



Article

High-Resolution Gridded Livestock Projection for Western China Based on Machine Learning

Xianghua Li ^{1,2,3}, Jinliang Hou ^{1,2,*} and Chunlin Huang ^{1,2}

- ¹ Key Laboratory of Remote Sensing of Gansu Province, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; lixianghua@lzb.ac.cn (X.L.); huangcl@lzb.ac.cn (C.H.)
 - ² Heihe Remote Sensing Experimental Research Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Zhangye 734000, China
 - ³ College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
- * Correspondence: jlhours@lzb.ac.cn

Abstract: Accurate high-resolution gridded livestock distribution data are of great significance for the rational utilization of grassland resources, environmental impact assessment, and the sustainable development of animal husbandry. Traditional livestock distribution data are collected at the administrative unit level, which does not provide a sufficiently detailed geographical description of livestock distribution. In this study, we proposed a scheme by integrating high-resolution gridded geographic data and livestock statistics through machine learning regression models to spatially disaggregate the livestock statistics data into 1 km × 1 km spatial resolution. Three machine learning models, including support vector machine (SVM), random forest (RF), and deep neural network (DNN), were constructed to represent the complex nonlinear relationship between various environmental factors (e.g., land use practice, topography, climate, and socioeconomic factors) and livestock density. By applying the proposed method, we generated a set of 1 km × 1 km spatial distribution maps of cattle and sheep for western China from 2000 to 2015 at five-year intervals. Our projected cattle and sheep distribution maps reveal the spatial heterogeneity structures and change trend of livestock distribution at the grid level from 2000 to 2015. Compared with the traditional census livestock density, the gridded livestock distribution based on DNN has the highest accuracy, with the determinant coefficient (R^2) of 0.75, root mean square error (RMSE) of 9.82 heads/km² for cattle, and the R^2 of 0.73, RMSE of 31.38 heads/km² for sheep. The accuracy of the RF is slightly lower than the DNN but higher than the SVM. The projection accuracy of the three machine learning models is superior to those of the published Gridded Livestock of the World (GLW) datasets. Consequently, deep learning has the potential to be an effective tool for high-resolution gridded livestock projection by combining geographic and census data.

Keywords: machine learning; livestock; spatialization; western China



Citation: Li, X.; Hou, J.; Huang, C. High-Resolution Gridded Livestock Projection for Western China Based on Machine Learning. *Remote Sens.* **2021**, *13*, 5038. <https://doi.org/10.3390/rs13245038>

Academic Editors: Stefano Pignatti, Jiali Shang, Liangxiu Han, Yanbo Huang and Wenjiang Huang

Received: 23 October 2021

Accepted: 8 December 2021

Published: 11 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

China's animal husbandry has unfolded rapidly in the past 40 years since the country's reform and opening up. It has become an unshakable leading economy in the agricultural and rural industries [1]. The development of animal husbandry requires a large number of grassland resources. At the same time, CH₄ and N₂O produced during livestock growth have become the main sources of agricultural greenhouse gas emissions [2,3]. Understanding the spatial distribution of livestock is of great significance for the effective utilization of grassland resources, protection of the ecological environment, and sustainable development of animal husbandry [4]. However, traditional livestock statistics are collected at the administrative unit level, mainly extracted from the "China Statistical Yearbook". Although census data can be regarded as the approximate "truth" within an administrative unit, it cannot provide enough detailed geographical descriptions of the spatial distribution

of livestock. In addition, census data cannot be shared and integrated with grid-based geographic data. The spatialization of census data refers to the projection of statistical values at the administrative level onto the regular grids of a specific scale [5,6]. Therefore, spatializing livestock statistical data and expressing the spatial distribution of livestock on a fine grid scale can be integrated with spatial ecological, social, and economic data on a grid scale to meet the needs of various spatial calculations, models, and analyses.

In 2007, the Food and Agriculture Organization of the United Nations released the world's first dataset of livestock spatialization data (named GLW1), which provided the first standardized global livestock density distribution map with a spatial resolution of 3 arc minutes (about 5 km at the equator); the time span of the dataset covers 2002 [7]. Robinson et al. (2014) further enhanced the GLW1 in terms of automated processing and data input; the global distribution maps of cattle, pigs, and chickens, and the partial distribution map of ducks with a resolution of 1 km in 2006 were obtained (namely GLW2) [8]. Nicolas et al. (2016) used the random forest and multi-layer linear regression to allocate the census data of African cattle and Asian chickens on the administrative unit scale to the grid [9]. The results show that the random forest always has better accuracy than the traditional stratified regression method. Consequently, Gilbert et al. (2018) used random forest regression instead of multi-layer linear regression to improve GLW1 and GLW2, the grid distribution maps of global cattle, buffalo, horse, sheep, goat, pig, chicken, duck, with a spatial resolution of 0.083° (about 10 km at the equator) in 2010, were obtained (namely GLW3) [10]. In addition to these global studies, several intercontinental or national, state/provincial, and other local-scale studies have also been published. For example, Neumann et al. (2009) disaggregated the livestock census data to the grid level in Europe using an expert-based and empirical statistical method [11]. Prosser et al. (2011) used an information-theoretic approach to produce the population distribution maps for chicken, ducks, and geese in the Chinese mainland at 1 km resolution [12]. Van Boeckel et al. (2011) constructed a stratified regression model between domestic duck densities and a set of agro-ecological explanatory variables to disaggregate domestic duck statistics to 1 km grid in Monsoon Asia [13]. Qiao et al. (2017) used the grid processing technology based on Clark negative exponential function model to analyze the spatial distribution pattern of livestock activities density in Xinyuan county [14]. In general, the existing livestock grid products have some defects in China, mainly affected by the spatial and temporal scales of livestock census data. For example, GLW1, GLW2, and GLW3 are produced mostly based on China's provincial and sub-provincial livestock statistics. GLW1 uses China's livestock statistics from the 1990s; GLW2 and GLW3 used livestock statistics in China in 2001, much earlier than the data product time.

In addition, the method used in constructing the relationship between livestock distribution density and environmental variables is also a key factor in obtaining high-precision livestock spatialization data. Multi-layer linear regression is one of the most basic and widely used regression algorithms in the research of livestock spatialization [6,7]. The stepwise regression algorithm based on Akaike Information Criterion (AIC) also has some applications [11]. Advanced machine learning technology, such as RF, provides new opportunities for developing livestock spatialization models [8]. GLW3, the latest version of gridded livestock of the world, which uses the RF regression method instead of the multi-layer linear regression method, has been proved to have much better accuracy. However, the machine learning methods currently used for livestock spatialization mainly focus on traditional machine learning methods. Compared with traditional machine learning, the more advanced deep learning methods have excellent feature learning capabilities and strong generalization capabilities and are more suitable for processing geographic data and complex system modeling [15]. Therefore, exploring the application of advanced deep learning technology in livestock spatialization and analyzing its application potential is a critical new task.

In general, the problems existing in the current researches on livestock spatialization are as follows: (1) the time of livestock statistical data is relatively backward, and its spatial

scale is rough, (2) the methods are traditional and lack the introduction of new methods, such as deep learning, which has made some attempts in population spatialization and achieved good results, and (3) there are currently three versions of livestock grid datasets (GLW1, GLW2, GLW3) corresponding to the livestock grid data in 2002, 2006, and 2010, respectively, however, the three versions of GLW differ with respect to the input data type, the predictor covariates, and modelling methods. It is discouraged their use for time-series analysis. Therefore, this study collects livestock statistical data with finer temporal and spatial resolution, discusses the performance of deep learning methods in livestock spatialization research, and aims to obtain high-precision, long-term series livestock grid data. Western China, with its vast territory, diverse climatic conditions, and rich grassland resources, is an essential base for developing animal husbandry in China. Therefore, this study selected six provinces in western China, including Shaanxi, Gansu, Ningxia, Xinjiang, Qinghai, and Tibet, as the study area. Thirteen environmental factors extracted from land cover, terrain, climate, and socioeconomic data are selected as prediction factors. In this study, a support vector machine, random forest, and deep neural network were used to develop livestock spatialization models to spatially disaggregate the livestock statistics data into $1 \text{ km} \times 1 \text{ km}$ spatial resolution from 2000 to 2015 at five-year intervals. Support vector machine, random forest, and deep neural network belong to shallow machine learning, ensemble learning, and deep learning, respectively. The shallow learning model can be regarded as the model with only one, two, or no hidden layers in the structure and has good nonlinear mapping capabilities in general. Compared with shallow learning, deep learning allows computational models composed of more processing layers to learn representations of data with multiple levels of abstraction. It has turned out to be very good at discovering intricate structures in high-dimensional data [16]. The following sections of the study are organized as follows. Section 2 provides an overview of the study area and the data used. Section 3 describes in detail the livestock spatialization scheme. We analyze and discuss the results of these analyses in Sections 4 and 5. Finally, in Section 6, we summarize our conclusions.

2. Study Area and Data

2.1. Study Area

Six provinces in western China, including Shaanxi Province, Gansu Province, Ningxia Hui Autonomous Region, Xinjiang Uygur Autonomous Region, Qinghai Province, and Tibet Autonomous Region, are selected as the study area (Figure 1). The geographical location is between $73^{\circ}30' \text{ E} \sim 111^{\circ}7' \text{ E}$ and $26^{\circ}50' \text{ N} \sim 49^{\circ}10' \text{ N}$, with an area of $4,308,500 \text{ km}^2$. The overall topography characteristics of the study area are high in the south and low in the north. The study area is the leading distribution area of natural pastures and an essential base for the development of animal husbandry in China. The grassland area accounts for 46.15% of the total area of the study area and is the primary land cover type. The proportions of the remaining classes in descending order are 37.60% of unused land, 6.65% of forest land, 5.67% of cultivated land, 3.42% of water area, and 0.50% of construction land.

