved understanding of surface-atmosphere interactions. Counterfactual experiments further validate the framework, showing physically consistent responses (e.g., increasing LAI reduces Dir_Vis albedo but increases Dir_Nir albedo due to vegetation absorption-scattering effects).

Several important directions remain. We will (1) evaluate spatial variations in performance and map SALF distributions to identify subgrid patterns and LCC-specific dynamics; (2) demonstrate scalability across spatial resolutions (~5-100 km), addressing a major limitation of traditional schemes; and (3) show interoperability across different land classification systems (IGBP, USGS, UMD), mitigating classification dependence in land-surface modeling.

Future work will also expand the framework to snow-covered albedo regimes, which requires distinct parameterization of surface-radiative dynamics. Applying the framework in real-time land surface modeling within coupled modeling systems such as CWRF will enable assessments of its impact on energy balance, feedbacks, and climate responses. Additionally, targeted expansions for specific LCCs, such as croplands—where crop type, phenology, and management strongly influence radiation dynamics—will further enhance realism and applicability.

By uniting physically informed machine learning with spatially adaptive corrections, this framework provides a scalable, interpretable, and operationally relevant

tool for representing land surface albedo in Earth system models from regional to global scales.

References

Abera, T. A., Heiskanen, J., Pellikka, P., Rautiainen, M., & Maeda, E. E. (2019). Clarifying the role of radiative mechanisms in the spatio-temporal changes of land surface temperature across the Horn of Africa. *Remote Sensing of Environment*, *221*, 210–224. https://doi.org/10.1016/j.rse.2018.11.024

Alibakhshi, S., Naimi, B., Hovi, A., Crowther, T. W., & Rautiainen, M. (2020). Quantitative analysis of the links between forest structure and land surface albedo on a global scale. *Remote Sensing of Environment*, *246*, 111854. https://doi.org/10.1016/j.rse.2020.111854

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., & Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, *5*, 180040. https://doi.org/10.1038/sdata.2018.40

Bala, G., Caldeira, K., Wickett, M., Phillips, T. J., Lobell, D. B., Delire, C., & Mirin, A. (2007). Combined climate and carbon-cycle effects of large-scale deforestation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(16), 6550–6555. https://doi.org/10.1073/pnas.0608998104

Baumgardner, M. F., Silva, L. F., Biehl, L. L., & Stoner, E. R. (1986). Reflectance properties of soils. In *Advances in Agronomy* (pp. 1–44). Elsevier. https://doi.org/10.1016/s0065-2113(08)60672-0

Betts, R. A. (2000). Offset of the potential carbon sink from boreal forestation by decreases in surface albedo. *Nature*, *408*(6809), 187–190. https://doi.org/10.1038/35041545

Bueno de Mesquita, C. P., Tillmann, L. S., Bernard, C. D., Rosemond, K. C., Molotch, N. P., & Suding, K. N. (2018). Topographic heterogeneity explains patterns of vegetation response to climate change (1972–2008) across a mountain landscape, Niwot Ridge, Colorado. *Arctic, Antarctic, and Alpine Research*, *50*(1), e1504492. https://doi.org/10.1080/15230430.2018.1504492

Chapin, F. S., 3rd, Sturm, M., Serreze, M. C., McFadden, J. P., Key, J. R., Lloyd, A. H., et al. (2005). Role of land-surface changes in arctic summer warming. *Science (New York, N.Y.)*, *310*(5748), 657–660. https://doi.org/10.1126/science.1117368

Chen, T., & Guestrin, C. (2016, March 8). *XGBoost: A Scalable Tree Boosting System*. *arXiv [cs.LG]*. Retrieved from http://arxiv.org/abs/1603.02754

Cierniewski, J., Karnieli, A., Kazmierowski, C., Krolewicz, S., Piekarczyk, J., Lewinska, K., et al. (2015). Effects of soil surface irregularities on the diurnal variation of soil broadband Blue-Sky albedo. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(2), 493–502. https://doi.org/10.1109/jstars.2014.2330691

Fang, H., Baret, F., Plummer, S., & Schaepman-Strub, G. (2019). An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Reviews of Geophysics (Washington, D.C.: 1985)*, *57*(3), 739–799. https://doi.org/10.1029/2018rg000608

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, *83*(1–2), 287–302. https://doi.org/10.1016/s0034-4257(02)00078-0

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2019). MODIS Collection 6 land cover products: Changes, impacts, and continuity. *Remote Sensing of Environment*, *222*, 183–197.

Galvdo, L. (1997). Relationships of spectral reflectance and color among surface and subsurface horizons of tropical soil profiles. *Remote Sensing of Environment*, *61*(1), 24–33. https://doi.org/10.1016/s0034-4257(96)00219-2

Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. (1965). Spectral properties of plants. *Applied Optics*, *4*(1), 11. https://doi.org/10.1364/ao.4.000011

Guan, X., Huang, J., Guo, N., Bi, J., & Wang, G. (2009). Variability of soil moisture and its relationship with surface albedo and soil thermal parameters over the Loess Plateau. *Advances in Atmospheric Sciences*, *26*(4), 692–700. https://doi.org/10.1007/s00376-009-8198-0

Halim, M. A., Chen, H. Y. H., & Thomas, S. C. (2019). Stand age and species composition effects on surface albedo in a mixedwood boreal forest. *Biogeosciences*, *16*(22), 4357–4375. https://doi.org/10.5194/bg-16-4357-2019

Hao, D., Wen, J., Xiao, Q., Wu, S., Lin, X., Dou, B., et al. (2018). Simulation and analysis of the topographic effects on snow-free albedo over rugged terrain. *Remote Sensing*, *10*(2), 278. https://doi.org/10.3390/rs10020278

Hao, D., Wen, J., Xiao, Q., Lin, X., You, D., Tang, Y., et al. (2019). Sensitivity of coarse-scale snow-free land surface shortwave albedo to topography. *Journal of Geophysical Research Atmospheres*, *124*(16), 9028–9045. https://doi.org/10.1029/2019jd030660

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society. Royal Meteorological Society (Great Britain)*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hill, M. J., Román, M. O., Schaaf, C. B., Hutley, L., Brannstrom, C., Etter, A., & Hanan, N. P. (2011). Characterizing vegetation cover in global savannas with an annual foliage clumping index derived from the MODIS BRDF product. *Remote Sensing of Environment*, *115*(8), 2008–2024. https://doi.org/10.1016/j.rse.2011.04.003

Hollinger, D. Y., Ollinger, S. V., Richardson, A. D., Meyers, T. P., Dail, D. B., Martin, M. E., et al. (2010). Albedo estimates for land surface models and support for a new paradigm based on foliage nitrogen concentration. *Global Change Biology*, *16*(2), 696–710. https://doi.org/10.1111/j.1365-2486.2009.02028.x

Hovi, A., Lindberg, E., Lang, M., Arumäe, T., Peuhkurinen, J., Sirparanta, S., et al. (2019). Seasonal dynamics of albedo across European boreal forests: Analysis of MODIS albedo and structural metrics from airborne LiDAR. *Remote Sensing of Environment*, *224*, 365–381. https://doi.org/10.1016/j.rse.2019.02.001

