mathematics

MDPI

Article

Error Correction Based Deep Neural Networks for Modeling and Predicting South African Wildlife–Vehicle Collision Data

Irene Nandutu [1,*], Marcellin Atemkeng [1,*], Nokubonga Mgqatsa [2], Sakayo Toadoum Sari [3], Patrice Okouma [1], Rockefeller Rockefeller [3], Theophilus Ansah-Narh [4], Jean Louis Ebongue Kedieng Fendji [5] and Franklin Tchakounte [1,6]

1 Department of Mathematics, Rhodes University, Grahamstown 6139, South Africa
2 Department of Zoology and Entomology, Rhodes University, Grahamstown 6139, South Africa
3 African Institute of Mathematical Sciences, Limbe P.O. Box 608, Cameroon
4 Radio Astronomy and Astrophysics Centre, Ghana Space Science and Technology Institute, GAEC, Accra P.O. Box LG 80 233, Ghana
5 Department of Computer Engineering, University Institute of Technology, University of Ngaoundéré, Ngaoundéré P.O. Box 454, Cameroon
6 Department of Mathematics and Computer Science, Faculty of Science, University of Ngaoundéré, Ngaoundéré P.O. Box 454, Cameroon
* Correspondence: irenenanduttu@gmail.com (I.N.); m.atemkeng@gmail.com (M.A.)

Abstract: The seasonal autoregressive integrated moving average with exogenous factors (SARIMAX) has shown promising results in modeling small and sparse observed time-series data by capturing linear features using independent and dependent variables. Long short-term memory (LSTM) is a promising neural network for learning nonlinear dependence features from data. With the increase in wildlife roadkill patterns, the SARIMAX-only and LSTM-only models would likely fail to learn the precise endogenous and/or exogenous variables driven by this wildlife roadkill data. In this paper, we design and implement an error correction mathematical framework based on LSTM-only. The framework extracts features from the residual error generated by a SARIMAX-only model. The learned residual features correct the output time-series prediction of the SARIMAX-only model. The process combines SARIMAX-only predictions and LSTM-only residual predictions to obtain a hybrid SARIMAX-LSTM. The models are evaluated using South African wildlife–vehicle collision datasets, and the experiments show that compared to single models, SARIMAX-LSTM increases the accuracy of a taxon whose linear components outweigh the nonlinear ones. In addition, the hybrid model fails to outperform LSTM-only when a taxon contains more nonlinear components rather than linear components. Our assumption of the results is that the collected exogenous and endogenous data are insufficient, which limits the hybrid model's performance since it cannot accurately detect seasonality on residuals from SARIMAX-only and minimize the SARIMAX-LSTM error. We conclude that the error correction framework should be preferred over single models in wildlife time-series modeling and predictions when a dataset contains more linear components. Adding more related data may improve the prediction performance of SARIMAX-LSTM.

Keywords: SARIMAX; SARIMAX-LSTM; modeling; LSTM; neural networks; wildlife; roadkill; endogenous variables; exogenous variables; time-series prediction; independent variables; dependent variables

MSC: 68T09; 68T07; 62-11; 68-11

check for updates

Citation: Nandutu, I.; Atemkeng, M.; Mgqatsa, N.; Toadoum Sari, S.; Okouma, P.; Rockefeller, R.; Ansah-Narh, T.; Ebongue Kedieng Fendji, J.L.; Tchakounte, F. Error Correction Based Deep Neural Networks for Modeling and Predicting South African Wildlife–Vehicle Collision Data. *Mathematics* **2022**, *10*, 3988. https://doi.org/10.3390/math10213988

Academic Editor: Alessandro Niccolai

Received: 18 September 2022
Accepted: 21 October 2022
Published: 27 October 2022

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

Copyright: © 2022 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

The number of wildlife roadkill on highways has increased significantly. In Europe alone, a total of 194 million birds and 29 million mammals are killed each year on roads, including many vulnerable species [1]. The Tropical Andes of North America have a

wildlife roadkill rate of approximately 7 animals per 100 km per day [2]. The number of wildlife roadkill has also been rampant on African highways, e.g., the Tarangire-Manyara ecosystem of Tanzania, which recorded a total of 101 roadkill in November 2013 for 10 consecutive days [3]. In South Africa, highway vertebrate road mortality records were captured from 2011 to 2014 and showed that at least two wildlife roadkill incidences occurred each day [4–6]. The latter studies showed sparsely distributed patterns of different dead species recorded usually on national and regional roads in the Greater Mapungubwe Transfrontier Conservation Area of South Africa.

