https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Open Access Earth System Science Data Discussions

Long-term Ruminant Livestock Distribution Datasets in Grazing Livestock Production Systems in China from 2000 to 2021 (CLRD-GLPS)

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 the sustainability of livestock systems, managing animal diseases, mitigating climate change risks, and controlling greenhouse gas emissions. In China, grazing ruminants are mostly distributed across the vast grasslands in semi-humid and alpine areas. However, existing datasets of gridded distribution of grazing ruminants in China do not distinguish grazing ruminants with other livestock production systems, nor capture their long-term and seasonal dynamics, and tend to overestimate grazing livestock distribution. This study uses the county-level data from the *Grassland Ecological Protection Subsidies* to differentiate grazing livestock from other forms of livestock rearing. Interpretable machine learning models were used to detect the seasonality of grazing pasture and map the **C**hina'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). The model's internal ten-fold cross-validation results (adjusted R^2) for cattle ranged from 0.850 to 0.952 and for sheep from 0.780 to 0.836. External validation using province-level livestock meat production data yielded Pearson correlation coefficients of 0.88-0.90 for cattle and 0.92-0.94 for sheep, respectively. The CLRD-GLPS datasets provide more detailed, gridded information on local livestock distribution than census-based data. Compared to actual census data and the GLW datasets, they better capture the spatial-temporal dynamics of livestock distribution. Spatially, the largest cattle numbers on seasonal pastures were in the south-eastern edge of the Qinghai-Tibet Plateau (QTP), while the largest sheep numbers were in north-eastern Qinghai and Xinjiang. Temporally (2000-2021), cattle numbers increased near the Three-River Source

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024


Open Access Earth System Science Data Discussions

National Park and Helan Mountains, while sheep numbers decreased on seasonal pastures on the QTP, with no significant changes on year-round pastures in Inner Mongolia. The datasets provide essential information for understanding the spatial-temporal dynamics of grazing ruminants and formulating relevant livestock management policies, among other applications. Additionally, the research framework developed in this study can serve as a new framework for creating livestock distribution datasets in 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., 2022a; 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. 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 livestock systems.

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). However, since most ruminant livestock depend on grasslands for grazing and often move seasonally, especially in regions like China that adopt a two-season transhumance system, the coarse resolution of GLW3 and its lack of consideration for seasonal livestock movements within local boundaries limit its applicability. This is particularly true 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). The high-resolution livestock maps enable more accurate tracking of livestock movements across different seasons (Ocholla et al., 2024). To address this gap in spatial resolution and the associated seasonal movements, 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.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

However, predicting long-term livestock distributions has become a challenge 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. Understanding and explaining the mechanisms behind spatial-temporal changes in livestock distribution are equally important. However, machine learning methods are often considered black-box models and cannot directly explain the mechanisms behind the data(Hassija et al., 2024). Interpretable machine learning (IML) methods can address this by extracting relevant knowledge from machine learning models regarding relationships present in the data or learned by the model (Murdoch et al., 2019). For example, feature importance scores provide insights into which features the model has identified as significant for specific outcomes and their relative importance (Breiman, 2001). Structural equation modelling (SEM), a statistical technique based on multivariate linear regression, can be used to test conceptual models based on cause-effect relationships, further enhancing our understanding of these mechanisms (Chernozhukov et al., 2024; Feuerriegel et al., 2024).

Apart from the above challenges in mapping long-term grazing livestock distribution, another key issue is that existing datasets often model livestock distribution directly based on census data, without distinguishing livestock numbers among various livestock production systems. Livestock production systems (LPS) have been previously classified into three basic types: grazing LPS, mixed farming LPS and landless LPS (Robinson et al., 2011). With China emerging as a significant new consumer market for ruminant products and having one of the largest expanses of remaining grasslands on Earth, these grasslands currently support a large number of ruminant livestock (Du et al., 2018). Previous studies have directly used census data to build models and to assign livestock numbers in census polygons for predicting livestock distribution in grazing LPS based on grassland (Gilbert et al., 2018; Zhan et al., 2023). However, as census data cannot distinguish grazing livestock from the total livestock, for China, it is difficult to separate ruminants distributed in western China from those in mixed-farming LPS. Ruminants in landless LPS are relatively few according to the statistical data (Ministry of Agriculture and Rural Affairs China, 2000). Previous studies on predicting livestock distribution have seldom differentiated between different LPS, often overlooking the variations in distribution patterns among these systems. Even when studies focus on grazing livestock within a single livestock production system, they typically use overall livestock numbers to predict the distribution of grazing livestock in grazing LPS, leading to an overestimation of the actual grazing livestock numbers.

This study aims to develop a long-term 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-2024-534
Preprint. Discussion started: 10 December 2024

Open Access Earth System Science Data Discussions

2 Data and method

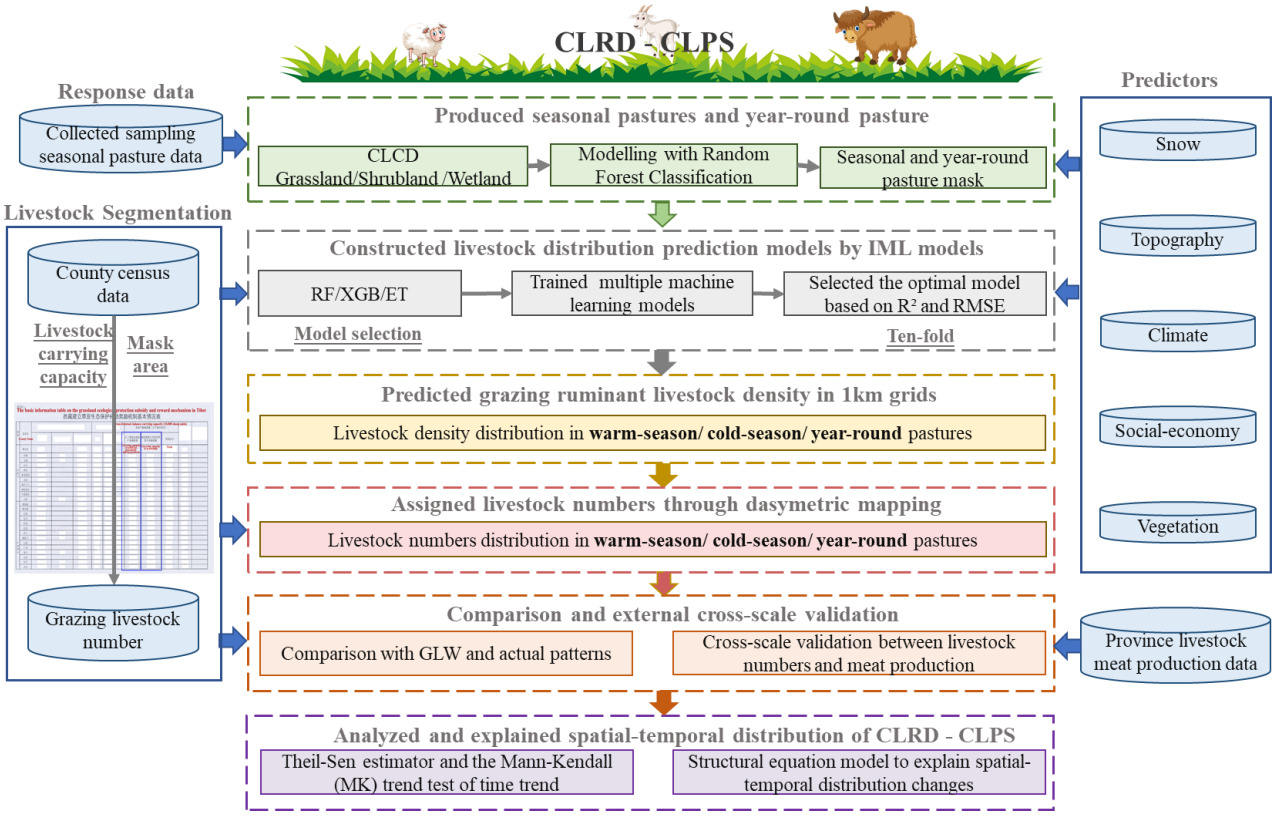


Figure 1: Framework of mapping grazing ruminant livestock distributions in China's grazing livestock production system.

In this study, there are six detailed steps to produce the CLRD-GLPS datasets and analyse their spatial-temporal distribution (Figure 1): (1) Preparation of data and variables, especially in segmenting grazing livestock from the whole livestock system, (2) preparation of a pasture mask suitable for livestock grazing and random forest classification modelling for predicting seasonal pastures, (3) construction of livestock distribution prediction models using machine learning models, (4) prediction of grazing ruminant livestock in 1 km grids and assignment livestock numbers within county boundary, (5) comparison with GLW and actual patterns, and external cross-validation between livestock numbers and meat production, and (6) analysis and explanation of the spatial-temporal distribution of grazing ruminant livestock.

2.1 Data

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

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

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 pig (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 and sheep species available. The units for GLW in 2010 and 2015 represent absolute livestock numbers (Gilbert et al., 2018), whereas in 2020, the data is provided as livestock density (heads per km^2).

For province-level validation, we collected animal husbandry production data from the China Statistical Yearbook (https://www.stats.gov.cn/english/Statisticaldata/yearbook/) spanning the years 2000 to 2021. This data set includes production figures for pork, beef, and mutton, with the unit of measurement expressed in ten thousand tons.

2.1.2 Mask data

The annual China Land Cover Dataset with a 30 m spatial resolution was utilized to create a suitable distribution mask for livestock (Yang and Huang, 2021). Land cover types including grassland, cropland, shrub, and wetland are considered suitable for grazing livestock. To generate a valid pasture boundary, we also obtained the boundaries of national nature reserves from National Nature Reserve Boundary Data published in the Resource and Environment Science and Data Center, Chinese Academy of Science (https://www.resdc.cn/data.aspx?DATAID=272). It includes 169 national nature reserves in China. The boundary of grazing ban regions was collected from the article (Sun et al., 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 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*.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

2.1.4 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), and snow cover data (Hall, D. K. and G. A. Riggs, 2016). 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 suitable grazing pastures and Segmentation of livestock population data in grazing LPS

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). Considering the impact of land use changes on grazing livestock distribution, the 30 m spatial resolution at five-year intervals China Land Cover Dataset from 2000 to 2020 were used to create a suitable grazing distribution mask. The grid data from the China Land Cover Dataset were resampled to a 1km spatial resolution, and only the pixels representing grassland, shrubland, and wetland land cover types were retained in the mask. To generate a valid pasture boundary, we used the boundaries of National Nature Reserves and the boundary of grazing ban regions, which are banned for livestock grazing (Figure A1).

To segment grazing ruminant livestock from whole ruminant livestock, we used the detailed tables on *Grassland Ecological Protection Subsides* (Table A2), which included data on livestock carrying capacities for grasslands and agricultural by-products across 74 counties from Tibet and Qinghai province. Based on such tables, we can get the average livestock carrying capacities between grazing LPS and mixed-farming LPS. Additionally, we use the livestock carrying capacities as weighting factors for different LPS types, where grazing LPS typically encompasses land uses such as grasslands, shrublands, and wetlands, whereas mixed-farming system primarily involves cropland. We calculated the area covered by each LPS type within each county with the data from China Land Cover Dataset for every five year, then multiply these areas by their corresponding livestock carrying capacities (in grazing LPS or other LPS) to compute a weighted area. This weighted area serves as the basis for allocating livestock across grazing and other LPS within the counties.

$$W_{Gi} = \frac{CC_G \cdot Area_{Gi}}{CC_G \cdot Area_{Gi} + CC_M \cdot Area_{Mi}} , \quad (1)$$

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

$$LN_{Gi} = LN_{Ti} \cdot W_{Gi} \ , \tag{2}$$

where W_{Gi} is weighted ratio for grazing LPS in county *i*; CC_G is average carrying capacity for grazing LPS, meanwhile CC_M is the same for mixed-farming LPS; $Area_{Gi}$ is the area for grazing LPS in the county *i*, and $Area_{Mi}$ is the same for mixed-farming LPS in the county *i*; LN_{Gi} is the livestock numbers for grazing LPS in the county *i*, and LN_{Ti} is the total livestock numbers in the county *i* from county-level statistical yearbook.

2.2.2 Classification for seasonal pastures and year-round pasture

Based on the suitable 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). 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 Construction of the livestock density distribution models and assignment of livestock numbers

This study employed interpretable machine learning techniques to develop distribution models for the ruminant livestock (cattle and sheep). Initially, the response variables required for the machine learning models are processed. Specifically, for different ruminant livestock types (cattle and sheep), statistical county-level livestock density was calculated by dividing the annual livestock numbers for grazing LPS (LN_{Gi}) by the area covered by a suitability mask for each county's single pasture types (warm-season pastures/cold-season pastures/year-round pastures). This density served as the response variable for constructing distribution models in the grazing LPS. Consequently, the study constructed a total of six livestock distribution

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

models: three models each for cattle and sheep within the single pasture types (warm-season pastures/cold-season pastures/year-round pastures), with livestock density as the response variable, measured in heads per square kilometer (head/km²).