Hu, S., & Boos, W. R. (2017). Competing effects of surface albedo and orographic elevated heating on regional climate. *Geophysical Research Letters*, *44*(13), 6966–6973. https://doi.org/10.1002/2016gl072441

Idso, S. B., Jackson, R. D., Reginato, R. J., Kimball, B. A., & Nakayama, F. S. (1975). The dependence of bare soil albedo on soil water content. *Journal of Applied Meteorology*, *14*(1), 109–113. https://doi.org/10.1175/1520-0450(1975)014<0109:tdobsa>2.0.co;2

Jin, Z., Qiao, Y., Wang, Y., Fang, Y., & Yi, W. (2011). A new parameterization of spectral and broadband ocean surface albedo. *Optics Express*, *19*(27), 26429–26443. https://doi.org/10.1364/OE.19.026429

Kala, J., Evans, J. P., Pitman, A. J., Schaaf, C. B., Decker, M., Carouge, C., et al. (2014). Implementation of a soil albedo scheme in the CABLEv1.4b land surface model and evaluation against MODIS estimates over Australia. *Geoscientific Model Development*, *7*(5), 2121–2140. https://doi.org/10.5194/gmd-7-2121-2014

Kim, J. E., Laguë, M. M., Pennypacker, S., Dawson, E., & Swann, A. L. S. (2020). Evaporative resistance is of equal importance as surface albedo in high-latitude surface temperatures due to cloud feedbacks. *Geophysical Research Letters*, *47*(4). https://doi.org/10.1029/2019gl085663

Leonardi, S., Magnani, F., Nolè, A., Van Noije, T., & Borghetti, M. (2015). A global assessment of forest surface albedo and its relationships with climate and atmospheric nitrogen deposition. *Global Change Biology*, *21*(1), 287–298. https://doi.org/10.1111/gcb.12681

Li, J., & Xiao, Z. (2020). Evaluation of the version 5.0 global land surface satellite (GLASS) leaf area index product derived from MODIS data. *International Journal of Remote Sensing*, *41*(23), 9140–9160. https://doi.org/10.1080/01431161.2020.1797222

Li, M., Wu, P., & Ma, Z. (2020). A comprehensive evaluation of soil moisture and soil temperature from third-generation atmospheric and land reanalysis data sets. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *40*(13), 5744–5766. https://doi.org/10.1002/joc.6549

Li, Y., Zhao, M., Mildrexler, D. J., Motesharrei, S., Mu, Q., Kalnay, E., et al. (2016). Potential and Actual impacts of deforestation and afforestation on land surface temperature. *Journal of Geophysical Research Atmospheres*, *121*(24), 14,372-14,386. https://doi.org/10.1002/2016jd024969

Li, Z., Yang, J., Gao, X., Yu, Y., Zheng, Z., Liu, R., et al. (2019). The relationship between surface spectral albedo and soil moisture in an arid Gobi area. *Theoretical and Applied Climatology*, *136*(3–4), 1475–1482. https://doi.org/10.1007/s00704-018-2577-3

Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., et al. (2021). The Global Land Surface Satellite (GLASS) Product Suite. *Bulletin of the American Meteorological Society*, *102*(2), E323–E337. https://doi.org/10.1175/bams-d-18-0341.1

Liang, X.-Z., Xu, M., Gao, W., Kunkel, K., Slusser, J., Dai, Y., et al. (2005). Development of land surface albedo parameterization based on Moderate Resolution Imaging Spectroradiometer (MODIS) data. *Journal of Geophysical Research: Atmospheres*, *110*(D11). https://doi.org/10.1029/2004JD005579

Liang, X.-Z., Choi, H., Kunkel, K., Dai, Y., Joseph, E., Wang, J., & Kumar, P. (2005). Surface boundary conditions for mesoscale regional climate models. *Earth Interactions*, *9*(18), 1–28. https://doi.org/10.1175/EI151.1

Liang, X.-Z., Xu, M., Yuan, X., Ling, T., Choi, H., Zhang, F., et al. (2012). Regional climate–Weather Research and Forecasting model. *Bulletin of the American Meteorological Society*, *93*, 1363–1387. https://doi.org/10.1175/BAMS-D-11-00180.1

Liu, H., Wang, B., & Fu, C. (2008). Relationships between surface albedo, soil thermal parameters and soil moisture in the semi-arid area of Tongyu, northeastern China. *Advances in Atmospheric Sciences*, *25*(5), 757–764. https://doi.org/10.1007/s00376-008-0757-2

Liu, H., Tu, G., & Dong, W. (2008). Three-year changes of surface albedo of degraded grassland and cropland surfaces in a semiarid area. *Science Bulletin*, *53*(8), 1246–1254. https://doi.org/10.1007/s11434-008-0171-2

Liu, Z., Chen, J. M., Jin, G., & Qi, Y. (2015). Estimating seasonal variations of leaf area index using litterfall collection and optical methods in four mixed evergreen–deciduous forests. *Agricultural and Forest Meteorology*, *209–210*, 36–48. https://doi.org/10.1016/j.agrformet.2015.04.025

Lobell, D. B., & Asner, G. P. (2002). Moisture effects on soil reflectance. *Soil Science Society of America Journal. Soil Science Society of America*, *66*(3), 722–727. https://doi.org/10.2136/sssaj2002.7220

Loew, A., van Bodegom, P. M., Widlowski, J.-L., Otto, J., Quaife, T., Pinty, B., & Raddatz, T. (2014). Do we (need to) care about canopy radiation schemes in DGVMs? Caveats and potential impacts. *Biogeosciences*, *11*(7), 1873–1897. https://doi.org/10.5194/bg-11-1873-2014

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., & Merchant, J. W. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, *21*(6–7), 1303–1330. https://doi.org/10.1080/014311600210191

Lucht, W., Schaaf, C. B., & Strahler, A. H. (2000). An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, *38*(2), 977–998. https://doi.org/10.1109/36.841980

Lukeš, P., Stenberg, P., & Rautiainen, M. (2013). Relationship between forest density and albedo in the boreal zone. *Ecological Modelling*, *261–262*, 74–79. https://doi.org/10.1016/j.ecolmodel.2013.04.009

Lukeš, P., Rautiainen, M., Manninen, T., Stenberg, P., & Mõttus, M. (2014). Geographical gradients in boreal forest albedo and structure in Finland. *Remote Sensing of Environment*, *152*, 526–535. https://doi.org/10.1016/j.rse.2014.06.023

Matthias, A. D., Fimbres, A., Sano, E. E., Post, D. F., Accioly, L., Batchily, A. K., & Ferreira, L. G. (2000). Surface roughness effects on soil albedo. *Soil Science Society of America Journal. Soil Science Society of America*, *64*(3), 1035–1041. https://doi.org/10.2136/sssaj2000.6431035x

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Nath, B., & Ni-Meister, W. (2021). The interplay between canopy structure and topography and its impacts on seasonal variations in surface reflectance patterns in the boreal region of Alaska—implications for surface radiation budget. *Remote Sensing*, *13*(16), 3108. https://doi.org/10.3390/rs13163108