Works on evaluating WVC spatial patterns to identify the priority of locations (hotspots) for mitigation strategies have been successful [7,8]. These hotspot locations usually record higher WVC incidences compared to non-priority sites [9]. The mitigation strategies implemented in these hotspots include intelligent systems [10–12] and traditional methods such as wildlife fencing, overpasses, and underpasses. The effectiveness of the mitigation measures on WVC has shown positive results [7,8]. However, higher rates of WVC continue to prompt researchers to predict and forecast future WVC cases using species distribution models (SDM) [13–16]. Species distribution models use wildlife roadkill locations of species and information on environmental conditions (e.g., temperature, precipitation, soil type, water depth, and land cover) to predict the wildlife roadkill risk distributions. The SDM has shown promising results in estimating the relationship between several environmental variables and forecasting where wildlife roadkill is likely to occur. However, the SDM faces some drawbacks, e.g., failing to forecast in areas where abiotic factors play a role in governing distributions, mainly because of a failure to capture the exogenous variables. In addition, the SDM records a high error rate when the data used for the prediction are small and spatially distributed. These challenges of the SDM are a strength of traditional and linear models such as the seasonal autoregressive integrated moving average with exogenous factors (SARIMAX-only), especially in the ability to predict from strongly stationary small data samples with limited sparse distributions and predict using endogenous variables and/or exogenous variables.

Previous studies have shown the potential of linear models, such as the SARIMAX-only, autoregressive integrated moving average (ARIMA-only), autoregressive moving average (ARMA-only), autoregressive (AR-only), and moving average (MA-only), to predict and forecast given their ability to build precise prediction models for a wide range of time-series data [17,18]. The SARIMAX-only is a powerful linear model for predicting exogenous variables [19,20]. However, with big data, these traditional models are prone to failure because they fail to learn from big data and the nonlinear components in the data. Recurrent neural networks (RNN), e.g., long short-term memory (LSTM-only), and optimization techniques draw their potential from the latter [21]. LSTM-only is a robust type of RNN that effectively learns long-term dependencies between items in a sequence and can also support large architectures during training. LSTM-only can mimic the behavior of randomly generated data over time via complex units with various components such as weights, activation functions, and backward and forward neuron connections. It is worth noting that AR-only, MA-only, ARMA-only, ARIMA-only, and SARIMAX-only can capture linear components and are effective on small stationary datasets. In contrast, LSTM-only can capture nonlinear details and are effective on big data. It is, therefore, imperative to investigate a type of model that can capture both linear and nonlinear components in big or small data regimes with exogenous variables. This work investigates the latter and proposes a hybrid architecture that captures the advantages of both the SARIMAX-only and LSTM-only models, which we refer to as SARIMAX-LSTM. The hybrid SARIMAX-LSTM uses the residual error components from the SARIMAX-only and corrects it using the LSTM-only to obtain corrected predicted time series. Hybrid models have proven to be effective by providing acceptable predictions for different indicators such as economic, financial, energy, demography, etc. [22–29]. However, these algorithms have not focused on predicting and forecasting wildlife roadkill, with only a few attempts made at predicting and forecasting fish migration using SARIMAX-only [30]. In addition, we find that the literature focuses less on SARIMAX-LSTM, ignoring research gaps in hybrid SDM-LSTM

in forecasting. Implementing a hybrid SDM-LSTM is possible due to emerging work on the hybrid PCA-LSTM model, which shows promising results [31,32].

