https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Open Access Earth System Science Data Discussions

CLRD-GLPS: A Long-term Seasonal Dataset of Ruminant Livestock Distribution in China's Grazing Production Systems (2000-2021) Using Stacking-based Interpretable Machine Learning

Ning Zhan [1,2,3,4], Tao Ye [1,2,3,4], Mario Herrero[5], Jian Peng[6,7], Weihang Liu [1,2,3,4], Heng Ma[8]

[1]State Key Laboratory of Earth Surface Processes and Resource Ecology (ESPRE), Beijing Normal University, Beijing, 100875, China
[2]Key Laboratory of Environmental Change and Natural Disasters, Ministry of Education, Beijing Normal University, Beijing 100875, China
[3]Academy of Disaster Reduction and Emergency Management, Ministry of Emergency Management and Ministry of Education, Beijing 100875, China
[4]Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China
[5]Department of Global Development, College of Agriculture and Life Sciences and Cornell Atkinson Center for Sustainability, Cornell University, Ithaca, NY 14850, USA.
[6]Xinjiang Uygur Autonomous Region Grassland Station, Urumqi, Xinjiang 830049, China
[7]Remote Sensing Monitoring Laboratory of Grassland Ecosystems in Arid Regions, Urumqi, Xinjiang 830049, China
[8]National Institute of Natural Hazards, Ministry of Emergency Management of China, Beijing, 100085, China

Correspondence to: Tao Ye (yetao@bnu.edu.cn)

Abstract.

Understanding the spatial-temporal distribution of grazing livestock is crucial for assessing livestock system sustainability, managing animal diseases, mitigating climate change risks, and controlling greenhouse gas emissions. In China, grazing ruminants are predominantly distributed across vast grasslands in semi-humid and alpine regions. However, existing gridded livestock distribution datasets fail to distinguish between grazing and other livestock production systems and do not simultaneously account for long-term and seasonal dynamics. This study introduces CLRD-GLPS, a comprehensive dataset mapping **C**hina's **r**uminant **l**ivestock **d**istribution in **g**razing **l**ivestock **p**roduction **s**ystems from 2000 to 2021. Our approach addresses limitations in existing datasets by integrating interpretable machine learning methods to segment grazing livestock from total livestock populations and generate seasonal grazing pastures with dynamic grazing suitability masks. We developed a stacking-based ensemble methodology that enhances predictive performance while providing insights into distribution mechanisms. The stacking ensemble models demonstrate robust performance through 5-fold cross-validation, with R^2 values ranging from 0.909 to 0.967 for cattle and 0.874 to 0.914 for sheep and goats. Validation results demonstrated the high accuracy of CLRD-GLPS across multiple spatial scales. At the county level, it strongly agreed with census data, effectively capturing grazing livestock distribution. City-level validation confirmed strong agreement (R^2 = 0.691–0.881), while grid-level validation using independent observations yielded R^2 = 0.79, further confirming the accuracy of fine-resolution predictions. The CLRD-GLPS dataset provides essential information for understanding grazing ruminant dynamics and developing

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Open Access Earth System Science Data Discussions

targeted livestock management policies. Furthermore, our methodological framework offers a template for creating similar livestock distribution datasets for other regions and livestock production systems.

1 Introduction

Livestock play an important role in global food systems, contributing 40% to the global agricultural gross domestic product. The global livestock sector is rapidly changing in response to growing demand for animal-source foods, employing over 1.3 billion people and supporting 600 million poor smallholder farmers in developing countries (Herrero et al., 2013; Thornton, 2010). Meanwhile, increasing livestock numbers contribute to a rise in greenhouse gas (GHG) emissions and places a significant burden on herders to gain access to the feed for livestock from natural resources (Gerber et al., 2013; Herrero et al., 2013). In the livestock sector, ruminant animals—such as cattle, sheep, and goats—occupy the largest land area worldwide compared to other livestock species, predominantly on grasslands (Pulina et al., 2017). Additionally, the lower feed use efficiency in ruminant than in monogastric livestock (such as pigs and poultry), has led to relatively higher GHG emissions intensities (Cheng et al., 2022; Knapp et al., 2014). Therefore, it is important to capture the spatial-temporal distribution of ruminant livestock for showcasing their role in studying sustainability (Michalk et al., 2019), managing disease (Li et al., 2024), mitigating climate change risks (Thornton et al., 2021), and especially in predicting GHG emissions (Uwizeye et al., 2020) associated with livestock production systems (LPS). Despite many efforts (Gilbert et al., 2018; Robinson et al., 2014), existing datasets often lack the spatial-temporal resolution and seasonal dynamics necessary for sustainability assessments and climate change impact studies in diverse LPS.

Existing of global ruminant livestock distribution maps, such as the Food and Agriculture Organization of the United Nations (FAO)'s Gridded Livestock of the World (GLW3) using machine learning methods with a spatial resolution of 10 km (Gilbert et al., 2018). Building upon these global datasets, tree-based models such as Random Forest (RF), Extra Trees Regressor (ET), and XGBoost (XGB) are widely employed in livestock distribution modelling and producing more high-resolution livestock distribution datasets in China (Li et al., 2021b; Zhan et al., 2023). The high-resolution livestock maps enable more accurate tracking of livestock movements across different seasons (Ocholla et al., 2024), Zhan *et al.* leveraged China's county-level livestock census data to generate cattle and sheep distribution data for the Qinghai-Tibet Plateau (QTP), particularly emphasizing seasonal variations with greater spatial resolution of 500m (Zhan et al. 2023). Additionally, the long-term distribution of livestock affects land use change and herd management, amongst others. The GDGI dataset provides the annual gridded grazing intensity data across the QTP from 1990 to 2020, offering valuable insights into long-term spatial and temporal variations in grazing pressure(Zhou et al., 2024) .

Despite these advancements, current livestock distribution datasets have several critical limitations. A key limitation is that existing datasets often model livestock distribution directly based on census data, without distinguishing livestock numbers

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

among various LPS. LPS have been previously classified into three basic types: grazing LPS, mixed farming LPS and landless LPS (Robinson et al., 2011). In China, a significant portion of ruminant livestock exists within the grazing LPS, particularly in western regions with extensive grasslands, while substantial numbers are also raised in mixed and landless systems throughout the country (Jiang and Wang, 2022). This distribution varies dramatically across regions, making spatial disaggregation particularly challenging. Previous studies have directly used total livestock census data to predict livestock distribution in grazing areas based on grassland suitability (Gilbert et al., 2018; Zhan et al., 2023), leading to systematic overestimation of actual grazing livestock numbers and misrepresentation of their spatial patterns. Recent advancements have attempted to address the distinction between pasture-based and crop-based livestock in China's grazing intensity assessment relies on NDVI-biomass conversion models (Wang et al., 2024a). A more comprehensive methodological framework is still needed to effectively distinguish livestock across different production systems.

Additionally, most ruminant livestock depend on grasslands for grazing and often move seasonally, especially in regions like QTP in China that adopt a two-season transhumance system (Zhan et al., 2023; Zhuang et al., 2019). Beyond these seasonal patterns, regions such as parts of Xinjiang utilize year-round pastures (Zheng, 2005). Existing datasets, such as GLW3, do not account for these regional differences in grazing patterns. Moreover, the coarse resolution and lack of consideration for seasonal livestock movements within local boundaries limit their applicability, particularly for studies focused on seasonal environmental stresses, such as heat stress and snow disasters (Thornton et al., 2021; Ye et al., 2021), and on seasonal grazing intensity (Fetzel et al., 2017). Furthermore, predicting long-term livestock distributions remains challenging, as there is currently no dataset that simultaneously meets the seasonal pattern and long-term series requirements for the diverse distribution patterns of livestock in grazing LPS.

The methodological approaches in current livestock distribution modelling also present challenges. Machine learning methods, which are commonly employed in creating livestock distribution datasets, are often considered "black-box" models that cannot directly explain the mechanisms behind the data. Interpretability remains a key challenge, as these models do not explicitly reveal the mechanisms driving spatial-temporal changes in livestock distribution (Hassija et al., 2024). This limitation hinders their use in understanding and explaining such changes, which is crucial for ecological and agricultural applications. Generalizability and stability are also critical concerns, as the performance of different ML methods varies across different datasets and spatial contexts.

Addressing these multifaceted challenges in livestock distribution datasets—including seasonal pattern identification, livestock production system differentiation, and methodological interpretability—requires an integrated and innovative approach. Interpretable machine learning (IML) techniques offer a promising solution by revealing the mechanisms driving model predictions and explaining the relationships between predictors and outcomes (Murdoch et al., 2019). For instance, feature importance scores and Shapley values help identify the key factors influencing predictions and their relative

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

significance(Breiman, 2001; Jiang et al., 2024), which can be applied to distinguish seasonal pastures and separate grazing livestock from the total livestock population. Furthermore, stacking ensemble machine learning, which combines multiple predictive models, has been shown to reduce bias while improving both accuracy and stability compared to individual ML models—critical attributes for generating reliable long-term spatial-temporal distribution data (Xu et al., 2024). Despite the potential benefits of stacking ensemble learning, no existing studies have systematically applied this approach to integrate the strengths of individual models for livestock distribution prediction. By leveraging stacking techniques, it is possible to enhance predictive performance while mitigating biases inherent to single-model approaches, ultimately addressing the key limitations identified in current livestock distribution datasets (Pavlyshenko, 2018).

This study aims to develop a long-term seasonal dataset mapping the distribution of grazing ruminant livestock in China from 2000 to 2021 (CLRD-GLPS), with a specific focus on the grazing LPS. The methodology for constructing this dataset involves addressing three key aspects. Firstly, long-term county-level statistical livestock data are collected and grazing ruminant livestock are identified within the total ruminant population. Secondly, grassland areas are differentiated into seasonal and year-round pastures using sampled seasonal pasture data. Lastly, the spatial-temporal distribution of ruminant livestock from 2000 to 2021 is predicted and explained using well-developed interpretable machine learning models and structural equation modelling.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

2 Data and methods

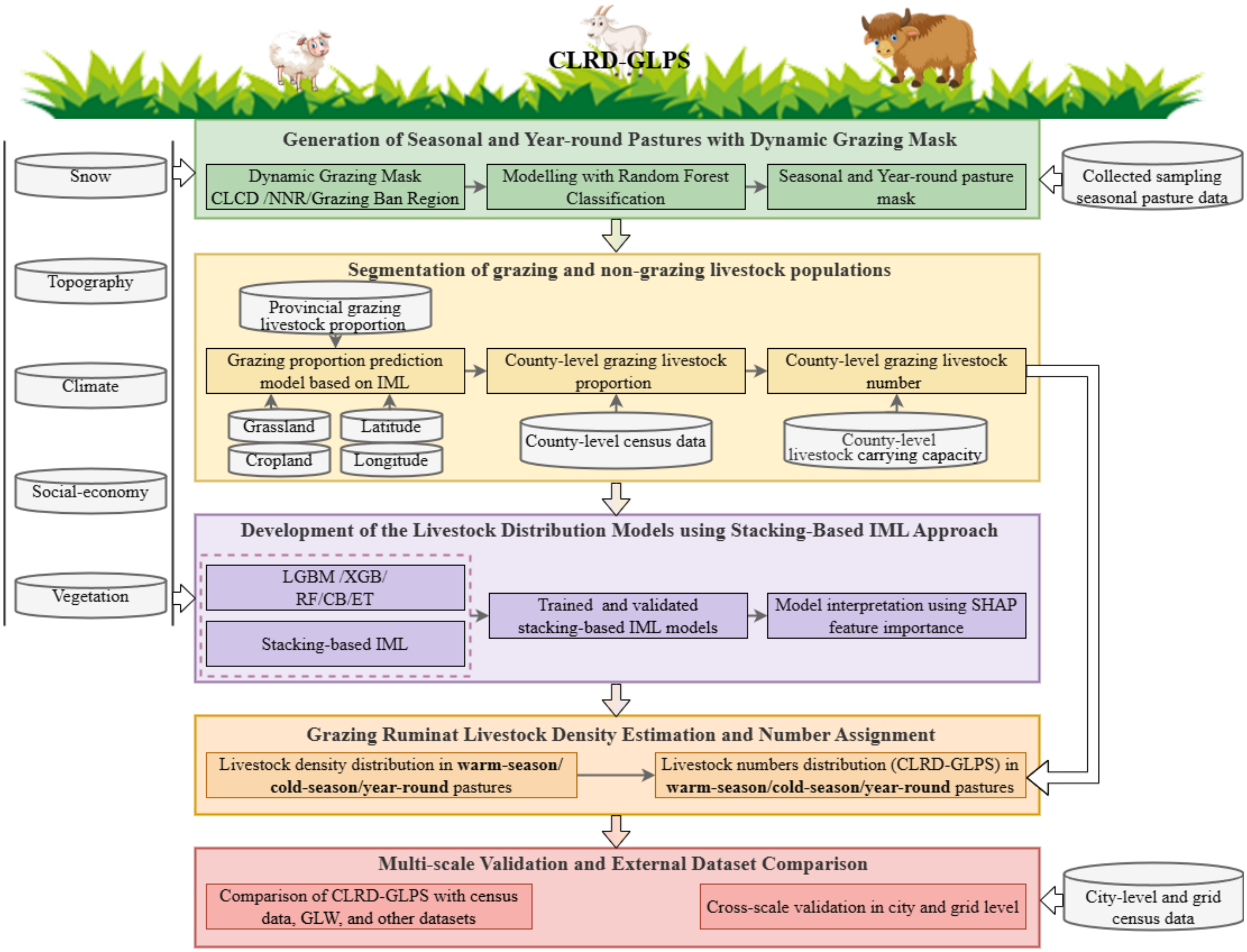


Figure 1: Framework of mapping grazing ruminant livestock distributions in China's grazing livestock production system.

In this study, we implemented five detailed steps to produce the CLRD-GLPS datasets (Figure 1): (1) Generation of seasonal and year-round pasture maps using dynamic grazing masks and Random Forest classification modelling, (2) livestock segmentation in grazing LPS by developing county-level grazing livestock proportion and number estimates, (3) development of stacking-based IML livestock distribution models integrating multiple algorithms (LGBM/XGB/RF/CB/ET), (4) estimation of livestock density in seasonal pastures and assignment of livestock numbers within county boundaries, and (5) multi-scale validation through comparison with census data, GLW, and other datasets.

2.1 Data

The data used for this study are categorized into four types: livestock data, mask data, pasture survey data, and environmental and socioeconomic data. All these categories are listed in Table A1, with detailed descriptions provided below.

2.1.1 Livestock data

County-level livestock data were collected from the livestock statistical yearbooks, encompassing information from 29 provinces across China (excluding Jiangsu, Fujian, Guangxi, Hong Kong, Macao, and Taiwan). These yearbooks provide the 2000-2021 year-end of cattle, sheep and goats (the data in some provinces of some years can't be found). In total, livestock numbers data are available for 16, 204 year·county.

For comparison, we downloaded the Gridded Livestock of the World (GLW) datasets for 2010, 2015, and 2020 from the FAO website (https://data.apps.fao.org/catalog/organization/gridded-livestock-of-the-world-glw), selecting the cattle, sheep and goats species available. The units for GLW in 2010 and 2015 represent absolute livestock numbers (Gilbert et al., 2018), whereas in 2020, the data are provided as livestock density (heads per km²). Additional livestock datasets were also collected for comparison, with detailed information recorded in Table A2 (Meng et al., 2023; Wang et al., 2024b; Zhou et al., 2024).

In addition, we further collected city-level livestock numbers from 2000 to 2021 from various provincial statistical yearbooks (Table A1), as well as grid-scale livestock density observation data recorded in previous literature and the *National Rural Fixed Observation Point Micro-household Survey* (NRFOP) data Table A3.

2.1.2 Mask data

The annual China Land Cover Dataset (2000-2021) with a 30 m spatial resolution was utilized to create a suitable distribution mask for livestock (Yang and Huang, 2021a). To generate a valid pasture boundary, we also obtained the boundaries of national nature reserves from the National Nature Reserve Boundary Data published by the Resource and Environment Science and Data Center, Chinese Academy of Sciences (https://www.resdc.cn/data.aspx?DATAID=272), which includes 169 national nature reserves established as of 2021 in China. The boundary of grazing ban regions in Tibet from 2004 to 2012 was collected from the article (Sun et al., 2020) ,while that in Inner Mongolia was collected in *Inner Mongolia Grassland Resources Ecological Monitoring Report (2016-2020)*. These regions are banned for livestock grazing.

2.1.3 Pasture survey Data

The sampled seasonal pasture data were collected for generating seasonal pasture of grazing livestock in the regions of Xinjiang, Tibet, and Qinghai. The warm-season, cold-season, and year-round pasture division maps for entire Xinjiang were obtained

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

from the *Xinjiang Autonomous Region Grassland Station*. The 1365 grassland survey sample locations of seasonal pasture of Qinghai Province were obtained from the *Qinghai Province Grassland Station*. For the Tibet Autonomous Region, the division maps of warm/cold-season pastures of 48 townships were obtained from *Zhada, Geji, Jilong, and Dingjie County Forestry and Grassland Administrations*.

2.1.4 Environmental and socioeconomic predictor data

Topography data such as digital elevation model (DEM) and slope were considered in the livestock density distribution model. Climate data, crucial for assessing the productivity of grasslands and other land cover types, also impacts the challenges posed by climatic conditions on livestock. From 2000 to 2021, we included monthly near-surface temperature and precipitation (Peng et al., 2019), snow cover data (Hall, D. K. and G. A. Riggs, 2016), and wind data (Hersbach et al., 2023). Vegetation productivity was represented by the normalized difference vegetation index (NDVI) data spanning from 2000 to 2021 (Didan, 2015). Additionally, the socioeconomic data comprised population distribution from 2000 to 2021(Oak Ridge National Laboratory, 2020) and travel time data from 2015 (Weiss et al., 2018).

2.2 Method

2.2.1 Preparation of dynamic suitability grazing mask

The suitable land cover types for grazing livestock in grazing LPS include grassland (Howard et al., 2012), shrubland (Sanz et al., 2017), and wetland (Burton et al., 2009). In the Qinghai-Tibet Plateau (QTP) region, all three land cover types—grassland, shrubland, and wetland—were used as suitable grazing mask areas (Zhan et al., 2023). Meanwhile, for other regions in China, only grassland was considered as a suitable grazing mask (Wang et al., 2024a). To account for the impact of land use changes on grazing livestock distribution, we utilized the China Land Cover Dataset (CLCD) with 30 m spatial resolution, available annually from 2000 to 2021. The CLCD data were resampled to 1 km spatial resolution, retaining only pixels corresponding to the suitable land cover types based on regional differences as described above.

To delineate valid pasture boundaries, we incorporated two key constraints: the boundaries of National Nature Reserves (NNRs) and designated grazing ban regions (GBRs), where livestock grazing is prohibited (Figure A1). For NNRs, we accounted for their establishment timelines by utilizing the National Ecological Protection Redline Database (2022) to mask grazing density maps according to each reserve's official designation year. For instance, grazing activity was prohibited in Qiangtang Reserve following its establishment in 1993. A detailed record of each NNR's establishment year is provided in Figure A2, with 82.2% of NNRs established before 2000. For GBRs, we integrated spatially explicit datasets that reflect region-specific restrictions. The fenced grazing ban regions in Tibet (2004–2012) were obtained from Sun *et al.* (2020), while grazing ban regions in Inner

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Mongolia (2016–2020) were derived from the *Inner Mongolia Grassland Resources Ecological Monitoring Report*. These temporal grazing bans were dynamically applied to the corresponding years in the grazing mask, ensuring consistency in restricted grazing areas over time.