Oleson, K. W., Bonan, G. B., Schaaf, C., Gao, F., Jin, Y., & Strahler, A. (2003). Assessment of global climate model land surface albedo using MODIS data. *Geophysical Research Letters*, *30*(8). https://doi.org/10.1029/2002gl016749

Pelosi, A., Terribile, F., D'Urso, G., & Chirico, G. (2020). Comparison of ERA5-Land and UERRA MESCAN-SURFEX reanalysis data with spatially interpolated weather observations for the regional assessment of reference evapotranspiration. *Water*, *12*(6), 1669. https://doi.org/10.3390/w12061669

Pinty, B., Roveda, F., Verstraete, M. M., Gobron, N., Govaerts, Y., Martonchik, J. V., et al. (2000). Surface albedo retrieval from Meteosat: 2. Applications. *Journal of Geophysical Research*, *105*(D14), 18113–18134. https://doi.org/10.1029/2000jd900114

Post, D. F., Fimbres, A., Matthias, A. D., Sano, E. E., Accioly, L., Batchily, A. K., & Ferreira, L. G. (2000). Predicting soil albedo from soil color and spectral reflectance data. *Soil Science Society of America Journal. Soil Science Society of America*, *64*(3), 1027–1034. https://doi.org/10.2136/sssaj2000.6431027x

Rago, M. M., Urretavizcaya, M. F., & Defossé, G. E. (2021). Relationships among forest structure, solar radiation, and plant community in ponderosa pine plantations in the Patagonian steppe. *Forest Ecology and Management*, *502*(119749), 119749. https://doi.org/10.1016/j.foreco.2021.119749

Ralhan, A., & Liang, X.-Z. (2025). Capturing Spatiotemporal and Subgrid Variability in Global Land Surface Albedo Parameterization [Data set]. Zenodo. https://doi.org/10.5281/ZENODO.16821420

Rechid, D., Raddatz, T. J., & Jacob, D. (2009). Parameterization of snow-free land surface albedo as a function of vegetation phenology based on MODIS data and applied in climate modelling. *Theoretical and Applied Climatology*, *95*(3–4), 245–255. https://doi.org/10.1007/s00704-008-0003-y

Roesch, A., Wild, M., Pinker, R., & Ohmura, A. (2002). Comparison of spectral surface albedos and their impact on the general circulation model simulated surface climate. *Journal of Geophysical Research*, *107*(D14). https://doi.org/10.1029/2001jd000809

Rotenberg, E., & Yakir, D. (2010). Contribution of semi-arid forests to the climate system. *Science (New York, N.Y.)*, *327*(5964), 451–454. https://doi.org/10.1126/science.1179998

Roxy, M. S., Sumithranand, V. B., & Renuka, G. (2010). Variability of soil moisture and its relationship with surface albedo and soil thermal diffusivity at Astronomical Observatory, Thiruvananthapuram, south Kerala. *Journal of Earth System Science*, *119*(4), 507–517. https://doi.org/10.1007/s12040-010-0038-1

Ryu, Y., Kang, S., Moon, S.-K., & Kim, J. (2008). Evaluation of land surface radiation balance derived from moderate resolution imaging spectroradiometer (MODIS) over complex terrain and heterogeneous landscape on clear sky days. *Agricultural and Forest Meteorology*, *148*(10), 1538–1552. https://doi.org/10.1016/j.agrformet.2008.05.008

Sadeghi, M., Babaeian, E., Tuller, M., & Jones, S. B. (2018). Particle size effects on soil reflectance explained by an analytical radiative transfer model. *Remote Sensing of Environment*, *210*, 375–386. https://doi.org/10.1016/j.rse.2018.03.028

Sanchez, P., Ahamed, S., Florence, Hempel, J., Jeroen Lagacherie, P., Mcbratney, A., et al. (2003). Digital soil map of the world (DSMW). *Science*, *325*.

Schaaf, C., & Wang, Z. (2021). MODIS/Terra+Aqua BRDF/AlbedoModel Parameters Daily L3 Global 0.05Deg CMG V061 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MCD43C1.061

Scherrer, D., & Körner, C. (2011). Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography*, *38*(2), 406–416. https://doi.org/10.1111/j.1365-2699.2010.02407.x

Scott, D., Menalda, P. H., & Brougham, R. W. (1968). Spectral analysis of radiation transmitted and reflected by different vegetations. *New Zealand Journal of Botany*, *6*(4), 427–449. https://doi.org/10.1080/0028825x.1968.10428581

Shuai, Y., Tuerhanjiang, L., Shao, C., Gao, F., Zhou, Y., Xie, D., et al. (2020). Re-understanding of land surface albedo and related terms in satellite-based retrievals. *Big Earth Data*, *4*(1), 45–67. https://doi.org/10.1080/20964471.2020.1716561

Sugathan, N., Biju, V., & Renuka, G. (2014). Influence of soil moisture content on surface albedo and soil thermal parameters at a tropical station. *Journal of Earth System Science*, *123*(5), 1115–1128. https://doi.org/10.1007/s12040-014-0452-x

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., & Friedl, M. A. (2019). Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment*, *222*, 183–194. https://doi.org/10.1016/j.rse.2018.12.013

Tian, Y., Zhang, Y., Knyazikhin, Y., Myneni, R. B., Glassy, J. M., Dedieu, G., & Running, S. W. (2000). Prototyping of MODIS LAI and FPAR algorithm with LASUR and LANDSAT data. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, *38*(5), 2387–2401. https://doi.org/10.1109/36.868894

Tsvetsinskaya, E. A., Schaaf, C. B., Gao, F., Strahler, A. H., & Dickinson, R. E. (2006). Spatial and temporal variability in Moderate Resolution Imaging Spectroradiometer–derived surface albedo over global arid regions. *Journal of Geophysical Research*, *111*(D20). https://doi.org/10.1029/2005jd006772

Wang, H., Wei, Z., Liu, Y., Liu, Y., & Li, X. (2021). Solar spectral albedo characteristics over a typical secondary evergreen broadleaf forest in the Lingnan area in China. *Theoretical and Applied Climatology*, *145*(3–4), 1075–1087. https://doi.org/10.1007/s00704-021-03688-9

Wang, H., Wei, Z., Huang, A., Li, X., Ma, L., & Guo, S. (2024). Validation of two newly developed albedo schemes based on the observations over the region with evergreen broadleaved forest in Southern China. *Journal of Geophysical Research Atmospheres*, *129*(1), e2023JD038682. https://doi.org/10.1029/2023jd038682

Wang, K., Wang, P., Liu, J., Sparrow, M., Haginoya, S., & Zhou, X. (2005). Variation of surface albedo and soil thermal parameters with soil moisture content at a semi-desert site on the western Tibetan Plateau. *Boundary - Layer Meteorology*, *116*(1), 117–129. https://doi.org/10.1007/s10546-004-7403-z

Wang, L., Zhu, H., Lin, A., Zou, L., Qin, W., & Du, Q. (2022). Evaluation of the ecological effects of vegetation cover change on solar radiation, surface temperature, and surface albedo. *Remote Sensing*, *14*(6).