2.2. Data and Preprocessing

The factors affecting the distribution of livestock are complex and changeable and can be divided into natural environmental and socioeconomic factors according to their attributes. A large number of studies have used environmental factors to predict the spatial distribution of livestock. For example, some researchers have pointed out that livestock grazing distribution is driven by spatial patterns of abiotic and biotic resources with primary abiotic factors, including topography and distance to water [17–19]. The International Livestock Research Institute (ILRI) used geospatial datasets on human population density, land cover, length of growing period (LGP), temperature and irrigation to estimate the distribution of livestock production systems in the developing world [20]. The Gridded Livestock of the World dataset uses environmental factors from anthropogenic, topography, climatic, etc., to spatialize the global

livestock [7,8,10]. Regarding the existing related researches, based on the principle of being able to be quantified by space, this study selected 13 environmental factors from four aspects: land use practice, topography, climate, and socioeconomic. The data used in this study include the high-resolution gridded geographic data and livestock statistics (Table 1). The geographic data is used to extract 13 environmental factors affecting livestock distribution and four suitability mask maps. Note that these variables may not be comprehensive, but they are representative and reflect different heterogeneous aspects related to livestock distribution. The county-level livestock statistics are regarded as the approximate “truth” within an administrative unit.

Table 1. The geographic data and livestock statistics used for generating gridded livestock distribution.

Type	Variables	Time ¹	Source	Initial Data Declaration
Environmental factors	Grassland coverage	2000–2015	Chinese Academy of Sciences Resource and Environmental Science Data Center (http://www.resdc.cn , accessed on 10 March 2021)	100 m
	Arable land coverage	2000–2015		100 m
	Forest land coverage	2000–2015		100 m
	Desert coverage	2000–2015		100 m
	NDVI	2000–2015	Geospatial Data Cloud (http://www.gscloud.cn , accessed on 19 March 2021)	500 m
	Elevation	2000	National Tibetan Plateau Data Center (http://data.tpdc.ac.cn , accessed on 21 December 2020)	1000 m
	Slope	2000		1000 m
	Daytime surface temperature	2000–2015	Geospatial Data Cloud (http://www.gscloud.cn , accessed on 19 March 2021)	1000 m
	Precipitation	2000–2015	National Tibetan Plateau Data Center (http://data.tpdc.ac.cn , accessed on 25 March 2021)	1000 m
	Distance to river	2000–2015	Open Street Map (https://www.openstreetmap.org , accessed on 7 April 2021)	shapefile
Travel time to major cities	2000, 2015	Nelson A. D. et al., D. J. Weiss et al.	1000 m	
Population grid data	2000–2015	Resource and Environment Science and Data Center (https://www.resdc.cn , accessed on 10 April 2021)	1000 m	
GDP grid data	2000–2015		1000 m	
Unsuitable areas	Permanent water	2000–2015	Chinese Academy of Sciences Resource and Environmental Science Data Center (http://www.resdc.cn , accessed on 10 March 2021)	100 m
	Urban cores	2000–2015	Resource and Environment Science and Data Center (https://www.resdc.cn , accessed on 10 April 2021)	1000 m
	Protected areas	2000–2015	World Database of Protected Areas (WDPA) (https://www.protectedplanet.net/country/CHN , accessed on 14 April 2021)	shapefile
	Pasture suitability	2005	United Nations Food and Agriculture Organization (https://data.apps.fao.org/map/catalog , accessed on 15 April 2021)	10,000 m
Census	Stock data of cattle	2000–2015	China Statistical Yearbooks (http://www.stats.gov.cn/tjsj/pcsj/ , accessed on 27 November 2020)	County
	Stock data of sheep	2000–2015		County

¹ Five-year intervals.

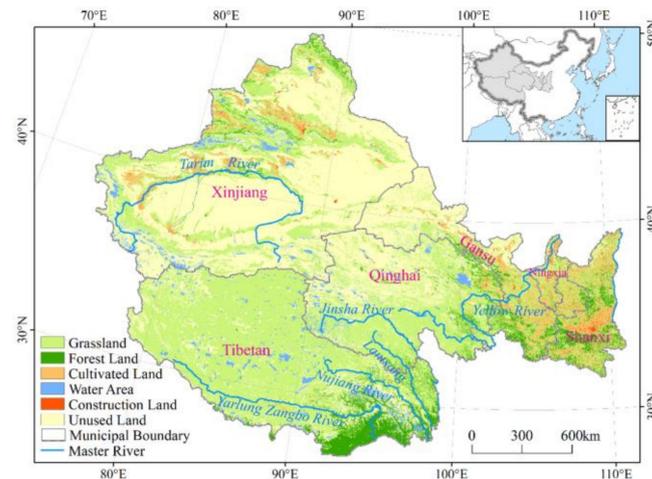


Figure 1. The location and land use and land cover of the study area.

2.2.1. The Gridded Geographic Data

The basic geographic data used in this study include China's land use and land cover data set (CNLULC) with 100 m spatial resolution from 2010 to 2015 [21,22], monthly normalized difference vegetation index (NDVI) composite product (MODND1M) with 500 m spatial resolution, digital elevation model [23], the monthly composite product of surface temperature (MODLT1M) and precipitation dataset with 1 km spatial resolution [24], open street map (OSM), city accessibility data [25,26], population [27] and gross domestic product (GDP) [28] with 1 km spatial resolution, world database of protected areas, and pasture suitability map [29]. Thirteen environmental factors from these essential geographic data, including grassland coverage, arable land coverage, forest land coverage, desert coverage, NDVI, elevation, slope, daytime surface temperature, precipitation, distance to river, travel time to major cities, population grid data, and GDP grid data were extracted. The coverage rate of grassland, arable land, forest land and desert refers to the percentage of grassland, arable land, forest land and desert per square kilometer extracted from CNLULC. The annual NDVI was calculated by synthesizing the maximum value of MODND1M. Similarly, the annual daytime surface temperature is calculated by "mean composition" using the MODLT1M. Annual precipitation is the sum of monthly precipitation. The distance to the river refers to the nearest Euclidean distance of each pixel to the nearby river. Travel time to major cities refers to the land travel time to the closest major city from each square kilometer of the pixel. In addition, for subsequent calculation and analysis, we unified spatial resolution (1 km) and coordinate system (Krasovsky ellipsoid coordinates and Albers projection, central longitude 105° E, and two standard latitude lines 25° N and 47° N, respectively). The above calculation and processing are all implemented based on Python's GDAL geographic data processing software package.

Suitability masking is an essential issue to consider during the modeling process. Firstly, the census livestock numbers used as the dependent variable in regression models are adjusted by eliminating areas that are very unsuitable for livestock distribution. Secondly, set the livestock density to 0 for areas that are very unsuitable areas for livestock survival [7]. In this study, we adopted a relatively conservative suitability mask way that only excludes permanent water (pixels covered by >50 percent of water), urban cores (areas where human population densities exceed 10,000 people km⁻²) [10], protected areas (areas by stringent conservation measures and tight regulation of human activity), and unsuitable site for pasture (areas with a pasture suitability index of 0). The remaining area after suitable mask in 2000, 2005, 2010, and 2015 accounted for 70.09%, 74.03%, 74.98%, and 74.51% of the total area of the study area, respectively.

2.2.2. Livestock Statistics

The livestock statistics we use are year-end stock data of cattle and sheep at the county level in six provinces, including Shaanxi, Gansu, Ningxia, Xinjiang, Qinghai, and Tibet in 2000, 2005, 2010, and 2015. These data are derived from the China Statistical Yearbooks (<http://www.stats.gov.cn/tjsj/pcsj/>, accessed on 27 November 2020) of 2001, 2006, 2011, and 2016, which generated a total of 1226 independent samples. We used 70% of the samples for model training and the remaining 30% for test sets to verify the model performance.

3. Methodology

3.1. Machine Learning Methods

Three machine learning methods, including support vector machine, random forest, and deep neural network, were selected to construct the livestock spatialization models. It was necessary to optimize the parameters of the machine learning models to improve the accuracy of models. Considering that the random search parameter is time-consuming, we used a simple trial and error method in the experiment to optimize the parameters in the machine learning model. We preset the possible value range of the parameter, then set the parameters in turn according to specific step size, and obtain the optimal parameters according to the model's performance. The following briefly describes these three machine learning methods and the parameter settings in this experiment.

3.1.1. Support Vector Machine

Support vector machine (SVM) is a machine learning method proposed by Vapnik [30]. It is divided into support vector classification (SVC) and support vector regression (SVR), which solve the classification and regression problems separately. The epsilon-support vector regression (ϵ -SVR) is used in this study. The purpose of ϵ -SVR is to find a regression equation that can fit all sample points and minimize the total variance between the sample points and the confidence interval of the regression equation [31]. Where C ($C > 0$) is the penalty factor that tunes the trade-off between the model generalization and error tolerance, and ϵ ($\epsilon > 0$) demonstrates the width of the insensitivity zone [32]. In practical application, when the C value is too large, the generalization ability of the SVR model will be reduced, which may lead to overfitting. ϵ -SVR uses kernel function to map the nonlinear problem in low dimensional feature space to the linear problem in higher-dimensional feature space. In this study, the most widely used radial basis kernel (RBF) function is selected as it is suitable for processing different samples and various dimensional problems and has nonlinear mapping capability, with C of 10 and ϵ of 0.01.

3.1.2. Random Forest

The random forest (RF) regression algorithm, first proposed by Breiman [33], is an integrated learning and data mining method composed of multiple decision trees. The essence of random forest regression is the collection of multiple independent regression decision trees. The construction process of the RF regression model is as follows. First, N training sets are generated using bootstrapping random sampling method, and a decision tree is generated based on the random subset of the predictor variables. Secondly, the average value of the prediction results of N decision trees is taken as the prediction result of RF. There are two crucial custom parameters in establishing the RF regression model, namely the number of decision trees (i.e., the number of training sets, N) in the random forest algorithm (it is also generally defined as `n_estimators`) and the number of features used when building the tree (we define it as `max_features`) [34]. Theoretically, the larger the value of `n_estimators`, the better the algorithm performance. However, the model error usually remains stable after a significant reduction with the increase of the number of decision trees. Therefore, the value of `n_estimators` usually takes the number of decision trees when the RF model error reaches stability in practical application. `Max_features` represent the number of randomly selected features. The smaller the `max_features`, the faster the

variance decreases, but the deviation increases. Generally, the value of `max_features` is set around one-third of the number of predictor variables [35]. In this study, `n_estimators` is set to 500, and `max_features` is set to 4.