In developing livestock distribution models within grazing LPS, factors including terrain, climate, vegetation, snow cover, and human activities were considered as predictors. The distribution patterns of livestock within grazing LPS and mixed-farming LPS differ significantly. In grazing LPS, the seasonal distribution of livestock was closely linked to the availability of grassland resources, environmental stress, and the activities of pastoralists. Specifically, terrain acts as a macro-control factor, climate determines the type and productivity of grasslands, and the productivity of grassland vegetation is crucial in determining the carrying capacity for livestock.

Interpretable machine learning models were utilized to predict the relationship between the response variable (livestock density) and various predictors. This study employed several types of machine learning algorithms, including Random Forest (RF) (Breiman, 2001), Extreme Gradient Boosting (XGBoost) (Friedman, 2001), and Extreme Randomized Trees (ET) (Geurts et al., 2006). Machine learning model training and predictions were implemented using Python 3.8.18, incorporating recursive feature elimination (Guyon and Elisseeff, 2003) and grid search algorithm (Bergstra and Bengio, 2012) to identify the optimal predictive variables and parameter combinations for each model. During the model training phase, the performance of these models was quantitatively assessed. This assessment was based on the error metrics adjusted R^2 and Root Mean Squared Error (RMSE) obtained through ten-fold cross-validation. These metrics helped select the model that exhibits the highest simulation accuracy and generalization capability. The model with the highest adjusted R^2 and the smallest RMSE was considered the optimal model and was then used to predict the distribution of livestock. Feature importance was introduced to identify the most importance feature in different interpretable machine learning methods (Breiman, 2001).

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).

2.2.4 Comparison and validation of grazing ruminant livestock distribution

To compare our results (CLRD-GLPS) with other livestock distribution maps and actual distribution patterns, we used the GLW datasets for 2010, 2015, and 2020, standardized to livestock density (heads/km²) as representatives of widely used global

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

datasets. Actual livestock distribution patterns were derived from county-level livestock numbers for grazing LPS (LN_{Gi}) in corresponding years, also turned to density (heads/km²) within the pasture mask (warm-season, cold-season, and year-round pastures) for each county-level polygon. First, we compared their livestock densities spatially. Second, we aggregated values to the county level and calculated the R^2 to assess validation accuracy.

To conduct external cross-validation and further ensure the robustness of our machine learning models, we used livestock meat production data at the provincial level. Before validation, livestock meat production data at the provincial level were adjusted using grazing weights to accurately reflect the livestock meat production from grazing LPS within each administrative unit. Predicted livestock distribution data was aggregated within the boundaries of corresponding administrative units. Pearson correlation coefficient (r) (Adler and Parmryd, 2010) was calculated to assess validation accuracy.

2.2.5 Time trend of grazing ruminant livestock distribution

Temporal trends in livestock distribution datasets and statistical county-level livestock density were analysed using the Theil-Sen estimator and the non-parametric Mann-Kendall (MK) trend test (Sen, 1968). For livestock distribution dataset, the Theil-Sen estimator provided the slope of the trend in livestock numbers (unit: heads/year), while the MK trend test assessed the significance of these trends at a grid scale. As for statistical county-level livestock density, the Theil-Sen estimator provided the slope of the trend in livestock density (unit: (heads/km^2)/year), while the MK trend test assessed the significance of these trends at a county scale. Notably, we only calculated trends for counties with statistical data that included at least five consecutive years without missing values.

2.2.6 Mechanism explanation using structural equation modelling

To further explain the mechanisms of the spatial-temporal changes in CLRD-GLPS, structural equation modelling (SEM) was employed. SEM is a multivariate data analysis method used to analyse complex relationships among constructs and indicators (Hair et al., 2021). Path diagrams can represent a structural equation model that addresses causal questions, where variables are depicted as nodes, causal paths are indicated by directed arrows, and causal parameters are denoted by symbols adjacent to the arrows (Chernozhukov et al., 2024).

In this analysis, we calculated the province-level mean livestock density based on the CLRD-GLPS results from 2000 to 2021 as the dependent variable in the SEM. The independent variables, calculated by province-level means for the corresponding years, included DEM, slope, GSpre, Wpre, GStmp, Wtmp, snow cover, NDVI, travel time, and POP (Table A1). Additionally, land use change was taken into account by incorporating the province-level ratio of suitable grazing pastures. The Comparative

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), Chi-Square, and Degrees of Freedom (df) were used to assess the model fit of the SEM. An excellent model fit is indicated by a CFI value between 0.9 and 1, an RMSEA value below 0.06, and a Chi-Square/df ratio below 3 with a p-value greater than 0.05 (Suhr, 2006).

3 Results

3.1 Pasture mask and grazing ruminant livestock density distribution

The ten-fold cross-validation results of the RFC model are shown in Figure A2, with the mean AUC of the ten cross-validations being 0.98, demonstrating the model's ability to predict seasonal pastures for livestock. Using the RFC model, we predicted the seasonal pastures in the five provinces of the Qinghai-Tibet Plateau. By combining these predictions with the suitable distribution mask and the seasonal and year-round pasture distribution maps of Xinjiang, we obtained a 1km seasonal and year-round pasture distribution mask for ruminant grazing livestock in a single livestock production system, as shown in Figure 2. Figure 2 depicts the mask for the year 2020, and similar masks were created for the years 2000, 2005, 2010, and 2015.

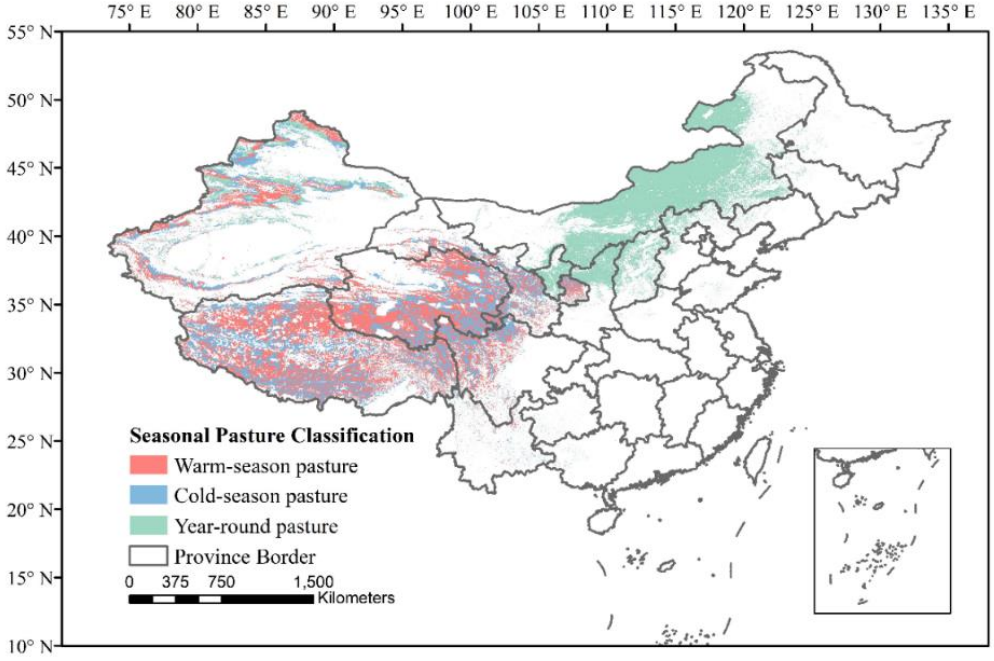


Figure 2: Seasonal and year-round pasture mask for livestock grazing in China (2020).

Based on the seasonal and year-round pasture masks, RF/XGB/ET models were utilized to predict the distribution of livestock in warm-season, cold-season, and year-round pastures respectively, covering all three pasture types with each model. The cattle model outperformed the sheep model, with R^2 values greater than 0.8 (Figure A3). Additionally, the seasonal pasture

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

model for cattle performed better than the year-round pasture model, with RMSE values of approximately 0.3 for the seasonal pasture model and about 0.7 for the year-round pasture model. For sheep, the R^2 values for all models ranged from 0.76 to 0.85. with the year-round pasture model having better R^2 values compared to the seasonal pasture model, but also a higher RMSE. This higher RMSE is due to the sample mean livestock density being higher in year-round pastures than in seasonal pastures.

The differences between the machine learning models are small, with the RF and ET models performing slightly better than the XGB model. Considering the similar performance of both RF and ET models in cross-validation, and the broader utilization of the random forest model in previous livestock prediction modelling, RF was chosen as the final model for predicting livestock density distribution (Gilbert et al., 2018; Zhan et al., 2023). The average of the ten-fold results for each grid was computed to determine the final predicted livestock density (Figure 3).

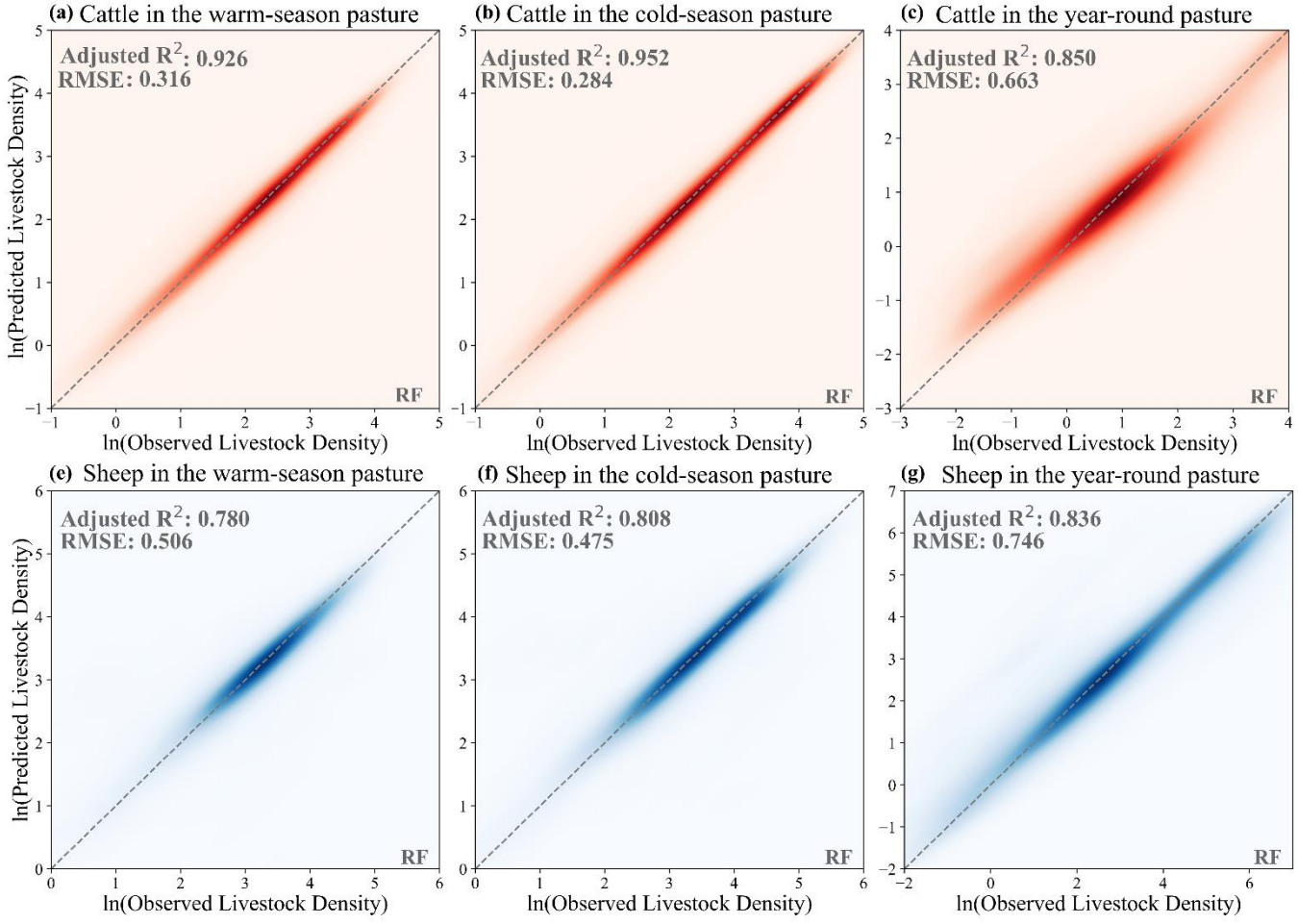


Figure 3: Kernel Density Estimation (KDE) plots of ten-fold cross-validation for predicting cattle (a-c) and sheep (e-g) distribution in warm-season, cold-season, and year-round pastures using the Random Forest models.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Open Access Earth System Science Data Discussions

3.2 Grazing ruminant livestock numbers distribution (CLRD-GLPS) and validation

Using the livestock density distribution predicted by the Random Forest model, we assigned county-level cattle and sheep numbers through dasymetric mapping. This approach resulted in the distribution of livestock numbers (CLRD-GLPS) across warm-season, cold-season, and year-round pastures. The CLRD-GLPS cover 22 years from 2000 to 2021, with results for every five years (2000, 2005, 2010, and 2015) shown in Figure A4 - Figure A9.