2.2.2 Generation of seasonal and year-round pastures

Based on the dynamic suitability grazing pastures, the distribution of seasonal pasture samples (warm-season pastures vs. cold-season pastures) was used to predict the seasonal pasture distribution across the entirety of China (Figure A1). Data collected from the scientific survey show that livestock in the QTP (Tibet, Qinghai, Sichuan, Yunnan, Gansu) follow seasonal grazing rules, grazing on cold-season pastures during the cold season and on warm-season pastures during the warm season. Although Xinjiang has seasonal pastures, there are also some areas with year-round pastures that can support grazing in either the cold or warm season. In other provinces of China, grazing typically occurs on year-round pastures without strict seasonal restrictions.

Therefore, this study predicted the seasonal pasture distribution separately for QTP, Xinjiang, and other provinces. For the QTP, we used a Random Forest Classification (RFC) model to predict seasonal pasture(warn-season/cold-season) (Breiman, 2001; Zhan et al., 2023). The Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) demonstrates the performance of the seasonal pasture prediction model (Negnevitsky, 2005). Detailed methods can be found in our previous study focused on livestock seasonal mapping (Zhan et al., 2023). In this study, we further integrated interpretable machine learning (IML) techniques into the RFC model by assessing feature importance (Breiman, 2001), allowing us to identify key environmental and socio-economic drivers of seasonal pasture distribution. In addition to topographic (DEM, slope), vegetation (NDVI), and climatic variables (GStmp, GSpre, Wtmp, Wpre), wind speed (wind) was also included as an important predictor. For Xinjiang, the distribution maps of seasonal and year-round pastures were directly converted into raster maps with a spatial resolution of 1 km, covering the entire region. In other provinces, pastures are all year-round.

2.2.3 Segmentation of grazing and non-grazing livestock populations

According to the *China Animal Husbandry and Veterinary Statistics Yearbook* (2000–2021), grazing livestock numbers at the end of each year are documented for pastoral and semi-pastoral regions. This dataset is available at the provincial (or autonomous regional) level and provides a comprehensive, long-term record (2000–2021) of grazing livestock distribution across China. By comparing the number of grazing ruminant livestock (cattle, sheep and goats) to the total ruminant livestock inventory for each province (or region), we derived the proportion of livestock within grazing livestock production systems (LPS) at the provincial scale.

To downscale the provincial-level grazing livestock proportions to counties with grazing LPS, we developed prediction models using Random Forest, an interpretable machine learning (IML) technique. We selected this method for its ability to capture

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

complex non-linear relationships between environmental factors and livestock distribution patterns, its effectiveness with spatial data, and its resistance to overfitting (Breiman, 2001).

For model development, we used a comprehensive set of predictors strongly associated with grazing proportion. The response variable was the proportion of grazing livestock at the provincial level from 2000 to 2021. Predictors included: (1) the proportion of grassland (calculated as the ratio of grassland area to total land area using CLCD), (2) the proportion of cropland (similarly derived from CLCD), (3) the Normalized Difference Vegetation Index (NDVI) of grassland and cropland (representing vegetation productivity), and (4) the geographic coordinates of the administrative centre (longitude and latitude). These variables were selected as they capture key ecological and land use factors commonly associated with grazing activities across different LPS.

We split the provincial-level data into training (80%) and validation (20%) sets. The IML model was optimized using hyperparameter tuning through a grid search approach with 5-fold cross-validation to identify the optimal combination of parameters (including number of trees, maximum depth, and minimum samples per leaf). Model performance was evaluated using coefficient of determination (R^2) and root mean square error (RMSE) metrics. To interpret model results, we employed SHAP (Shapley Additive Explanations) values (Sundararajan and Najmi, 2020), which provide transparent insights into how each predictor contributes to the model's predictions.

To evaluate the accuracy of these predictions, we validated the county-level grazing livestock proportions using empirical data on livestock carrying capacities from 74 counties participating in the *Grassland Ecological Protection Subsidy and Reward Program* (Table A3). This dataset, recorded during the implementation of the subsidy program, captures actual grazing intensity, including instances of overgrazing, thereby reflecting real-world grazing conditions. Notably, this dataset records the total grazing livestock population, including both cattle and sheep (sheep and goats). Therefore, for validation purposes, we converted the predicted grazing proportions of cattle and sheep (sheep and goats) into the overall grazing proportion of total livestock at the county level, using a standard conversion factor where one head of cattle is equivalent to five standard sheep units. Model performance was assessed using the R^2 and RMSE.

Finally, we used the predicted county-level grazing livestock proportions—specifically for cattle and sheep (sheep and goats)—to estimate the number of grazing livestock within grazing LPS at the county level for each year from 2000 to 2021:

$$LN_{Gij} = LN_{Tij} \cdot P_{Gij} \ , \tag{1}$$

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

where LN_{Gij} represents the number of grazing livestock in grazing LPS for county *i* in year *j*, and LN_{Tij} is the total livestock population in county *i* in year *j* as reported in the county-level statistical yearbook, and P_{Gij} the predicted proportion of grazing livestock in grazing LPS for county *i* in year *j*.

2.2.4 Development of the livestock distribution models and assignment of livestock numbers

This study employed stacking-based interpretable machine learning (IML) techniques to develop distribution models for the grazing ruminant livestock (cattle,sheep and goats). We constructed six livestock distribution models: three each for cattle and sheep (sheep and goats) corresponding to the three pasture types (warm-season, cold-season, and year-round pastures).

For our models, county-level livestock density served as the response variable, measured in heads per square kilometre (heads/km²). For counties with a single pasture type, we calculated this density by simply dividing annual livestock numbers for grazing LPS (LN_{Gi}) by the area covered by the suitability mask for that pasture type. For counties with multiple pasture types (particularly in Xinjiang where cold-season, warm-season, and year-round pastures coexist), we employed a specific methodology. Since year-round pastures are accessible during both seasonal periods, we calculated livestock density for warm-season models by dividing the county's total livestock census by the combined area of warm-season and year-round pastures. Similarly, for cold-season models, we divided the same county's livestock census by the combined area of cold-season and year-round pastures. For the year-round pasture model specifically, we used the average of these two density calculations. This approach accounts for the overlapping functional role of year-round pastures in our density estimations while maintaining consistency with established seasonal grazing patterns.

In developing these models, we considered multiple environmental and anthropogenic factors as predictors (Gilbert et al., 2018; Zhan et al., 2023). The distribution patterns of livestock within grazing LPS are primarily influenced by topography (acting as a macro-control factor), climate (determining grassland type and productivity), vegetation productivity (affecting carrying capacity), and pastoralist activities. Detailed information on all predictor variables is provided in Table A1.

Our modelling approach utilized a stacking-based IML framework to enhance prediction accuracy and stability. We split the dataset into training (80%) and testing (20%) sets to ensure proper model evaluation. The stacking architecture employed a two-layer structure with base learners in the first layer and a meta-model in the second (Figure A3). For base learners, we selected five machine learning regression models: Random Forest Regressor (RF) (Breiman, 2001), Extra Trees Regressor (ET) (Geurts et al., 2006), XGBoost Regressor (XGB) (Friedman, 2001), LightGBM Regressor (LGBM) (Ke et al., 2017), and CatBoost Regressor (CB) (Prokhorenkova et al., 2018). These models were chosen for their strong fitting capabilities and robustness in handling nonlinear relationships. Each base model was trained using 5-fold cross-validation on the training data,

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

generating predictions that served as meta-features for the second layer. For the second layer, we compared performance across base models and selected the one with highest R^2 and lowest RMSE as the meta-model for final ensemble prediction.

To ensure the robustness of the model, we employed a separate validation process using the 20% testing data that was not used during model training. The evaluation metrics included the coefficient of R^2 and RMSE, which were used to assess the fit and predictive accuracy of the models. The validation results of all base models and the stacking model were statistically analysed and compared to identify the best-performing model. Additionally, SHAP values were utilized to interpret the feature importance of each base model in the stacking ensemble, providing valuable insights for subsequent model optimization and result analysis(Sundararajan and Najmi, 2020).

The optimal models selected in this study were converted into grid-based weights within the pasture mask (warm-season pastures, cold-season pastures, year-round pastures) for each county-level polygon to assign county-level livestock numbers for grazing LPS (LN_{Gi}) using the dasymetric mapping method (Mennis, 2009). The final distribution maps of livestock numbers (CLRD-GLPS) were produced with units per grid cell in heads (as each grid cell is 1 km, this can be considered heads/km²).

2.2.5 Multi-scale validation and external dataset comparison

To compare our results (CLRD-GLPS) with other livestock distribution maps and actual distribution patterns, we used the GLW, LHGI, GDGI, Meng datasets for 2000, 2010, 2015, and 2020, standardized to livestock numbers (SSUs) as representatives of widely used global datasets (one cattle is equivalent to five Standard Sheep Units (SSUs)). Detailed information about these datasets is provided in Table A2. Actual (census) livestock distribution patterns were derived from county-level livestock numbers for grazing LPS (LN_{Gi}) in corresponding years for each county-level polygon. We aggregated values to the county level and calculated the R^2 and RMSE to assess validation accuracy.

To conduct rigorous external cross-validation and further ensure the robustness of our results, we performed additional multi-scale validation at both city and grid levels. For city-level validation, we aggregated CLRD-GLPS results to the city administrative level as predicted values, and then compared these with city-level census data adjusted by provincial grazing proportion to obtain observed grazing livestock numbers in the Grazing LPS. At the grid level, we validated our results against independent livestock density observations from previous literature, as well as calculated livestock densities from the National Rural Fixed Observation Point Micro-household Survey data (NRFOP), which provided valuable ground-truth information on livestock numbers and pasture areas for grazing and semi-grazing villages. Grid-level validation incorporated combined

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

livestock densities (cattle, sheep, and goats) expressed in SSUs. For all validation approaches, we calculated the R^2 and RMSE to comprehensively assess accuracy and reliability across different spatial scales.

3 Results

3.1 Seasonal and year-round pastures generated based on dynamic suitable grazing mask

With the dynamic CLCD datasets, National Nature Reserves (NNRs), and designated grazing ban regions (GBRs), we produced dynamic suitable grazing masks for 2000 to 2021. Based on these dynamic suitable grazing masks, we established seasonal pasture masks using an interpretable machine learning (IML) approach with Random Forest Classification (RFC) models. The RFC model demonstrated excellent performance with a mean AUC of 0.98 across ten-fold cross-validation (Figure A4 (a)), indicating high reliability in predicting seasonal pastures for livestock grazing.

The feature importance analysis revealed that topographic factors were the primary drivers of seasonal pasture distribution, with slope being the most influential factor (15% importance) followed by elevation (DEM, 12%) (Figure A4 (b)). Vegetation condition (NDVI, 11%) and socioeconomic factors (Travel Time, 11%) also significantly contributed to the model's predictive power. Climate variables collectively accounted for approximately 20% of the classification power, with growing season precipitation (GSpre) and summer temperature (GSmp) being the most important climate factors.

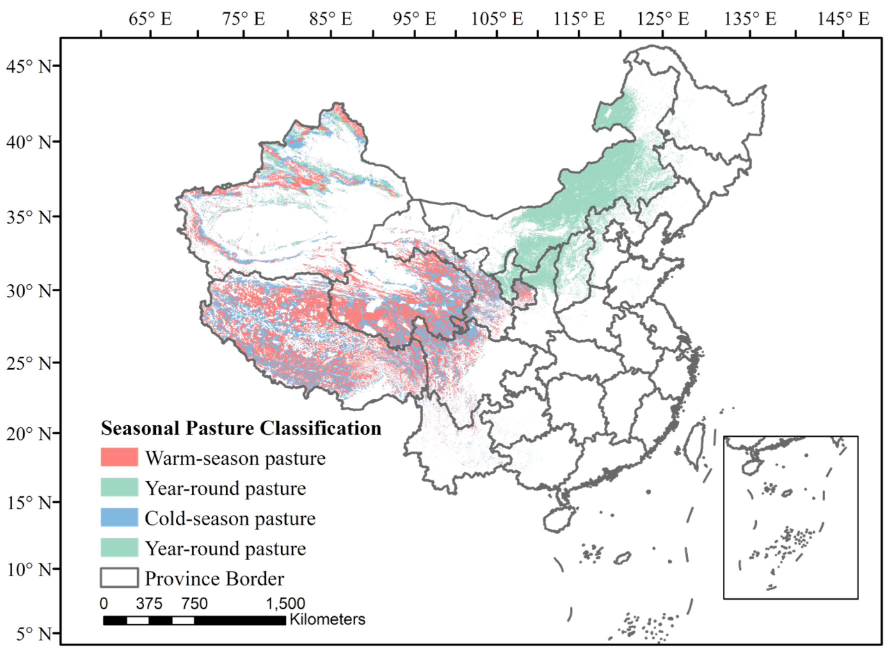


Figure 2: Seasonal and year-round pasture mask for livestock grazing in China (2021).

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Using this validated RFC model, we predicted seasonal pastures across the five provinces of the Qinghai-Tibet Plateau. By integrating these predictions with the dynamic suitability mask and the seasonal and year-round pasture distribution maps of Xinjiang, we generated 1-km resolution seasonal and year-round pasture distribution masks for grazing ruminant livestock in grazing production systems for each year from 2000 to 2021, with the 2021 distribution shown in Figure 2.

The temporal dynamics of seasonal pasture types across China from 2000 to 2021 illustrates in Figure A5. The total pasture area exhibited a slight declining trend, primarily driven by changes in grassland area, the establishment of NNRs, and the implementation of the GBR policy (Figure A5 (a)). Throughout the study period, warm-season pastures consistently occupied approximately 30% of the total area, cold-season pastures dominated at around 35%, while year-round pastures maintained approximately 35% of the total pasture area (Figure A5 (b)). The implementation of grazing ban policies in Tibet (2004-2012) and Inner Mongolia (2016-2020) coincided with minimal changes in the absolute area of each seasonal pasture type. The proportional composition of seasonal pastures remained relatively stable despite policy interventions, suggesting limited effectiveness of grazing bans in altering seasonal utilization patterns of grazing ecosystems.

3.2 Grazing ruminant livestock segmentation in grazing livestock production system

We developed interpretable machine learning models to predict grazing livestock proportions for cattle and sheep (sheep and goats) at the provincial level. The 10-fold cross-validation results demonstrated high predictive performance for both models, with R^2 values of 0.933 for both cattle and sheep (sheep and goats), and RMSE values of 0.084 and 0.081, respectively (Figure 3 (a)-(b)). The SHAP value distribution revealed distinct feature importance for livestock proportions. For cattle, the primary determinants were cropland area, longitude, and grassland area, while for sheep and goats, the key factors were grassland area, cropland area, and longitude (Figure 3 (c)-(d)). Notably, the SHAP analysis consistently showed that larger grassland areas, smaller cropland areas, and higher longitudes corresponded to higher livestock proportions.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

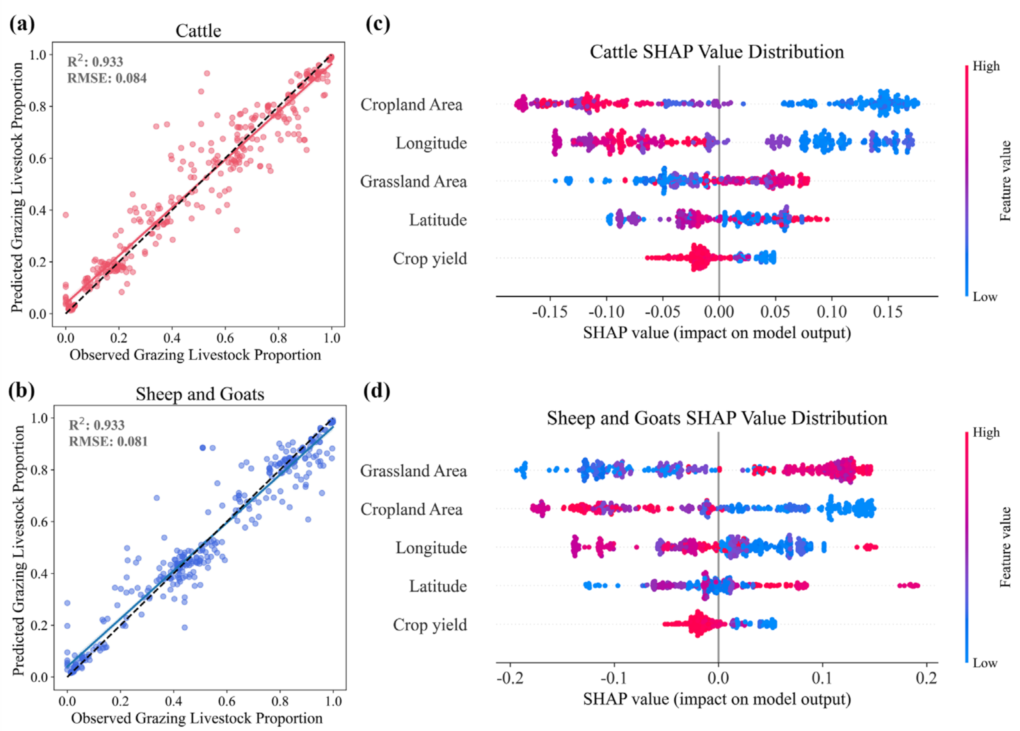


Figure 3: Model performance and feature importance analysis for grazing ruminant livestock segmentation model in China. (a–b) Cross-validation results of the grazing livestock proportion prediction models. (c-d) SHAP value distributions illustrating the relative importance of features in the grazing livestock proportion models. (d) County-level validation of razing ruminant livestock segmentation model using existing livestock carrying capacity data.