The contributions of this work are:

- A literature review of SDM models is discussed and a comprehensive survey of the SDM models in terms of their performances, strengths, and limitations is provided in Appendix A.
- We present a mathematical framework using matrix formalization to describe the AR-only, ARIMA-only, MA-only, ARMA-only, and SARIMAX-only linear models. The matrix formalization can improve computational capabilities compared to their functional form.
- We propose an error correction modeling approach based on SARIMAX-only and LSTM-only to capture the linear and nonlinear components in wildlife roadkill patterns. The error correction model is trained on daily wildlife roadkill cases from 2011 to 2012 recorded in South Africa. The LSTM-only, SARIMAX-only, and SARIMAX-LSTM models incorporate the three ecological seasons of South Africa in predicting WVC.

The study checked the possibility of improving the prediction performance on limited taxonomic WVC datasets. The taxonomic data contains endogenous variables (incidences of daily roadkill events) and exogenous data (related to the event). Further, exploratory analysis of time-series and exogenous data is conducted to develop effective prediction and modeling of time series using the SARIMAX-only, LSTM-only, and SARIMAX-LSTM-based architectures. The model results show acceptable performance compared to the SDM in predicting sparsely distributed small and/or big datasets. As per our knowledge, this is the first study incorporating exogenous features from the taxonomic WVC dataset in the wildlife conservation domain, which contributes to mitigating WVC.

The paper's organization is as follows. Section 2 discusses the SDM models used to model wildlife roadkill. Section 3 discusses the mathematical frameworks for the AR-only, MA-only, ARIMA-only, ARMA-only, SARIMAX-only, and LSTM-only models and the proposed error correction mathematical frameworks. Section 4 describes the data collection process and evaluation metrics. Section 5 discusses the data analysis, data transformation and stationarity, and parameter selection. Section 6 provides the model results, Section 7 discusses the significance of the results, and Section 8 concludes the study and suggests future directions.

2. Related Work

Roadkill risk is determined along an urban–rural gradient using datasets such as the wildlife roadkill citizen science database and data from other organizations, e.g., the Cardiff University otter project (https://www.cardiff.ac.uk/otter-project accessed on 15 September 2022) and national biodiversity network Atlas (https://registry.nbnatlas.org/public/show/dp205 accessed on 15 September 2022) [33]. The results from the study determined the correlations in the data by helping to assess how the roadkill risk varied along an urban–rural gradient for British wildlife. The outcomes stated the abiotic factors that underlie the roadkill risk and determine if the roadkill risk is a function of species distribution. The study is significant since roadkill risk was positively associated with road density. However, the latter study experienced weaknesses such as high costs and difficulty in implementing road mitigation strategies in dense urban road networks due to increased roadkill. A logistic regression model was implemented in [13] to predict hotspots of snake road mortality. The study used a dataset for snake road mortality in southeastern Ohio for over 15 years (i.e., from 2003 to 2018). The results showed that the average number of roadkill per survey was highest in September 2011 and lowest in April 2003. High roadkill numbers were more prevalent on state routes and low roadkill numbers were recorded more on township routes. The study was substantial in associating high-risk areas with high-traffic roads throughout the study area including in regions not well covered by the surveys. It also strongly showed that the proportion of pasture, the route type, and the distance to Athens were essential in predicting road mortality. The work was significant

since it informed management and conservationists about species occurrence and potential risk population persistence. However, the study equally experienced some weaknesses such as the collection of sparse data for some species leading to the poor performance of the model. The work in [14] forecasted potential hotspots of bird causalities using generalized linear mixed models (GLMM) and two datasets (roadkill events and road density). The results showed a relative frequency of bird roadkill of zero or one, cumulative risk of bird mortality rates, road density, and elaborations of potential avian roadkill risk by combining the data on road density and the cumulative roadkill risk of bird communities. The study's strength was the ability to show a pair-wise relationship between road density and the cumulative risk of roadkill in avian species assemblages for the whole European continent. However, the work registered some weaknesses, e.g., size bias in road surveys. The size bias was not only generated from the anticipated over-representation of small birds compared to larger ones but also viewed from the underestimation of larger birds in roadkill foot surveys. Larger birds were quickly detected and removed from roads by road maintenance crews, drivers, or scavengers, which was not the case with small birds.