3.1.3. Deep Neural Network

Deep neural network (DNN) is strictly defined as a fully connected deep neural network in this study. The calculation process of DNN can be divided into two stages: forward propagation and backpropagation. In forward propagation, the DNN randomly initializes the parameters of the neural network. The value of each hidden layer neuron is the weighted sum of the activation value of the previous layer neuron and the weight of the current layer, and then it is activated by a nonlinear activation function. During the backpropagation stage, the DNN quantifies the difference between the calculated output of the training samples and the actual value through a loss function. When the difference is greater than the given threshold, DNN performs backpropagation, gradually adjusts weights and bias of the network until the loss is less than the threshold. Finally, the final training results are output [36,37]. The initial setup of the DNN in this study consists of three fully connected neural networks, each with 64 neurons. The DNN adopted the rectified linear unit (ReLU) activation function, an Adam optimizer, a learning efficiency of 0.01, and a discard ratio of 0.5, and the models were trained 2000 times.

3.2. Livestock Density Estimation Models

Firstly, the livestock census dataset and land cover, terrain, climate, and socioeconomic dataset were preprocessed, including suitability masking, unified coordinate system, and spatial resolution. Then, 13 environmental factors were extracted from the pre-processed land cover, topography, climate, and socioeconomic databases, with a spatial resolution of 1 km × 1 km. Perform regional statistics and average values to obtain the mean values of environmental factors in counties as independent variables for model construction. For the model dependent variable, we calculated the livestock density of each county, then converted it to $\log_{10}(n + 1)$ values to normalize the distribution of the dependent variable. Based on the above independent variables and dependent variables, we obtained a total of 1226 samples (counties), of which 70% were used to train the model and 30% were used to verify the model's accuracy. The SVM, RF, and DNN based regression models are constructed on the county scale. The basic hypothesis of this study is that there is a robust statistical relationship between livestock density and these environmental predictors at the county-level scale, which in turn could be used to disaggregate livestock census data spatially [11,12]. We apply the trained models to the grid level to obtain livestock density data with a spatial resolution of 1 km based on this assumption. To maintain better consistency between the number of livestock predicted by the developed machine learning models and the census data, we further fine-tuned the estimated results. Finally, the livestock density data were compared with all county-level livestock statistics data to verify the accuracy of the livestock spatialization. The overall process is shown in Figure 2.

3.2.1. Livestock Density Estimation

First, we established SVM, RF, and DNN models at the county level. Thirteen environmental factors of grassland coverage, arable land coverage, forest land coverage, desert coverage, NDVI, elevation, slope, daytime surface temperature, precipitation, distance to river, travel time to major cities, population grid data, and GDP grid data are aggregated to the county level. With the above 13 county-level average values are used as independent predictor variables and the logarithmic value of the county-level livestock census with base ten as the dependent variable, three different livestock spatialization models are trained based on SVM, RF, and DNN, respectively. Then, we apply the trained model to the 1 km grid scale to obtain the livestock density distribution with a spatial resolution of 1 km. It should be noted that the SVR and RF based livestock spatialization models are constructed by invoking the relevant functions in the scikit-learn machine learning library, while the

DNN based livestock spatialization model is developed by using Keras deep learning library based on the TensorFlow platform.

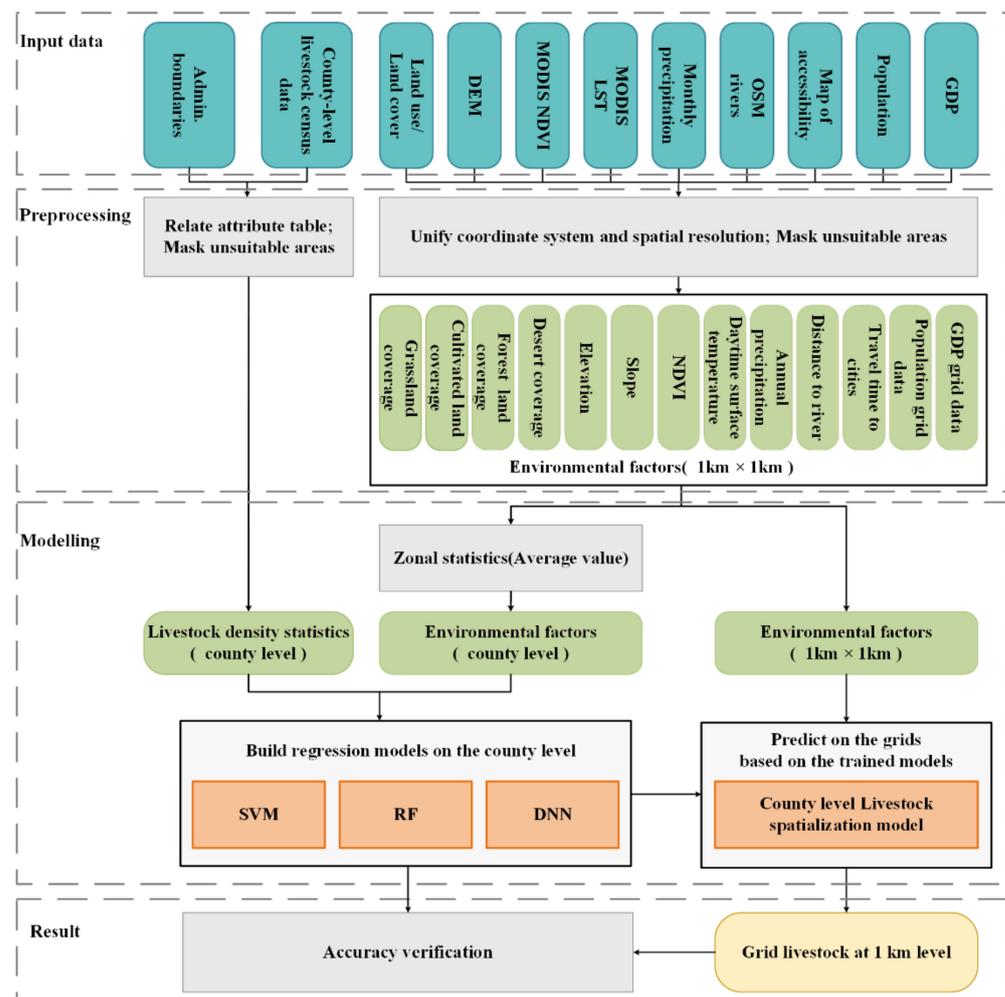


Figure 2. Flowchart of the livestock spatialization process.

3.2.2. Livestock Density Adjustment

The potential assumption is that the relationship between the environmental factors and livestock density is identical at the county and grid scales. However, there are obvious differences in the distribution characteristics of environmental factor values at the two scales, which will inevitably lead to errors when the models established at the county scale are directly applied to the grid scale. Since the model used to simulate the gridded livestock is established based on the average factor value and the county-level livestock density, the actual livestock density distribution needs to be controlled by the total livestock of each county-level administrative region [8,10,38]. The specific method calculates the difference between the number of livestock estimated by the model and the census data on the municipal scale, obtains the corresponding adjustment coefficient, and uses the adjustment coefficient to redistribute the estimated values on all grids in this municipality [39]. Therefore, the adjusted livestock density distribution on a grid is Equation (1):

$$A_i = P_i \times \frac{A_j}{P_j}, \quad (1)$$

where i represents a grid and j represents a municipal administrative district. A_i is the adjusted value of the grid i , and P_j is the corresponding predicted value of the grid i

before adjustment. A_j stands for the statistical value of livestock in municipal administrative district j , and P_j stands for the total predicted gridded livestock of this municipal administrative district.

3.2.3. Performance Evaluation

Since the regression models with continuous dependent variables are constructed, two commonly used performance indicators, coefficient of determination (R^2) and root mean square error (RMSE), are used to evaluate the performance of the regression models constructed in this study. Their respective formulas are Equations (2) and (3):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \quad (3)$$

where y_i is the statistical livestock density of county i , \bar{y} is an average statistical livestock density of counties, \hat{y}_i is the model's predicted value for county i , and n is the number of samples. It can be seen from the formula that R^2 can be negative. Generally speaking, if the predicted value of the developed model is precisely equal to the true value without any error, then R^2 is 1. If the explanatory power of the developed model is equivalent to that \bar{y} , then R^2 is 0. If the explanatory power of the developed model is worse than that \bar{y} , R^2 is less than 0.

4. Results

4.1. Gridded Livestock Distribution Maps

Figures 3 and 4 show cattle and sheep distribution of western China estimated by the three machine models in 2000, 2005, 2010, and 2015. The overall trends of the cattle and sheep distribution derived by three livestock spatialization machine learning models are similar and generally consistent with the statistical data. Obviously, the mapping results give the detailed spatial distribution characteristics of livestock, which the census data cannot describe. The cattle are mainly concentrated in the southeast and northwest areas of the study area, showing two northeast-southwest distribution belts (Figure 3). The high cattle densities are found in central Shaanxi, southeastern Gansu, and the northern and southern ends of the Ningxia. Cattle are also dense in southeastern Qinghai, central to northeastern Tibet, and western Xinjiang. It can be seen from the spatial distribution of sheep in Figure 4 that the number of sheep in the study area significantly exceeds the number of cattle, which is consistent with the actual situation (i.e., census data). Figure 4 shows there are denser sheep in the east, northwest, and southwest part of the study area, while there are sparser sheep in the middle part. The northern Shaanxi, entire Ningxia, Hexi Corridor of Gansu Province, eastern Qinghai, and western Xinjiang have a dense sheep population. The sheep of Tibet are mainly compact in the central and southern regions. The regions where livestock gathers are mostly grassland and arable land resource-rich areas. These resources provide better natural conditions for the survival of livestock.

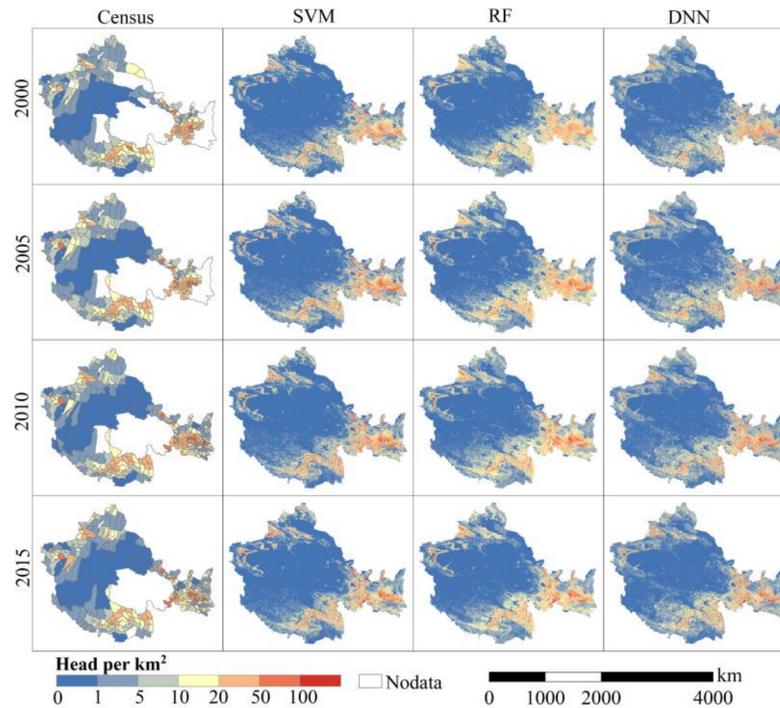


Figure 3. Cattle distribution in six provinces of Western China. The first column is the cattle distribution density obtained from the county level census, and the second to fourth columns are the density at the 1 km scale estimated by SVR, RF, and DNN, respectively. The first to fourth rows indicate 2000, 2005, 2010, and 2015, respectively.