To validate the CLRD-GLPS, we first compared the CLRD-GLPS (Figure 4 (a)–(b)) and GLW (Figure 4 (c)–(d)) datasets separately with county-level census data. The validation results showed an R^2 of 0.984 for cattle and 0.999 for sheep in the CLRD-GLPS, while the R^2 for the GLW was only 0.003 for cattle and 0.291 for sheep. These findings demonstrate that CLRD-GLPS more accurately reflect the actual total livestock numbers in grazing livestock production systems at the county level.

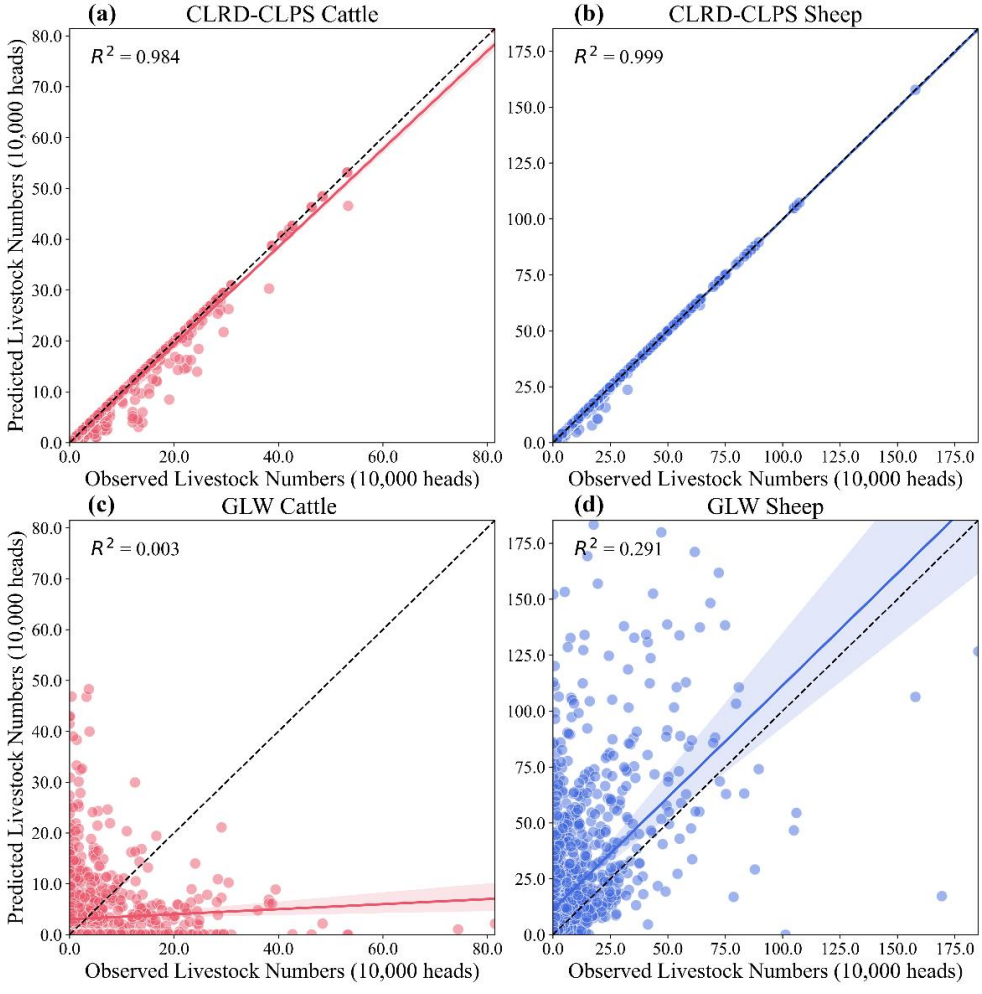
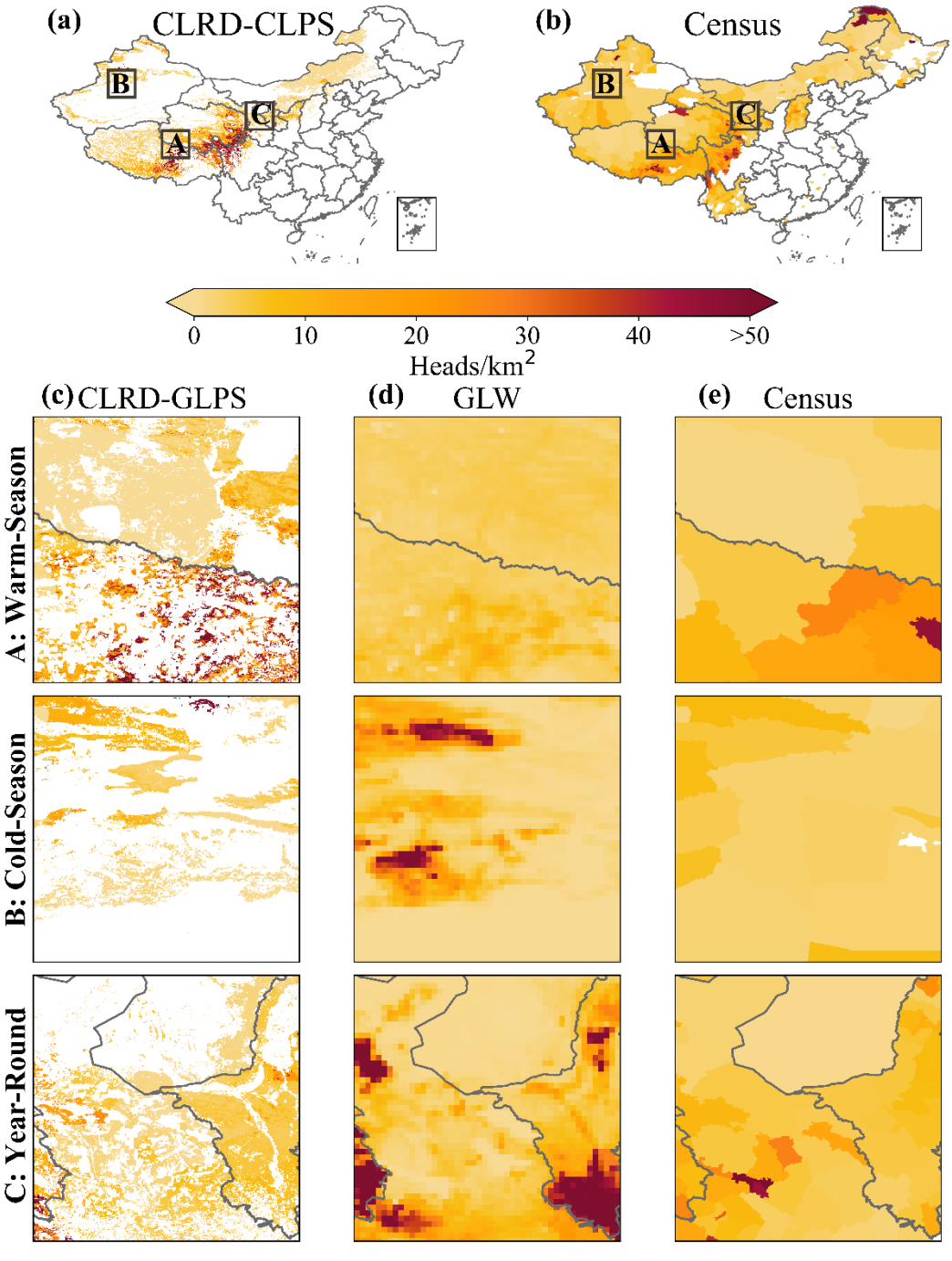


Figure 4: Validation of livestock numbers across all pasture types at the county level in 2010, 2015, and 2020, comparing CLRD-GLPS ((a)–(b)) and GLW ((c)–(d)) datasets separately with census data.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

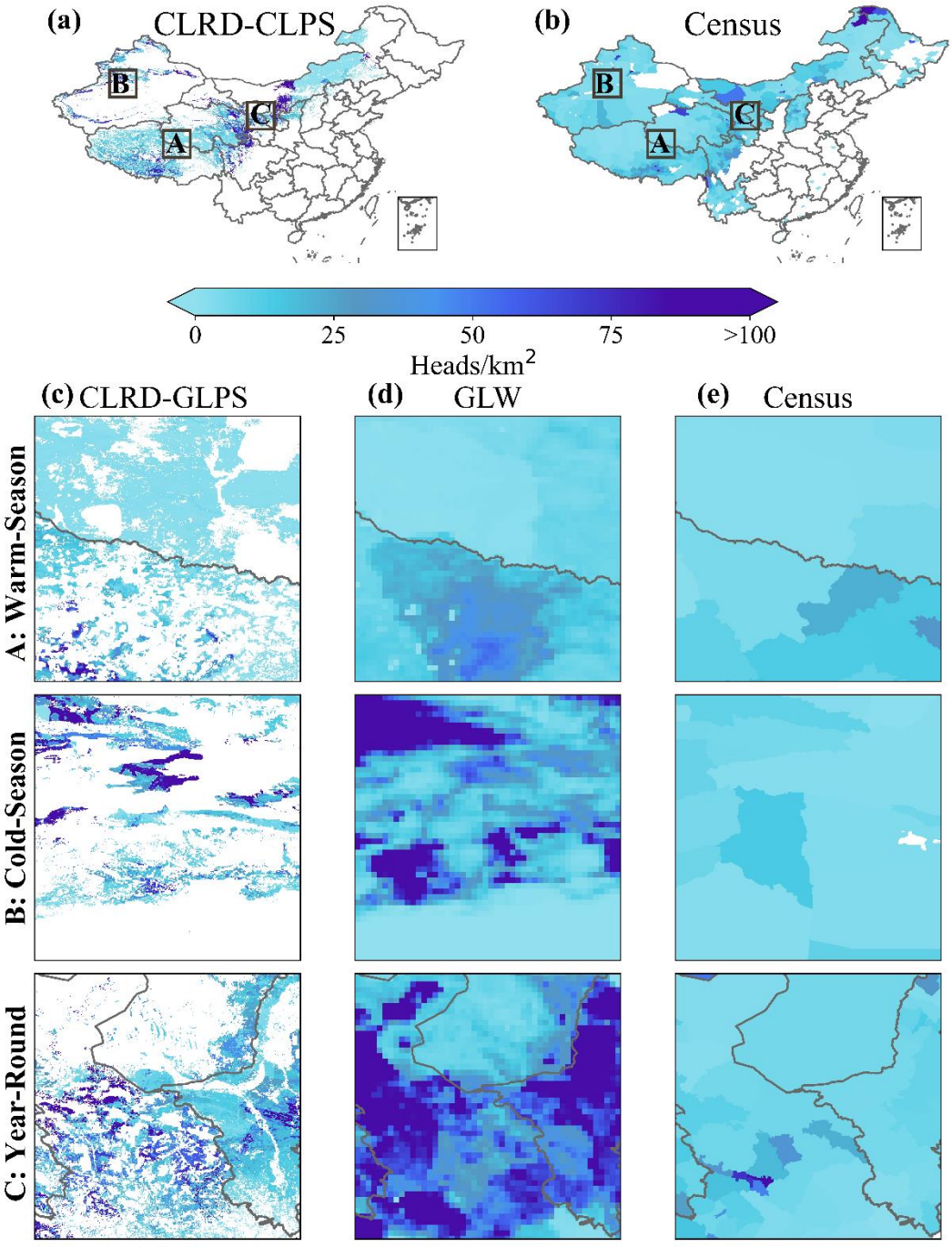


315

Figure 5: Comparison of CLRD-GLPS with GLW and actual census cattle density distribution in 2020. (a) CLRD-GLPS prediction map of cattle density across all pastures in China using Random Forest models; (b) actual census map of cattle density in warm-season pastures; zonal cattle density distribution for CLRD-GLPS (c), GLW (d), and census (e) across different pasture types.