Bobcat (*Lynx rufus*) roadkill, land-cover, and road-trait variables datasets were analyzed to predict the vehicle-strike hotspots and mortality rates using logistic regression, spherical semivariogram, and additive algorithms [15]. The results showed that bobcat roadkill are positively related to the proportion of forests and roadways with higher traffic and four lanes, e.g., interstate highways and some United States roads. However, this work found higher bobcat mortality on roadways with high traffic and four lanes. Habitat suitability and landscape connectivity were compared in [16] to determine the roadkill risk at a regional scale for European badger species (*Meles meles*). The data were composed of data on living animals from camera trapping and roadkill data from a citizen science initiative. The dataset was input into the SDM models, e.g., the habitat suitability model (HSM) and landscape connectivity model (LCM). Both the HSM and LCM models were used as predictors with road characteristics to fit a roadkill risk model. This study showed that landscape connectivity is more important than habitat suitability in determining the roadkill risk for the Eurasian badger. It also showed that higher risk values frequently occur in hillside areas and valleys with large urban surfaces. In addition, the study found that the density of regional roads is an essential variable in determining roadkill. The study's strength was that both the HSM and LCM models determined the roadkill risk models. However, the roadkill risk model was weak as it exhibited a poor-to-no relationship with the distance from urban areas and the density of highways.

In summary, we have discussed the SDM models used to predict wildlife roadkill with a comparative study that highlights each study's weaknesses, strengths, and performances, as shown in Appendix A.

3. Materials and Methods

3.1. Autoregressive, MA-Only, ARMA-Only, ARIMA-Only, and SARIMAX-Only

Autoregressive is a time-series model that uses the observations $\mathbf{X}_{t-1}$ from the previous time steps as the input for a regression equation to predict the value $\mathbf{X}_t$ at the next time step. Here, we represent the observed data in matrix formalism $\mathbf{X}_t$ of size $m \times n$. Using the terminology in [34], an AR model in matrix formalism and for p previous observations can be calculated as

$$\mathbf{X}_t = \mathbf{E}_t + \sum_{i}^{p} \mathbf{A}_i \mathbf{X}_{t-1} \mathbf{B}_i, \tag{1}$$

where $\mathbf{A}$ of size $m \times m$ and $\mathbf{B}$ of size $n \times n$ are the matrices containing the autoregressive coefficients. The matrix $\mathbf{E}_t$ represents the white noise during observation t. We refer the

reader to [34] for a detailed interpretation of **A** and **B**. Rather than using the past values of the forecast variable in a regression, an MA model uses past forecast errors:

$$\mathbf{X}_t = \mathbf{E}_t + \sum_{i}^{q} \mathbf{C}_i \mathbf{E}_{t-1} \mathbf{D}_i. \tag{2}$$

The entries of the matrices **C** and **D** are now the model's parameters for the q previous observations. The entries in $\mathbf{E}_t$ are not the observed values but are the white noise error coefficients from the forecasting at time t, thus making Equation (2) non-causal to a regression in its actual sense.

The ARMA-only model implements the AR and MA models in a single regression workflow, i.e., it explains a time series based on its past values and the forecast error coefficients for predictions. The ARMA-only can be modeled in matrix formalism as

$$\mathbf{X}_t = \mathbf{E}_t + \sum_{i}^{p} \mathbf{A}_i \mathbf{X}_{t-1} \mathbf{B}_i + \sum_{i}^{q} \mathbf{C}_i \mathbf{E}_{t-1} \mathbf{D}_i. \tag{3}$$

The ARIMA-only is modeled in matrix formalism as

$$\mathbf{\Delta^d X}_t = \mathbf{E}_t + \sum_{i}^{p} \mathbf{A}_i \mathbf{\Delta^d X}_{t-1} \mathbf{B}_i + \sum_{i}^{q} \mathbf{C}_i \mathbf{E}_{t-1} \mathbf{D}_i, \tag{4}$$

where $\Delta^d = 1 - L$ is the difference operator with L the lag operator and d the order of integration.