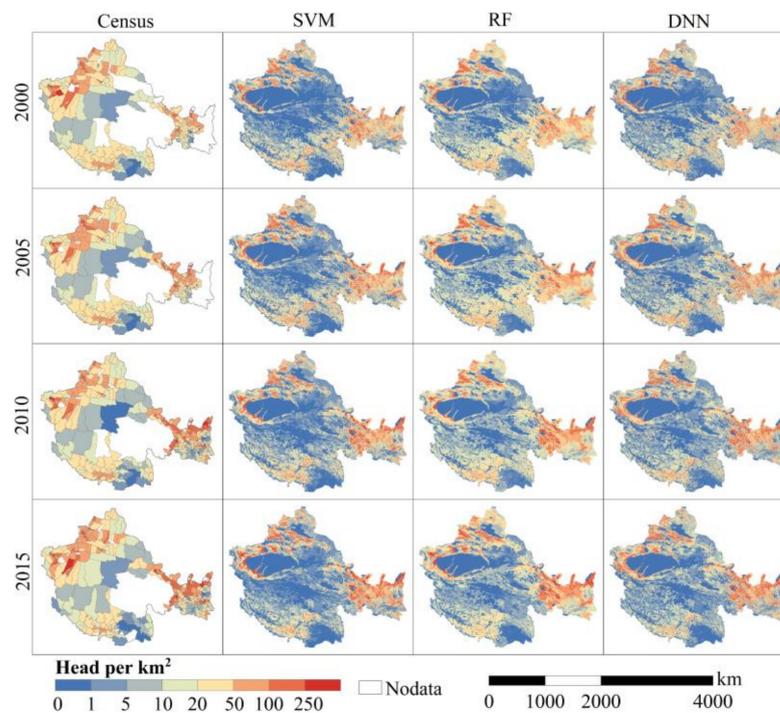


Figure 4. Sheep distribution in six provinces of western China. The first column is the sheep distribution density obtained from the county level census, and the second to fourth columns are the density at the 1 km scale estimated by SVR, RF, and DNN, respectively. The first to fourth rows indicate 2000, 2005, 2010, and 2015, respectively.

In order to further explore the weak differences between the three machine learning models base livestock spatialization, we randomly selected two small local regions (We refer it as regions A and B) and enlarged them to show the details of the spatial distribution of cattle and sheep, as shown in Figures 5 and 6. The highest concentrations of cattle and sheep are found in cultivated land and grassland. Cultivated land corresponds to agro-pastoral production systems, where agricultural waste can provide rations for herbivorous livestock, thereby promoting cattle and sheep breeding. Furthermore, the grazing area in agro-pastoral production systems is small. Thus, there is a relatively high distribution density of cattle and sheep on the cultivated land. The grassland corresponds to pastoral production systems. Pastoral production systems have a large number of forage resources, providing high-quality forage and broad activity space for cattle and sheep, so a large number of cattle and sheep are raised. Simultaneously, due to the vast activity space of pastoral production systems, when the total number of cattle and sheep is about the same, the density of cattle and sheep on the grassland may be slightly lower than that of the cultivated land. Other land-use types, such as forest land, construction land, and unused land, are difficult to provide a suitable living environment for cattle and sheep. Thus, there are few cattle and sheep distributed on them. The above livestock distribution law is consistent with existing research conclusions [40,41]. In addition, compared with traditional statistical data, the gridded livestock data has more obvious granularity and more prominent texture, which can better reflect the details of the spatial distribution of cattle and sheep. The mapping results of the three models have very similar morphological distributions. The spatial detail features DNN based spatialization results are more prominent than the other two models, which is more in line with the natural livestock distribution in the complex surface environment.

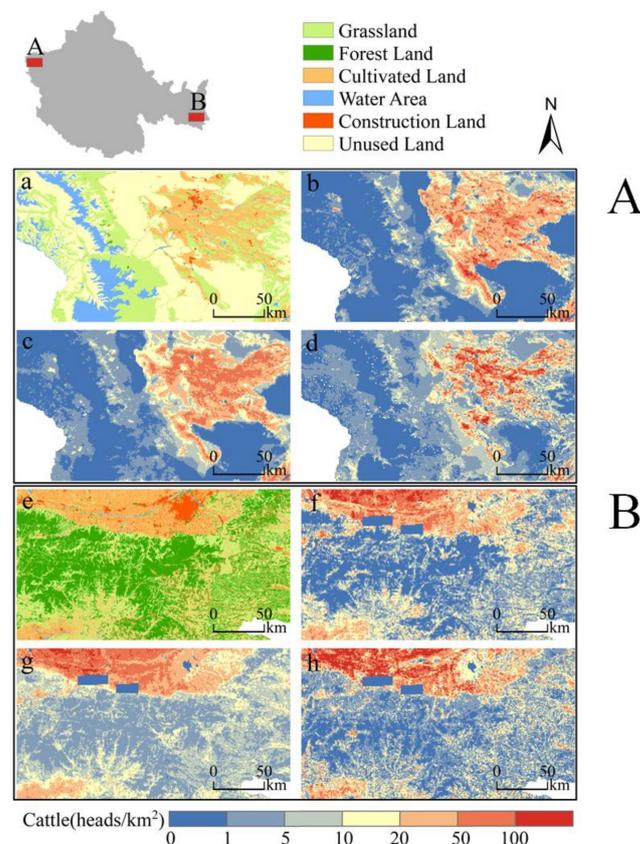


Figure 5. Enlarged spatial detail distribution of cattle for two randomly selected small local regions A and B. (a–d) are the land use situation and the spatialized cattle results of the SVM, RF, and DNN models of region A in 2015. (e–h) are the land use situation and the spatialized cattle results of the SVM, RF, and DNN models of region B in 2015.

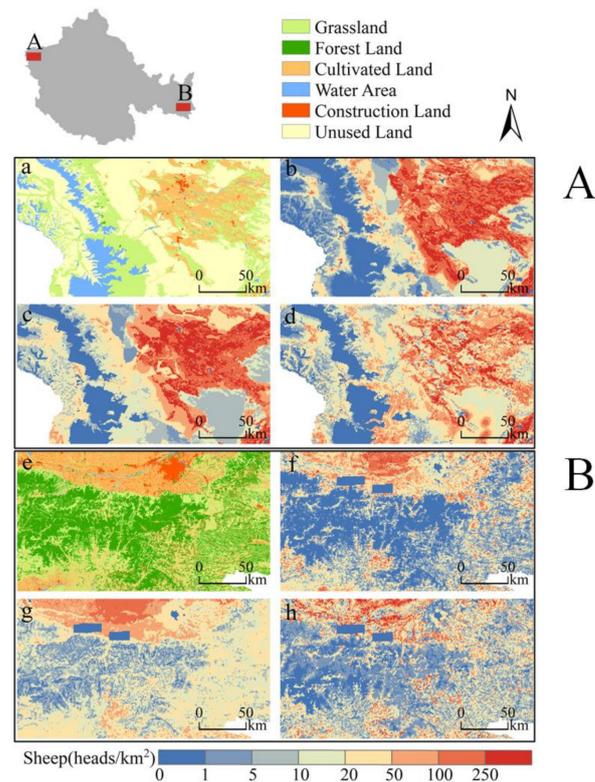
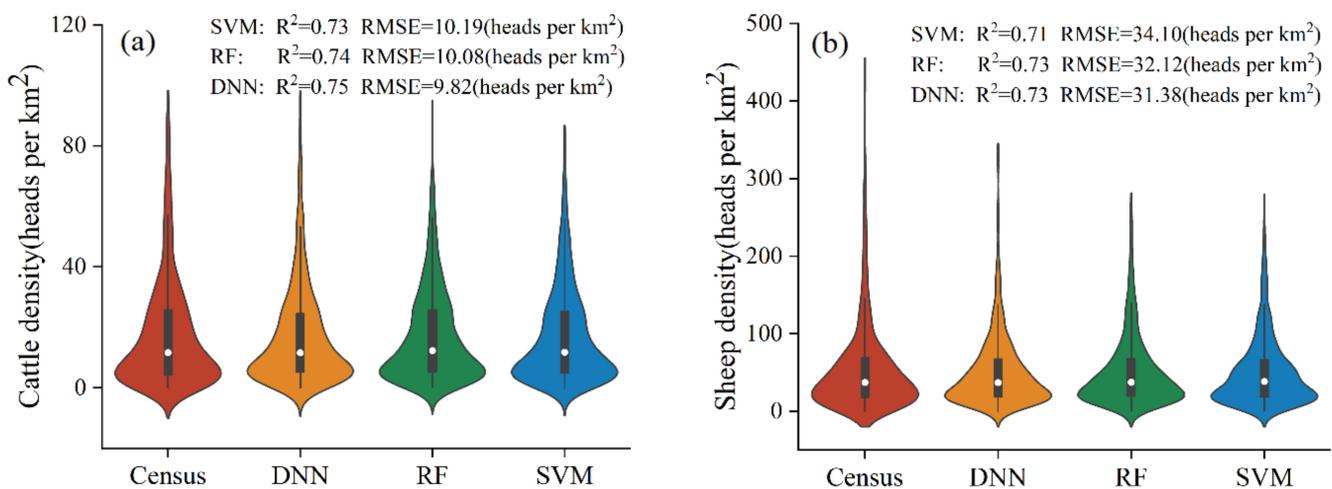


Figure 6. Enlarged spatial detail distribution of sheep for two randomly selected small local regions A and B. (a–d) are the land use situation and the spatialized sheep results of the SVM, RF, and DNN models of region A in 2015. (e–h) are the land use situation and the spatialized sheep results of the SVM, RF, and DNN models of region B in 2015.