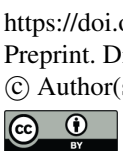
https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Open Access Earth System Science Data Discussions



320 **Figure 6: Similar to Figure 5, but for sheep.**

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024



Furthermore, we compared the spatial distribution of cattle (Figure 5) and sheep (Figure **6**) in CLRD-GLPS with the GLW and actual census distributions across different types of pastures in 2020. Overall, the results of CLRD-GLPS are closer to the actual census data and provide more detailed information into livestock movements across different pastures compared to GLW in grazing livestock production systems. In seasonal pastures, cattle are most numerous in the southeast of the Qinghai-Tibet Plateau (QTP), with numbers decreasing as they extend into Xinjiang (Figure 5). In contrast, sheep are most numerous in the north-eastern part of Qinghai Province and Xinjiang (Figure 6). In year-round pastures, the livestock density is lower than that in cold-season or warm-season pastures, as grazing can occur year-round. However, for sheep, there is still a relatively high density of livestock distribution in Xinjiang and the western part of Inner Mongolia (Figure 6). Based on the localized distribution maps for 2020, the detailed distributions of cattle and sheep are better illustrated in the three main regions ((c) - (e) in Figure 5 and Figure 6). Based on the future importance scores derived from the interpretable machine learning (RF model), the distribution of cattle and sheep in seasonal pastures is primarily influenced by vegetation growth and terrain (Figure A10 and Figure A11). In year-round pastures, however, travel time emerges as a more crucial factor influencing cattle and sheep distribution (Figure A10 and Figure A11). Additionally, the GLW datasets are overall overestimated compared to the CLRD-GLPS datasets and actual census data.

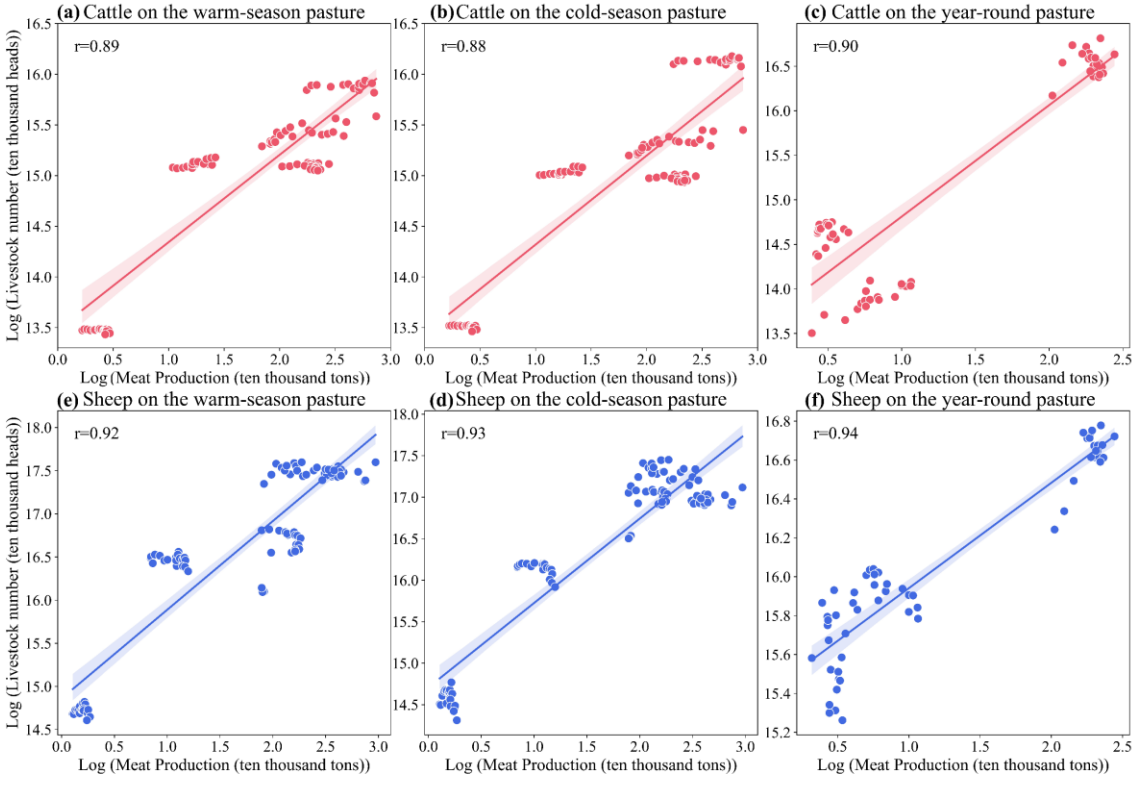
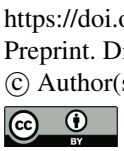


Figure 7: External cross-validation for cattle (a-c) and sheep (e-g) in the CLRD-GLPS datasets across warm-season, cold-season, and year-round pastures by comparing the results with livestock meat production.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

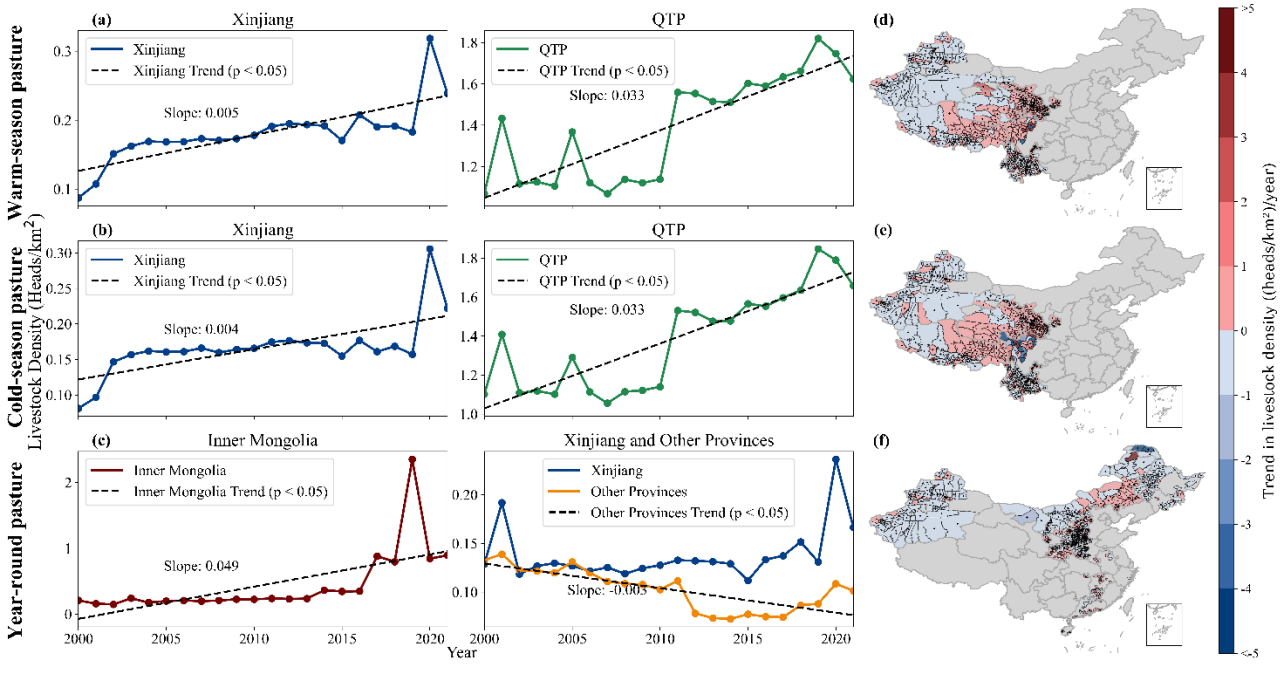


Open Access Earth System Science Data Discussions

Finally, external cross-scale validations of the CLRD-GLPS results were conducted using livestock meat production data. Livestock meat production data at the provincial level were adjusted using weights to accurately reflect the figures for single livestock production systems within province unit. Each point in Figure 7 represents the livestock meat production (in ten thousand tons) from 2000 to 2021 in the provinces (excluding those where grazing livestock constitute less than 5% of the whole livestock) covered by cold-season, warm-season, or year-round pastures, corresponding to the predicted livestock numbers for the respective years and pastures. Although due to provincial differences, the scatter plots appear as dispersed clusters, all external cross-validation results demonstrate good validation accuracy, indicating effective capture of spatial-temporal variations in livestock distribution. The validation accuracy (Pearson correlation coefficient) for cattle ranges from 0.88 to 0.90, while for sheep, it ranges from 0.92 to 0.94.

3.3 Analysis of temporal trends and mechanisms in the CLRD-GLPS

Based on the main regional distribution of grazing livestock in China, the overall time series analysis is divided into Xinjiang, the five provinces of the Qinghai-Tibet Plateau (Tibet, Qinghai, Gansu, Sichuan, and Yunnan), Inner Mongolia, and other provinces. We calculated the county-level average density for grazing ruminant livestock distribution datasets in the four regions and plotted the time series. For cattle, the density in the seasonal pastures of the QTP and Xinjiang shows a slow significant increase. In year-round pastures, cattle density in Inner Mongolia displays a significant increasing trend, while in Xinjiang it shows a non-significant fluctuating trend, and in other provinces, a significant decreasing trend is observed (Figure 8 (c)). When comparing these results to the statistical county-level cattle density trends, we found consistent conclusions (Figure 8 (d)-(f)).



https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Open Access Earth System Science Data Discussions

Figure 8: Time series of grazing cattle distribution datasets in Xinjiang, the Qinghai-Tibet Plateau, Inner Mongolia, and other provinces for warm, cold, and year-round pastures ((a)-(c)). Temporal trends in statistical county-level cattle density, analysed using the Theil-Sen estimator, are marked with black crosses for significant trends ($p < 0.05$) using the Mann-Kendall test ((d)-(f)).

For sheep, in the seasonal pastures of Xinjiang, some counties show a significant increase while others show a significant decrease (Figure 9 (d) and (e)), leading to an overall trend that shows only minor significant increases (Figure 9 (a) and (b)). In contrast, seasonal pastures in the QTP exhibit a significant overall decreasing trend (Figure 9 (a) and (b)). Comparison with statistical county-level data reveals that most counties in the QTP show significant reductions, with some counties experiencing declines of more than 20 heads per km² per year. In Inner Mongolia, year-round pastures show an overall upward trend, although it is not significant (Figure 9 (c)). For year-round pastures in Xinjiang and other provinces, there are slightly increasing trend changes (Figure 9 (c) and (f)).

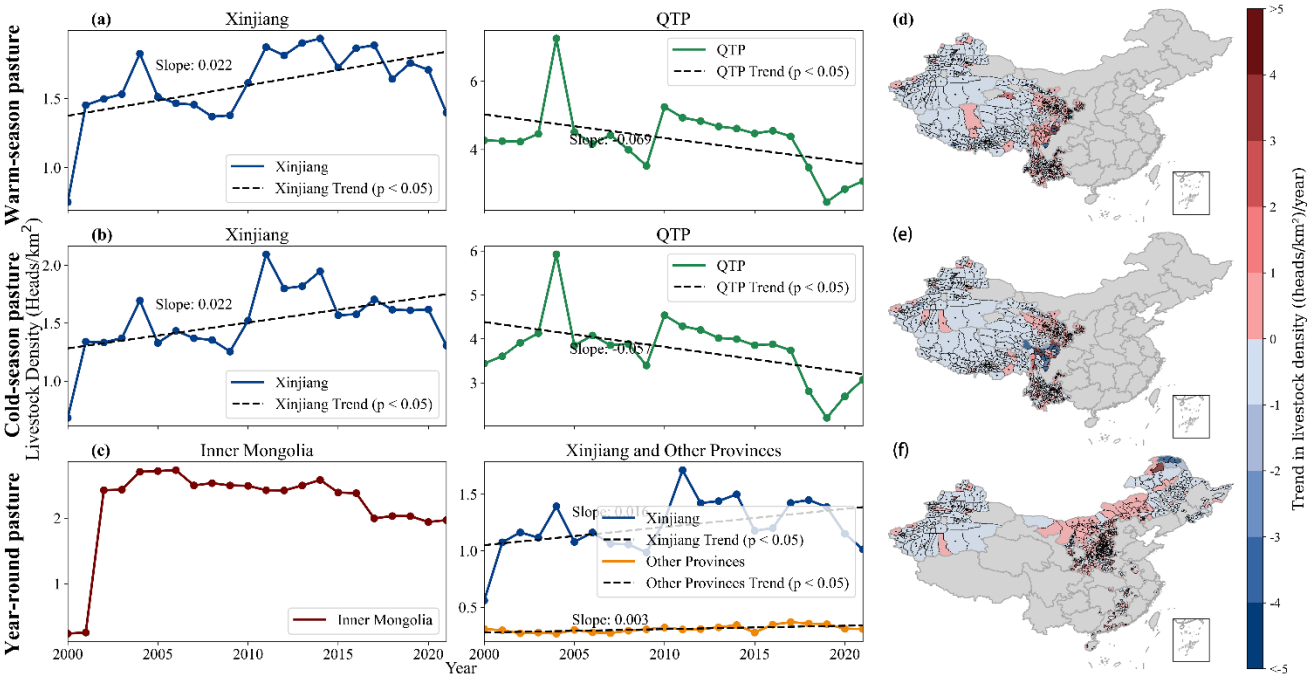


Figure 9: Similar to Figure 8, but for sheep.