Wang, S., & Davidson, A. (2007). Impact of climate variations on surface albedo of a temperate grassland. *Agricultural and Forest Meteorology*, *142*(2–4), 133–142. https://doi.org/10.1016/j.agrformet.2006.03.027

Wang, Z., Schaaf, C. B., Sun, Q., Shuai, Y., & Román, M. O. (2018). Capturing rapid land surface dynamics with Collection V006 MODIS BRDF/NBAR/Albedo (MCD43) products. *Remote Sensing of Environment*, *207*, 50–64. https://doi.org/10.1016/j.rse.2018.02.001

Wei, Z., Hu, J., Dong, W., Chen, G., Wu, Z., Wei, X., & Zheng, Z. (2016). Basic observations and diurnal variation of key meteorological variables on clear days in the Phoenix Mountain area of Zhuhai. *Chinese Journal of Atmospheric Sciences/Daqi Kexue*.

Wen, J., You, D., Han, Y., Lin, X., Wu, S., Tang, Y., et al. (2022). Estimating surface BRDF/albedo over rugged terrain using an extended multisensor combined BRDF inversion (EMCBI) model. *IEEE Geoscience and Remote Sensing Letters : A Publication of the IEEE Geoscience and Remote Sensing Society*, *19*, 1–5. https://doi.org/10.1109/lgrs.2022.3143197

Xiao, D., Tao, F., & Juana, P. (2011). Research progress on surface albedo under global change. *Advances in Earth Science*, *26*(11), 1217–1224.

Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., & Song, J. (2016). Long-Time-Series Global Land Surface Satellite Leaf Area Index Product Derived From MODIS and

AVHRR Surface Reflectance. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, *54*(9), 5301–5318. https://doi.org/10.1109/TGRS.2016.2560522

Yang, F., Mitchell, K., Hou, Y.-T., Dai, Y., Zeng, X., Wang, Z., & Liang, X.-Z. (2008). Dependence of land surface albedo on solar zenith angle: Observations and model parameterization. *Journal of Applied Meteorology and Climatology*, *47*(11), 2963–2982. https://doi.org/10.1175/2008jamc1843.1

Yang, J., Shuai, Y., Duan, J., Xie, D., Zhang, Q., & Zhao, R. (2022). Impact of BRDF Spatiotemporal Smoothing on Land Surface Albedo Estimation. *Remote Sensing*, *14*(9), 2001. https://doi.org/10.3390/rs14092001

Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., et al. (2006). MODIS leaf area index products: from validation to algorithm improvement. *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society*, *44*(7), 1885–1898. https://doi.org/10.1109/tgrs.2006.871215

Zhang, X., Jiao, Z., Zhao, C., Qu, Y., Liu, Q., Zhang, H., et al. (2022). Review of Land Surface Albedo: Variance Characteristics, Climate Effect and Management Strategy. *Remote Sensing*, *14*(6), 1382. https://doi.org/10.3390/rs14061382

Zhang, X. Z. (2012). The responses of surface albedo to climatic changes in xilin gol grassland. *Geographical Research*, *13*(3), 374–382.

Zhao, Z. Z., Zhao, K., Xu, J. B., Xiao, Z.-F., Cui, J.-F., & Hong, Z. (2014). Spatial-temporal changes of surface albedo and its relationship with climate factors in the source of three rivers region. *Arid Zone Research*, (06).

Zheng, G., Yang, H., Lei, H., Yang, D., Wang, T., & Qin, Y. (2018). Development of a physically based soil albedo parameterization for the Tibetan plateau. *Vadose Zone Journal: VZJ*, *17*(1), 1–21. https://doi.org/10.2136/vzj2017.05.0102

Zheng, Z., Dong, W., Li, Z., Zhao, W., Hu, S., Yan, X., et al. (2015). Observational study of surface spectral radiation and corresponding albedo over Gobi, desert, and bare loess surfaces in northwestern China: Surface spectral radiation and albedo. *Journal of Geophysical Research Atmospheres*, *120*(3), 883–896. https://doi.org/10.1002/2014jd022516

Zheng, Z., Wei, Z., Wen, Z., Dong, W., Li, Z., Wen, X., et al. (2017). Inclusion of solar elevation angle in land surface albedo parameterization over bare soil surface. *Journal of Advances in Modeling Earth Systems*, *9*(8), 3069–3081. https://doi.org/10.1002/2017MS001109

Zhu, J., & Zeng, X. (2015). Comprehensive study on the influence of evapotranspiration and albedo on surface temperature related to changes in the leaf area index. *Advances in Atmospheric Sciences*, *32*(7), 935–942. https://doi.org/10.1007/s00376-014-4045-z

Supplemental Information

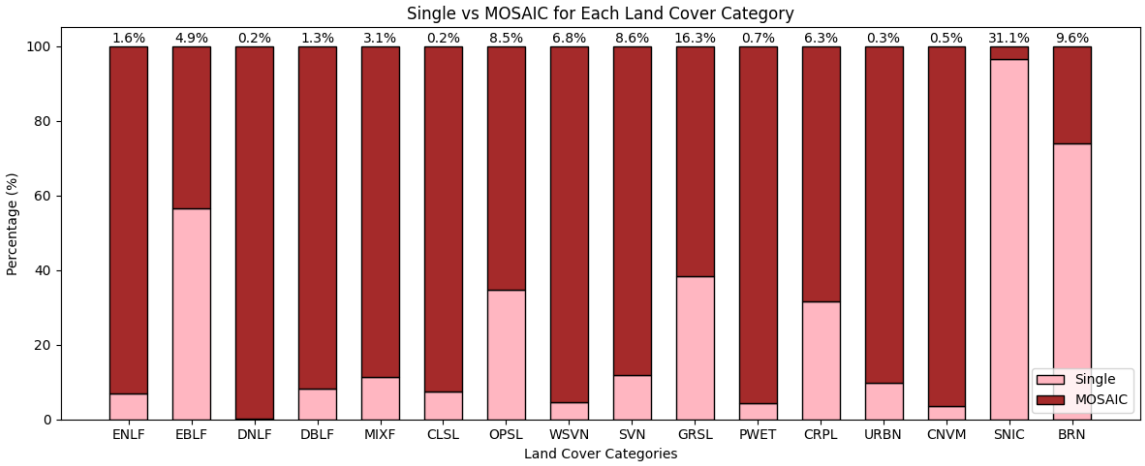

Figure S1. The plots display the fraction of the grids belonging to the Pure and Mosaic types. Each bar is labeled with the overall percentage of land surface covered by the given LCC.