Table 2 summarizes R^2 and $RMSE$ of estimated distribution density of cattle and sheep for each machine learning based livestock spatialization model. In general, the errors of the three models are within acceptable limits. It can be seen from Table 2 that the DNN model has the highest accuracy on the county scale, with its R^2 exceeding 0.95 and $RMSE$ is significantly lower than the other two models for both cattle and sheep on the training set. On the test set, the performance of all three machine learning models has some degradation, but DNN is still superior to the other two, with its R^2 exceeding 0.73, and $RMSE$ is the smallest of the three models. The accuracy of the RF model is slightly lower than that of the DNN model, and the SVM model performs the worst. In addition, to analyze the estimation performance of the three machine learning models on the 1 km grid scale, we further aggregate the prediction results on the 1 km grid scale to the county scale, and compare them with the census data, as shown in Figure 7. The livestock distribution density estimated by the three machine learning models is very consistent with the census data, which shows that the three machine learning models have good robustness and can provide a stable estimation of livestock distribution density on the grid scale. Moreover, the performance of the three models can still be ranked as DNN > RF > SVM. However, there is no remarkable performance difference between them. For example, the R^2 of DNN is 0.75 for cattle and 0.73 for sheep, but the R^2 of RF and SVR also reaches 0.74 and 0.73 for cattle, 0.73 and 0.71 for sheep, respectively. In terms of different species, the estimation accuracy of gridded cattle distribution density is higher than that of sheep. The cattle and sheep distribution density has concentrated in 0–20 heads per km^2 and 0–100 heads per km^2 , respectively. The distribution density of cattle has a slight peak at 20–40 head/ km^2 .

Table 2. Accuracy of the models on the county scale.

Species	Model	Training Set		Test Set	
		R^2	RMSE	R^2	RMSE
Cattle	SVM	0.50	14.86	0.54	13.21
	RF	0.92	5.82	0.74	9.57
	DNN	0.95	4.73	0.75	8.98
Sheep	SVM	0.55	43.38	0.52	52.65
	RF	0.93	19.59	0.72	34.58
	DNN	0.96	14.71	0.73	33.97

**Figure 7.** Accuracy of the livestock spatialization results. The distribution density of (a) cattle and (b) sheep estimated by the model on the 1 km scale was aggregated to the county scale and compared with the census data.

In short, the spatialization results of the DNN model are better than the RF and SVM model in all accuracy indicators. The possible reason is that the deep learning model can automatically extract features, actively mine the relationship between features, and has a better nonlinear fitting ability. The RF also achieves a good result. Since the RF is the integration of multiple decision trees, and random attribute selection is introduced in the training process of decision trees, which effectively alleviates the over-fitting problem that is prone to occur in traditional machine learning algorithms. SVR relies more on artificially extracted features as a shallow machine learning algorithm. When the feature is not representative enough, the problem of under-fitting is prone to occur.

4.2. Spatiotemporal Changes of Livestock

We take the livestock distribution on the 1 km grid scale estimated by the DNN model as the benchmark. We further analyze the temporal and spatial distribution changes of livestock from 2000 to 2015. Figure 8 shows the characteristics of the spatiotemporal change of cattle from 2000 to 2015 at five-year intervals. The histogram in each subgraph represents the statistical value of cattle in each province in the corresponding year. Overall, the distribution of cattle in the study area showed a general trend of increasing first and then decreasing. Specifically, the spatiotemporal change map of cattle from 2000 to 2005 shows an increasing trend in almost the entire study area. It is consistent with the increase of each province in the corresponding statistical data histogram (note that there is a lack of data in Qinghai in 2005). From 2005 to 2010, the decline of cattle in the study area initially appeared, for example, Shaanxi and Ningxia. In addition, the number of cattle in the central and western regions of Xinjiang has significantly reduced, which is in line with Xinjiang's sustainable development of animal husbandry requirements. The decline of cattle in west Qinghai is the most obvious, a possible reason for this was speculated to be the "Three

Rivers Source Ecological Protection Project”, which was implemented by Qinghai province in 2005 to reduce grazing and prohibition in crucial areas. From 2010 to 2015, the regions where the number of cattle has decreased have further expanded, especially the Qinghai-Tibet Plateau. The main reason is that the Qinghai-Tibet Plateau is an essential barrier to ecological security. China implemented the “Qinghai-Tibet Plateau Regional Ecological Construction and Environmental Protection Plan” and issued the “Opinions on Improving the Policy of Returning Pasture to Grassland” in 2011. However, there are still small areas where cattle have increased in the southwestern part of Qinghai Province. According to relevant data, this region is the main gathering area of village-level settlements in Qinghai Province. With the policy changes, the livestock production pattern has shifted from the traditional grassland animal husbandry mainly to farming, grassland, and suburban animal husbandry, which may lead to the denseness of livestock near residential areas.

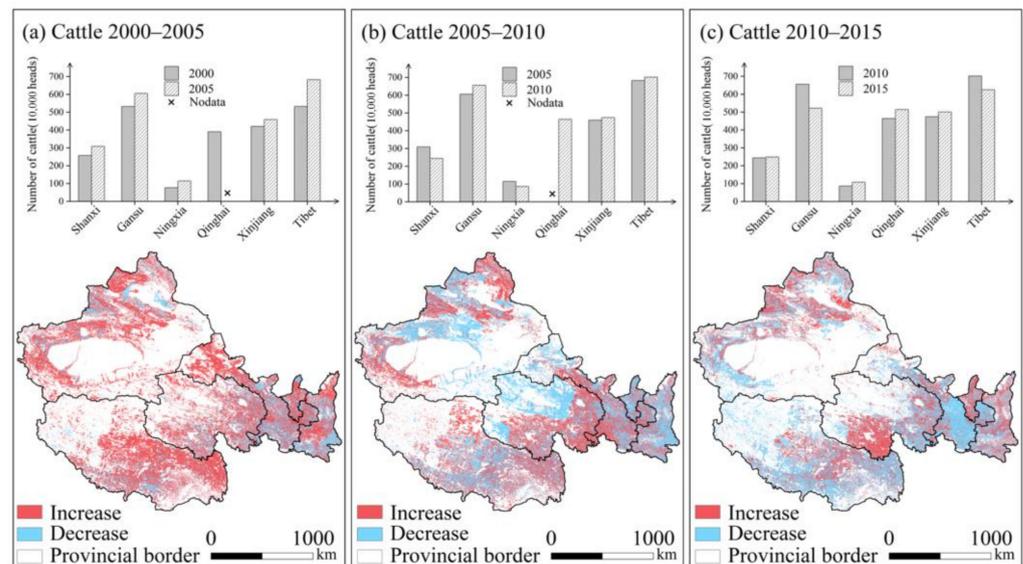


Figure 8. Spatiotemporal changes of cattle based on the DNN estimation. (a–c) represent 2000 to 2005, 2005 to 2010, and 2010 to 2015, respectively. The bar charts show the statistical value of cattle in each province.

Figure 9 shows the characteristics of the spatiotemporal change of sheep from 2000 to 2015 at five-year intervals. Overall, the distribution of sheep shows a trend of increasing first, then decreasing, and it tends to stabilize finally in the study area. From 2000 to 2005, the number of sheep shows an increasing trend in most areas of the study area. However, the distribution of sheep in Qinghai Province has decreased significantly. This phenomenon was mainly affected by the “Three Rivers Source Ecological Protection Project” in 2005. From 2005 to 2010, the distribution of sheep decreased in most areas, most notably in Xinjiang and Qinghai. After 2005, driven by the comparative benefits of the breeding industry and planting, the size of forest and fruit in Xinjiang has expanded rapidly. The planting area of crops and forage has been significantly reduced. Due to the shortage of forage in the agricultural area, the supply of forage for cattle and sheep cannot be guaranteed, and the breeding cost is increasing year after year. Therefore, there is a trend to reduce the breeding scale. The decrease of sheep in Qinghai Province can be attributed to the “Three Rivers Source Ecological Protection Project” implemented in 2005. From 2010 to 2015, China began to focus on the ecological protection of the Qinghai-Tibet Plateau. The “Plan for Ecological Construction and Environmental Protection of the Qinghai-Tibet Plateau” may be led to a significant decrease of sheep on the Qinghai-Tibet Plateau.

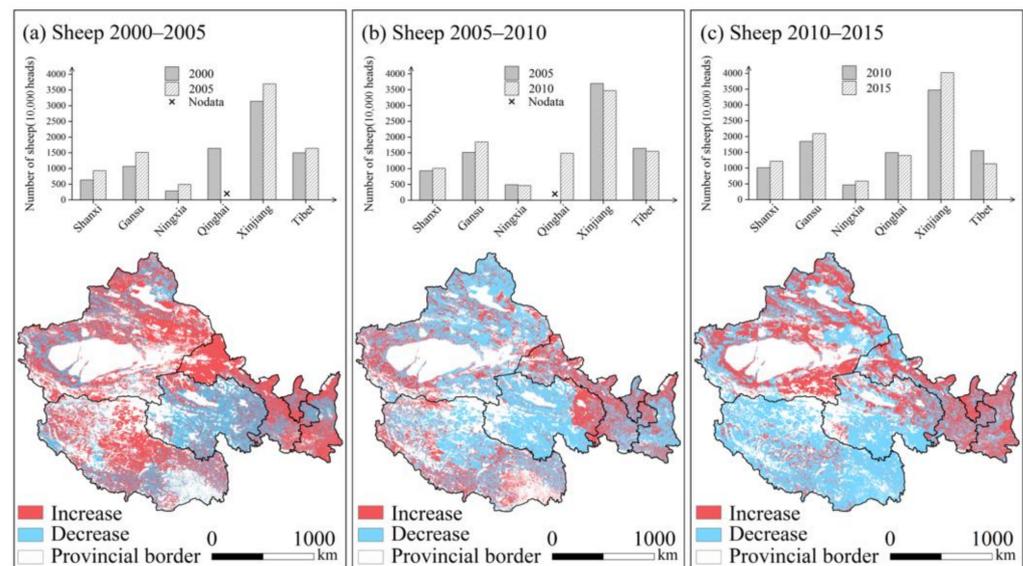


Figure 9. Spatiotemporal changes of sheep based on the DNN estimation. (a–c) represent 2000 to 2005, 2005 to 2010, and 2010 to 2015, respectively. The bar charts show the statistical value of cattle in each province.