The SARIMAX-only model is an extension of the ARIMA-only model. It further captures exogenous variables, a technique not implemented in the AR, MA, ARMA, and ARIMA models. The SARIMXA-only model explains a time series based on its past seasons and the forecast error coefficients $\mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_{t-s}$ for predictions. We model the SARIMAX-only in matrix formalism by adding the exogenous variable $\mathbf{S}_t$ on the right-hand side, with s as the previous seasonal time lag:

$$\mathbf{\Delta^d X}_t = \mathbf{S}_t + \mathbf{E}_t + \sum_{i}^{p} \mathbf{A}_i \mathbf{\Delta^d X}_{t-s-p} \mathbf{B}_i + \sum_{i}^{q} \mathbf{C}_i \mathbf{E}_{t-s-q} \mathbf{D}_i, \tag{5}$$

where s is the seasonal time lag.

3.2. Long Short-Term Memory and Optimization via Gradient Descent

A family of neural networks that is efficient for processing sequential data is RNNs, and LSTMs are a type of RNN that is suitable for learning long-term dependencies from sequential data including time series. LSTM comes with the idea of self-loops that produce paths where the gradient can flow for long durations. The memory block component is the key part that enhances the LSTM's capability to model these long-term dependencies. The memory block is a recurrently connected sub-network that contains memory cells and gates. The purpose of the memory cell is to remember the temporal state of the neural network and the gates are a set of multiplicative units that control the pattern of the flow of information [35]. This is summarized mathematically as follows: for an input vector $\mathbf{x}_t$, the forget gate is a nonlinear representation that determines the information that should be deleted from the cell state:

$$f_t = \sigma\left(\mathbf{W}_f^x \mathbf{x}_t + \mathbf{W}_f^h \mathbf{h}_{t-1} + \mathbf{b}_f\right), \tag{6}$$

where $\mathbf{W}_f^x$ and $\mathbf{W}_f^h$ are the weight matrices for the forget gate that, respectively, connect to the input vector and the previous short-term state $\mathbf{h}_{t-1}$. The bias term of the forget gate is represented here by $\mathbf{b}_f$ and σ is an element-wise sigmoid activation function. The input gate is an independent gate that decides the values to update:

$$i_t = \sigma\left(\mathbf{W}_i^x \mathbf{x}_t + \mathbf{W}_i^h \mathbf{h}_{t-1} + \mathbf{b}_i\right), \tag{7}$$

where $\mathbf{b}_i$ is the bias term of the input gate; $\mathbf{W}_i^x$ and $\mathbf{W}_i^h$ are the weight matrices for the input gate that connect to the input vector and the previous short-term state, respectively. The hyperbolic tangent activation function *tanh* is used to compute a vector of new candidate values to be added to the state; we refer to this process as the new candidate network:

$$\widetilde{\mathbf{C}}_t = tanh\left(\mathbf{W}_c^x \mathbf{x}_t + \mathbf{W}_c^h \mathbf{h}_{t-1} + \mathbf{b}_c\right). \tag{8}$$

Here, $\mathbf{b}_c$ is the bias term of the new candidate network; $\mathbf{W}_c^x$ and $\mathbf{W}_c^h$ are the weight matrices for the new candidate network that connect to the input vector and the previous short-term state, respectively. The current cell state is then computed from an element-wise product of the forget gate by the previous cell state, which is then added to an element-wise product of the input gate by the new candidate network:

$$\mathbf{C}_t = f_t \cdot \mathbf{C}_{t-1} + i_t \cdot \widetilde{\mathbf{C}}_t. \tag{9}$$

The part of the cell state to output is then decided by the output gate, as defined in Equation (10), and the overall LSTM is the result of an element-wise product of the output gate by the cell gate squashed by the hyperbolic tangent activation, as shown in Equation (11).

$$o_t = \sigma\left(\mathbf{W}_o^x \mathbf{x}_t + \mathbf{W}_o^h \mathbf{h}_{t-1} + \mathbf{b}_o\right) \tag{10}$$

$$\mathbf{h}_t = o_t \cdot tanh(\mathbf{C}_t), \tag{11}$$

where $\mathbf{b}_o$ is the bias term of the output gate; $\mathbf{W}_o^x$ and $\mathbf{W}_o^h$ are the weight matrices for the output gate that connect to the input vector and the previous short-term state respectively.