Using provincial-level models, we predicted cattle and sheep (sheep and goats) proportions for counties with grazing LPS from 2000 to 2021. The 2021 predictions (Figure 4 (a)-(b)) highlighted regions with high grazing livestock proportions, particularly in Tibet, Qinghai, Xinjiang, and Inner Mongolia, where proportions exceeded 0.6. Slight variations were observed, with cattle proportions marginally higher than sheep and goats in these regions. Change analysis between 2000-2021 (Figure 4(c)-(d)) revealed a predominant decreasing trend in grazing livestock proportions across China, with 56.7% of counties showing decreased cattle proportions and 60.0% showing decreased sheep (sheep and goats) proportions. Simultaneously, notable increases were observed in 40.6% of counties for cattle and 37.6% for sheep and goats, primarily in specific regions of western China. To validate the county-level predictions, we compared our results with existing county-level livestock carrying capacity datasets. The validation (Figure A6) yielded a robust R^2 of 0.800 and a low RMSE of 0.023, indicating high prediction accuracy. Finally, we applied the annual county-level grazing livestock proportions to segment grazing livestock numbers within grazing LPS using county-level livestock census data from 2000 to 2021.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

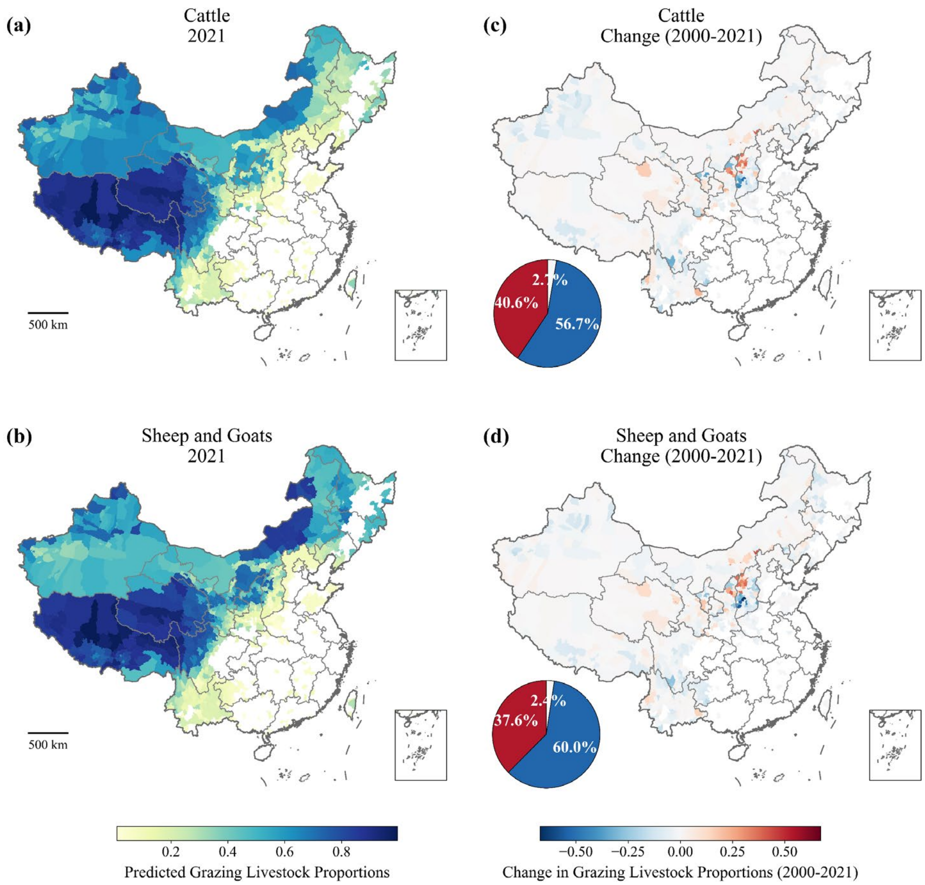


Figure 4: Spatial distribution and temporal changes in grazing livestock proportions across China. (a-b) Cattle (sheep and goats) proportions in 2021; (c-d) Change in cattle (sheep and goats) proportions between 2000-2021 with pie chart showing proportion of counties experiencing increase (red) or decrease (blue).

3.3 Grazing ruminant livestock density model developed and livestock number assigned

The stacking-based interpretable machine learning (IML) approach yielded highly accurate predictions for grazing ruminant livestock density across different pasture types. Figure 5 presents the kernel density estimation plots comparing predicted versus observed livestock densities from 5-fold cross-validation. For cattle, the models demonstrated excellent performance across all pasture types, with R^2 values of 0.961, 0.967, and 0.909 for warm-season, cold-season, and year-round pastures, respectively. The corresponding RMSE values were 0.312, 0.292, and 0.586, indicating high prediction accuracy, particularly for seasonal pastures. The models for sheep and goats also showed strong predictive capability, with R^2 values of 0.900 for warm-season, 0.914 for cold-season, and 0.874 for year-round pastures. The RMSE values ranged from 0.419 to 0.475 across these categories. While these models performed slightly below the cattle models, they still provided reliable spatial distribution predictions.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

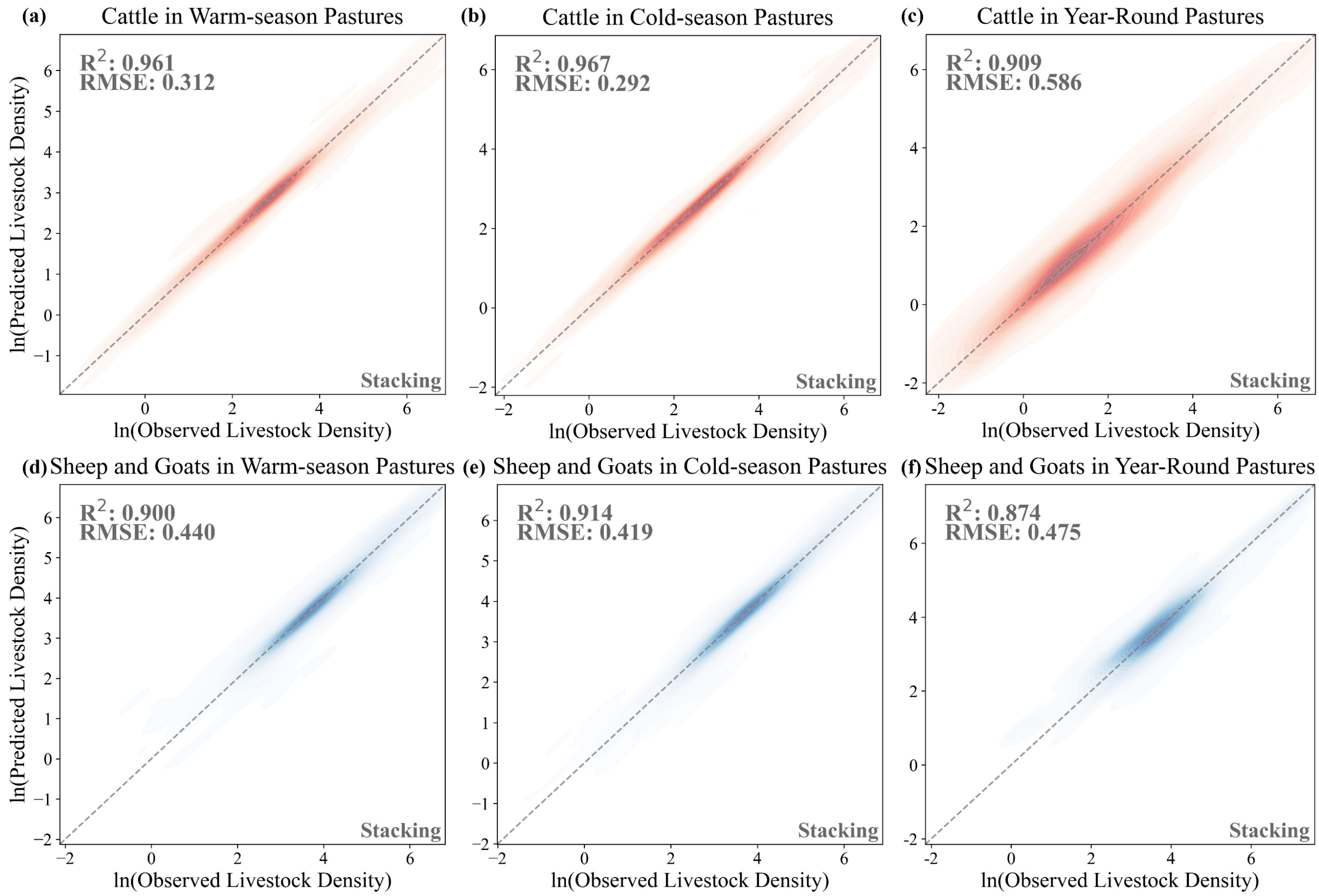


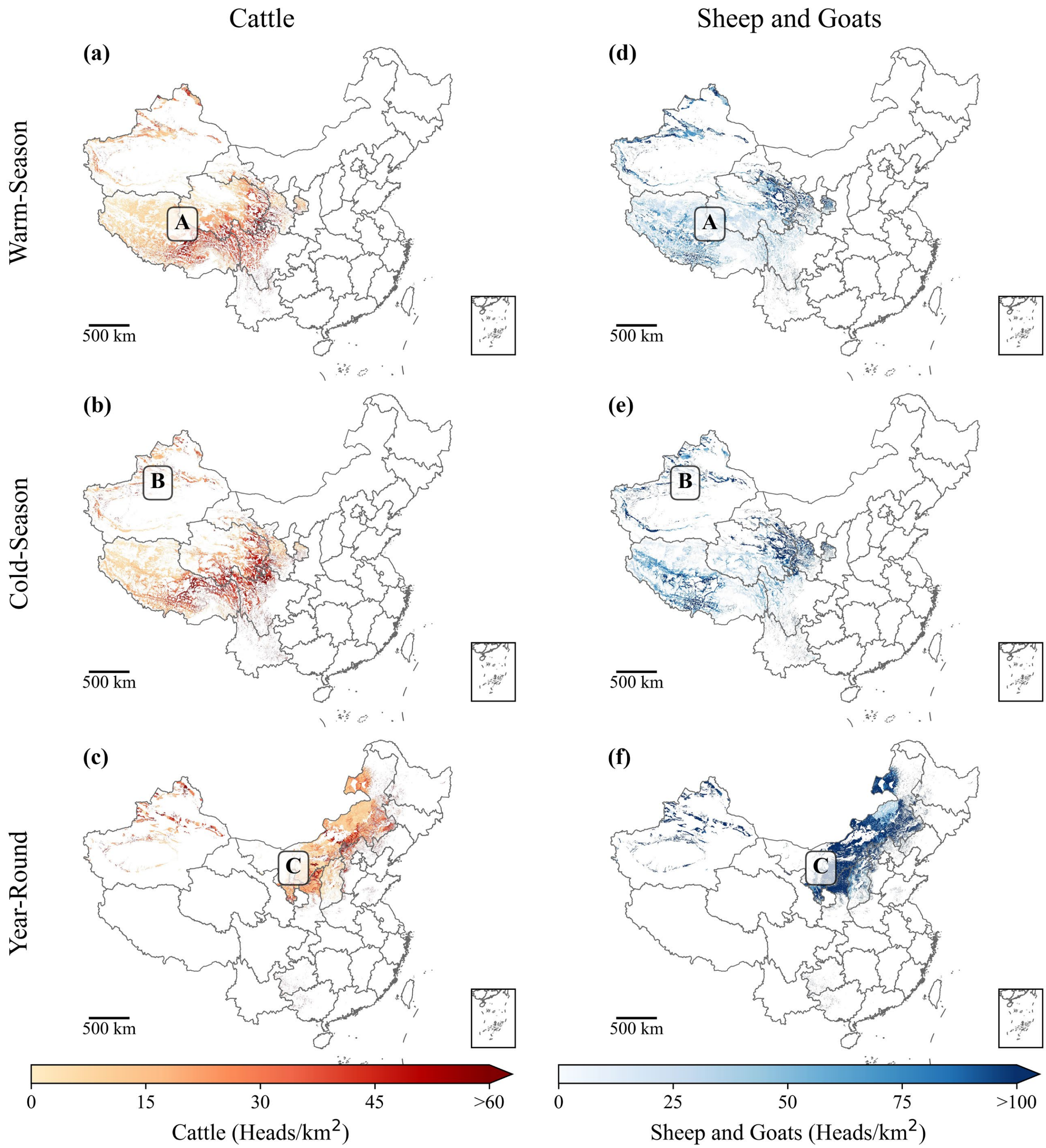
Figure 5: Kernel Density Estimation (KDE) plots of model performance on the independent 20% test data for predicting cattle (a-c) and sheep (e-g) distribution in warm-season, cold-season, and year-round pastures using the stacking-based IML models.

The best-performing algorithms within the stacking framework varied by livestock type and pasture category. Extra Trees (ET) emerged as the optimal meta-model for both cattle and sheep (sheep and goats) in seasonal pastures (warm-season and cold-season), whereas XGBoost and CatBoost delivered superior results for cattrle and sheep (sheep and goats) in year-round pastures, respectively (Meta-Model contribution in Figure A7 - Figure A12). This variability suggests distinct underlying patterns in livestock distribution across different grazing environments. Using the livestock density distribution predicted by the best-performing stacking-based IML models, we assigned county-level cattle and sheep (sheep and goats) numbers through dasymetric mapping. This approach resulted in the distribution of livestock numbers (CLRD-GLPS) across warm-season, cold-season, and year-round pastures. In seasonal pastures, cattle are most densely distributed in the southeast of the Qinghai-Tibet Plateau (QTP), with numbers decreasing as they extend into Xinjiang (Figure 6 (a)-(b)). In contrast, sheep and goats are most numerous in the north-eastern part of Qinghai Province and Xinjiang (Figure 6 (d)-(e)). In year-round pastures, the cattle density is lower than that in cold-season or warm-season pastures, as grazing can occur year-round (Figure 6 (c)). However,

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Earth System Science Data
Open Access Discussions

for sheep and goats, there is still a relatively high density of livestock distribution in Xinjiang and the western part of Inner Mongolia (Figure 6 (f)). The CLRD-GLPS cover 22 years from 2000 to 2021.



400 **Figure 6: CLRD-GLPS distributions of cattle ((a)–(c)) and sheep and goats ((d)–(f)) across warm-season, cold-season, and year-round pastures in 2020. The boxes labelled A, B, and C represent magnified regions of seasonal pastures in different areas.**

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

SHAP feature importance analysis revealed distinctive drivers of livestock distribution patterns. For cattle density distribution, vegetation indicators (NDVI) and topographical features (DEM) emerged as the most influential factors (Figure A7 - Figure A9). In contrast, sheep and goat density distributions were primarily influenced by snow cover, elevation (DEM), and annual mean temperature (Tmpmean_year), representing snow conditions, topography, and climate factors, respectively (Figure A10 - Figure A12).

3.4 Multi-scale Validation of CLRD-GLPS

3.4.1 Spatial distribution validation

To validate the CLRD-GLPS, we examined detailed regional comparisons of cattle (Figure 7) and sheep and goats (Figure 8) distributions across different pasture types in 2020. In these magnified regions (A: warm-season, B: cold-season, and C: year-round pastures), CLRD-GLPS results demonstrate closer alignment with actual census data compared to GLW datasets. The CLRD-GLPS provides more detailed spatial information on livestock distribution patterns within these specific pasture regions, capturing local variations that GLW misses.

Quantitative comparison of regions A, B, and C further validates these visual observations (Figure 7 and Figure 8). For cattle distribution, our analysis reveals that GLW underestimates densities, with an average of 3.42 heads/km² compared to census data of 9.37 heads/km², representing a 63.5% underestimation. In contrast, CLRD-GLPS provides much closer estimates at 7.86 heads/km², reducing the deviation from census data to only 16.1%. For sheep and goats, GLW shows considerable overestimation with an average of 51.07 heads/km² compared to census data of 33.78 heads/km², representing a 51.2% overestimation. CLRD-GLPS demonstrates better accuracy with an estimate of 44.86 heads/km². Overall, these numerical comparisons demonstrate CLRD-GLPS's superior accuracy in capturing livestock distribution patterns within different pasture systems, providing a more reliable foundation for livestock management and policy development.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

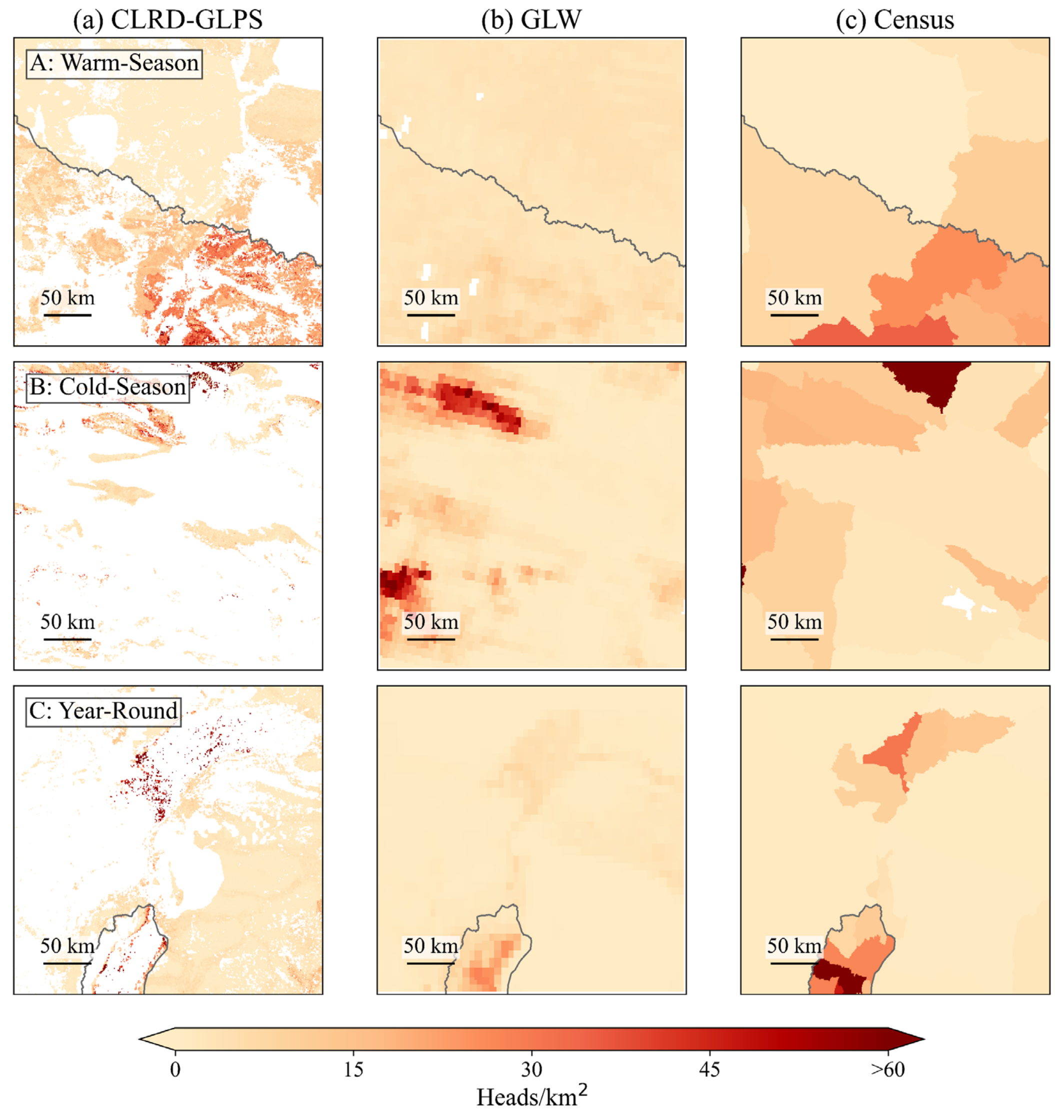


Figure 7: Comparison of CLRD-GLPS with GLW and actual census cattle density distribution in 2020. Detailed comparison in three regions corresponding to Figure 6 (A: warm-season pastures, B: cold-season pastures, C: year-round pastures) showing CLRD-GLPS (a), GLW (b), and census data (c).