The Theil-Sen estimator and Mann-Kendall test were used in each pixels of livestock distribution datasets. Our results indicated an increase trend for cattle in most of the area in all types of pastures during 2000-2021, with some trends showing statistical significance. For cattle, seasonal pastures with significant increases are primarily concentrated near the source of the Three-River Source National Park and some areas in Xinjiang (Figure A12). In year-round pastures, there is a significant increasing trend in cattle density near the Helan Mountains, with the grids showing the greatest change averaging an increase of over 10 heads per year. For sheep, in seasonal pastures, there was a significant decreasing trend from 2000 to 2021, especially in the

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

QTP (Figure A13). In the year-round pastures of Inner Mongolia, sheep density shows an upward trend, though the overall change is not statistically significant.

We explored the mechanisms underlying the spatial-temporal distribution in the CLRD-GLPS datasets using a structural equation modelling (SEM) approach. The results of the six distinct SEM models for cattle and sheep across different seasonal pasture types indicate optimal model fit, as evidenced by a CFI value between 0.9 and 1, an RMSEA value below 0.06, and a Chi-Square/df ratio below 3 with a p-value greater than 0.05 (Figure 10). The cattle density shows a significant increasing trend over the years across all seasonal pasture types, with a positive relationship between year and cattle density (Figure 10 (a) - (c)). The year variable positively influences cattle density trend indirectly through its effect on NDVI in year-round pastures. Additionally, precipitation significantly affects cattle density. However, its impact varies across different seasons. Increased precipitation during the growing season often leads to higher cattle density, while increased precipitation in winter typically results in a decrease in cattle density (Figure 10 (a) - (c)).

In contrast, the negative relationship between year and sheep density indicates a declining trend in sheep density, particularly in cold-season pastures, where the decline is significant (Figure 10 (d) - (f)). The primary factor influencing the trend in sheep density is also NDVI. Precipitation during the growing season and snow cover significantly impact sheep density positively by affecting NDVI. Additionally, factors such as temperature during the growing season and winter precipitation have a significant negative effect on sheep density, while land use changes positively affect sheep density in warm-season and year-round pastures.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

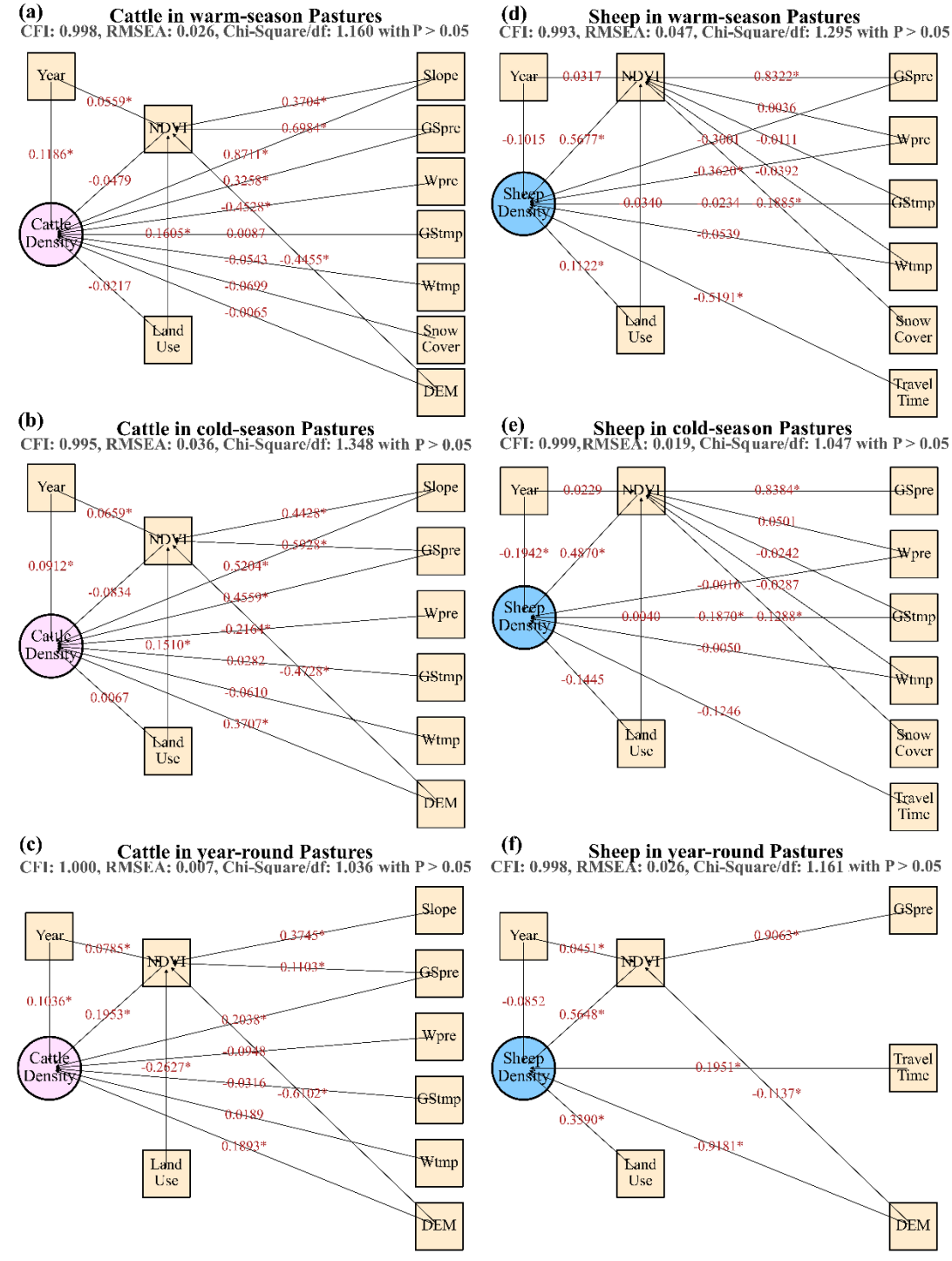


Figure 10: Path diagram of the structural equation models illustrating the factors affecting cattle density ((a)-(c)) and sheep density ((d)-(f)) in different seasonal pastures.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

4 Discussion

4.1 Improvements of livestock distribution modelling framework

In this study, we developed a new set of long-term livestock distribution data focusing on grazing livestock system in China (CLRD-GLPS). 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 livestock production systems from the total livestock count across all livestock production systems. Second, it expands the classification of grazing pastures to include seasonal types, specifically warm-season, cold-season, and year-round pastures across China, based on observed seasonal pastures and interpretable machine learning methods. Finally, using interpretable machine learning and structural equation modelling, we explore and explain the spatial-temporal distribution patterns of grazing ruminant livestock.

Livestock density varies significantly across different livestock production systems due to differences in feed sources, climate, and human activities (Thornton, 2002). In East Asia, including China, Mongolia, and Korea, grazing LPS account for as much as 39.54% of the total area of all LPS (Kruska et al., 2003). The most common methods used for predicting livestock distribution in previous studies involve modelling the census statistics of livestock numbers (density) as response data in multivariate linear regression or machine learning models, and then allocating the census statistics of livestock numbers within grid cells in administrative units (Gilbert et al., 2018; Robinson et al., 2014). According to the livestock census statistics from the yearbook in China, only the total livestock number of the entire county can be recorded. Apart from pastoral counties, where there are both grazing and captive livestock, most counties have a mix of both. Consequently, relying solely on census statistics fails to capture the spatial variations brought about by different LPS within an administrative unit. This limitation leads to overestimation in predicting the number of grazing livestock in previous studies focusing on grazing livestock distribution. To address this issue, our study segregates grazing livestock from the total livestock count using county-level carrying capacity data. This approach allows for better control of livestock numbers, aiming to maintain them at their actual levels as much as possible.

Seasonal distribution of grazing livestock in grazing LPS is a crucial issue. Influenced by policies, seasonal and year-round pastures are used concurrently within China's pasture-based grazing LPS. A crucial step prior to modelling livestock distribution in these grazing LPS is the delineation of seasonal and year-round pastures. By utilizing policy documents from the Chinese government and a relatively comprehensive sample of seasonal pastures, this study scientifically delineated seasonal and year-round pastures with the assistance of machine learning classification models. 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 and topography are the primary factors influencing distribution in both cold-season and warm-season pastures, whereas travel time, representing pastoral activities, becomes more crucial in year-round pastures (Figure A10); for sheep, topography features prominently as the main determinant

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

in seasonal pastures, while grass quality and travel time become the primary factors shaping distribution in year-round pastures (Figure A11).

4.2 Long-term trends and mechanisms in livestock distribution and their implications for livestock management

This study predicted trends in livestock density changes at a 1 km grid scale between 2000-2021. We found that while the interannual variation in average livestock density per province is small, at the grid scale, there is significant variation in livestock numbers, showing considerable movement of animals using grazing resources (Figure 8, Figure 9, Figure A12 and Figure A13). The highest changes show significant increases or decreases exceeding 20 heads per year, highlighting substantial spatial differences in livestock numbers changes. According to the SEM results, NDVI is a primary factor influencing changes in the spatial-temporal distribution of livestock, generally having a positive effect on livestock trends. This is likely because NDVI positively correlates with ground vegetation biomass, meaning that high NDVI values indicate abundant forage resources (Borowik et al., 2013). Additionally, precipitation during the growing season positively affects livestock distribution, likely because moderate increases in rainfall in arid regions support forage production and available water for livestock (Baumgard et al., 2012). In contrast, increased winter precipitation, especially in cold regions, may negatively impact livestock density due to cold stress effects on livestock (Godde et al., 2021; Young, 1983).

From a policy perspective, since the release of the *Opinions on Promoting the Development of Animal Husbandry* in 1999, China's animal husbandry industry has entered a new stage of development, characterized by significant improvements in livestock survival rates and production efficiency compared to previous years (Liu et al., 2016). Coupled with the rising demand for animal-source food in China, this explains the overall increase in livestock numbers. However, for grazing ruminant livestock, the increasing trend after 2000 is not significant in some areas, and there are even declining trends in certain periods. This is related to the rapid urbanization and economic development in China, as livestock farming has shifted from grazing to mixed farming and intensive feeding (Bai et al., 2018). Many labour forces have been diverted to the secondary and tertiary industries, and remaining herders have transitioned to mixed feeding with feed, leading to a decline in grazing ruminant livestock. Nevertheless, we still observe an increasing trend in livestock density in areas such as near the Three-River Source National Park, which is related to the fenced grazing policies implemented in the QTP (Sun et al., 2020). A series of measures including grazing bans and rotational grazing have restored grassland productivity, leading to an increase in the carrying capacity for livestock.

These findings underscore the critical importance of predicting long-term livestock distribution at the grid scale. Concurrently, existing livestock studies also emphasize the necessity of dynamic livestock distribution dataset, a gap that this dataset partially addresses. Over extended periods, this dataset enables investigation into the risks of grazing livestock by climate change (Thornton et al., 2021), evaluation of grazing potential and overgrazing within the constraints of grassland resources (Fetzel et al., 2017), and analysis of greenhouse gas emissions from ruminant livestock (Cheng et al., 2022b). Livestock managers can

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

utilize these insights to effectively manage livestock distribution, thereby mitigating climate change impacts, enhancing crop-livestock integration, and reducing environmental pollution—all vital for the sustainability of livestock production systems.

4.4 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, land use changes significantly affected livestock distribution. Although we have utilized the CLCD land cover dataset, which provides the best trade-off in terms of time (every five years) and spatial resolution (30 meters), having access to annual land cover datasets could potentially enhance the accuracy of this study's results on an interannual uncertainties. Lastly, in the process of segmenting grazing livestock from the overall livestock population, the livestock carrying capacity sample data used primarily from the QTP. Despite the QTP accounting for over 50% of the national grassland area (Li et al., 2021), obtaining sample data from more provinces (such as Inner Mongolia) could further reduce uncertainties in this study's findings.

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.14093125(Zhan et al., 2024). 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 head per pixel (or heads/km^2).

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.

Competing interests

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

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

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

Adler, J. and Parmryd, I.: Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient, Cytometry Part A, 77A, 733–742, https://doi.org/10.1002/cyto.a.20896, 2010.

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

Bai, Z., Ma, W., Ma, L., Velthof, G. L., Wei, Z., Havlík, P., Oenema, O., Lee, M. R. F., and Zhang, F.: China's livestock transition: Driving forces, impacts, and consequences, Science Advances, 4, eaar8534, https://doi.org/10.1126/sciadv.aar8534, 2018.