The predicted time series is then obtained by passing Equation (11) through the forward propagation, which is then squashed by some activation, ϕ:

$$\widehat{y} = \phi\left(\mathbf{W}_y \mathbf{h}_t + \mathbf{b}_y\right). \tag{12}$$

The choice of ϕ depends on the nature of the problem. Here, $\mathbf{b}_y$ and $\mathbf{W}_y$ are the bias term and the weight matrice for the forward propagation. Let, $\mathcal{J}_{\mathbf{W},\mathbf{b}}$ be the cost function with parameters $\mathbf{W}$ and $\mathbf{b}$ defined to minimize the error for the prediction in Equation (12). In this notation, $\mathbf{W}$ and $\mathbf{b}$ represent all the weight matrices and bias parameters for the forget gate, the input gate, the new candidate network, the output gate, and the forward propagation, respectively. The gradient at the time period t of $\mathcal{J}_{\mathbf{W},\mathbf{b}}$ with respect to each of the parameters $\mathbf{W}_k$ and $\mathbf{b}_k$ is given by

$$s_{t,k} = \Delta \mathcal{J}_{\mathbf{W},\mathbf{b}}. \tag{13}$$

The stochastic gradient descent [36] will then minimize $\mathcal{J}_{\mathbf{W},\mathbf{b}}$ by updating each of the parameters $\mathbf{W}_k$ and $\mathbf{b}_k$ in the reverse direction of $s_{t,k}$:

$$\mathbf{W}, \mathbf{b}\Big|_{t+1,k} = \mathbf{W}, \mathbf{b}\Big|_{k,j} - \alpha s_{t,k}, \tag{14}$$

where α is the learning rate that provides the step size toward the local minimum.

3.3. Error Correction Mathematical Framework

In this section, we present the mathematical framework of the proposed approach. Note that the mathematical framework provides a general framework for correcting the prediction errors from any linear model using the LSTM in time-series data. One can decompose time-series data into its linear, x_t^l, and nonlinear x_t^n components including the noise $\epsilon_\mathbf{t}$ present in the data:

$$\mathbf{x}_t = \mathbf{x}_t^l + \mathbf{x}_t^n + \epsilon_\mathbf{t}. \tag{15}$$

Assuming that the time series is passed through a filter ψ to filter out the noise and/or outliers, we have

$$\psi(\mathbf{x}_t) = \mathbf{x}_t^l + \mathbf{x}_t^n. \tag{16}$$

If we give the filtered time-series data to a linear predictor $\mathcal{L}$ to predict only the linear component, then

$$\widehat{\mathbf{x}}_t^l = \mathcal{L}(\psi(\mathbf{x}_t)), \tag{17}$$

which subtracts from the filtered time-series data the residual error from the linear prediction:

$$\mathcal{R}_t = \psi(\mathbf{x}_t) - \widehat{\mathbf{x}}_t^l. \tag{18}$$

Equation (18) is nonlinear and now represents the input vector of the nonlinear predictor $\mathcal{N}$

$$\widehat{\mathbf{x}}_t^n = \mathcal{N}(\mathcal{R}_t). \tag{19}$$

The result of the prediction or forecast of $\mathbf{x}_t$ can now be written as

$$\widehat{\mathbf{x}}_t = \widehat{\mathbf{x}}_t^l + \widehat{\mathbf{x}}_t^n + \epsilon_{\mathbf{t}}. \tag{20}$$

The noise is not modeled but is present in both $\mathbf{x}_t$ and $\widehat{\mathbf{x}}_t$.