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

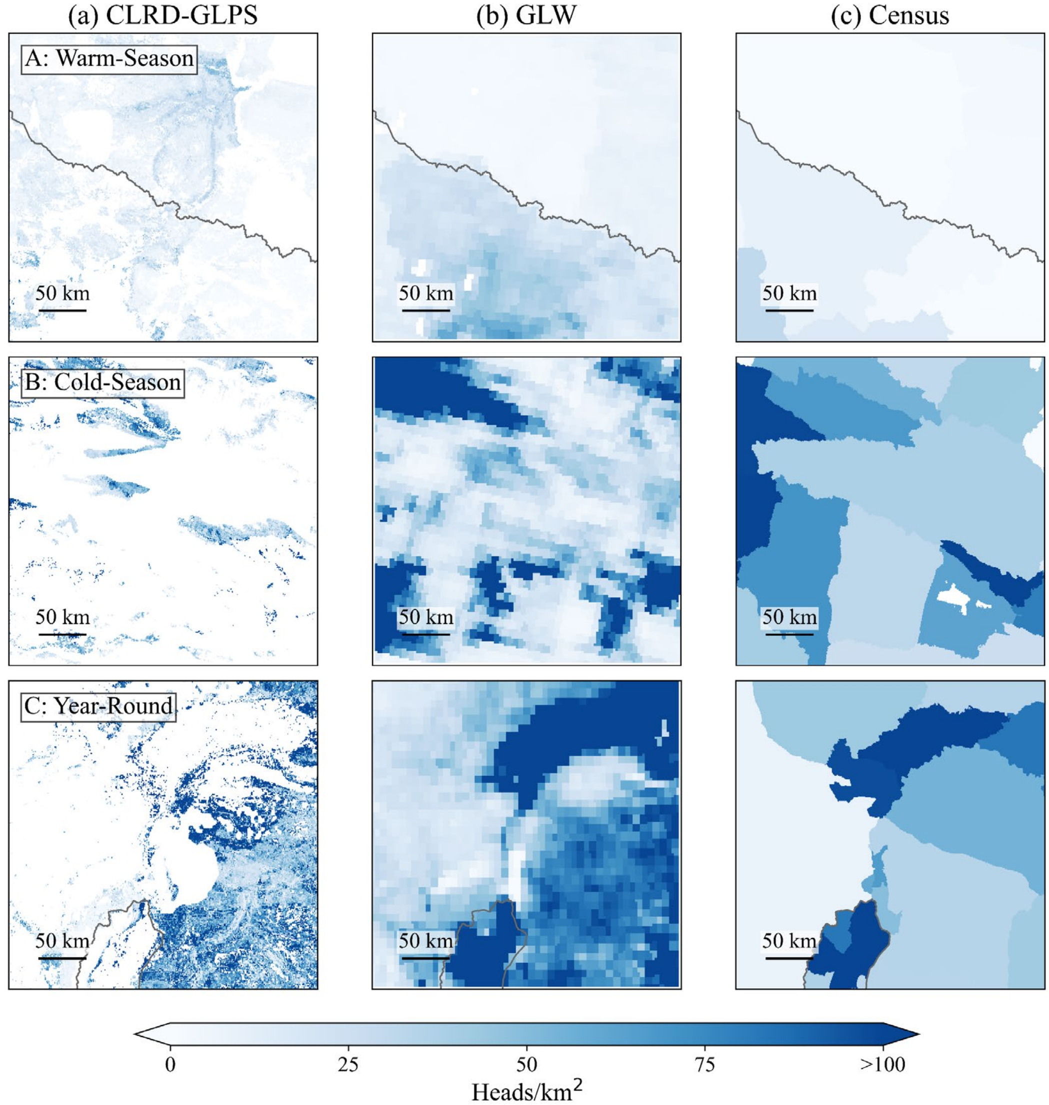


Figure 8: Similar to Figure 7, but for sheep and goats.

3.4.2 Quantitative validation at multiple scales

Quantitative validations were conducted by comparing CLRD-GLPS with existing livestock distribution datasets (GLW, GDGI, LHGI, and Meng) for benchmark years 2000, 2005, 2010, 2015, and 2020 across three spatial scales.

At the county level, following the dasymetric mapping process to generate CLRD-GLPS, our validation showed high consistency (R^2 ranging from 0.990 to 0.999) for both cattle and sheep (sheep and goats) distributions (Figure 9 (a)-(h)). This high correspondence was expected since county-level census data served as constraints for our downscaling approach. In contrast, other livestock datasets showed substantially lower agreement with observed data, with R^2 values ranging from 0.445

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Earth System Science Data Discussions
Open Access

to 0.826, and RMSE values from 19.24 to 41.95 (10,000 heads) (Figure 9 (i)-(l)). Our CLRD-GLPS dataset therefore appears to better capture the spatial distribution patterns of grazing ruminant livestock in grazing LPS compared to previous approaches.

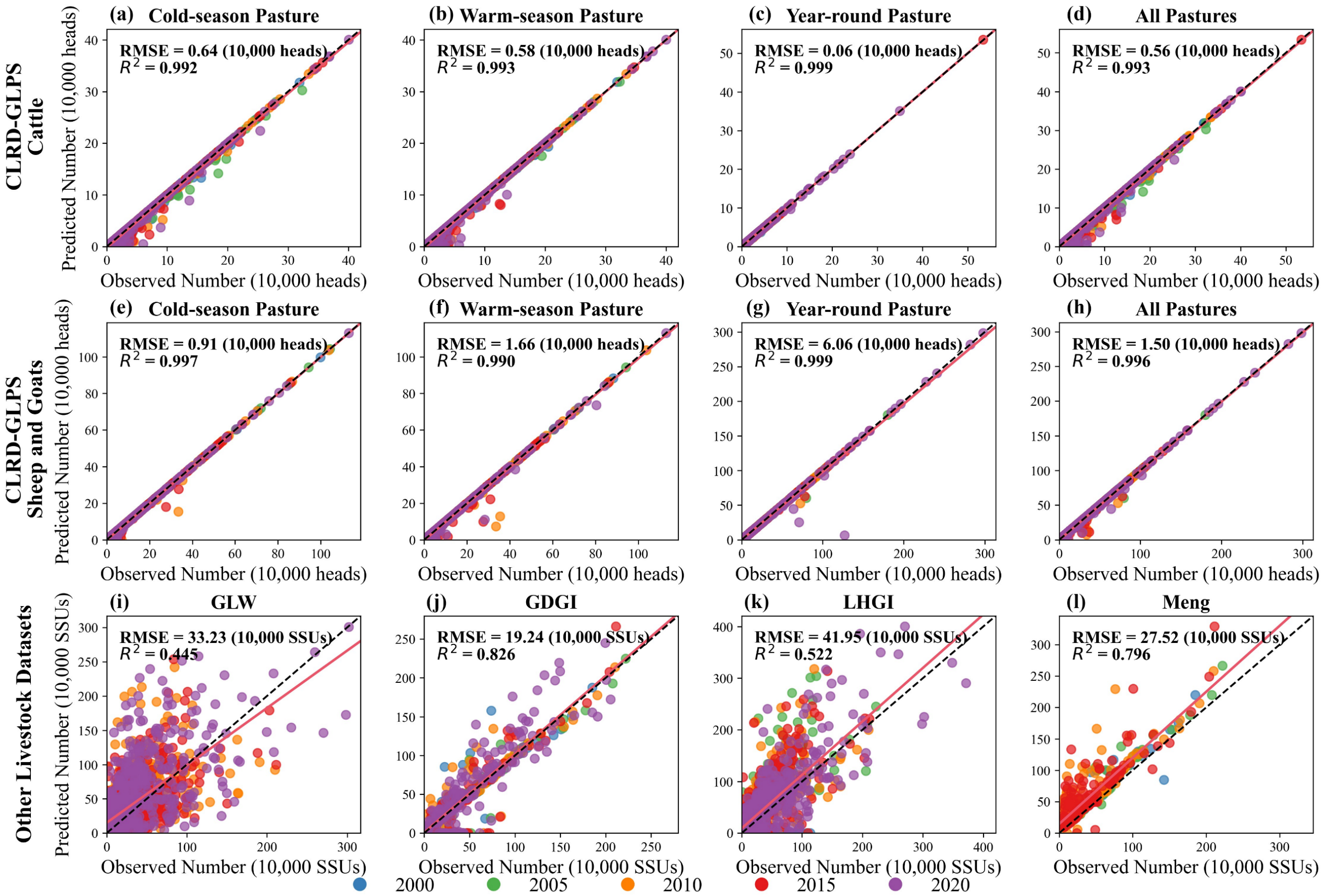
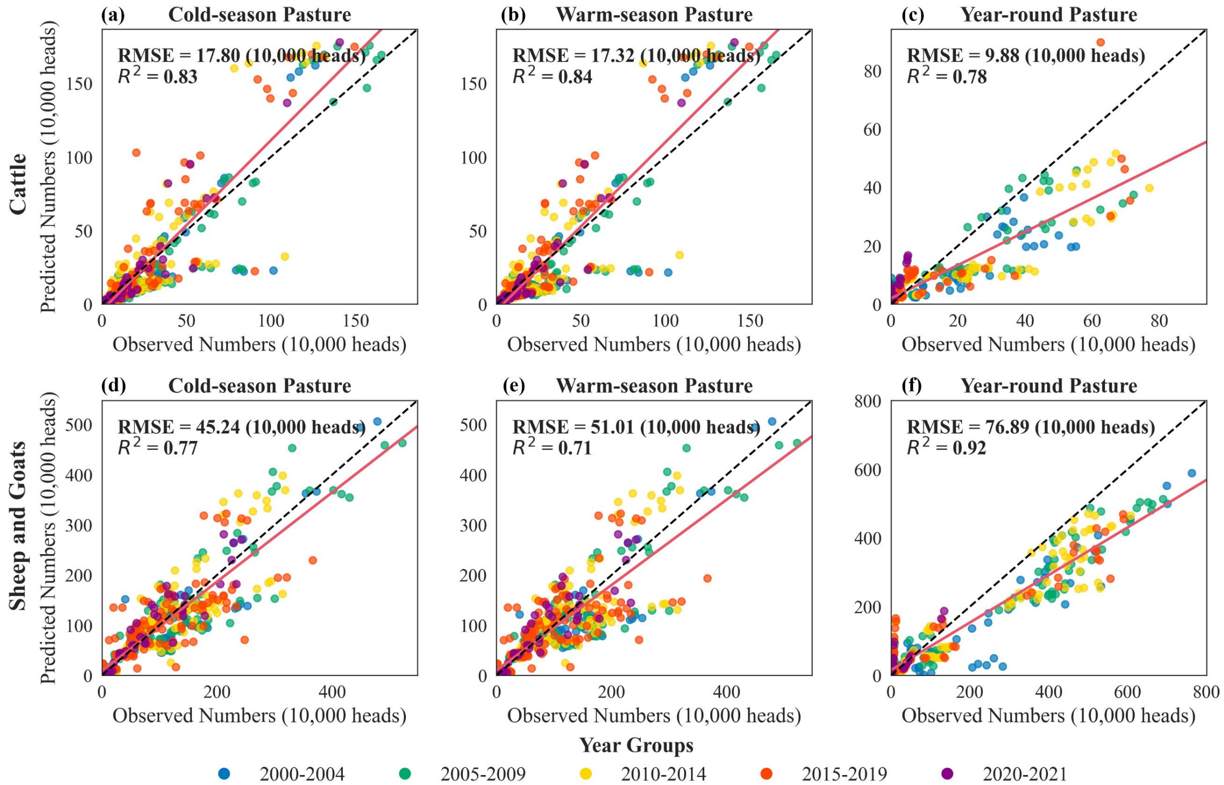


Figure 9: Validation of cattle ((a)–(d)) / sheep and goats ((e)–(h)) numbers from CLRD-GLPS across all pasture types, comparing GLW (i), GDGI (j), GLW (k), and Meng (l) datasets separately with census data at the county level in 2000, 2005, 2010, 2015, and 2020.

The city-level validation between CLRD-GLPS and observed grazing livestock numbers across warm-season, cold-season, and year-round pastures demonstrated robust performance, with R² values ranging from 0.71 to 0.92, and RMSE from 9.88 to 76.89 (10,000 heads) (Figure 10 (a)-(f)). This confirms that our methodology effectively preserved the spatial patterns at administrative levels higher than the county. Further validation of CLRD-GLPS at the grid level using independently collected point observations of grazing livestock yielded an RMSE of 33.3 (SSUs/km²) and an R² of 0.79, indicating strong agreement between our predicted grid-level densities and observed values (Figure 11 (a)-(b)). These multi-scale validation results collectively demonstrate that CLRD-GLPS maintains high reliability across different spatial resolutions, from administrative units to fine-resolution grids.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025



455 **Figure 10: Validation of cattle ((a)–(c)) / sheep and goats ((d)–(f)) numbers from CLRD-GLPS at city-level across all pasture types from 2000 to 2021.**

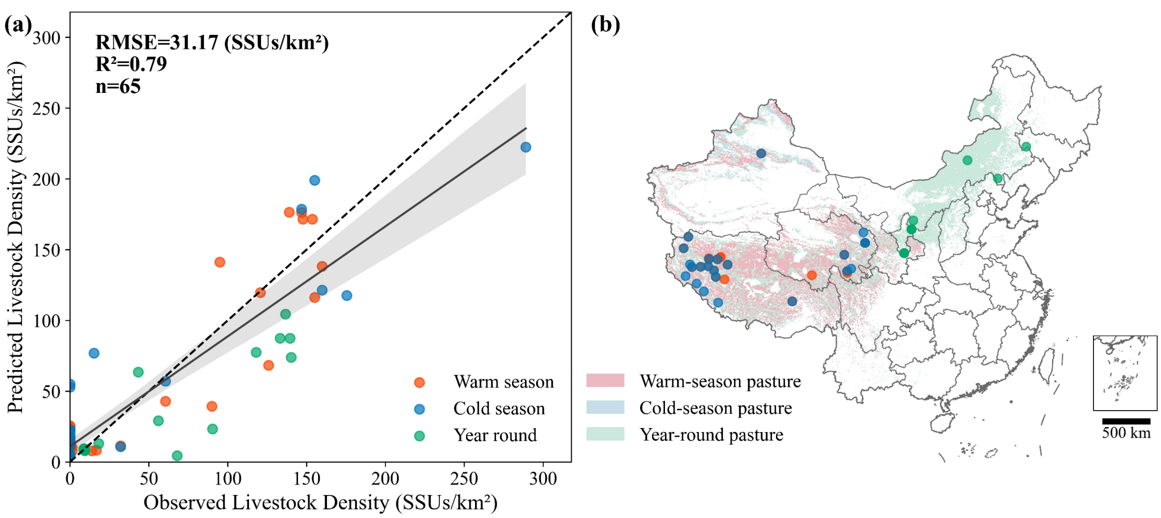


Figure 11: Validation of livestock numbers across all pasture types at the grid level from 2000 to 2021 (a), and the spatial distribution of sampling points across different seasonal pastures in China (b). Livestock numbers are expressed in Sheep Stock Units (SSUs) to
460 standardize mixed cattle, sheep, and goat counts in the validation dataset.

4 Discussion

4.1 Improvements of livestock distribution modelling framework

Our development of CLRD-GLPS represents several important advancements in livestock distribution modelling compared to existing approaches. The main advancements in livestock distribution modelling in this study include three aspects. First, this study is the first to segment ruminant livestock within grazing LPS from the total livestock count across all LPS based on interpretable machine learning methods. Second, it expands the classification of grazing pastures to include seasonal types, specifically warm-season, cold-season, and year-round pastures across China, with dynamic suitability grazing mask. Finally, using stacking-based interpretable machine learning approach, we explore and explain the spatial-temporal distribution patterns of grazing ruminant livestock.

Livestock density varies significantly across different LPS. While this variation is well-documented (Thornton, 2002), previous livestock distribution models have largely failed to account for differences between grazing and non-grazing livestock within the same administrative units. In China, where grazing LPS account for 39.54% of the total LPS area (Kruska et al., 2003), this methodological limitation leads to substantial overestimation of grazing livestock numbers when relying solely on county-level census statistics(Wang et al., 2024b). Our quantitative analysis reveals the magnitude of this overestimation problem. Based on our grazing ruminant livestock segmentation prediction results (Figure 4), the average grazing livestock proportion in 2021 across counties with grazing systems ranges from 20% to 80%. This means previous approaches would overestimate grazing livestock numbers by 25-400% (calculated as (1/proportion – 1) × 100%). For cattle, with an average grazing proportion of 0.58, the mean overestimation reaches 72.3%. For sheep and goats (average proportion 0.61), the overestimation averages 63.9%.

To address this issue, our machine learning approach successfully segments grazing livestock from the total livestock count with high accuracy. The model validation demonstrates excellent performance (R^2 of 0.933 for both cattle and sheep/goats models at provincial level, and R^2 of 0.800 at county level), confirming the reliability of our segmentation method. By applying this approach, we effectively reduced the systematic overestimation that has plagued previous grazing livestock distribution studies. The spatial heterogeneity in grazing proportions revealed by our analysis reflects the fundamental competition between different LPS across China. Our SHAP analysis (Figure 3 (c)-(d)) demonstrates that the trade-off between cropland and grassland availability serves as the primary determinant of grazing livestock proportions, representing the classic competition between agricultural and pastoral land uses. This competition manifests differently across China's geographic gradient, with eastern regions showing lower grazing proportions due to intensive crop production and more developed mixed-farming and landless LPS (Jiang and Wang, 2022). The significant influence of geographic factors (longitude and latitude) in our models captures broader socioeconomic and policy dimensions that shape regional specialization in LPS. These findings highlight the

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

critical importance of accurately determining grazing proportions before modelling grazing ruminant livestock distribution, as failure to do so leads to significant spatial and number biases.

Leveraging interpretable machine learning methods, our findings reveal notable differences between year-round pastures and seasonal pastures in the factors influencing livestock distribution and livestock density. For cattle, grass quality (NDVI) is the primary factor influencing distribution in both warm-season and cold-season pastures, with NDVI consistently showing the highest importance (approximately 0.6-0.8 on the importance scale) across all models. In contrast, for year-round pastures, topography (DEM) emerges as the dominant factor (with values around 0.6-0.7), while NDVI becomes a secondary influence (approximately 0.4-0.5) (Figure A7 - Figure A9). For sheep and goats, topography (DEM) and snow cover prominently feature as the main determinants in warm-season pastures (with DEM reaching 0.4 and snow cover up to 0.5 in some models), while in cold-season pastures, a more balanced influence of topographical features (slope, DEM) and snow cover shapes distribution. Most notably, in year-round pastures for sheep and goats, climate factors (particularly Tmpmean_year) become the dominant influence (reaching values of 0.4-0.5) (Figure A10 - Figure A12). These differences highlight the complex interaction between environmental factors and human activities in different grazing systems, providing valuable insights for targeted management strategies.

4.2 Methodological advantages of the stacking-based IML approach

In this study, we introduced stacking-based interpretable machine learning (IML) to livestock distribution modelling. Unlike previous studies that relied on single machine learning algorithms such as Random Forest or Extra Trees for livestock distribution prediction, our stacking ensemble approach demonstrates significant advantages in both prediction accuracy and model robustness.