Baumgard, L. H., Rhoads, R. P., Rhoads, M. L., Gabler, N. K., Ross, J. W., Keating, A. F., Boddicker, R. L., Lenka, S., and Sejian, V.: Impact of Climate Change on Livestock Production, in: Environmental Stress and Amelioration in Livestock Production, edited by: Sejian, V., Naqvi, S. M. K., Ezeji, T., Lakritz, J., and Lal, R., Springer, Berlin, Heidelberg, 413–468, https://doi.org/10.1007/978-3-642-29205-7_15, 2012.

Bergstra, J. and Bengio, Y.: Random Search for Hyper-Parameter Optimization, n.d.

Borowik, T., Pettorelli, N., Sönnichsen, L., and Jędrzejewska, B.: Normalized difference vegetation index (NDVI) as a predictor of forage availability for ungulates in forest and field habitats, Eur J Wildl Res, 59, 675–682, https://doi.org/10.1007/s10344-013-0720-0, 2013.

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, 2022a.

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, 2022b.

Chernozhukov, V., Hansen, C., Kallus, N., Spindler, M., and Syrgkanis, V.: Applied Causal Inference Powered by ML and AI, https://doi.org/10.48550/arXiv.2403.02467, 4 March 2024.

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.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Du, Y., Ge, Y., Ren, Y., Fan, X., Pan, K., Lin, L., Wu, X., Min, Y., Meyerson, L. A., Heino, M., Chang, S. X., Liu, X., Mao, F., Yang, G., Peng, C., Qu, Z., Chang, J., and Didham, R. K.: A global strategy to mitigate the environmental impact of China's ruminant consumption boom, Nat Commun, 9, 4133, https://doi.org/10.1038/s41467-018-06381-0, 2018.

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.

Feuerriegel, S., Frauen, D., Melnychuk, V., Schweisthal, J., Hess, K., Curth, A., Bauer, S., Kilbertus, N., Kohane, I. S., and van der Schaar, M.: Causal machine learning for predicting treatment outcomes, Nat Med, 30, 958–968, https://doi.org/10.1038/s41591-024-02902-1, 2024.

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.

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.

Godde, C. M., Mason-D'Croz, D., Mayberry, D. E., Thornton, P. K., and Herrero, M.: Impacts of climate change on the livestock food supply chain; a review of the evidence, Global Food Security, 28, 100488, https://doi.org/10.1016/j.gfs.2020.100488, 2021.

Guyon, I. and Elisseeff, A.: An Introduction to Variable and Feature Selection, n.d.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., and Ray, S.: An Introduction to Structural Equation Modeling, in: Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook, edited by: Hair Jr., J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., and Ray, S., Springer International Publishing, Cham, 1–29, https://doi.org/10.1007/978-3-030-80519-7_1, 2021.

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.

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.

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, 2021.

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.

Liu, R., Xu, F., Liu, Y., Wang, J., and Yu, W.: Spatio-temporal characteristics of livestock and their effects on pollution in China based on geographic information system, Environ Sci Pollut Res, 23, 14183–14195, https://doi.org/10.1007/s11356-016-6576-6, 2016.

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.

Ministry of Agriculture and Rural Affairs China: China animal husbandry and veterinary yearbook 2000-2022, 2000.

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.

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.

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.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

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.

Robinson, T. P., William Wint, G. R., Conchedda, G., Van Boeckel, T. P., Ercoli, V., Palamara, E., Cinardi, G., D'Aietti, L., Hay, S. I., and Gilbert, M.: Mapping the global distribution of livestock, PLoS ONE, 9, https://doi.org/10.1371/journal.pone.0096084, 2014.

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.

Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau, Journal of the American Statistical Association, 1968.

Suhr, D.: The basics of structural equation modeling, Presented: Irvine, CA, SAS User Group of the Western Region of the United States (WUSS), 1–19, 2006.

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.

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.

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.

Yang, J. and Huang, X.: The 30m 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, 2021.

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.

Young, B.: Ruminant cold stress: Effect on production, Journal of Animal Science, 57, 1601–1607, 1983.

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.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

630 Zhan, N., Ye, T., Herrero, M., Peng, J., Liu, W., and Ma, H.: Long-term Ruminant Livestock Distribution Datasets in Grazing Livestock Production Systems in China from 2000 to 2021 (CLRD-CLPS) (version 1), https://doi.org/10.5281/zenodo.14093125, 2024.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Open Access Earth System Science Data Discussions

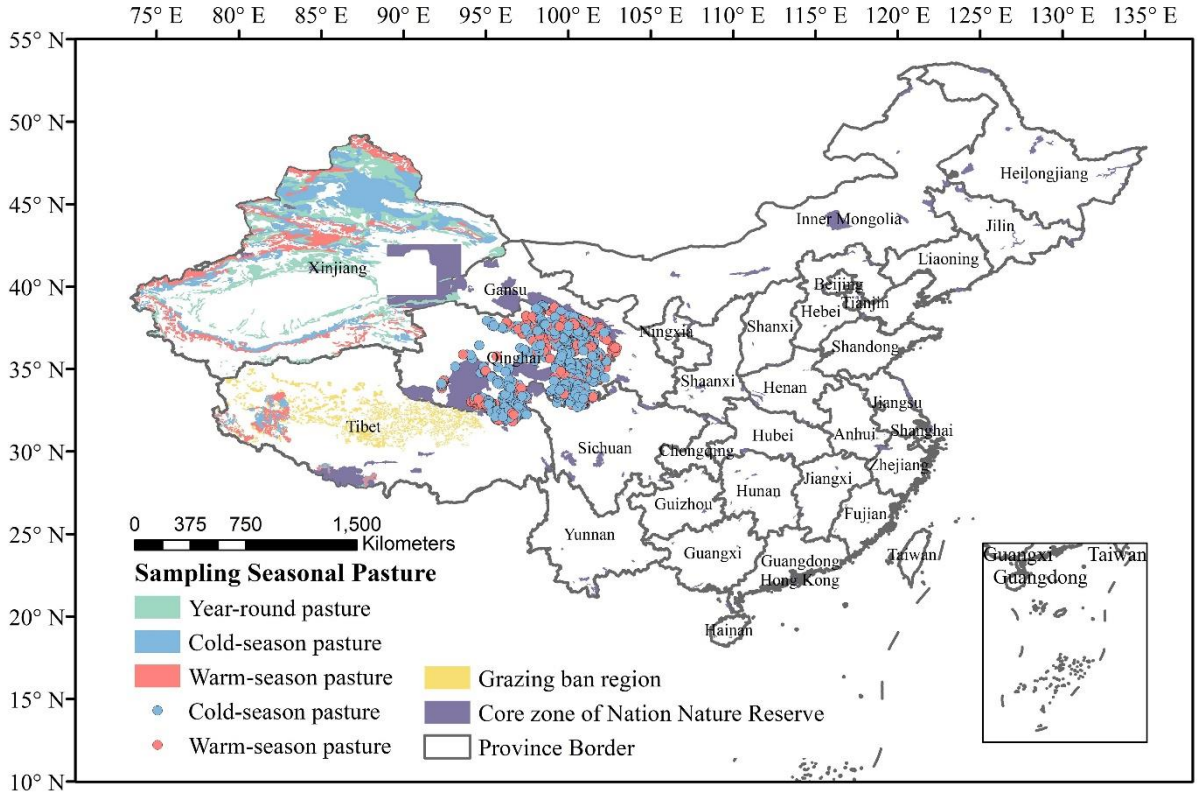
635 **Appendices**

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)
	Township-level Livestock Numbers	Livestock (cattle/sheep/pigs) year-end stock numbers for 2000-2021 at the township level	County-level Agriculture and Pastoral Bureaus
	Township-level Insured Livestock Numbers	Number of insured livestock (cattle/sheep) at township level in 2020	People's Insurance Company of China Property and Casualty, Tibet Branch
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)
	ET	Average total evapotranspiration for 2000-2021	MODIS Terra MOD16A2H

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Data Type	Variable Name	Variable Description	Data Source
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)



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

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Open Access Earth System Science Data Discussions

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				Grass-livestock balance carrying capacity (10,000 sheep units) 草畜平衡载畜量（万个绵羊单位）				
					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								

645 **Table A2: The sample basic information table on the grassland ecological protection subsidy and reward mechanism in Tibet.**

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

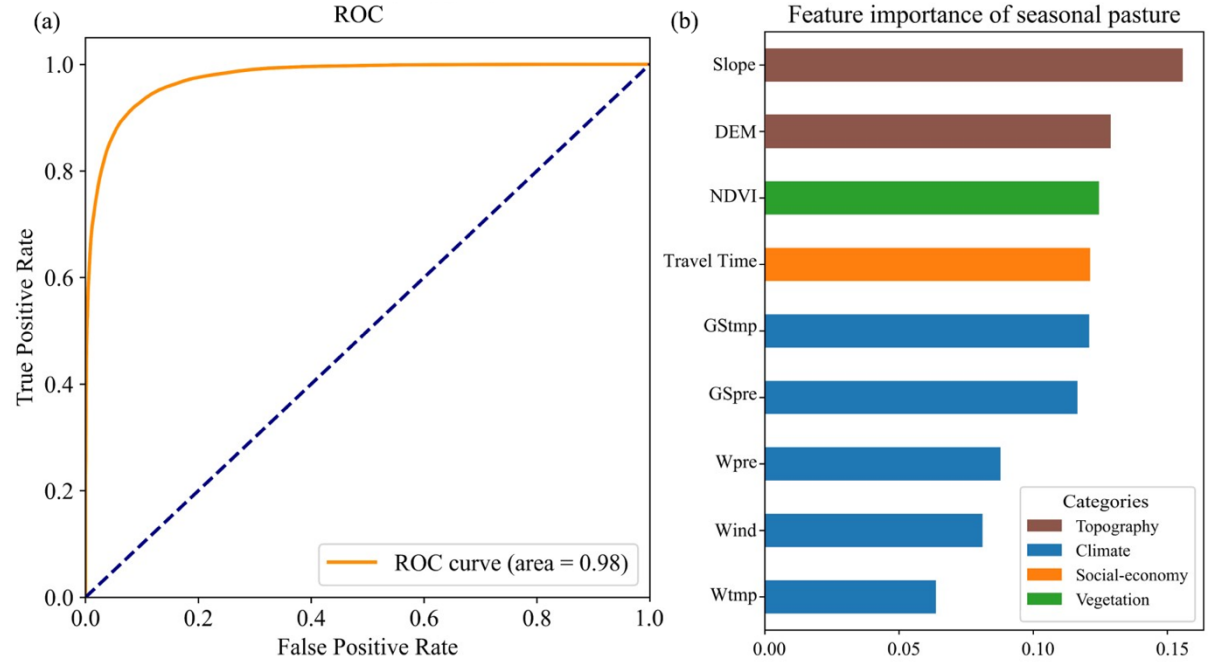


Figure A2: Results of ten-fold cross-validation (a) and feature importance (b) for the Random Forest Classification model.