5. Discussion

5.1. Comparison with the Open Access Gridded Livestock Datasets

In order to further verify the effectiveness and reliability of the three models we developed, the mapping results of the three models were compared with two open-access gridded livestock datasets (i.e., GLW2 and GLW3). The China region in the livestock grid data of GLW2 and GLW3 is produced based on the livestock statistics data in 2001. Therefore, we use the livestock grid data of the same year calculated by our research for comparison. We did not compare with the GLW1 database since the livestock statistical data used to produce the dataset was from the 1990s, not within our research period. The scatter diagram of Figure 10 shows that R^2 of cattle for GLW2 and GLW3 are -1.16 and -0.41 , this is significantly lower than the accuracy of the three models developed in this study (R^2 exceeds 0.7), when the distribution density values of them are aggregated to the county scale and compared with the census data. Although the accuracy of GLW3 is higher than that of GLW2, it is still difficult to accurately describe the spatial distribution of livestock in six provinces in western China.

Similarly, gridded distribution density aggregated to the county scale and census data for sheep are compared in Figure 11. What needs special attention here is that the sheep and goats in GLW2 and GLW3 are independent, while some statistical data are combined. Therefore, the sheep and goat data in GLW2 and GLW3 are added to calculate the R^2 . Although the performance of GLW3 is greatly improved compared with GLW2, with its R^2 can reach 0.5 , it is still significantly inferior to the three machine learning models developed in this study.

In terms of spatial distribution, although the distribution results of five grid livestock are consistent with the census data in the overall spatial distribution trend, there are still some obvious differences between GLW2/GLW3 and the three models developed in this study (Figures 12 and 13). From Figure 12, GLW2 and GLW3 describe the spatial distribution of cattle very roughly, with obvious administrative boundaries, which is unreasonable in practice. In addition, GLW2 and GLW3 overestimated the distribution of cattle in the southern part of Tibet and the eastern part of Qinghai Province, which was inconsistent with the statistics. As shown in Figure 13, the spatial distribution of sheep in the southwest of GLW2 is quite different from the statistical data, while GLW3 sheep is much more consistent. Due to spatial resolution limitations, their distribution patterns are still very rough, with blocky distributions. In general, from the perspective of the visual

effects of the spatial distribution of livestock, the livestock spatial distribution structure obtained by the three spatial models in this study is more stable and reasonable. It is more in line with the actual livestock distribution in the complex surface environment.

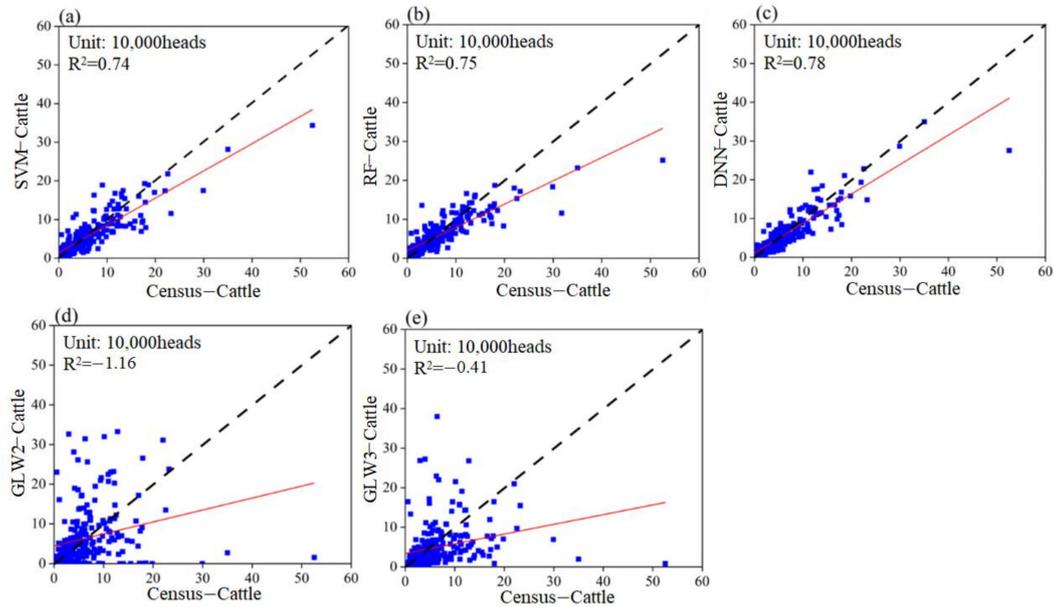


Figure 10. The scatter diagram of grided distribution density aggregated to the county scale and census data for cattle. (a) SVR; (b) RF; (c) DNN; (d) GLW2; and (e) GLW3. The red line is the linear regression line, and the dotted line is the 1:1 line.

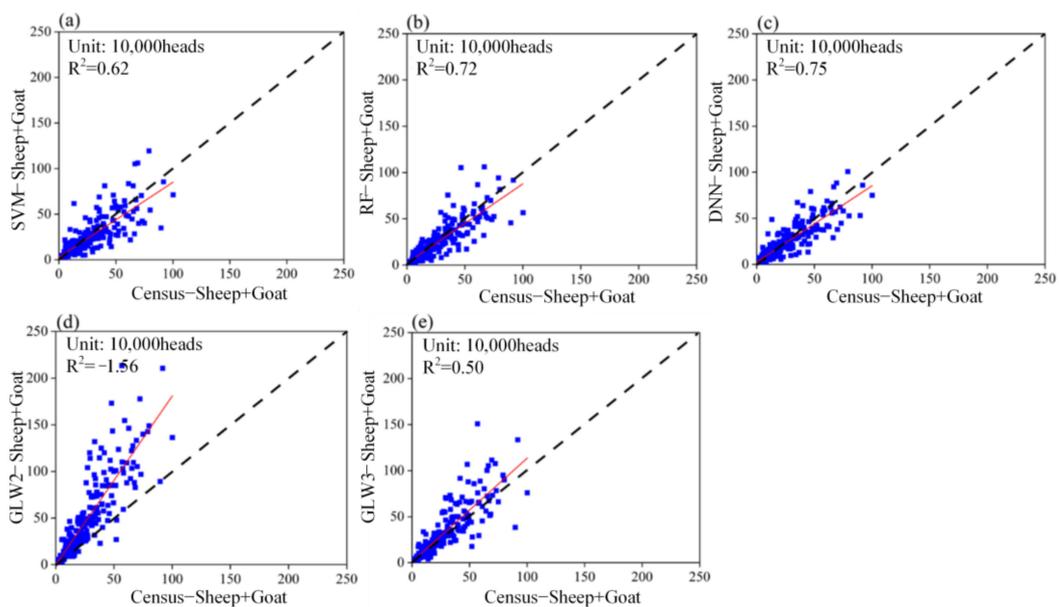


Figure 11. The scatter diagram of grided distribution density aggregated to the county scale and census data for sheep. (a) SVR; (b) RF; (c) DNN; (d) GLW2; and (e) GLW3. The red line is the linear regression line, and the dotted line is the 1:1 line.

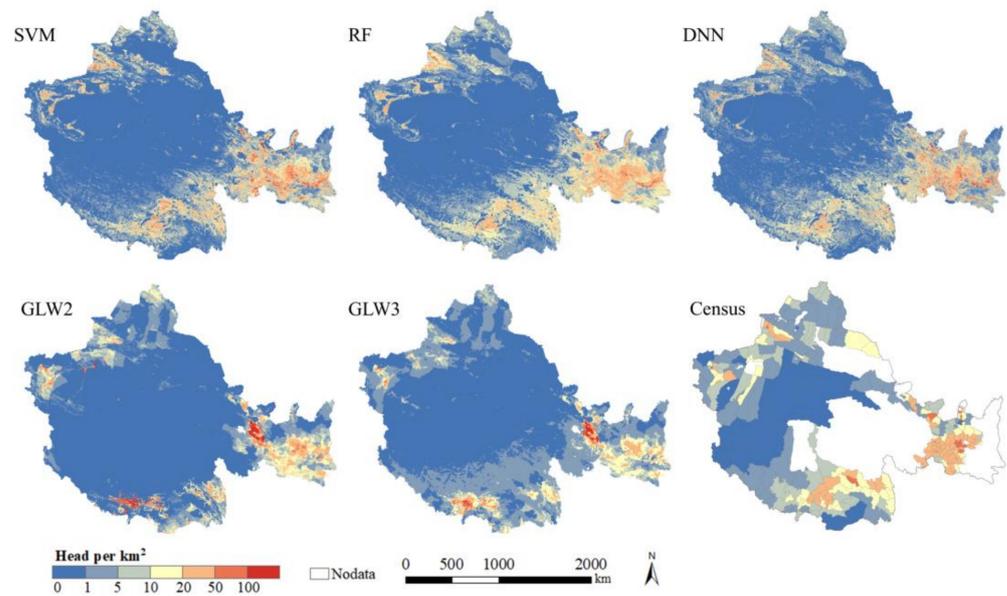


Figure 12. The spatial distribution density of cattle in 2000.

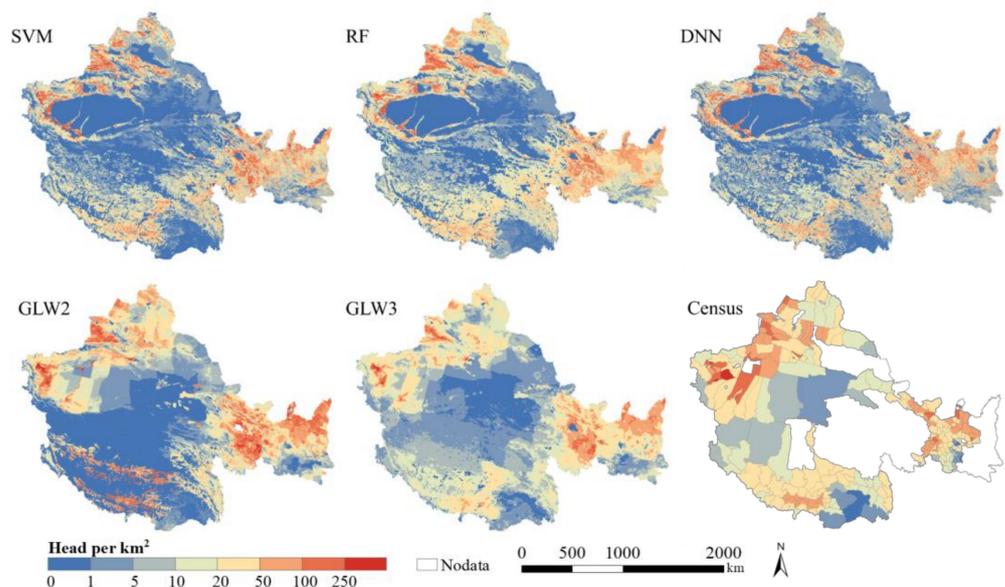


Figure 13. The spatial distribution density of sheep in 2000.