The comparative analysis of model performance revealed that our stacking-based IML approach consistently outperformed individual base models. For cattle distribution prediction, the stacking ensemble achieved up to 4.2% improvement in R^2 values over the best-performing individual model (from R^2=0.926 in RF to R^2=0.967 in stacking) for cold-season pastures, and 6.2% improvement (from R^2=0.856 in RF to R^2=0.909 in stacking) for year-round pastures (Figure A13). These improvements are particularly valuable across different pasture types within grazing LPS, where single-model approaches often struggle to maintain consistent performance. Similar enhancements were observed for sheep (sheep and goats) distribution prediction, with the stacking approach providing more stable predictions across diverse geographical conditions and reducing overfitting tendencies common in single model approaches (Figure A14).

Beyond improved accuracy, the stacking-based approach offers several additional advantages. First, it effectively mitigates the inherent biases of individual algorithms. For instance, in regions with complex topography such as the south-eastern QTP, RF and XGB models produced inconsistent estimates of livestock density, while the stacking ensemble provided more

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

balanced predictions. Second, through SHAP value analysis across different base models, we gained more comprehensive insights into the driving factors of livestock distribution patterns than would be possible from a single model perspective. This revealed, for example, that while vegetation productivity (NDVI) is consistently important across all models for cattle distribution in warm-season pastures, its relative importance varies substantially (15-28% contribution) depending on the algorithm used, highlighting the value of ensemble approaches in robust feature importance assessment (Figure A7).

Together, the stacking-based IML approach provide an innovation way that can be adapted for creating similar livestock distribution datasets in other regions and LPS. By combining the strengths of multiple algorithms while maintaining interpretability, our approach advances the field of livestock distribution modelling and provides more reliable data for sustainability assessments, climate change impact studies, and policy development.

4.3 Uncertainties and limitations

This study still faces certain limitations. Firstly, our historical county-level livestock statistical data did not manage to cover every county and year, due to the availability or accessibility of the data. While each model ensures a minimum of 12,000 training samples, collecting more statistical data in the future may reduce errors introduced by model response data and expand the analysis across more years. Secondly, the dynamic grazing suitable mask is crucial. Although we utilized the annual 30m CLCD dataset, a more comprehensive coverage of grazing ban regions across both space and time would further improve the accuracy of this study. Lastly, in segmenting grazing livestock from the overall livestock population, the validation data on livestock carrying capacity mainly come from the QTP. While the QTP accounts for over 50% of China's grassland area (Li et al., 2021a), incorporating samples from more provinces, such as Inner Mongolia, would further enhance the validation of our grazing proportion predictions.

5 Data availability

China's **l**ong-term annual **r**uminant livestock **d**istribution in **g**razing **l**ivestock **p**roduction **s**ystems from 2000 to 2021 (CLRD-GLPS) is accessible on Zendo at the following link: https://doi.org/10.5281/zenodo.15347430 (Zhan et al., 2025). The datasets include cattle and sheep distributions in warm-season, cold-season, and year-round pastures, organized in corresponding folders. Each folder contains 22 GeoTIFF files from 2000 to 2021, with a 1km resolution (0.00083° at the equator) and units in heads per pixel (or heads/km²).

Author contributions

NZ and TY conceptualized the paper and developed the methodology. JP and HM provided the base data. NZ and TY produced the dataset. NZ and TY prepared the paper with contributions from MH, JP, WL and HM.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Competing interests

The contact author has declared that none of the authors has any competing interests.

Financial support

This study was supported by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP, Grant No. 2019QZKK0906), and the Program of Introducing Talent to Universities (111 Project, Grant No. BP0820003).

References

MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 6: https://nsidc.org/data/mod10a1/versions/6, last access: 30 January 2024.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1007/978-3-030-62008-0_35, 2001.

Burton, E. C., Gray, M. J., Schmutzer, A. C., and Miller, D. L.: Differential Responses of Postmetamorphic Amphibians to Cattle Grazing in Wetlands, Journal of Wildlife Management, 73, 269–277, https://doi.org/10.2193/2007-562, 2009.

Cheng, L., Zhang, X., Reis, S., Ren, C., Xu, J., and Gu, B.: A 12% switch from monogastric to ruminant livestock production can reduce emissions and boost crop production for 525 million people, Nat Food, 3, 1040–1051, https://doi.org/10.1038/s43016-022-00661-1, 2022.

Didan, K.: MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set], https://doi.org/10.5067/MODIS/MOD13Q1.006, 2015.

Dong, Q.-M., Zhao, X.-Q., Wu, G.-L., and Chang, X.-F.: Optimization yak grazing stocking rate in an alpine grassland of Qinghai-Tibetan Plateau, China, Environmental Earth Sciences, 73, 2497–2503, https://doi.org/10.1007/s12665-014-3597-7, 2015.

Fetzel, T., Havlik, P., Herrero, M., and Erb, K. H.: Seasonality constraints to livestock grazing intensity, Global Change Biology, 23, 1636–1647, https://doi.org/10.1111/gcb.13591, 2017.

Fischer, G., Nachtergaele, F., Prieler, S., Van Velthuizen, H., Verelst, L., and Wiberg, D.: Global agro-ecological zones assessment for agriculture (GAEZ 2008), IIASA, Laxenburg, Austria and FAO, Rome, Italy, 10, 2008.

Friedman, J. H.: Greedy function approximation: a gradient boosting machine, Annals of statistics, 1189–1232, 2001.

Gerber, P. J., Steinfeld, H., Henderson, B., Mottet, A., Opio, C., Dijkman, J., Falcucci, A., and Tempio, G.: Tackling climate change through livestock: a global assessment of emissions and mitigation opportunities., Tackling climate change through livestock: a global assessment of emissions and mitigation opportunities., 2013.

Geurts, P., Ernst, D., and Wehenkel, L.: Extremely randomized trees, Mach Learn, 63, 3–42, https://doi.org/10.1007/s10994-006-6226-1, 2006.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T. P., Vanwambeke, S. O., Wint, G. R. W., and Robinson, T. P.: Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010, Scientific Data, 5, 1–11, https://doi.org/10.1038/sdata.2018.227, 2018.

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., and Hussain, A.: Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence, Cogn Comput, 16, 45–74, https://doi.org/10.1007/s12559-023-10179-8, 2024.

Herrero M., Grace D., Njuki J., Johnson N., Enahoro D., Silvestri S., and Rufino M. C.: The roles of livestock in developing countries, animal, 7, 3–18, https://doi.org/10.1017/S1751731112001954, 2013.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1940 to present, https://doi.org/10.24381/cds.adbb2d47, 2023.

Howard, K. S. C., Eldridge, D. J., and Soliveres, S.: Positive effects of shrubs on plant species diversity do not change along a gradient in grazing pressure in an arid shrubland, Basic and Applied Ecology, 13, 159–168, https://doi.org/10.1016/j.baae.2012.02.008, 2012.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org), 15, 5, 2008.

Jiang, S., Sweet, L., Blougouras, G., Brenning, A., Li, W., Reichstein, M., Denzler, J., Shangguan, W., Yu, G., Huang, F., and Zscheischler, J.: How Interpretable Machine Learning Can Benefit Process Understanding in the Geosciences, Earth's Future, 12, e2024EF004540, https://doi.org/10.1029/2024EF004540, 2024.

Jiang, X. and Wang, L.: Grassland-based ruminant farming systems in China: Potential, challenges and a way forward, Anim Nutr, 10, 243–248, https://doi.org/10.1016/j.aninu.2022.04.007, 2022.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Advances in Neural Information Processing Systems, 2017.

Knapp, J. R., Laur, G. L., Vadas, P. A., Weiss, W. P., and Tricarico, J. M.: *Invited review:* Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions, Journal of Dairy Science, 97, 3231–3261, https://doi.org/10.3168/jds.2013-7234, 2014.

Kruska, R. L., Reid, R. S., Thornton, P. K., Henninger, N., and Kristjanson, P. M.: Mapping livestock-oriented agricultural production systems for the developing world, Agricultural Systems, 77, 39–63, https://doi.org/10.1016/S0308-521X(02)00085-9, 2003.

Li, M., Liu, S., Sun, Y., and Liu, Y.: Agriculture and animal husbandry increased carbon footprint on the Qinghai-Tibet Plateau during past three decades, Journal of Cleaner Production, 278, 123963, https://doi.org/10.1016/J.JCLEPRO.2020.123963, 2021a.

Li, X., Hou, J., and Huang, C.: High-Resolution Gridded Livestock Projection for Western China Based on Machine Learning, 1–21, 2021b.

Li, Y., McIntyre, K. M., Rasmussen, P., Gilbert, W., Chaters, G., Raymond, K., Jemberu, W. T., Larkins, A., Patterson, G. T., Kwok, S., Kappes, A. J., Mayberry, D., Schrobback, P., Acosta, M. H., Stacey, D. A., Huntington, B., Bruce, M., Knight-Jones, T., and Rushton, J.: Rationalising development of classification systems describing livestock production systems for disease burden analysis within the Global Burden of Animal Diseases programme, Research in Veterinary Science, 168, 105102, https://doi.org/10.1016/j.rvsc.2023.105102, 2024.

Meng, N., Wang, L., Qi, W., Dai, X., Li, Z., Yang, Y., Li, R., Ma, J., and Zheng, H.: A high-resolution gridded grazing dataset of grassland ecosystem on the Qinghai–Tibet Plateau in 1982–2015, Sci Data, 10, 68, https://doi.org/10.1038/s41597-023-01970-1, 2023.

Mennis, J.: Dasymetric mapping for estimating population in small areas, Geography Compass, 3, 727–745, https://doi.org/10.1111/j.1749-8198.2009.00220.x, 2009.

Michalk, D. L., Kemp, D. R., Badgery, W. B., Wu, J., Zhang, Y., and Thomassin, P. J.: Sustainability and future food security—A global perspective for livestock production, Land Degradation & Development, 30, 561–573, https://doi.org/10.1002/ldr.3217, 2019.

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B.: Definitions, methods, and applications in interpretable machine learning, Proceedings of the National Academy of Sciences, 116, 22071–22080, https://doi.org/10.1073/pnas.1900654116, 2019.

Negnevitsky, M.: Artificial intelligence: a guide to intelligent systems, Pearson education, 2005.

Ning Zhan, Liu, W., Ye, T., Li, H., Chen, S., and Ma, H.: High-resolution livestock seasonal distribution data on the Qinghai-Tibet Plateau in 2020, Sci Data, 10, 142, https://doi.org/10.1038/s41597-023-02050-0, 2023.

Oak Ridge National Laboratory: LandScan Global Population Database 2019, 2020.

Ocholla, I. A., Pellikka, P., Karanja, F. N., Vuorinne, I., Odipo, V., and Heiskanen, J.: Livestock detection in African rangelands: Potential of high-resolution remote sensing data, Remote Sensing Applications: Society and Environment, 33, 101139, https://doi.org/10.1016/j.rsase.2024.101139, 2024.

Pavlyshenko, B.: Using Stacking Approaches for Machine Learning Models, in: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 255–258, https://doi.org/10.1109/DSMP.2018.8478522, 2018.

Peng, D., Huang, J., Li, C., Liu, L., Huang, W., Wang, F., and Yang, X.: Modelling paddy rice yield using MODIS data, Agricultural and Forest Meteorology, 184, 107–116, https://doi.org/10.1016/j.agrformet.2013.09.006, 2014.

Peng, S., Ding, Y., and Li, Z.: 1-km monthly temperature and precipitation dataset for China from 1901–2017, Earth System Science Data, 11, 1931–1946, https://doi.org/10.5194/essd-2019-145, 2019.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A.: CatBoost: unbiased boosting with categorical features, in: Advances in Neural Information Processing Systems, 2018.

Pulina, G., Francesconi, A. H. D., Stefanon, B., Sevi, A., Calamari, L., Lacetera, N., Dell'Orto, V., Pilla, F., Ajmone Marsan, P., Mele, M., Rossi, F., Bertoni, G., Crovetto, G. M., and Ronchi, B.: Sustainable ruminant production to help feed the planet, Italian Journal of Animal Science, 16, 140–171, https://doi.org/10.1080/1828051X.2016.1260500, 2017.

Robinson, T. P., Thornton, P. K., Francesconi, G. N., Kruska, R. L., Chiozza, F., Notenbaert, A. M. O., Cecchi, G., Herrero, M., Epprecht, M., Fritz, S., You, L., Conchedda, G., and See, L.: Global livestock production systems, FAO and ILRI, 2011.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Sanz, M. J., Vente, J. de, Chotte, J.-L., Bernoux, M., G. Kust, I., Ruiz, M., Almagro, J.-A., Alloza, R., Vallejo, V., Castillo, A., Hebel, and Akhtar-Schuster, M.: . Sustainable Land Management contribution to successful land-based climate change adaptation and mitigation. A Report of the Science-Policy Interface., 2017.

Sun, J., Liu, M., Fu, B., Kemp, D., Zhao, W., Liu, G., Han, G., Wilkes, A., Lu, X., Chen, Y., Cheng, G., Zhou, T., Hou, G., Zhan, T., Peng, F., Shang, H., Xu, M., Shi, P., He, Y., Li, M., Wang, J., Tsunekawa, A., Zhou, H., Liu, Y., Li, Y., and Liu, S.: Reconsidering the efficiency of grazing exclusion using fences on the Tibetan Plateau, Science Bulletin, 65, 1405–1414, https://doi.org/10.1016/J.SCIB.2020.04.035, 2020.

Sundararajan, M. and Najmi, A.: The Many Shapley Values for Model Explanation, in: Proceedings of the 37th International Conference on Machine Learning, International Conference on Machine Learning, 9269–9278, 2020.

Thornton, P.: Mapping poverty and livestock in the developing world, ILRI (aka ILCA and ILRAD), 2002.

Thornton, P., Nelson, G., Mayberry, D., and Herrero, M.: Increases in extreme heat stress in domesticated livestock species during the twenty-first century, Global Change Biology, 27, 5762–5772, https://doi.org/10.1111/gcb.15825, 2021.

Thornton, P. K.: Livestock production: recent trends, future prospects, Philosophical Transactions of the Royal Society B: Biological Sciences, 365, 2853–2867, https://doi.org/10.1098/rstb.2010.0134, 2010.

Uwizeye, A., de Boer, I. J. M., Opio, C. I., Schulte, R. P. O., Falcucci, A., Tempio, G., Teillard, F., Casu, F., Rulli, M., Galloway, J. N., Leip, A., Erisman, J. W., Robinson, T. P., Steinfeld, H., and Gerber, P. J.: Nitrogen emissions along global livestock supply chains, Nat Food, 1, 437–446, https://doi.org/10.1038/s43016-020-0113-y, 2020.

Wang, D., Peng, Q., Li, X., Zhang, W., Xia, X., Qin, Z., Ren, P., Liang, S., and Yuan, W.: A long-term high-resolution dataset of grasslands grazing intensity in China, Sci Data, 11, 1194, https://doi.org/10.1038/s41597-024-04045-x, 2024a.

Wang, D., Peng, Q., Li, X., Zhang, W., Xia, X., Qin, Z., Ren, P., Liang, S., and Yuan, W.: A long-term high-resolution dataset of grasslands grazing intensity in China, Sci Data, 11, 1194, https://doi.org/10.1038/s41597-024-04045-x, 2024b.

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T. C. D., Howes, R. E., Tusting, L. S., Kang, S. Y., Cameron, E., Bisanzio, D., Battle, K. E., Bhatt, S., and Gething, P. W.: A global map of travel time to cities to assess inequalities in accessibility in 2015, Nature, 553, 333–336, https://doi.org/10.1038/nature25181, 2018.

Xu, X., Chen, F., Wang, B., Harrison, M. T., Chen, Y., Liu, K., Zhang, C., Zhang, M., Zhang, X., Feng, P., and Hu, K.: Unleashing the power of machine learning and remote sensing for robust seasonal drought monitoring: A stacking ensemble approach, Journal of Hydrology, 634, 131102, https://doi.org/10.1016/j.jhydrol.2024.131102, 2024.

Yang, J. and Huang, X.: The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019, Earth System Science Data, 13, 3907–3925, https://doi.org/10.5194/essd-13-3907-2021, 2021a.

Yang, J. and Huang, X.: The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019, Earth System Science Data, 13, 3907–3925, https://doi.org/10.5194/essd-13-3907-2021, 2021b.

Ye, T., Liu, W., Chen, S., Chen, D., Shi, P., Wang, A., and Li, Y.: Reducing livestock snow disaster risk in the Qinghai–Tibetan Plateau due to warming and socioeconomic development, Science of The Total Environment, 151869, https://doi.org/10.1016/j.scitotenv.2021.151869, 2021.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Yu, H., Liu, B., Wang, G., Zhang, T., Yang, Y., Lu, Y., Xu, Y., Huang, M., Yang, Y., and Zhang, L.: Grass-livestock balance based grassland ecological carrying capability and sustainable strategy in the Yellow River Source National Park, Tibet Plateau, China, J. Mt. Sci., 18, 2201–2211, https://doi.org/10.1007/s11629-020-6087-2, 2021.

Zhan, N., Liu, W., Ye, T., Li, H., Chen, S., and Ma, H.: High-resolution livestock seasonal distribution data on the Qinghai-Tibet Plateau in 2020, Sci Data, 10, 142, https://doi.org/10.1038/s41597-023-02050-0, 2023.

Zhan, N., Ye, T., Herrero, M., Peng, J., Liu, W., and Ma, H.: CLRD-GLPS: A Long-term Seasonal Dataset of Ruminant Livestock Distribution in China's Grazing Production Systems (2000-2021) Using Stacking-based Interpretable Machine Learning (2), https://doi.org/10.5281/zenodo.15347430, 2025.

Zheng, J.: Research on the Development of the Sheep Industry in Xinjiang, Doctoral Dissertation, Xinjiang Agricultural University, 2005.

Zhou, J., Niu, J., Wu, N., and Lu, T.: Annual high-resolution grazing-intensity maps on the Qinghai–Tibet Plateau from 1990 to 2020, Earth System Science Data, 16, 5171–5189, https://doi.org/10.5194/essd-16-5171-2024, 2024.

Zhuang, M., Gongbuzeren, Zhang, J., and Li, W.: Community-based seasonal movement grazing maintains lower greenhouse gas emission intensity on Qinghai-Tibet Plateau of China, Land Use Policy, 85, 155–160, https://doi.org/10.1016/j.landusepol.2019.03.032, 2019.