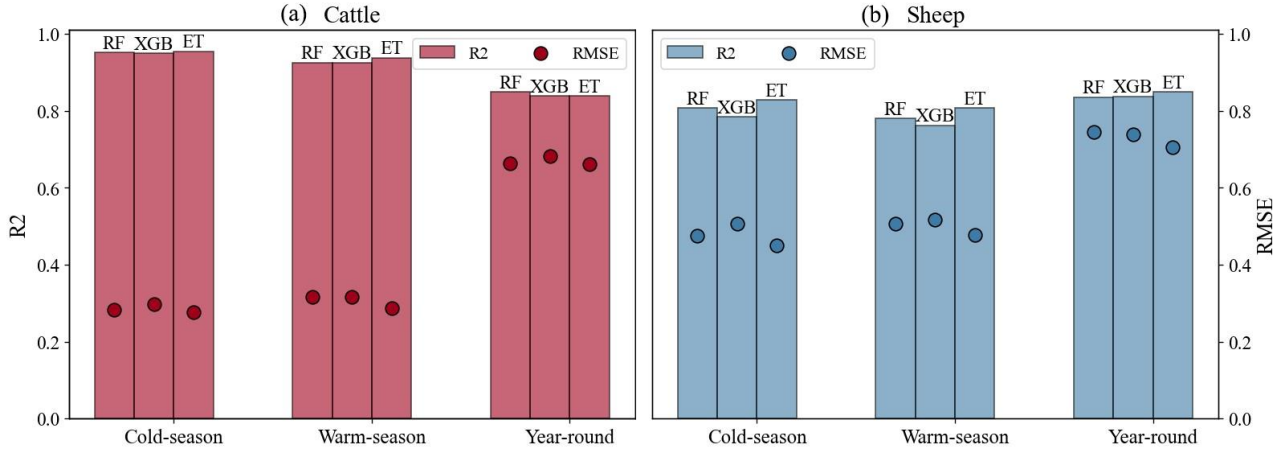


Figure A3: Results of ten-fold cross-validation for cattle (a) and sheep (b) distribution prediction models in cold-season, warm-season, and year-round pastures.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

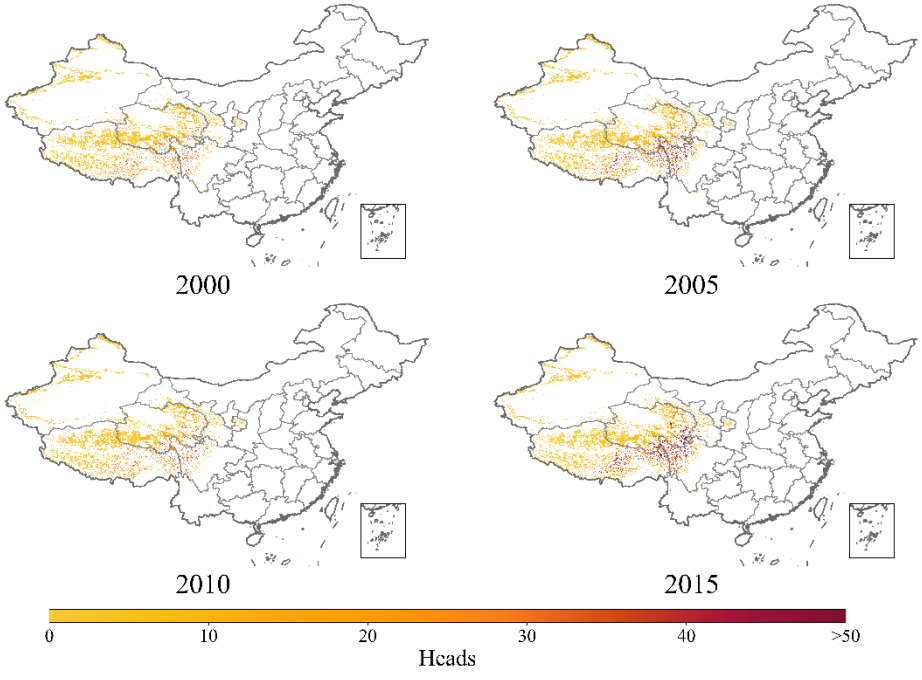


Figure A4: CLRD-GLPS prediction results for cattle in warm-season pastures.

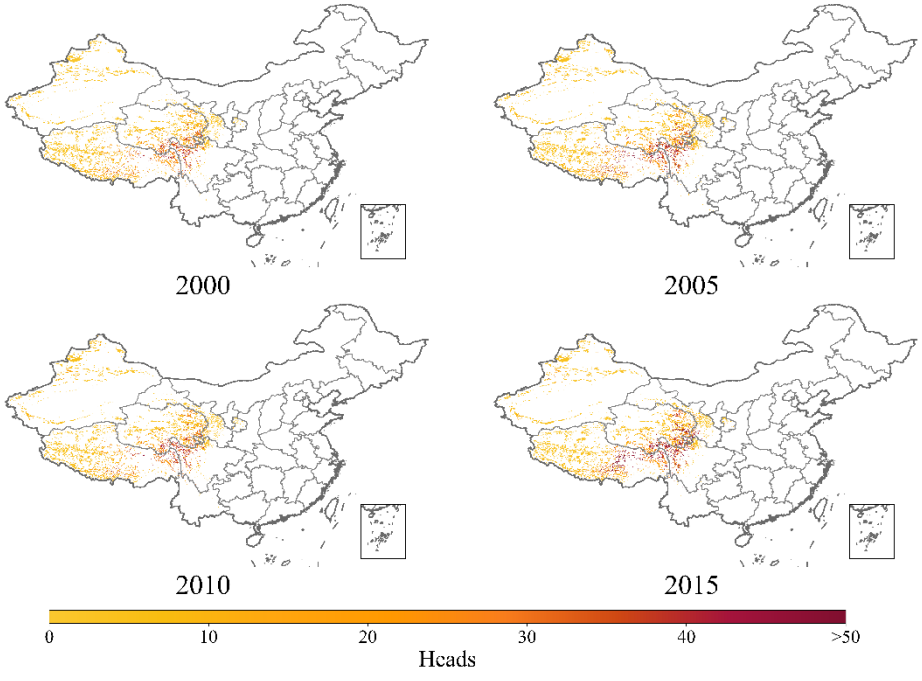


Figure A5: CLRD-GLPS prediction results for cattle in cold-season pastures.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

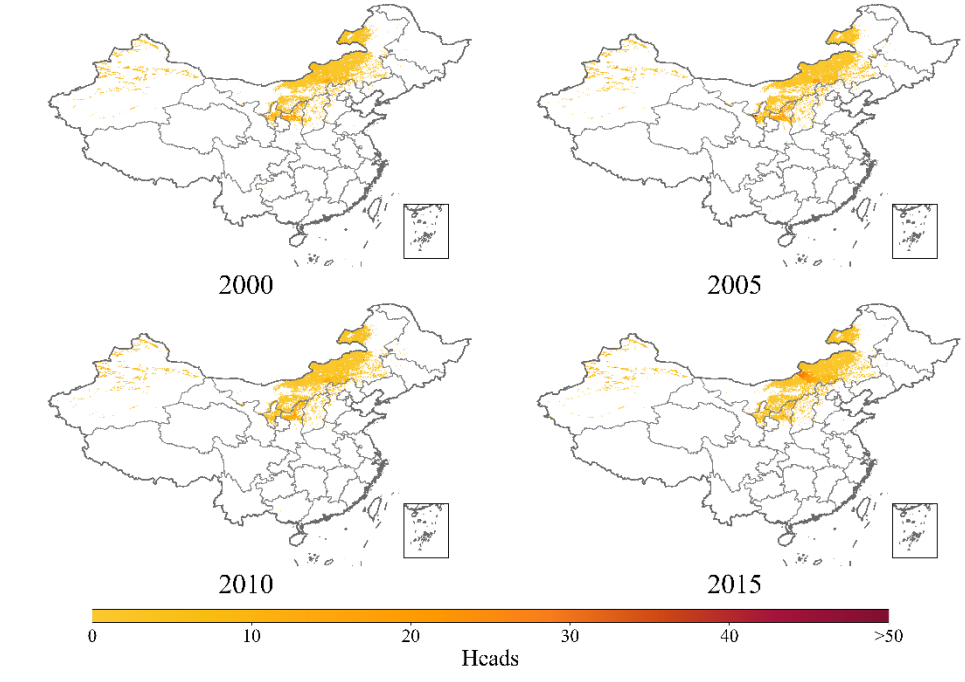


Figure A6: CLRD-GLPS prediction results for cattle in year-round pastures.

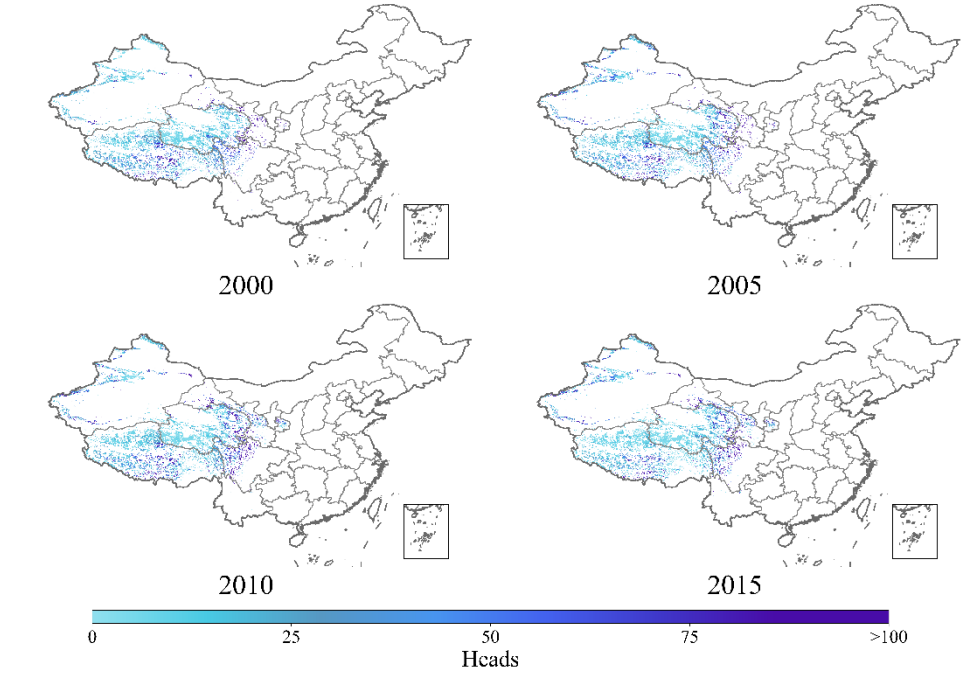
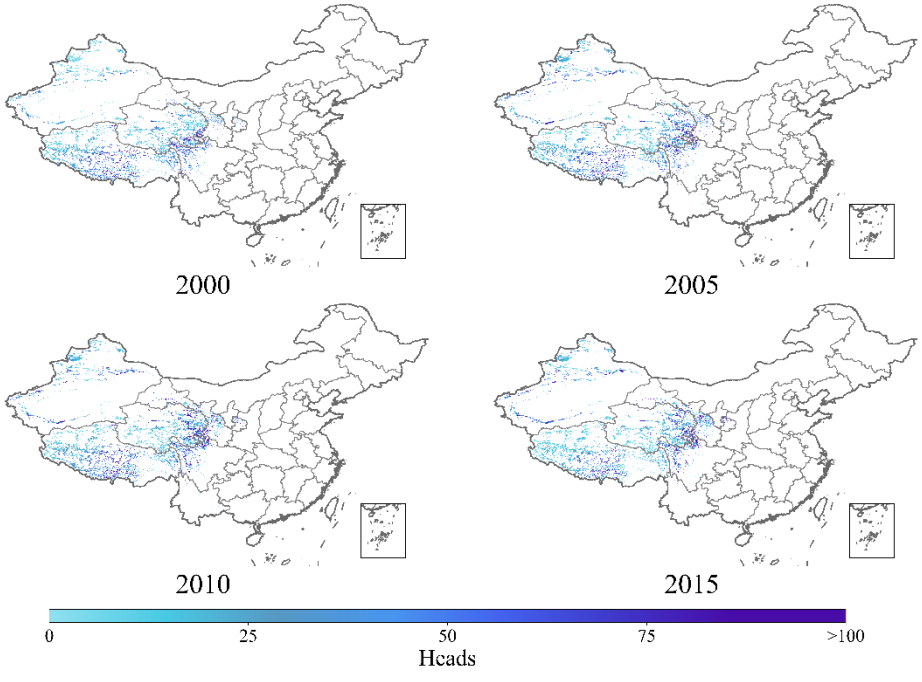


Figure A7: CLRD-GLPS prediction results for sheep in warm-season pastures.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

Open Access Earth System Science Data Discussions



660

Figure A8: CLRD-GLPS prediction results for sheep in cold-season pastures.