4. Data Description and Evaluation Metrics

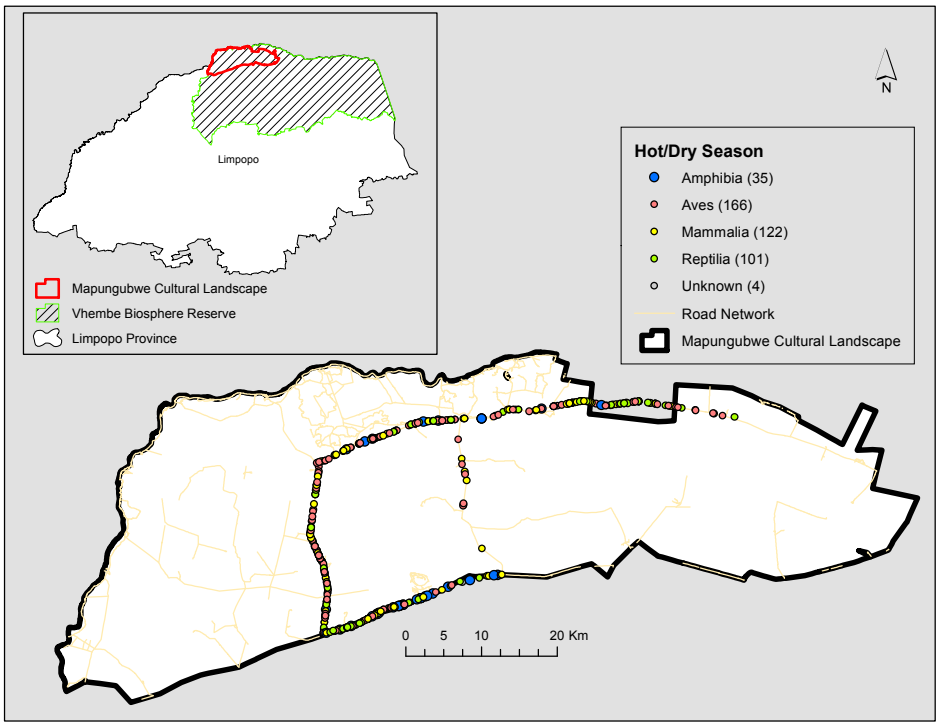
4.1. Data Collection

We used published daily wildlife roadkill data previously collected from 2 October 2011 to 4 July 2012 by the Endangered Wildlife Trust (EWT) (https://www.ewt.org.za/ accessed on 5 May 2022), a wildlife conservation organization in Johannesburg, South Africa. These data were collected from Mapungubwe Cultural Landscape, Vhembe Biosphere Reserve, and Limpopo Province in South Africa. From these data, we extracted the date, genus, class, and individual count of species to generate twelve datasets that contained six different taxa such as Mammalia, Aves, Reptilia, *Lepus saxatilis*, *Caprimulgus*, and *Numida meleagris*, as shown in Table 1. It was important to split the data into different taxa because animals are biological species grouped into meaningful biological and evolutionary units. The groups are associated with different animal behaviors regarding mating, migration, and foraging. These species possess different characteristics, thus splitting the data into different taxa helped the model extract meaningful information and minimized bias while making predictions of WVC cases. Each taxon was selected because it had the highest total count of roadkill. Further, the original samples were re-sampled to generate the daily data by counting the total number of wildlife roadkill events registered daily. Zero entries observed in the data mean no roadkill was recorded. These details were represented as a multivariate time series for making model predictions. We observed that some species were killed more than others over the two-year period during the three ecological seasons, with Aves taking the lead, as shown in Table 1. The datasets used were not equally distributed, with many zero values, especially for *Caprimulgus* and *Numida meleagris*. Non-stationary datasets were accurately predicted using LSTM-only and failed to make an accurate prediction using SARIMAX-only. The latter imposed a non-stationarity test since the proposed error prediction model draws its potential from the output of the SARIMAX-only. To test the non-stationarity properties, such as weak seasonality and trends (i.e., changing levels), in the dataset, the augmented Dickey–Fuller test (ADF), the autocorrelation function (ACF), and the partial autocorrelation function (PACF) were implemented and are discussed in Section 5.1.