Zou, J., Luo, C., Xu, X., Zhao, N., Zhao, L., and Zhao, X.: Relationship of plant diversity with litter and soil available nitrogen in an alpine meadow under a 9-year grazing exclusion, Ecological Research, 31, 841–851, https://doi.org/10.1007/s11284-016-1394-3, 2016.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Open Access Earth System Science Data Discussions

Appendices

Table A1: List of datasets used in this study

Data Type	Variable Name	Variable Description	Data Source
Livestock Data	County-level Livestock Numbers	Livestock (cattle/sheep/pigs) year-end stock numbers for 2000-2021 in 29 provinces	China and Provincial Statistical Yearbooks (partially from Provincial Statistical Bureaus)
	City-level Livestock Numbers	Livestock (cattle/sheep/pigs) year-end stock numbers for 2000-2021 in 29 provinces	China and Provincial Statistical Yearbooks (partially from Provincial Statistical Bureaus)
	Grid-level observed Livestock Density	Livestock density at grid-level (countryside-level) from 2000 to 2021	Previous literature and the National Rural Fixed Observation Point Micro-household Survey
Mask Data	China Land Cover Dataset (CLCD)	30-meter resolution land cover dataset of China for the years 2000, 2005, 2010, 2015, and 2020	(Yang and Huang, 2021b)
	National Nature Reserves	Boundaries of the core areas of national nature reserves	Chinese Academy of Sciences Resource and Environmental Sciences and Data Center
	Grazing ban Areas	Boundaries of fenced pastures where grazing is prohibited	(Sun et al., 2020)
Pasture Data	Seasonal Pasture Sample Data	Seasonal pasture location data and distribution maps	Qinghai Province Grassland Station; County-level Forestry and Grassland Bureaus of Tibet Autonomous Region; Xinjiang Autonomous Region Grassland Station
Topographic Data	DEM	Digital Elevation Model	NASA Shuttle Radar Topography Mission (Jarvis et al., 2008)
	Slope	Slope	(Fischer et al., 2008)
Climate Data	Presum_year	Average annual cumulative precipitation for 2000-2021	(Peng et al., 2014)
	Tmpmean_year	Average annual mean temperature for 2000-2021	(Peng et al., 2014)
	GSpre	Average cumulative precipitation during the grass growth season (April-October) for 2000-2021	(Peng et al., 2014)
	Wpre	Average cumulative precipitation during winter (November-March) for 2000-2021	(Peng et al., 2014)
	GStmp	Average mean temperature during the grass growth season (April-October) for 2000-2021	(Peng et al., 2014)
	Wtmp	Average mean temperature during winter (November-March) for 2000-2021	(Peng et al., 2014)

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Data Type	Variable Name	Variable Description	Data Source
	Wind	Average 10-meter wind speed for 2000-2021	ECMWF (European Centre for Medium-Range Weather Forecasts)
Snow Data	Snow depth	Average snow depth during winter (November-March) for 2000-2021	Spatial-temporal Tier 3 Environmental Big Data Platform
	Snow Cover	Average number of snow cover days during winter (November-March) for 2000-2021	US National Snow and Ice Data Center
Vegetation Data	GPP	Maximum Gross Primary Productivity for 2000-2021	MOD17A3H v006
	NPP	Maximum Net Primary Productivity for 2000-2021	MOD17A3H v006
	NDVI	Maximum Normalized Difference Vegetation Index for 2000-2021	MOD13Q1
Socioeconomic Data	Travel time	Shortest travel time to cities with at least 50,000 people in 2015	(Weiss et al., 2018)
	POP	LandScan 1-km Global population for 2000-2021	(Oak Ridge National Laboratory, 2020)

Table A2: Information of different livestock distribution datasets.

Datasets	Method	Region	Resolution	Samples	Season	LPS	Suitability mask	Species
LHGI Wang et al. (2024)	RF	Western China	1980-2000 Annual 0.1° 2001-2022 Annual 0.0025°	No clearly number	No	Yes	Grassland	Livestock (cattle, goats, sheep, horses, donkeys, and camels)
GDGI Zhou *et* al. (2024)	ET RF GB KNN SVM	Qinghai-Tibet Plateau	1990-2020 Annual 100m	182 counties, and 4998 independent records	No	No	Suitable pasture: Grassland, Altitude<5600, slope<40%, Population density<0.5/hm2	Livestock (cattle, sheep, horse, and mule)
Meng *et* al. (2023)	CSF RF	Qinghai-Tibet Plateau	1982-2015 Annual 0.083°	242 counties	No	No	Unsuitable pasture: Non-grassland, Altitude>5600, NNR, Urban area	Livestock (cattle, yaks, horses, donkeys, mules, camels, goats, and sheep)

715

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Table A3: List of observed grid-level grazing density data used in this study

Mean grazing density (SSUs/km2)	Season	Time	Longitude	Latitude	Source
13.8	Cold_season	2021	84.5724	33.5086	(Zhou et al., 2024)
16.8	Cold_season	2021	84.5764	33.5173	(Zhou et al., 2024)
1.3	Cold_season	2021	83.2922	33.1169	(Zhou et al., 2024)
32.1	Cold_season	2021	83.3174	33.1207	(Zhou et al., 2024)
155.2	Cold_season	2010	93.5	30.45	(Zhou et al., 2024)
0	Cold_season	2021	80.3398	34.6949	(Meng et al., 2023)
0	Cold_season	2021	85.5507	32.9491	(Meng et al., 2023)
0	Cold_season	2021	83.3801	32.4206	(Meng et al., 2023)
0	Cold_season	2021	81.3056	32.1976	(Meng et al., 2023)
0	Cold_season	2021	81.1381	31.0308	(Meng et al., 2023)
0	Cold_season	2021	82.5525	30.5754	(Meng et al., 2023)
0	Cold_season	2021	83.5271	30.0013	(Meng et al., 2023)
0	Cold_season	2021	81.5937	32.0056	(Meng et al., 2023)
0	Cold_season	2021	84.5005	31.5758	(Meng et al., 2023)
0	Cold_season	2021	93.8996	35.4687	(Meng et al., 2023)
0	Cold_season	2021	99.0157	35.2812	(Meng et al., 2023)
175.5	Cold_season	2005	101.213	37.48	(Zou et al., 2016)
15.2	Cold_season	2018	99.934	34.009	(Yu et al., 2021)
60.6	Cold_season	2018	99.461	33.757	(Zhou et al., 2024)
146.856465	Cold_season	2004	101.4512	36.49412	NRFOP
120.678685	Cold_season	2001	101.4512	36.49412	NRFOP
159.9051008	Cold_season	2000	101.4512	36.49412	NRFOP
410.5294118	Cold_season	2003	88.05187	44.15674	NRFOP
147.6139979	Cold_season	2002	101.4512	36.49412	NRFOP
153.7366548	Cold_season	2002	101.4512	36.49412	NRFOP
416.2280702	Cold_season	2001	88.05187	44.15674	NRFOP
1.4	Warm_season	2021	83.2653	33.1603	(Zhou et al., 2024)
0	Warm_season	2021	84.2821	33.2489	(Meng et al., 2023)
0	Warm_season	2021	82.5718	32.2185	(Meng et al., 2023)
0	Warm_season	2021	84.1376	32.1522	(Meng et al., 2023)

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

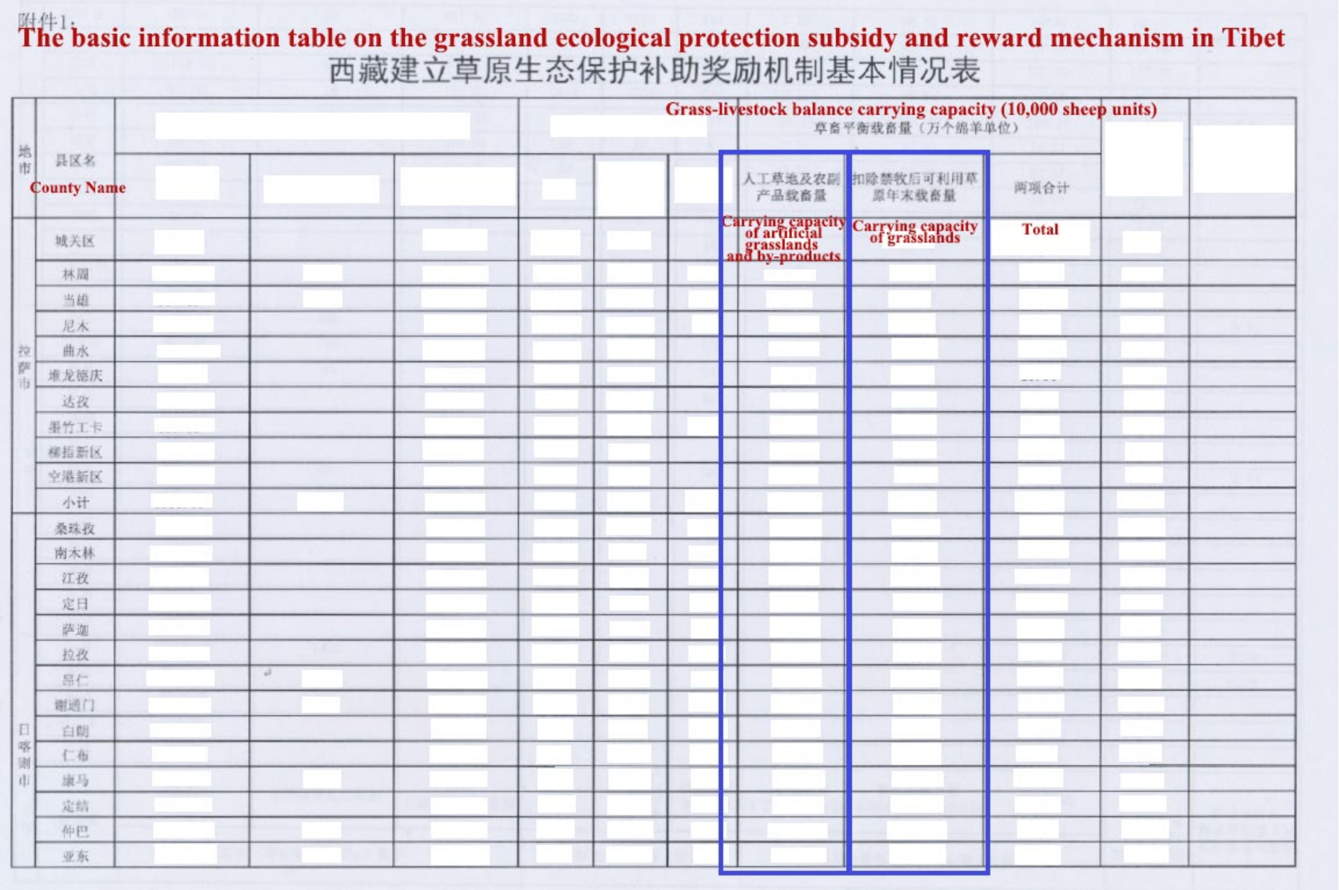
Open Access Earth System Science Data Discussions

Mean grazing density (SSUs/km2)	Season	Time	Longitude	Latitude	Source
0	Warm_season	2021	84.5005	31.5758	(Meng et al., 2023)
126	Warm_season	1998-2000	99.506	33.573	(Dong et al., 2015)
60.6	Warm_season	2018	99.461	33.757	(Zhou et al., 2024)
62	Warm_season	2012	87.078	43.878	(Zhou et al., 2024)
95	Warm_season	2012	87.085	43.888	(Zhou et al., 2024)
146.856465	Warm_season	2004	101.4512	36.49412	NRFOP
410.5294118	Warm_season	2003	88.05187	44.15674	NRFOP
147.6139979	Warm_season	2002	101.4512	36.49412	NRFOP
153.7366548	Warm_season	2002	101.4512	36.49412	NRFOP
120.678685	Warm_season	2001	101.4512	36.49412	NRFOP
416.2280702	Warm_season	2001	88.05187	44.15674	NRFOP
159.9051008	Warm_season	2000	101.4512	36.49412	NRFOP
56	Year_round	2010	107.369	38.641	NRFOP
68	Year_round	2010	114.84	44.0178	NRFOP
90.39077514	Year_round	2004	118.5405	41.96392	NRFOP
43.30666667	Year_round	2004	122.9213	44.52178	NRFOP
8.584105442	Year_round	2004	107.1001	37.81548	NRFOP
8.912958371	Year_round	2003	107.1001	37.81548	NRFOP
136.625	Year_round	2003	106.1724	35.58533	NRFOP
83.72837924	Year_round	2002	118.5405	41.96392	NRFOP
43.46153846	Year_round	2002	121.4012	42.41777	NRFOP
9.518465353	Year_round	2002	107.1001	37.81548	NRFOP
139.5	Year_round	2002	106.1724	35.58533	NRFOP
81.99871877	Year_round	2002	118.5405	41.96392	NRFOP
47.91208791	Year_round	2002	121.4012	42.41777	NRFOP
9.196137283	Year_round	2002	107.1001	37.81548	NRFOP
133.25	Year_round	2002	106.1724	35.58533	NRFOP
96.15631006	Year_round	2001	118.5405	41.96392	NRFOP
43.24175824	Year_round	2001	121.4012	42.41777	NRFOP
9.213101918	Year_round	2001	107.1001	37.81548	NRFOP
140.25	Year_round	2001	106.1724	35.58533	NRFOP

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Mean grazing density (SSUs/km2)	Season	Time	Longitude	Latitude	Source
84.11274824	Year_round	2000	118.5405	41.96392	NRFOP
18.21956846	Year_round	2000	107.1001	37.81548	NRFOP
118	Year_round	2000	106.1724	35.58533	NRFOP

The ordinary words in the sample table of *Grassland Ecological Protection Subsides* are in Chinese (Table A2), the red words are translated into English, and the blue-framed column contains the data used in this study. Additionally, as this data is not permitted for publication, all numbers in the table are masked.



附件1:
The basic information table on the grassland ecological protection subsidy and reward mechanism in Tibet
西藏建立草原生态保护补助奖励机制基本情况表

地市	县区名 County Name	草畜平衡载畜量（万个绵羊单位） Grass-livestock balance carrying capacity (10,000 sheep units): 人工草地及农副产品载畜量 Carrying capacity of artificial grasslands and by-products	扣除禁牧后可利用草原年末载畜量 Carrying capacity of grasslands	两项合计 Total
拉萨市	城关区			
	林周			
	当雄			
	尼木			
	曲水			
	堆龙德庆			
	达孜			
	墨竹工卡			
	柳梧新区			
	空港新区			
	小计			
日喀则市	桑珠孜			
	南木林			
	江孜			
	定日			
	萨迦			
	拉孜			
	昂仁			
	谢通门			
	白朗			
	仁布			
	康马			
	定结			
	仲巴			
	亚东			

Table A3: The sample basic information table on the grassland ecological protection subsidy and reward mechanism in Tibet.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Open Access Earth System Science Data Discussions

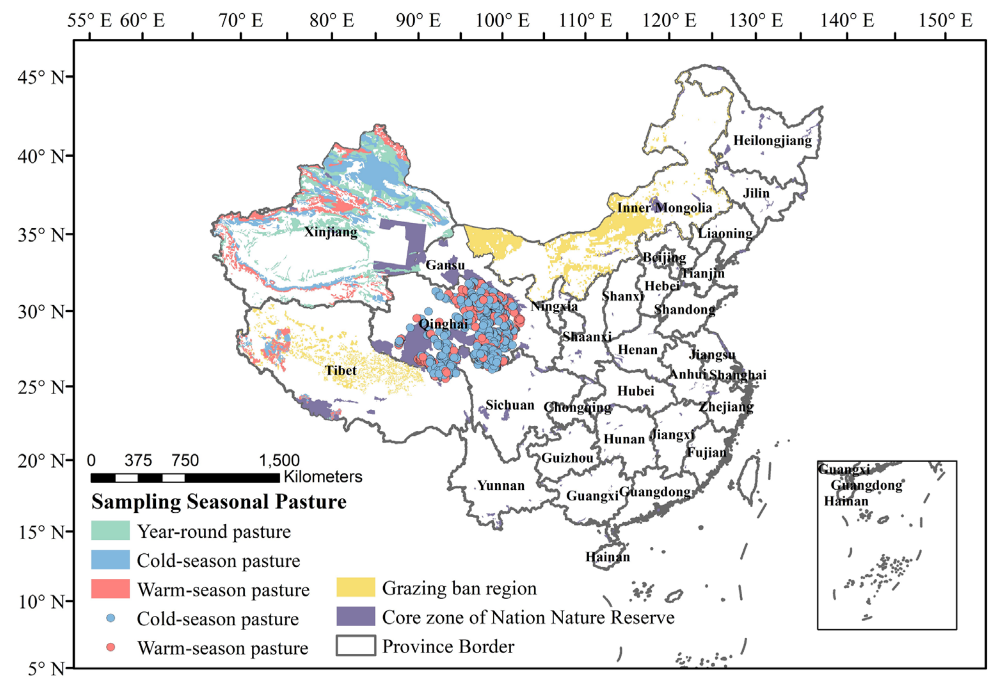


Figure A1: Sampling seasonal pasture mask and unsuitable mask for livestock grazing.

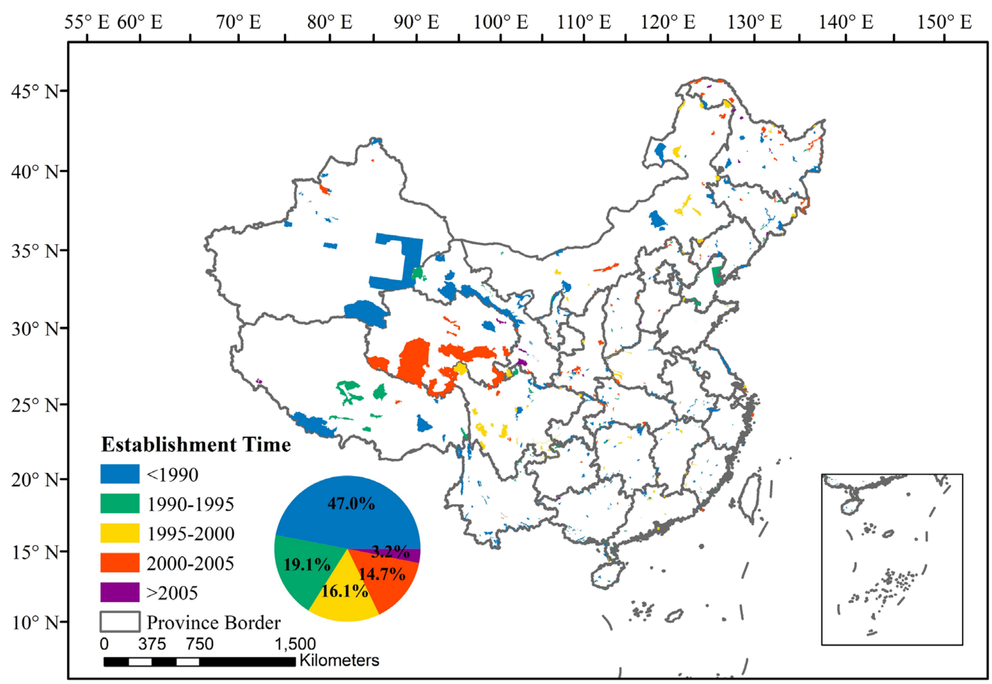


Figure A2: Establishment time distribution of National Nature Reserves in China.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Open Access Earth System Science Data Discussions