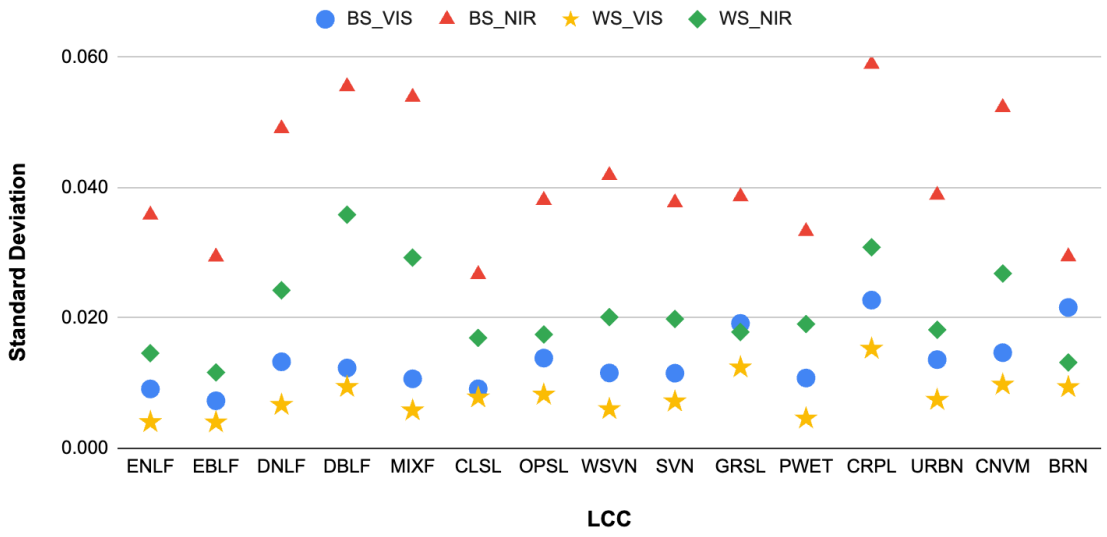

Figure S2. Scatter plot showing the average standard deviation of observed albedo for each LCC type and albedo parts.

Figure S3. Scatter plot showing the change in percentage points of R^2 (R^2_{SALF} - R^2_{XGB}) with the inclusion of SALF for each LCC and albedo parts.

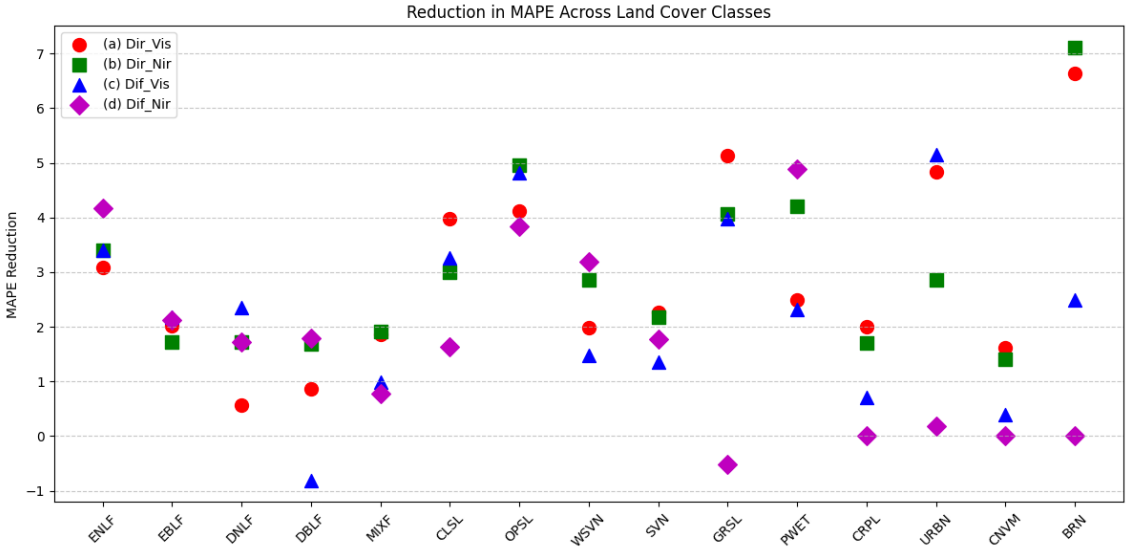

Figure S4. Scatter plot showing the change in percentage points of MAPE ($MAPE_{XGB}$ - $MAPE_{SALF}$) with the inclusion of SALF for each LCC and albedo parts.

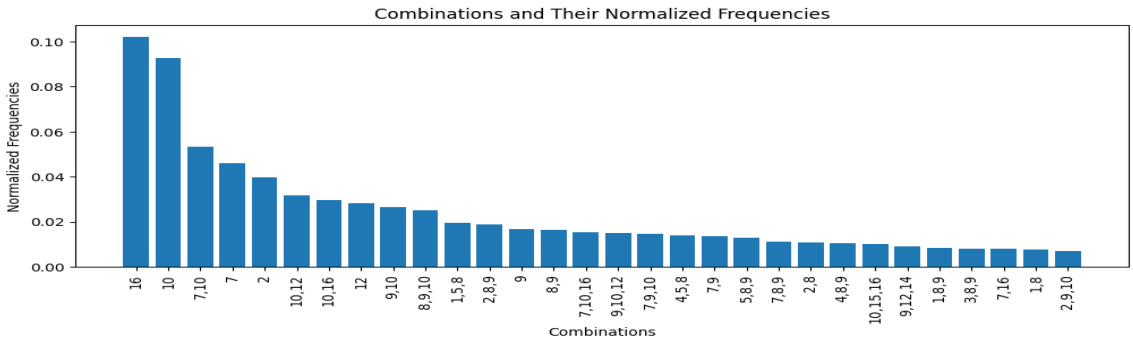

Figure S5: Frequency distribution showing the top 50 groups of Mosaic grids, highlighting the co-occurrence of certain LCCs with each other.

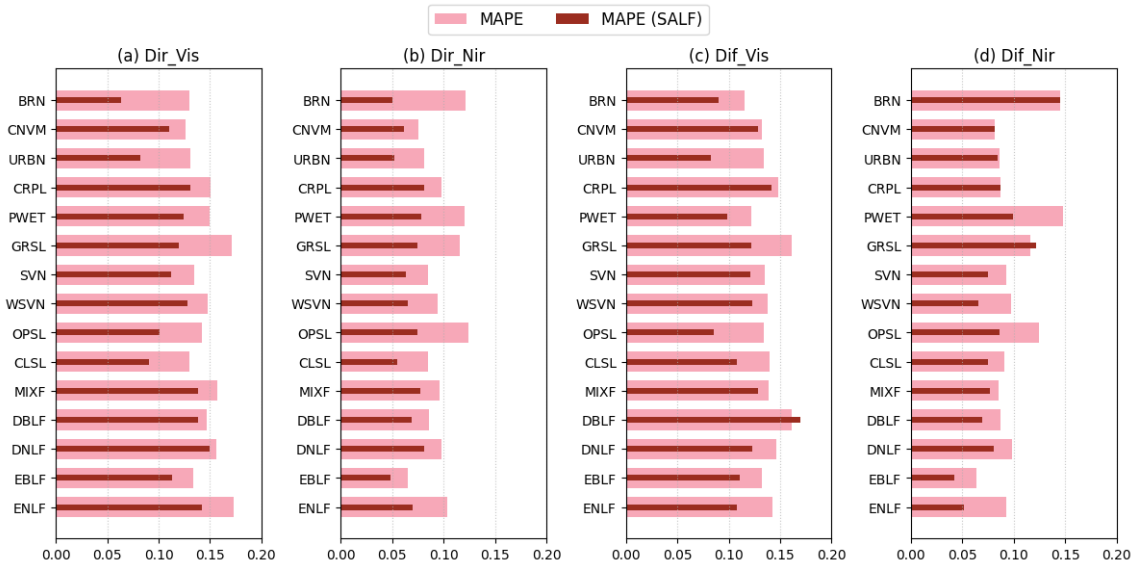

Figure S6: Comparison of changes in MAPE with the inclusion of SALF (solid dark red bars) for each LCC. The four panels represent different albedo parts: (a) Dir_Vis; (b) Dir_Nir; (c) Dif_Vis; and (d) Dif_Nir. Metrics are computed over all available grids, including both Pure and Mosaic grid types.