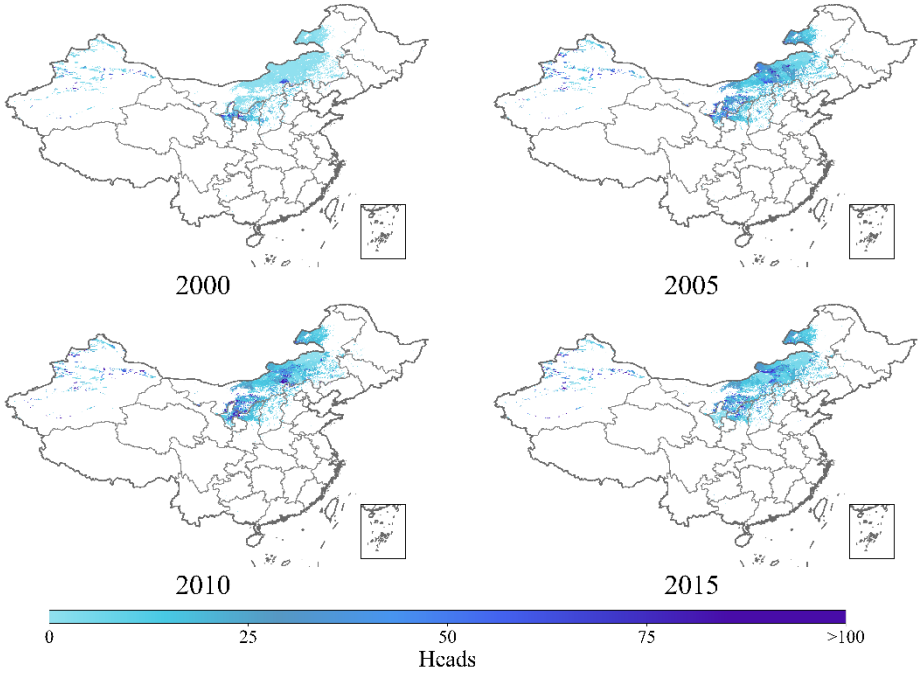
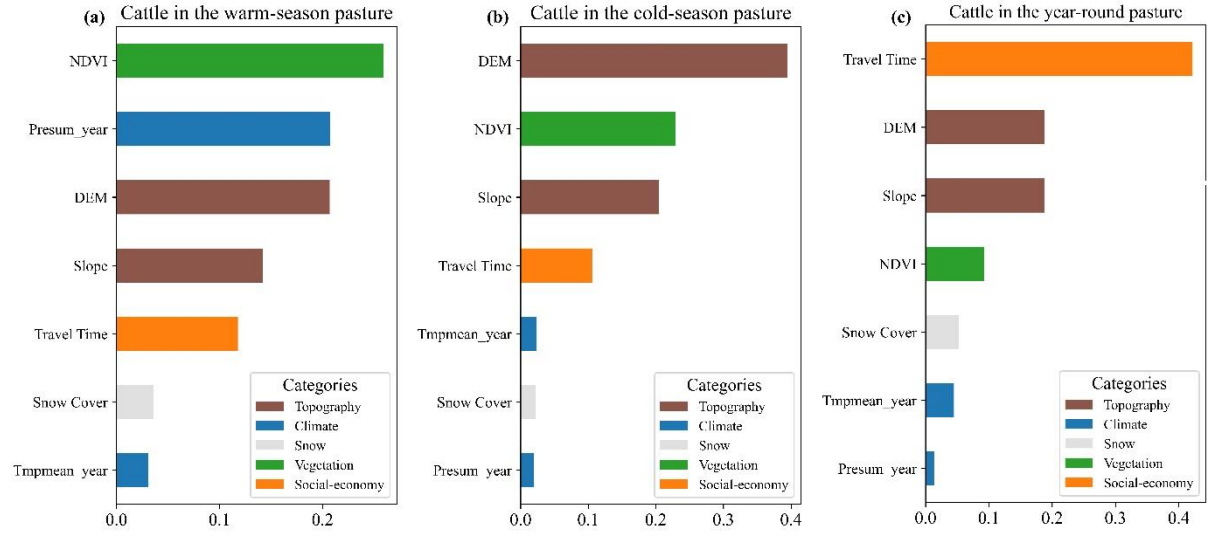


Figure A9: CLRD-GLPS prediction results for sheep in year-round pastures.

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024



665

Figure A10: Random forest feature importance of cattle distribution prediction in warm-season (a), cold-season (b), and year-round (c) pastures.

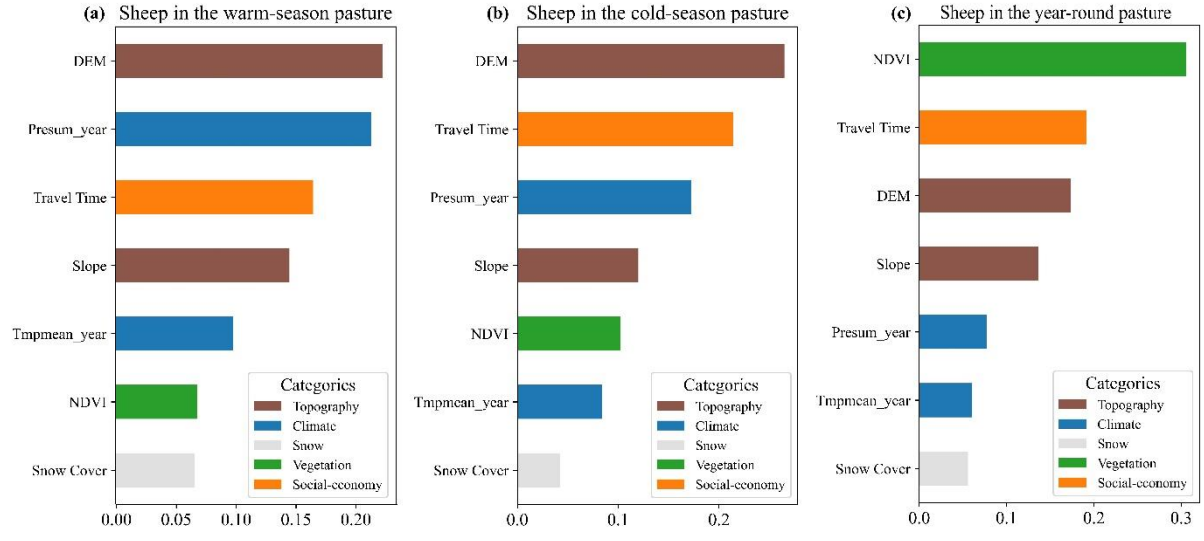


Figure A11: Similar to Figure A10, but for sheep.

670

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

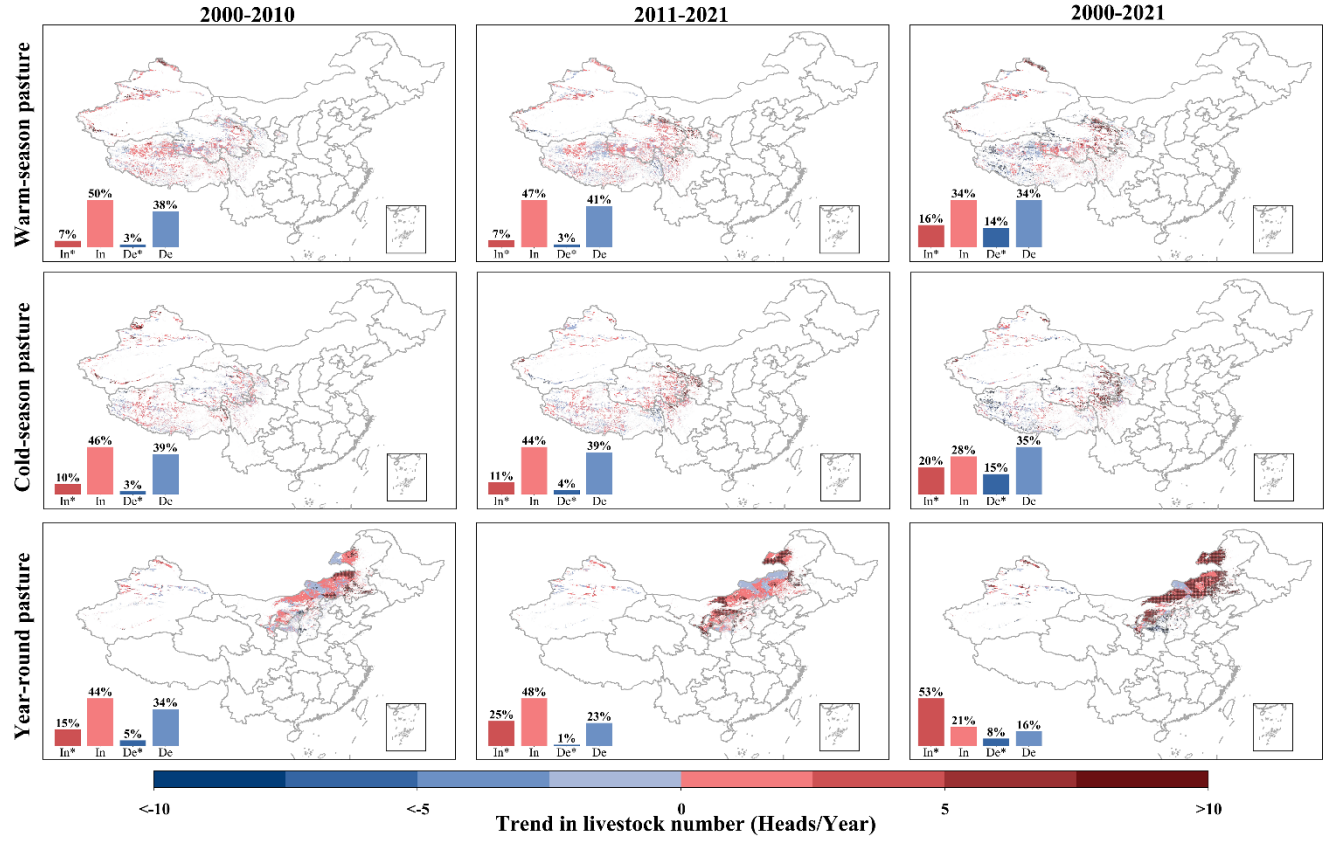


Figure A12: Temporal trend in grazing cattle spatial distribution in warm-season, cold-season, and year-round pastures tested by the Theil-Sen estimator. The black crosses in the panels indicate a significant trend with p < 0.05, tested by the Mann-Kendall test. The bar charts in the panels show the percentage of significant increases (In*), increases (In), significant decreases (De*), and decreases (De).

https://doi.org/10.5194/essd-2024-534
Preprint. Discussion started: 10 December 2024

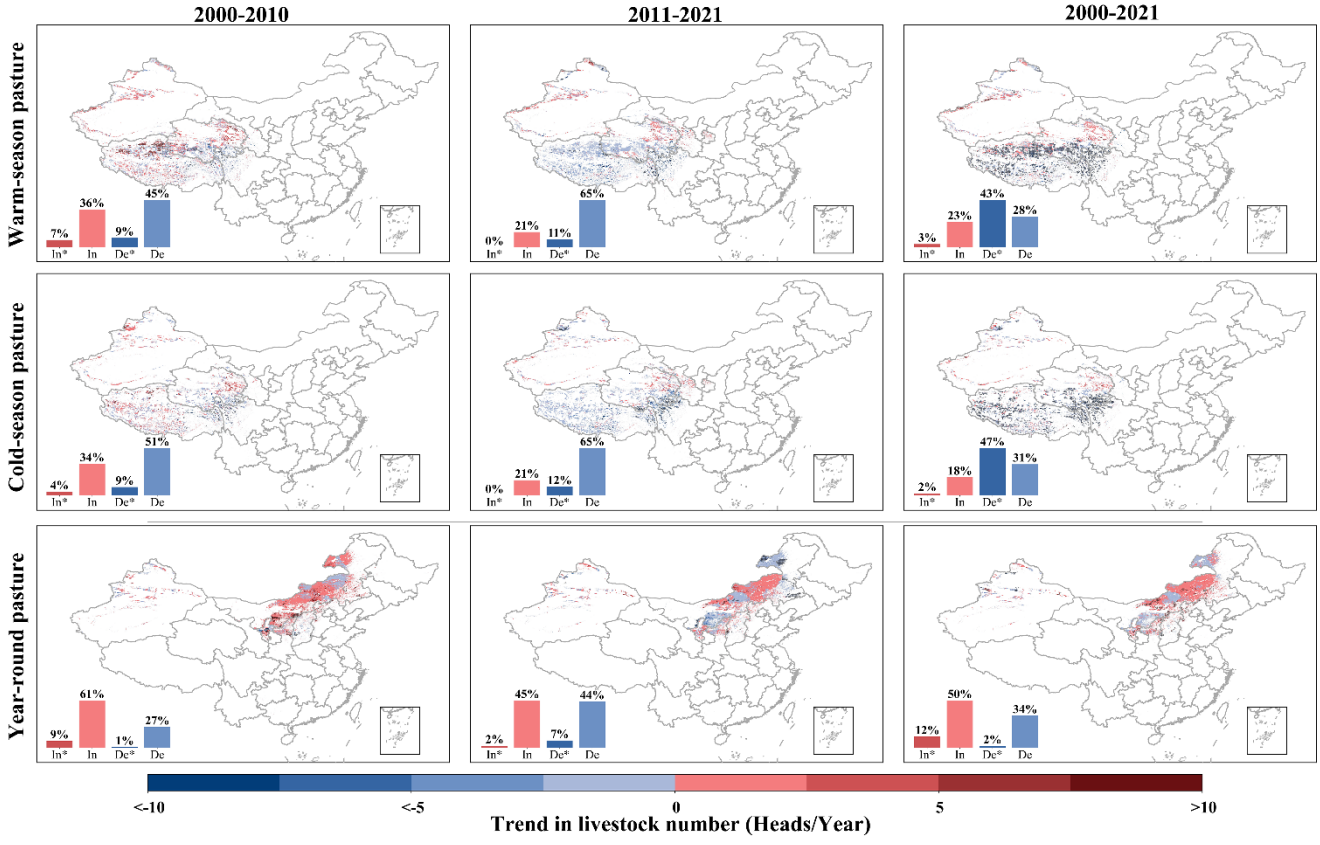


Figure A13: Similar to Figure A12, but for sheep.