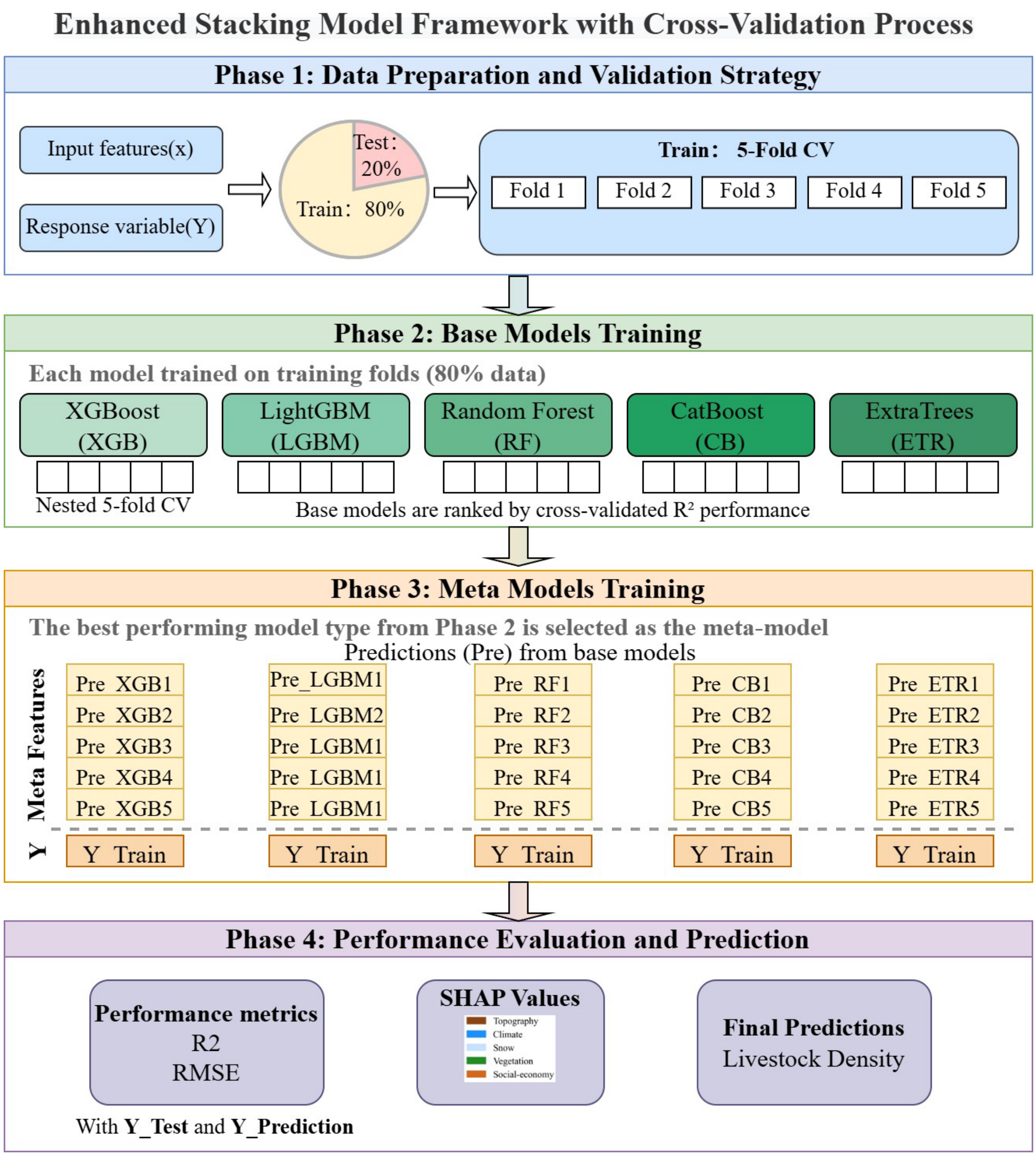


Figure A3: Structure of the stacking-based interpretable machine learning (IML) model for livestock density distribution

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

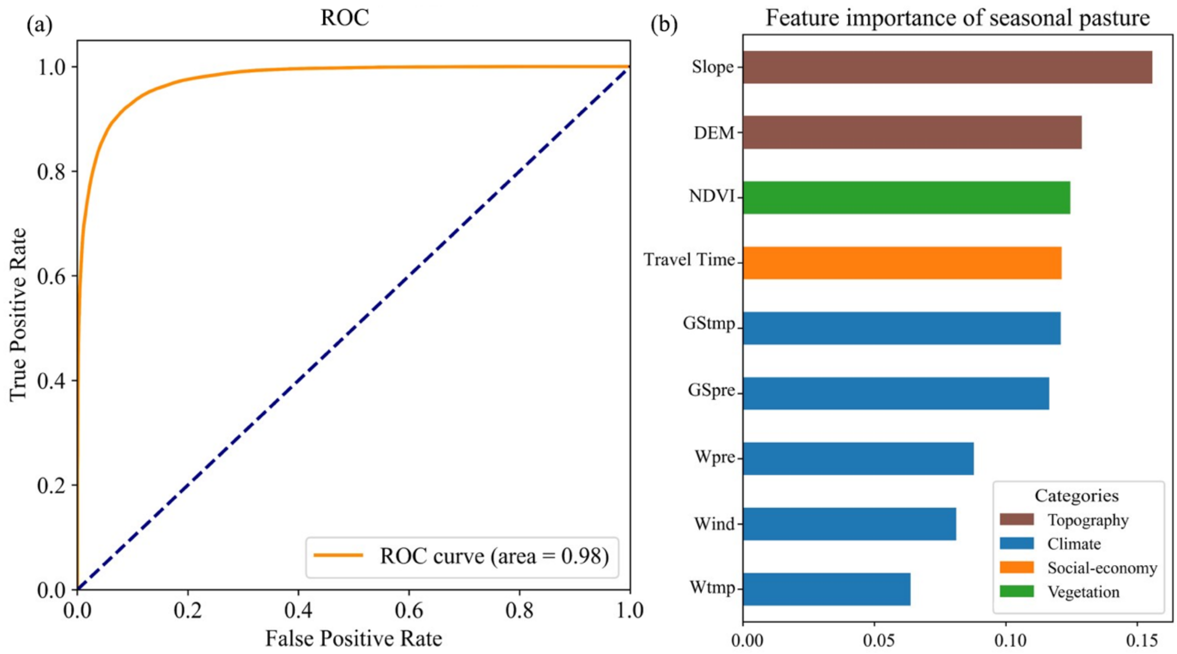


Figure A4: Results of ten-fold cross-validation (a) and feature importance (b) for the Random Forest Classification model.

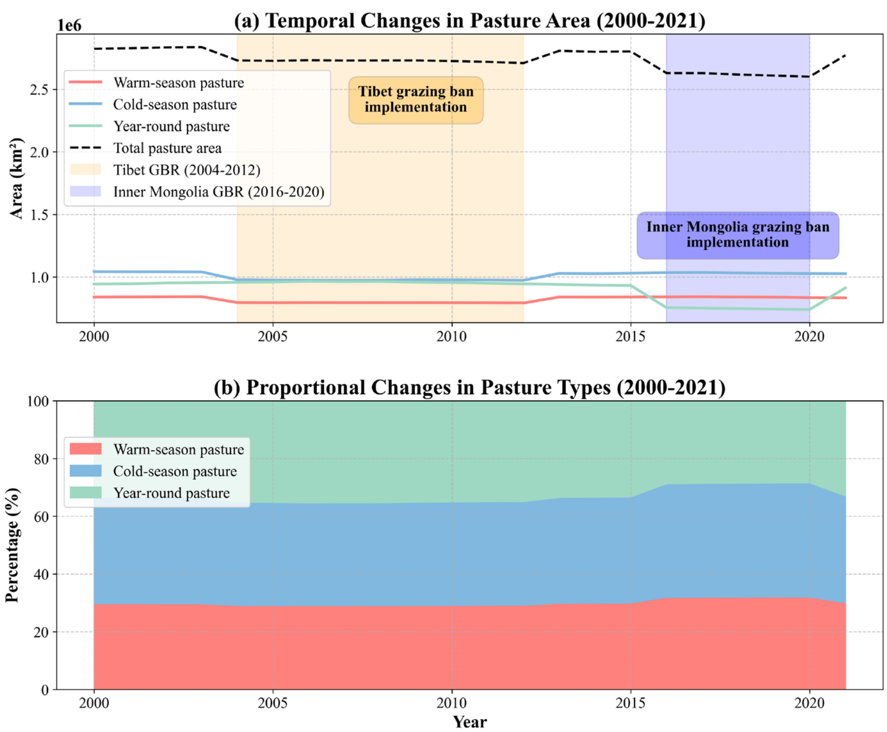


Figure A5: Temporal changes in seasonal pastures (2000-2021). (a) Absolute area (km²) of warm-season, cold-season, and year-round pastures. (b) Proportional composition (%) of the three pasture types.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

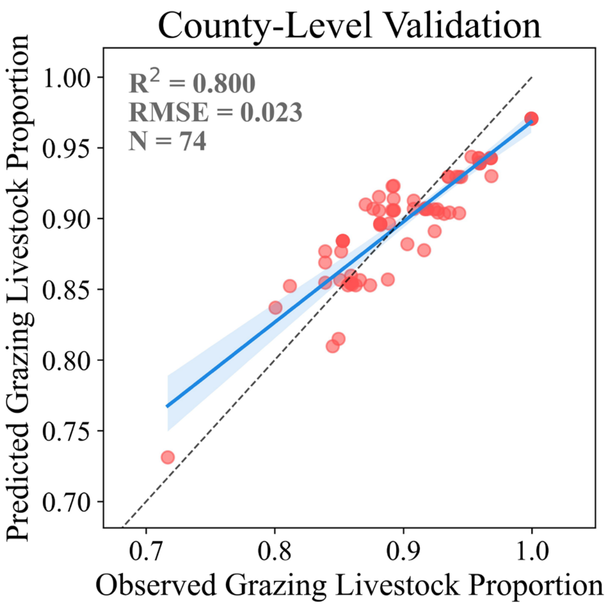
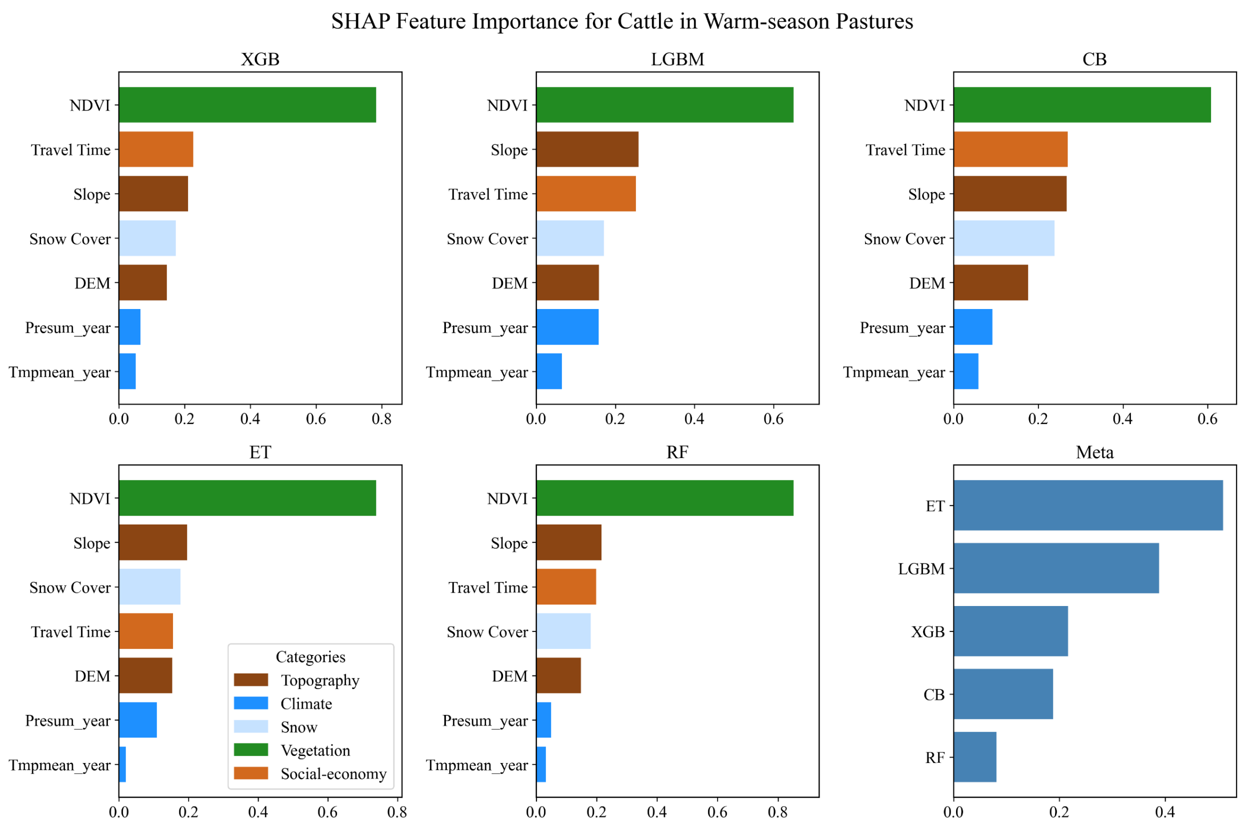


Figure A6: County-level validation of grazing ruminant livestock segmentation model using existing livestock carrying capacity data.



740 **Figure A7: SAHP feature importance and Meta-Model contributions of cattle distribution prediction in warm-season pastures.**

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

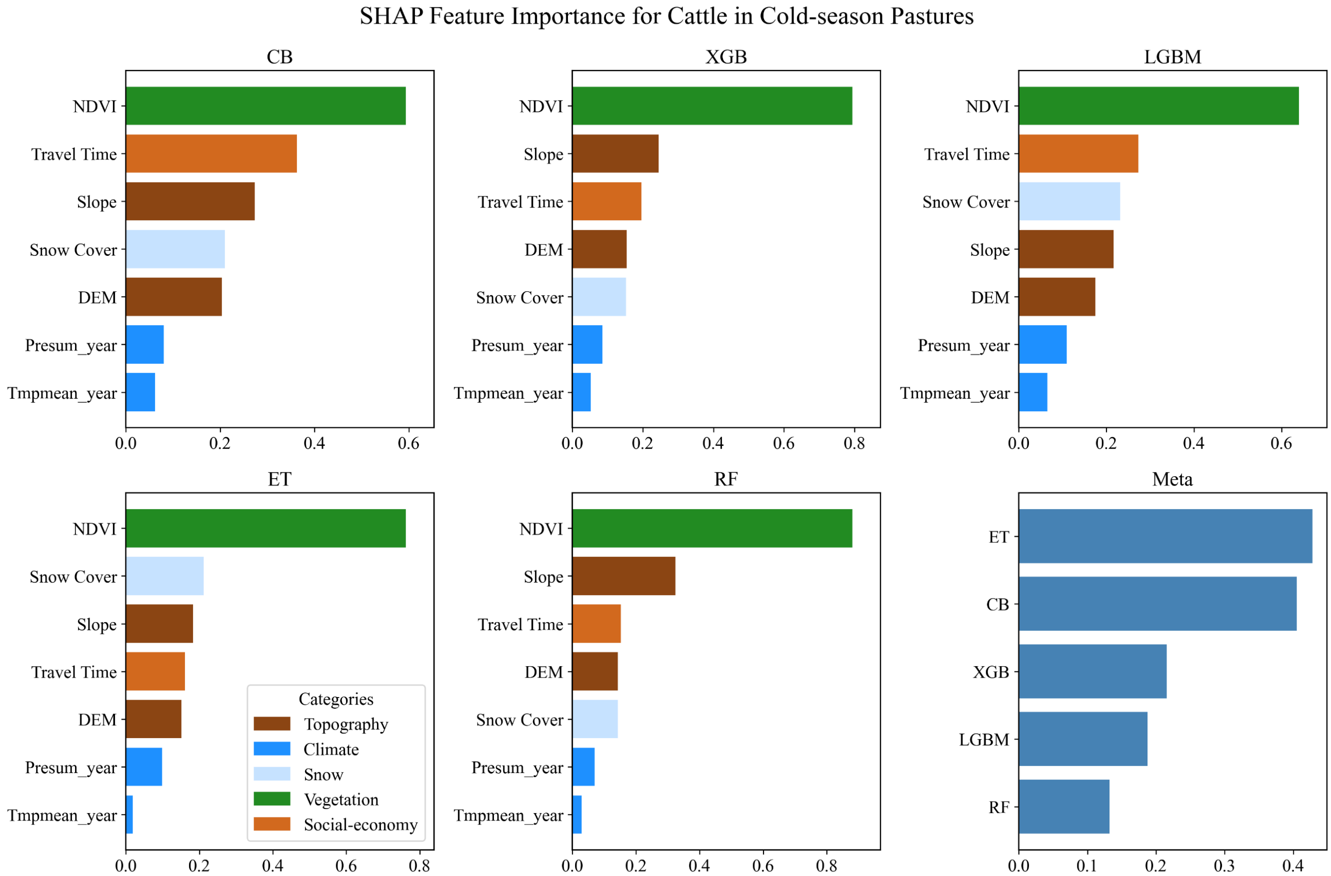
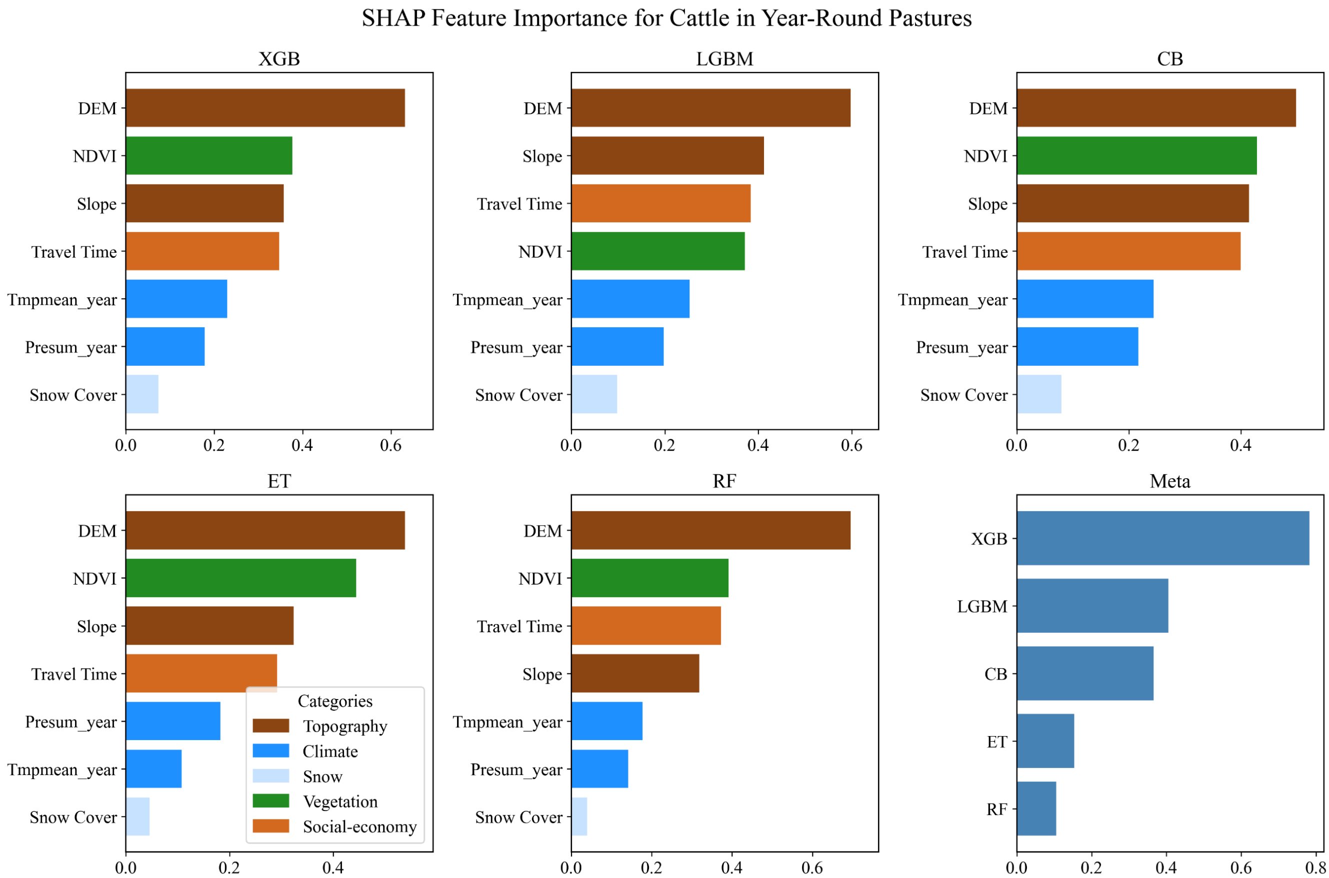