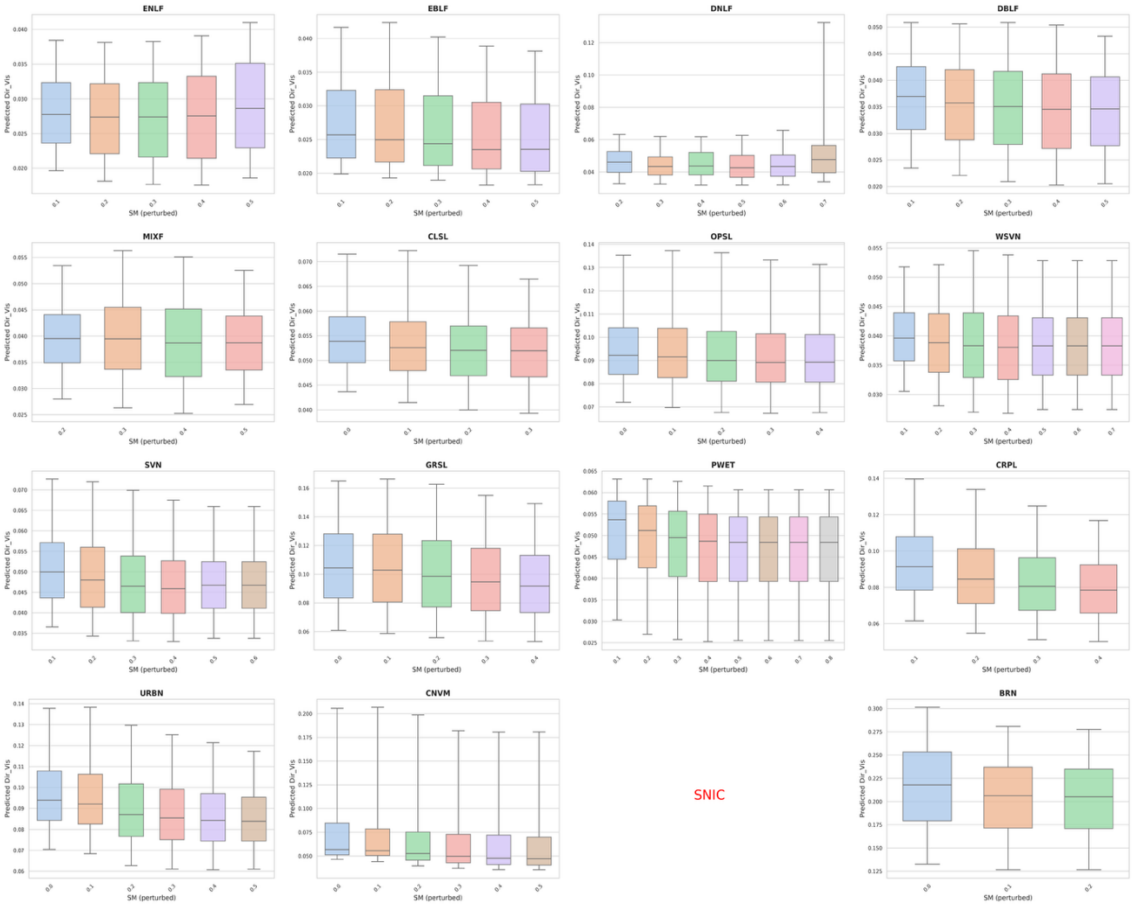

Figure S7: Counterfactual analysis of predicted Dir_Vis albedo to key biophysical parameters across 15 LCCs. Each panel shows the response of the XGBoost model predictions to perturbations in SM for Dir_Vis albedo, with boxplots representing the distribution of predicted values.

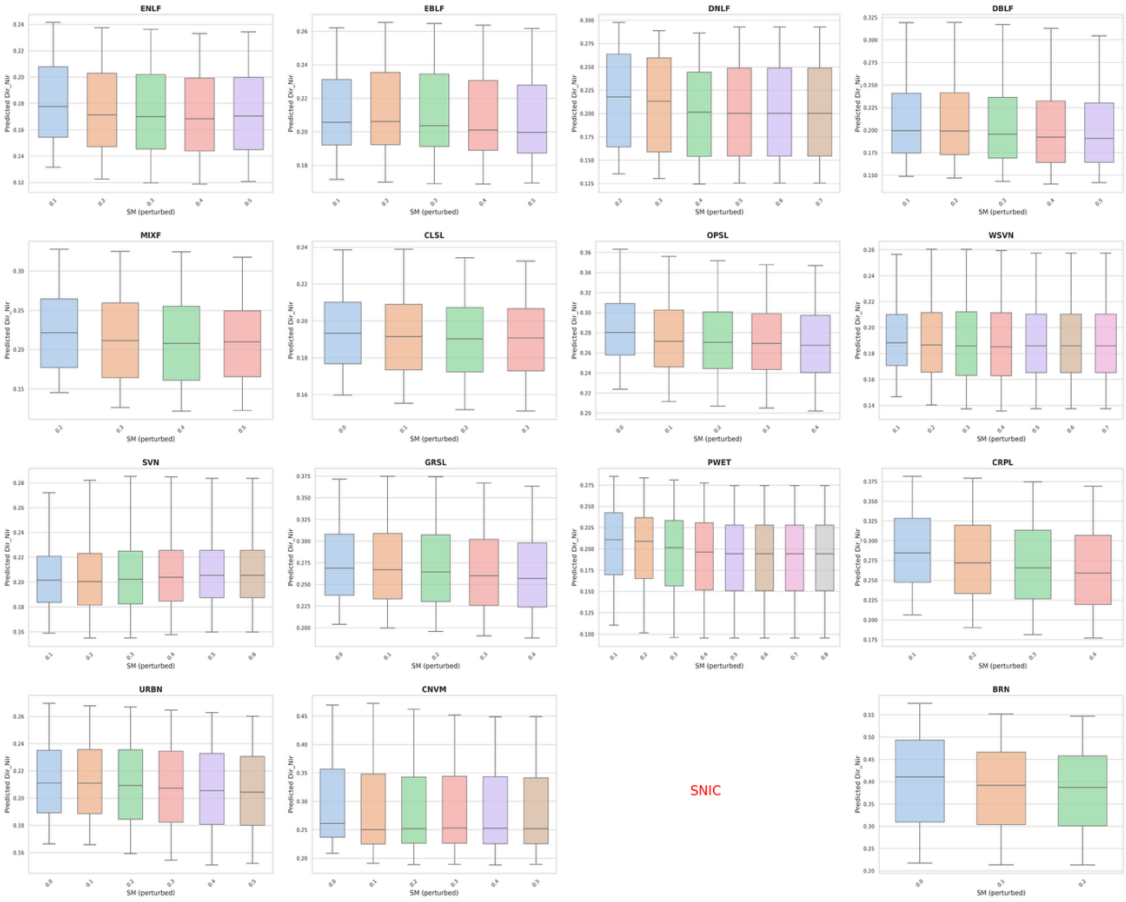

Figure S8: Counterfactual analysis of predicted Dir_Nir albedo to key biophysical parameters across 15 LCCs. Each panel shows the response of the XGBoost model predictions to perturbations in SM for Dir_Nir albedo, with boxplots representing the distribution of predicted values.