However, the above differences are understandable. Since GLW data is global-scale data, its primary purpose is to portray the detailed information of the spatial distribution of livestock on a large scale and find livestock distribution laws, which have a wide range of application values and crucial guiding significance. The scope of this research is only six provinces in western China. The research scale is smaller, and the statistical data used is more detailed, which is conducive to improving the model's accuracy. Robinson et al. (2014) use different levels of livestock statistics to build spatial models and prove that the finer the scale of statistical data used to establish the model, the better the estimation result [7].

5.2. The Reasonableness of the Hypothesis

This study is based on a hypothesis that a similar causal relationship exists between livestock density and environmental factors on different scales. Such a similar assumption

has been widely used in applying social factors spatialization [38,42,43]. However, as far as the actual situation is concerned, the impact of environmental factors on the spatial distribution of livestock is not the same at different scales. Using the model trained on a coarse scale to predict the distribution of livestock on a fine scale has a certain degree of uncertainty, which may lead to a large estimation error. One should note that the high spatial resolution of the output does not necessarily represent the ground truth value of livestock at the same resolution (i.e., 1 km), but only reflects the potential distribution of livestock on a 1 km × 1 km grid. The follow-up plan considers using physical guided methods to better develop the research on the spatialization of statistical data.

5.3. Selection and Contribution of Environmental Factors

To explore the influence of selected environmental factors on the distribution of livestock, we have designed two parts of work: (1) correlation analysis of environmental factors and density of cattle and sheep (Figure 14); and (2) important analysis of each factor by random forest (Figure 15). It can be seen that when the significance level is less than 0.05, the selected factor has a specific correlation with the density of cattle and sheep and has the potential to predict the spatial distribution of livestock. There is a positive correlation between cattle density and population grid data, arable land coverage, and NDVI, and a negative correlation with desert coverage and the distance to the river. There is a positive correlation between sheep density and arable land coverage, daytime surface temperature, population, and GDP grid data. In contrast, it has a significant negative correlation with forest land coverage and slope. This is because areas with high arable land coverage and NDVI provide abundant resources for livestock activities, while natural conditions in regions with high desert coverage, forest coverage, and slopes are not suitable for livestock activities. The environmental factors that have a more significant impact on the spatial distribution of cattle are population grid data, NDVI, and arable land coverage. The environmental factors that have a more significant effect on the spatial distribution of sheep are arable land coverage, forest land coverage, and population grid data. It can be seen that population density and arable land coverage are the most critical environmental factors affecting the distribution of livestock density. The result is reasonable since the livestock is mostly dense around human activity areas, which is the difference between livestock and wild animals.

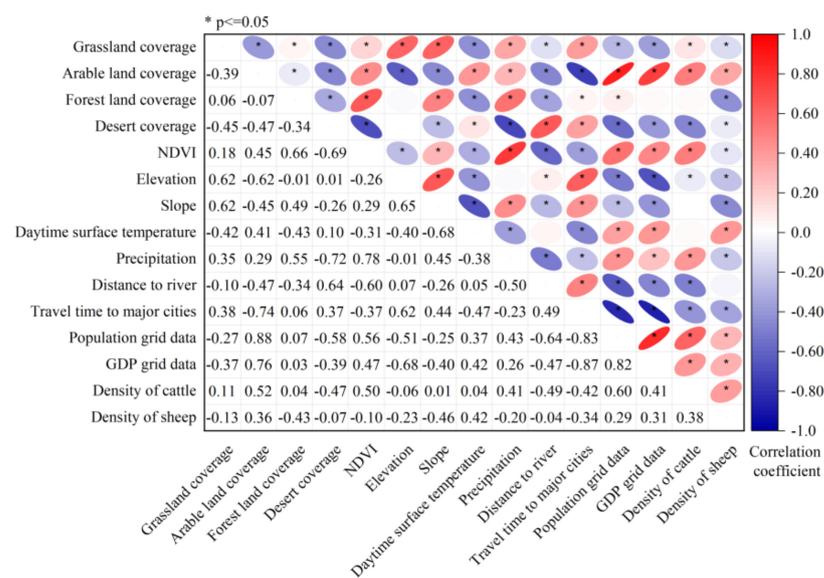


Figure 14. Correlation between environmental factors and density of cattle and sheep. The shape direction of the ellipse in the upper triangular area represents the positive or negative of the correlation, the color is the level of the corresponding correlation, and the lower triangular area is the value of the corresponding correlation coefficient.

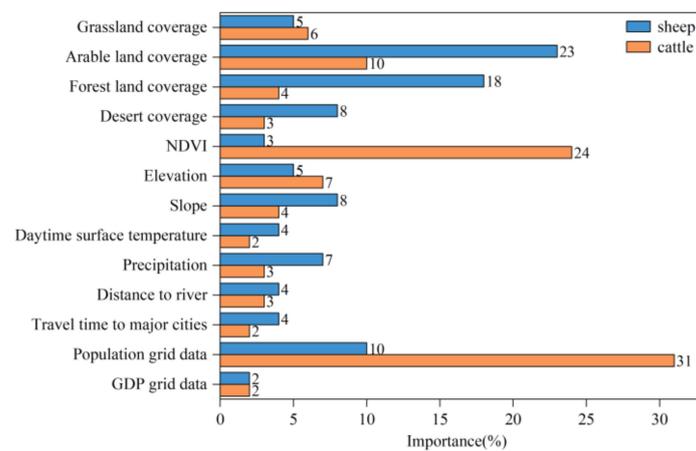


Figure 15. Importance of environmental factors influencing the spatial distribution of cattle and sheep.

6. Conclusions

The importance of livestock spatialization stems from various studies' demands for fine-grained livestock spatial distribution data. In recent years, livestock grid data have been applied to many aspects as primary data. Livestock grid data have been widely involved in the rational use of natural resources, such as assessing grass-livestock balance based on livestock grid data [44], estimating oxygen consumption of livestock [45], and quantifying water use for animal husbandry [46]. There are also specific applications in environmental impact assessment research, such as quantifying methane emissions based on livestock grid data [47]. In addition, the spatialization of livestock data provides the possibility to assess the risk of infectious diseases. Some scholars have evaluated high-risk areas of bluetongue virus outbreak based on livestock grid data [48]. Therefore, spatialization technology for livestock is of great significance to research the rational use of natural resources, the environmental and ecological protection, the risk assessment of zoonotic diseases, and the sustainable development of animal husbandry.

Taking the spatialization of cattle and sheep distribution in six provinces in Western China in 2000, 2005, 2010, and 2015 as an example, this study selects thirteen environmental factors from terrain, climate, land use, and social economy as predictor variables, and county-level livestock statistics as the response variable. Using three machine learning models to effectively integrate these grid geographic data with animal husbandry statistical data, the distribution density of cattle and sheep on 1 km grid scale was obtained. This study proves that the accuracy of livestock density data with a resolution of 1 km in six provinces in western China based on three machine learning models is much superior to the existing open-access dataset, which is more in line with the actual livestock spatial distribution in a complex surface environment. The overall accuracy of the three livestock spatialization models is ranked as DNN > RF > SVM. The DNN model can thoroughly mine various characteristics of factors affecting the spatial distribution of livestock and then characterize the complex nonlinear relationship between variables. It can better highlight environmental details, and a relatively higher precision is achieved in downscaling livestock data. The livestock grid data produced in this study for 2000, 2005, 2010, and 2015 can provide detailed data support for rational use of resources, environmental impact assessment, and sustainable development of animal husbandry.

The highlight of this research is to explore the applicability of deep machine learning in the study of livestock spatialization. The results prove that the machine learning methods, especially the new deep learning methods, have great potential in the research of livestock spatialization. However, this study still has some obvious deficiencies, such as the fact that the actual verification of the spatialization results should collect livestock distribution data on the same scale of 1 km to test the model prediction value. Due to the limitation of the data, we only aggregated the derived 1 km livestock grid distribution map to the county-level administrative unit and compared it with the statistical data. This verification method is

somewhat simple and crude. In the future, we hope to obtain more detailed livestock data, such as township-level livestock statistics or household-based livestock statistics, to better verify the model's accuracy. In addition, we will further explore and introduce some more appropriate environmental factors and more effective deep learning algorithms into the study of livestock spatialization.

Author Contributions: Conceptualization, X.L. and J.H.; methodology, X.L., J.H. and C.H.; software, X.L.; validation, X.L. and C.H.; investigation, X.L.; resources, X.L.; data curation, X.L.; writing—original draft preparation, X.L.; writing—review and editing, J.H. and C.H.; visualization, X.L.; supervision, J.H. and C.H.; funding acquisition, C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Science Foundation of China under grants (42130113), the Strategic Priority Research Program of the Chinese Academy of Sciences under grants XDA19040504, and the Basic Research Innovative Groups of Gansu province, China (Grant No. 21JR7RA068).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the website given in the manuscript.