Figure A8: SAHP feature importance and Meta-Model contributions of cattle distribution prediction in cold-season pastures.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025



745 **Figure A9: SAHP feature importance and Meta-Model contributions of cattle distribution prediction in year-round pastures.**

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Earth System Science Data
Open Access
Discussions

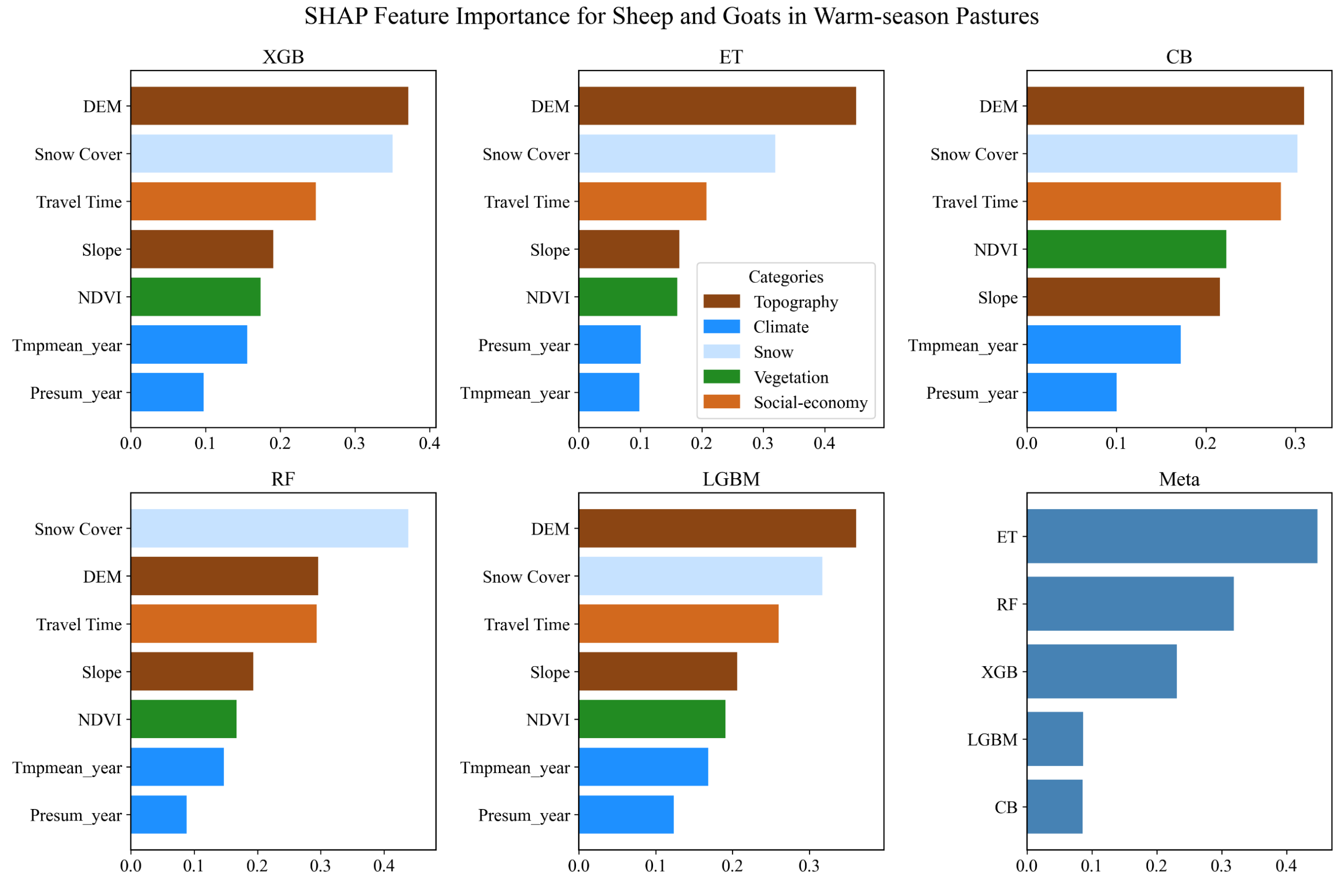


Figure A10: SAHP feature importance and Meta-Model contributions of sheep and goats distribution prediction in warm-season pastures.

750

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

Open Access Earth System Science Data Discussions

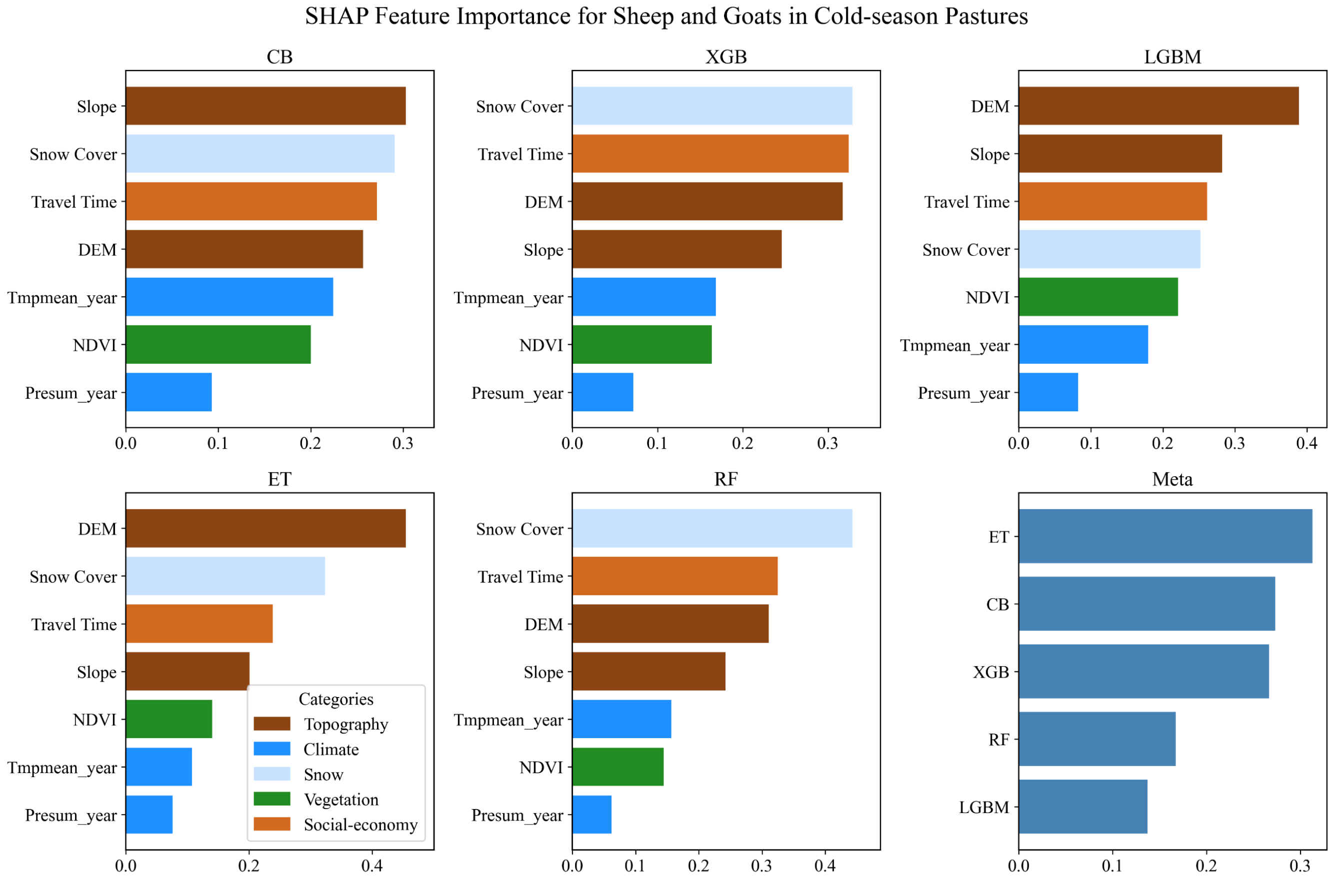


Figure A11: SAHP feature importance and Meta-Model contributions of sheep and goats distribution prediction in cold-season pastures.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

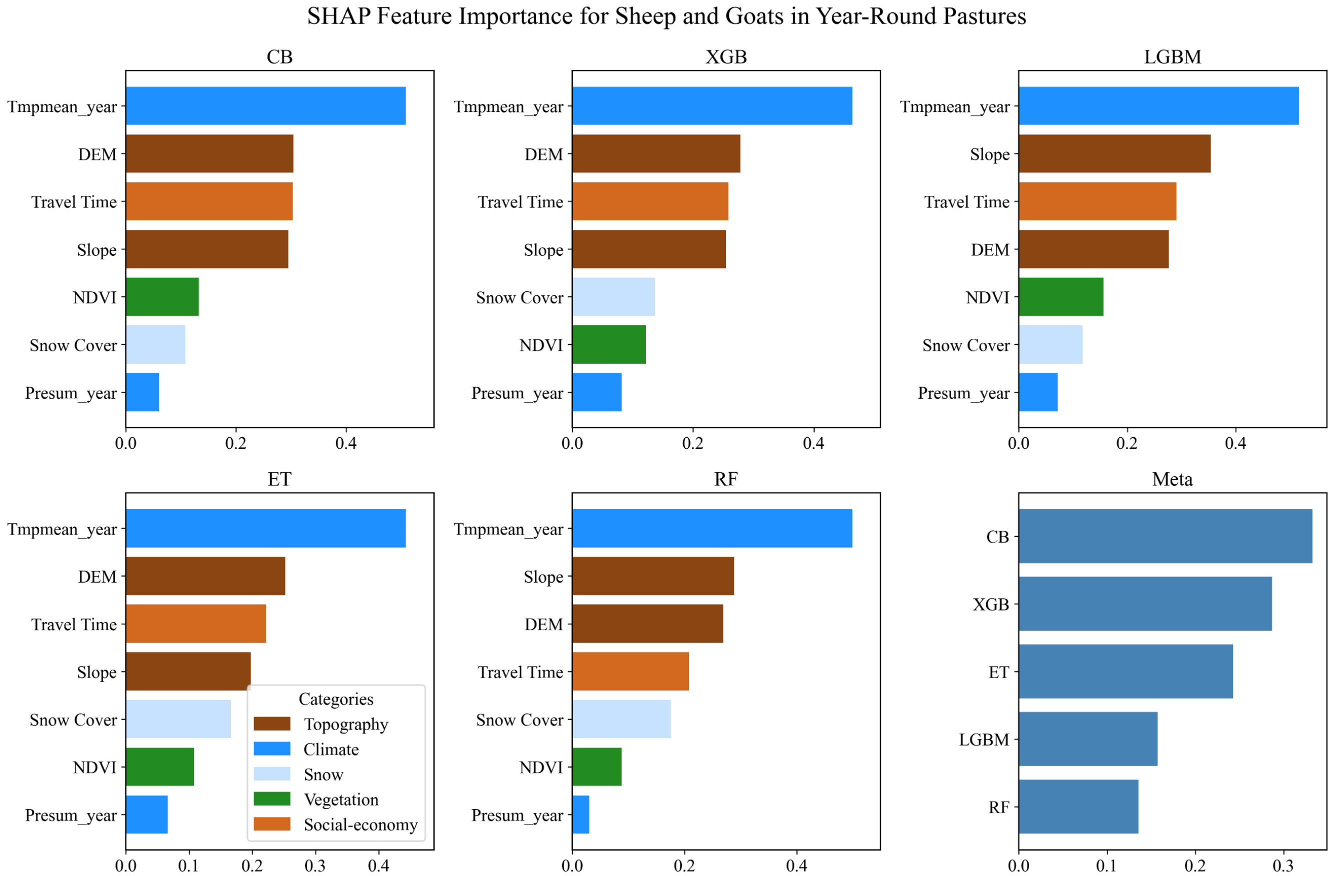


Figure A12: SAHP feature importance and Meta-Model contributions of sheep and goats distribution prediction in year-round pastures.

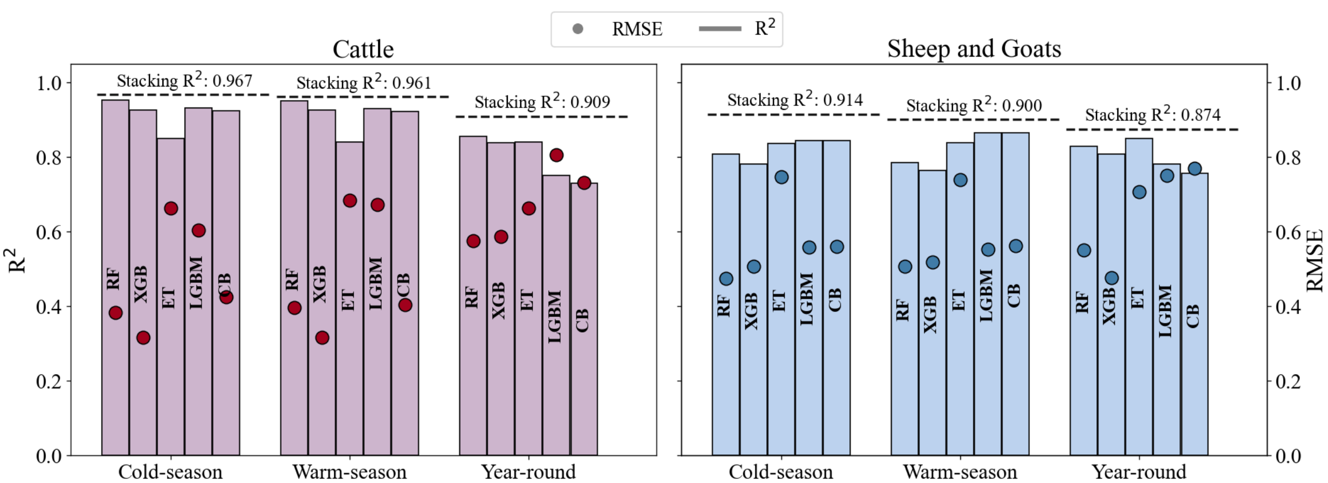


Figure A13: Comparison between individual model and stacking model for cattle and sheep (sheep and goats) distribution prediction in cold-season, warm-season, and year-round pastures.

https://doi.org/10.5194/essd-2025-263
Preprint. Discussion started: 4 June 2025

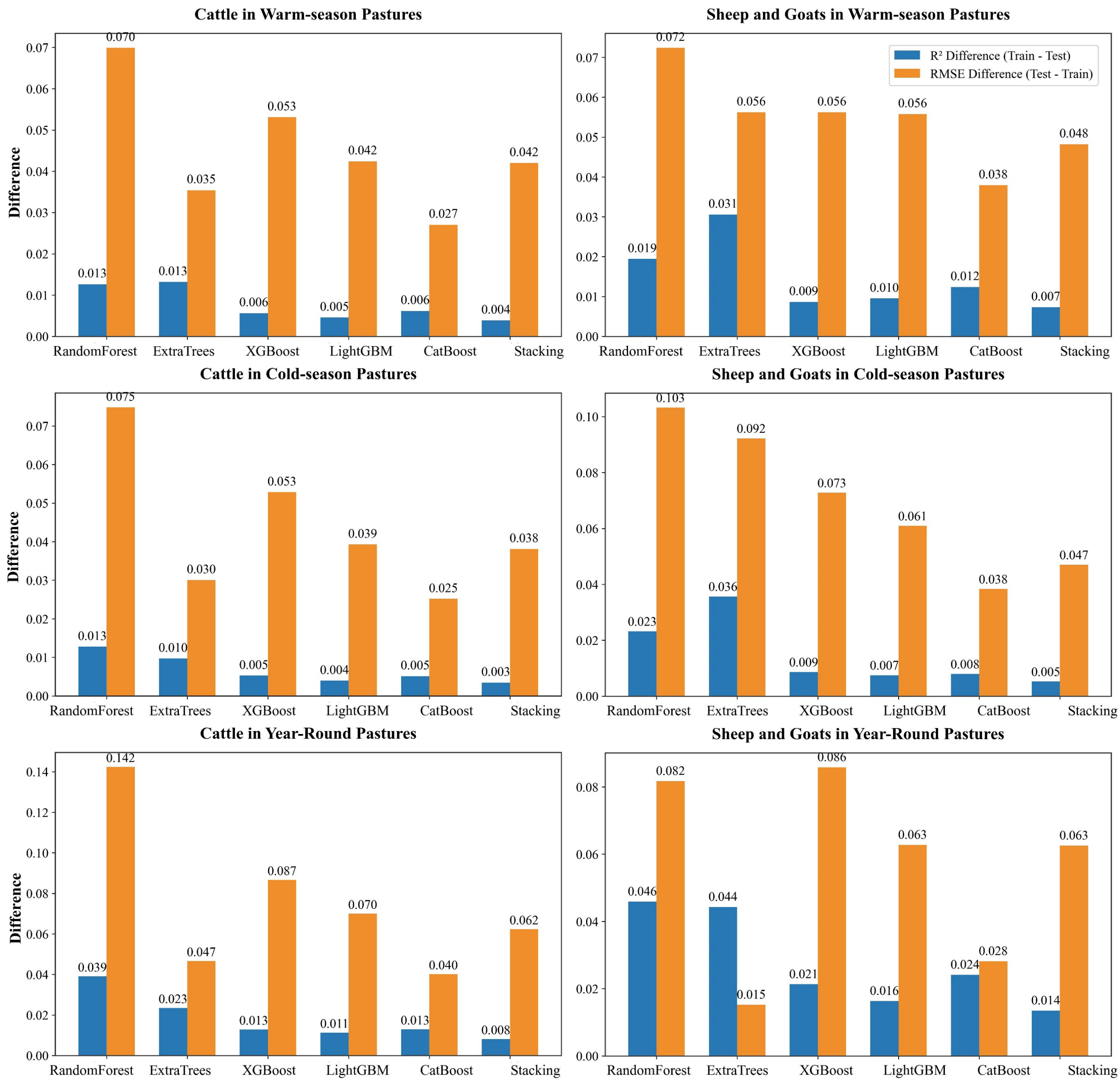


Figure A14: Overfitting assessment of cattle and sheep (sheep and goats) across all pasture types.

765