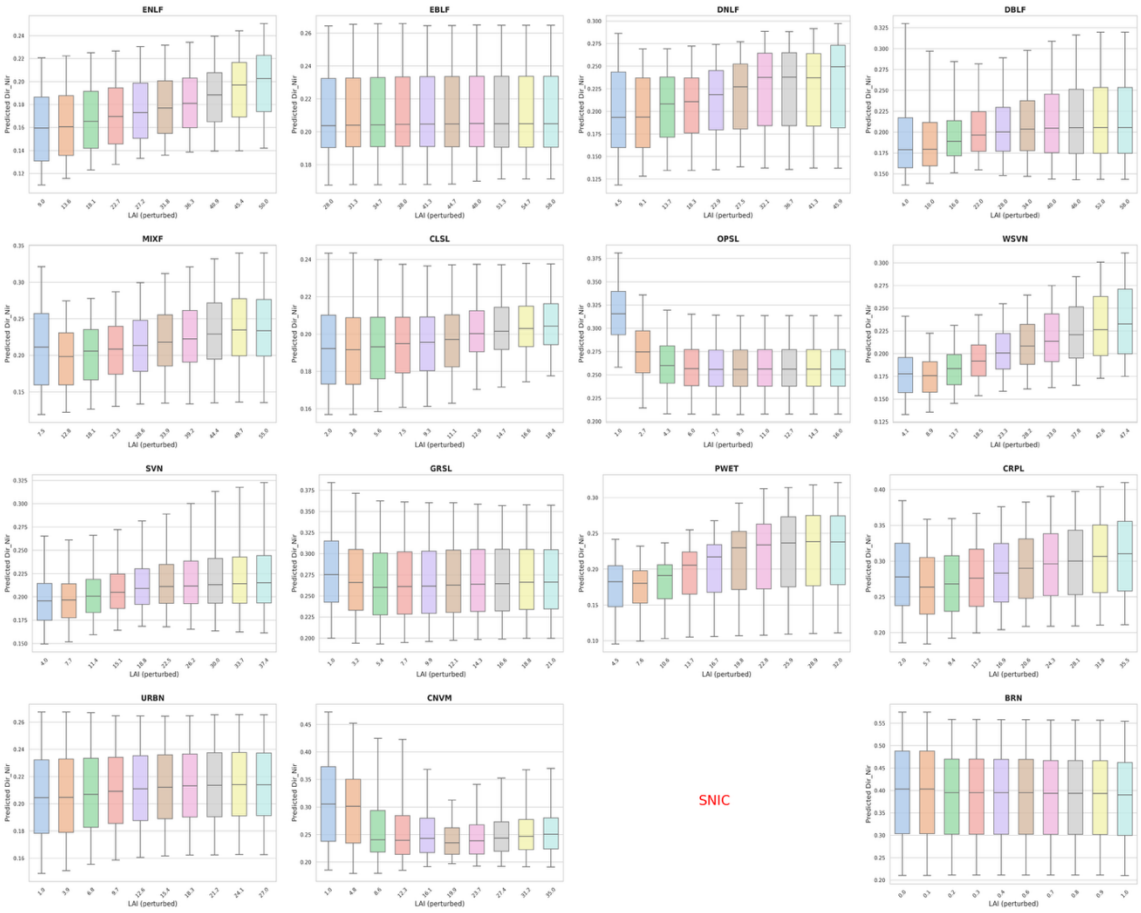

Figure S9: Counterfactual analysis of predicted Dir_Nir albedo to key biophysical parameters across 15 LCCs. Each panel shows the response of the XGBoost model predictions to perturbations in LAI for Dir_Nir albedo, with boxplots representing the distribution of predicted values.

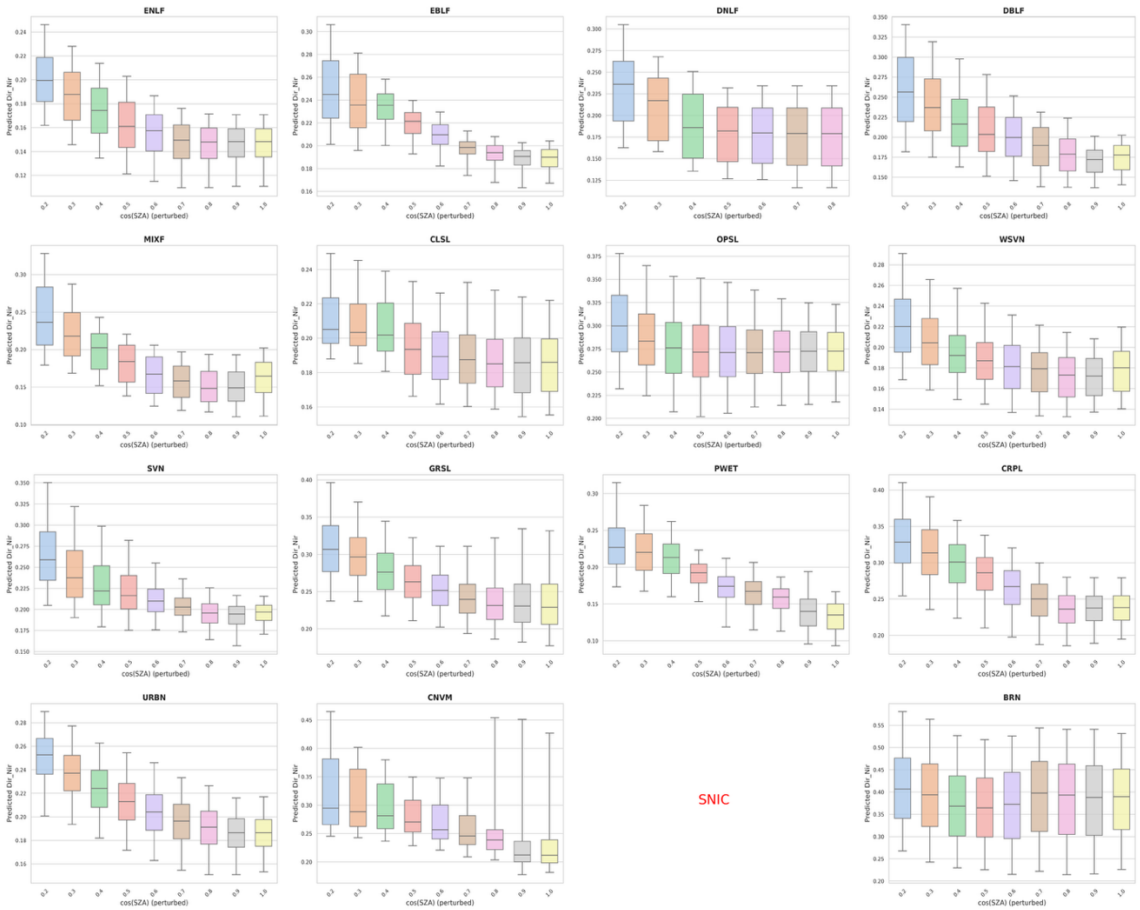

Figure S10: Counterfactual analysis of predicted Dir_Nir albedo to key biophysical parameters across 15 LCCs. Each panel shows the response of the XGBoost model predictions to perturbations in cos(SZA) for Dir_Nir albedo, with boxplots representing the distribution of predicted values.