Acknowledgments: We are grateful for the use of data in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, M. China's livestock industry development: Achievements, experiences and future trends. *Chin. Issues Agric. Econ.* **2018**, *8*, 60–70.
2. Olesen, J.; Schelde, K.; Weiske, A.; Weisbjerg, M.; Asman, W.; Djurhuus, J. Modelling greenhouse gas emissions from European conventional and organic dairy farms. *Agric. Ecosyst. Environ.* **2006**, *112*, 207–220. [\[CrossRef\]](#)
3. Monteny, G.-J.; Bannink, A.; Chadwick, D. Greenhouse gas abatement strategies for animal husbandry. *Agric. Ecosyst. Environ.* **2006**, *112*, 163–170. [\[CrossRef\]](#)
4. Steinfeld, H.; Gerber, P.; Wassenaar, T.D.; Castel, V.; Rosales, M.; Rosales, M.; de Haan, C. *Livestock's Long Shadow: Environmental Issues and Options*; Food & Agriculture Organization: Rome, Italy, 2006.
5. Dong, N.; Yang, X.; Cai, H. Research progress and perspective on the spatialization of population data. *J. Geo-Inf. Sci.* **2016**, *18*, 1295–1304. [\[CrossRef\]](#)
6. Goodchild, M.F.; Anselin, L.; Deichmann, U. A framework for the areal interpolation of socioeconomic data. *Environ. Plan. A* **1993**, *25*, 383–397. [\[CrossRef\]](#)
7. Wint, W.; Robinson, T. *Gridded Livestock of the World 2007*; FAO: Roma, Italy, 2007.
8. Robinson, T.P.; Wint, G.W.; Conchedda, G.; Van Boeckel, T.P.; Ercoli, V.; Palamara, E.; Cinardi, G.; D'Aiotti, L.; Hay, S.I.; Gilbert, M. Mapping the global distribution of livestock. *PLoS ONE* **2014**, *9*, e96084. [\[CrossRef\]](#)
9. Nicolas, G.; Robinson, T.P.; Wint, G.W.; Conchedda, G.; Cinardi, G.; Gilbert, M. Using random forest to improve the downscaling of global livestock census data. *PLoS ONE* **2016**, *11*, e0150424. [\[CrossRef\]](#)
10. Gilbert, M.; Nicolas, G.; Cinardi, G.; Van Boeckel, T.P.; Vanwambeke, S.O.; Wint, G.W.; Robinson, T.P. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci. Data* **2018**, *5*, 180227. [\[CrossRef\]](#)
11. Neumann, K.; Elbersen, B.S.; Verburg, P.H.; Staritsky, I.; Pérez-Soba, M.; de Vries, W.; Rienks, W.A. Modelling the spatial distribution of livestock in Europe. *Landsc. Ecol.* **2009**, *24*, 1207–1222. [\[CrossRef\]](#)
12. Prosser, D.J.; Wu, J.; Ellis, E.C.; Gale, F.; Van Boeckel, T.P.; Wint, W.; Robinson, T.; Xiao, X.; Gilbert, M. Modelling the distribution of chickens, ducks, and geese in China. *Agric. Ecosyst. Environ.* **2011**, *141*, 381–389. [\[CrossRef\]](#)
13. Van Boeckel, T.P.; Prosser, D.; Franceschini, G.; Biradar, C.; Wint, W.; Robinson, T.; Gilbert, M. Modelling the distribution of domestic ducks in Monsoon Asia. *Agric. Ecosyst. Environ.* **2011**, *141*, 373–380. [\[CrossRef\]](#)
14. Qiao, Y.; Zhu, H.; Shao, X.; Zhong, H. Research on Gridding of Livestock Spatial Density Based on Multi-Source Information. *China Sci. Technol. Res. Guide* **2017**, *49*, 53–59. [\[CrossRef\]](#)
15. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#)
16. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
17. Bailey, D.W.; Stephenson, M.B.; Pittarello, M. Effect of terrain heterogeneity on feeding site selection and livestock movement patterns. *Anim. Prod. Sci.* **2015**, *55*, 298–308. [\[CrossRef\]](#)

18. Zhou, L.; Xiong, L.-Y. Natural topographic controls on the spatial distribution of poverty-stricken counties in China. *Appl. Geogr.* **2018**, *90*, 282–292. [[CrossRef](#)]
19. Raynor, E.; Gersie, S.; Stephenson, M.; Clark, P.; Spiegel, S.; Boughton, R.; Bailey, D.; Cibils, A.; Smith, B.; Derner, J. Cattle grazing distribution patterns related to topography across diverse rangeland ecosystems of North America. *Rangel. Ecol. Manag.* **2021**, *75*, 91–103. [[CrossRef](#)]
20. Kruska, R.; Reid, R.S.; Thornton, P.K.; Henninger, N.; Kristjanson, P.M. Mapping livestock-oriented agricultural production systems for the developing world. *Agric. Syst.* **2003**, *77*, 39–63. [[CrossRef](#)]
21. Liu, J.; Kuang, W.; Zhang, Z.; Xu, X.; Qin, Y.; Ning, J.; Zhou, W.; Zhang, S.; Li, R.; Yan, C.; et al. Spatiotemporal characteristics, patterns, and causes of land-use changes in China since the late 1980s. *J. Geogr. Sci.* **2014**, *24*, 195–210. [[CrossRef](#)]
22. Liu, J.; Liu, M.; Tian, H.; Zhuang, D.; Zhang, Z.; Zhang, W.; Tang, X.; Deng, X. Spatial and temporal patterns of China's cropland during 1990–2000: An analysis based on Landsat TM data. *Remote Sens. Environ.* **2005**, *98*, 442–456. [[CrossRef](#)]
23. National Tibetan Plateau Data Center. Available online: <http://www.tpdc.ac.cn/zh-hans/data/12e91073-0181-44bf-8308-c50e5bd9a734/> (accessed on 21 December 2020).
24. Peng, S. *1-km Monthly Precipitation Dataset for China (1901–2017)*; A Big Earth Data Platform for Three Poles; National Tibetan Plateau Data Center: Beijing, China, 2020.
25. Nelson, A. *Travel Time to Major Cities: A Global Map of Accessibility*; Global Environment Monitoring Unit, Joint Research Centre of the European Commission: Enschede, The Netherlands, 2008.
26. Weiss, D.J.; Nelson, A.; Gibson, H.S.; Temperley, W.; Peedell, S.; Lieber, A.; Hancher, M.; Poyart, E.; Belchior, S.; Fullman, N.; et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **2018**, *553*, 333–336. [[CrossRef](#)] [[PubMed](#)]
27. Resource and Environment Science and Data Center. China Population Spatial Distribution Kilometer Grid Dataset. Available online: <https://www.resdc.cn/DOI/DOI.aspx?DOIid=32> (accessed on 10 April 2021).
28. Resource and Environment Science and Data Center. China GDP Spatial Distribution Kilometer Grid Data Set. Available online: <https://www.resdc.cn/DOI/doi.aspx?DOIid=33> (accessed on 10 April 2021).
29. Van Velthuisen, H. *Mapping Biophysical Factors that Influence Agricultural Production and Rural Vulnerability*; Food & Agriculture Organization: Rome, Italy, 2007.
30. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1999.
31. Thomas, S.; Pillai, G.; Pal, K. Prediction of peak ground acceleration using ϵ -SVR, ν -SVR and Ls-SVR algorithm. *Geomat. Nat. Hazards Risk* **2017**, *8*, 177–193. [[CrossRef](#)]
32. Pasolli, L.; Notarnicola, C.; Bruzzone, L. Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1080–1084. [[CrossRef](#)]
33. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
34. Scornet, E. Tuning parameters in random forests. *ESAIM Proc. Surv.* **2017**, *60*, 144–162. [[CrossRef](#)]
35. Fang, K.; Wu, J.; Zhu, J.; Xie, B. Research review of random forest method. *Stat. Inf. Forum* **2011**, *26*, 32–38.
36. Merkel, G.D.; Povinelli, R.J.; Brown, R.H. Short-term load forecasting of natural gas with deep neural network regression. *Energies* **2018**, *11*, 2008. [[CrossRef](#)]
37. Kuwata, K.; Shibasaki, R. Estimating crop yields with deep learning and remotely sensed data. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 858–861.
38. Zhao, S.; Liu, Y.; Zhang, R.; Fu, B. China's population spatialization based on three machine learning models. *J. Clean. Prod.* **2020**, *256*, 120644. [[CrossRef](#)]
39. Wang, Y.; Huang, C.; Zhao, M.; Hou, J.; Zhang, Y.; Gu, J. Mapping the Population Density in Mainland China Using NPP/VIIIRS and Points-Of-Interest Data Based on a Random Forests Model. *Remote Sens.* **2020**, *12*, 3645. [[CrossRef](#)]
40. Cecchi, G.; Wint, W.; Shaw, A.; Marletta, A.; Mattioli, R.; Robinson, T. Geographic distribution and environmental characterization of livestock production systems in Eastern Africa. *Agric. Ecosyst. Environ.* **2010**, *135*, 98–110. [[CrossRef](#)]
41. Ganskopp, D.; George, M.; Bailey, D.; Borman, M.; Surber, G.; Harris, N. Factors and practices that influence livestock distribution. *Univ. Calif. Div. Agric. Nat. Resour.* **2007**, *8217*, 20.
42. Lloyd, C.T.; Sorichetta, A.; Tatem, A.J. High resolution global gridded data for use in population studies. *Sci. Data* **2017**, *4*, 170001. [[CrossRef](#)]
43. Hollings, T.; Robinson, A.; van Andel, M.; Jewell, C.; Burgman, M. Species distribution models: A comparison of statistical approaches for livestock and disease epidemics. *PLoS ONE* **2017**, *12*, e0183626. [[CrossRef](#)]
44. Piipponen, J.; Jalava, M.; de Leeuw, J.; Rizayeva, A.; Godde, C.; Herrero, M.; Kummu, M. Global assessment of grassland carrying capacities and relative stocking densities of livestock. *Earth Space Sci. Open Arch.* **2021**. [[CrossRef](#)]
45. Liu, X.; Huang, J.; Huang, J.; Li, C.; Ding, L.; Meng, W. Estimation of gridded atmospheric oxygen consumption from 1975 to 2018. *J. Meteorol. Res.* **2020**, *34*, 646–658. [[CrossRef](#)]
46. Leng, G.; Hall, J.W. Where is the Planetary Boundary for freshwater being exceeded because of livestock farming? *Sci. Total Environ.* **2021**, *760*, 144035. [[CrossRef](#)]
47. Zhang, L.; Tian, H.; Shi, H.; Pan, S.; Qin, X.; Pan, N.; Dangal, S.R. Methane emissions from livestock in East Asia during 1961–2019. *Ecosyst. Health Sustain.* **2021**, *7*, 1918024. [[CrossRef](#)]
48. El Moustaid, F.; Thornton, Z.; Slamani, H.; Ryan, S.J.; Johnson, L.R. Predicting temperature-dependent transmission suitability of bluetongue virus in livestock. *Parasites Vectors* **2021**, *14*, 382. [[CrossRef](#)]