Models	Key Variables	Parameterization Equations	References
CLM5 (CESM2 - NCAR) NorESM2 (Norway)	• Soil color class (20 classes) • Soil moisture (θ_l) • Solar zenith angle (μ) (for vegetation only) • LAI and SAI • PFT • Two-stream RTM • direct & diffuse • Vis & Nir	**Soil albedo:** Direct beam: $\alpha^{\mu}_{soi,\Lambda} = \alpha^{\mu}_{dry,\Lambda} + \Delta(\theta_1)$ Diffuse radiation: $\alpha_{soi,\Lambda} = \alpha_{dry,\Lambda} + \Delta(\theta_1)$ Both includes soil wetness adjustment: $\Delta = 0.11 - 0.40\theta_1 \quad (\text{if } \theta_1 > 0)$ Lookup tables for each soil color class for dry and saturated conditions **Vegetation albedo:** - Two-stream RTM - PFT-based lookup tables - Separate albedo values for stem and leaf reflectance	(Cenlin He et al., 2024; Lawrence et al., 2019)
JULES (HadGEM/ UKESM)	• PFT (5 categories) • LAI • Soil color • Solar zenith angle (μ) (for vegetation only) • Spectral albedo model (two-stream RTM) • direct & diffuse • Vis & Nir	Two options for the albedo scheme 1. Bulk albedo) $\alpha = \alpha_b \cdot e^{-kL} + \alpha_\infty \cdot (1 - e^{-kL})$ -α_b: Soil albedo (spatially varying with soil color). - α_∞: Maximum canopy albedo for dense vegetation. - k=0.5: Extinction coefficient. - L: Leaf Area Index (LAI). Lookup tables for each PFT 2. Two-stream RTM	Best et al. (2011) Sellers (1985) JULES Technical Note
Noah-MP (WRF)	• Soil moisture (θ_{soil}) • LAI and SAI • Solar zenith angle (μ) • Sunlit fraction of the canopy • direct & diffuse • Vis & Nir	**Bare soil albedo:** $C_w = 0.11 - 0.4 \cdot \theta_{soil}$ $\alpha_{dir} = \alpha_{sat} + C_w$ if $\alpha_{sat} + C_w < \alpha_{dry}$ otherwise α_{dry} $\alpha_{dif} = \alpha_{dir}$ Lookup tables for different soil classes, dry, diffuse, and albedo parts **Vegetation albedo:** Two-stream RTM - Absorbed flux depends on vegetation structure, solar zenith angle, and leaf optical properties - Three gap-fraction options: 1. Modified two-stream (3D canopy structure) 2. Grid average (gap = 0) 3. Vegetated fraction (gap = $1-f_{veg}$).	(C. He et al., 2023; Niu et al., 2011; Z.-L. Yang et al., 2011)
HTESSEL (EC-Earth)	• Vegetation fraction ($C_{eff,L/H}$) • Soil texture (7 types) • Soil color (9 types) • LAI	**Bare soil albedo:** $\alpha_{bs} = \alpha_{soil} * [1 - C_{eff,L} - C_{eff,H}]$ time-dependent soil albedo anomaly $\alpha_{soil} = \alpha_{soil,R}(i,j) + \alpha_{anomaly}(SM(t), texture, color)$ $\alpha_{soil,R}$= constant background map	(Balsamo et al., 2009)

	• Soil Moisture • Land categories (7 types) • direct & diffuse • Vis & Nir	Lookup tables for each albedo part **Vegetation albedo:** $\alpha_{veg} = \alpha_{vegL} * C_{eff,L} + \alpha_{vegH} * C_{eff,H}$ $C_{eff,L/H}$ = effective cover for low (L) and high (H) vegetation	
JSBACH (MPI-ESM)	•PFT-specific albedo (10 types) • LAI • direct & diffuse • Vis & Nir	1**. Regression framework:** $\alpha_{vis}(t) = f_{cover}(t) \cdot \alpha_{leaf,vis} + \left(1 - f_{cover}(t)\right) \cdot \alpha_{soil,vis}$ $\alpha_{nir}(t) = f_{cover}(t) \cdot \alpha_{leaf,nir} + \left(1 - f_{cover}(t)\right) \cdot \alpha_{soil,nir}$ - $f_{cover} \approx f_{apar} \cdot \left(1 - \alpha_{leaf,vis}\right)$ for Vis range. - Static maps were derived using MODIS white-sky albedo (MOD43C1) and fapar (MOD15A2) data (2001-2004) at 0.25° resolution **2. Regression Coefficients:** - Visible (Vis): $\alpha_{soil,vis} = b$ (intercept) $\alpha_{leaf,vis} = \frac{a + b - 1}{1 + a}$ (slope a) - Near-Infrared (Nir): $\alpha_{soil,nir} = \frac{\alpha_{nir} - f_{cover} \cdot \alpha_{leaf,\,nir}}{1 - f_{cover}}$ **3. Special Cases:** - Low variability: If $\Delta(\alpha_{vis}) < 0.02$ *or* $\Delta(f_{apar}) < 0.05$, $\alpha_{soil} = \alpha_{leaf}$ - Tropical adjustment: $\alpha_{soil,nir} = 0.15$ (dense tropical forest proxy) - High latitudes: Add 0.01-0.04 to albedo depending on latitude (e.g., +0.04 beyond 70°)	(Otto et al., 2011)
ORCHIDEE (IPSL-CM6A)	•PFT-specific albedo (15 types) • Soil Color (9 types) • LAI • Vis & Nir	1**. Bare soil and Vegetation albedo for PFT patches:** $\text{frac}_{bs} = \sum_{pft=1}^{13} \text{frac}_{max,pft} - \sum_{pft=2}^{13} \text{frac}_{pft}$ $\text{alb}_{veg} = \text{frac}_{bs} \cdot \text{alb}_{bs} + \sum_{pft=2}^{13} \text{frac}_{pft} \cdot \text{alb}_{leaf,pft}$ **2. Composite grid albedo:** $\text{albedo} = \text{frac}_{veg} \cdot \text{alb}_{veg} + \text{frac}_{nobio} \cdot \text{alb}_{nobio}$ - Non-biological surfaces (bare soil) based on lookup tables and soil type	

Table S1: Overview of key variables and surface albedo parametrization approaches in state-of-the-art land surface models (LSMs)