The data showed that different taxa were killed in the three different ecological seasons (Figure 1); during the hot–dry season, amphibians (35), mammals (122), reptiles (101), birds (166), and four unknown wildlife species were killed. In the hot–wet season, amphibians (7), mammals (96), reptiles (127), birds (216), and three unknown species died from vehicle

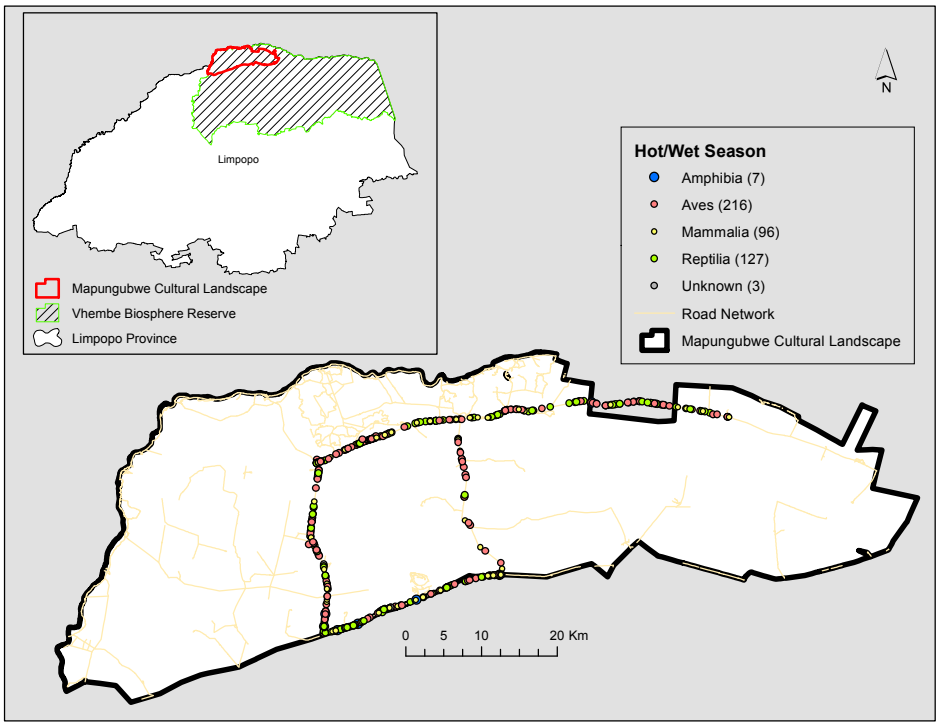
collisions. Finally, we observed a similar death rate in the cold–dry season for mammals (82), reptiles (2), birds (65), and one unknown species. Other similar datasets and research related to [4] in South Africa are [5,6].

There were 1027 wildlife roadkills recorded in the three ecological seasons, comprising amphibians (4.08%), reptiles (22.39%), birds (43.52%), mammals (29.21%), and unknown taxa (0.77%). Of these taxa, birds and mammals were more likely to be killed on the roads compared to other taxa. These results could be attributed to the behavioral patterns (e.g., sociality, home range, searching for mates and food, etc.) of birds and mammals, which make them susceptible to vehicle collisions when crossing roads near protected areas (see [3]). These data show that the probability of roadkill for amphibians was higher during the hot–dry season compared to the hot–wet and cold–dry seasons. Amphibians are ectotherms, thus they are most active during warm weather conditions, and the use of roads during this season might increase as some species might use roads for thermoregulation or when migrating to foraging or breeding sites (see [3,37]). The probability of roadkill for reptiles, birds, and mammals was higher during both the hot–dry and hot–wet seasons compared to the cold–dry season; these trends can be explained by changes in the activity patterns of reptiles, birds, and mammals during cold–dry seasons to minimize metabolic costs given that food resources are scarce during this season [38–40].

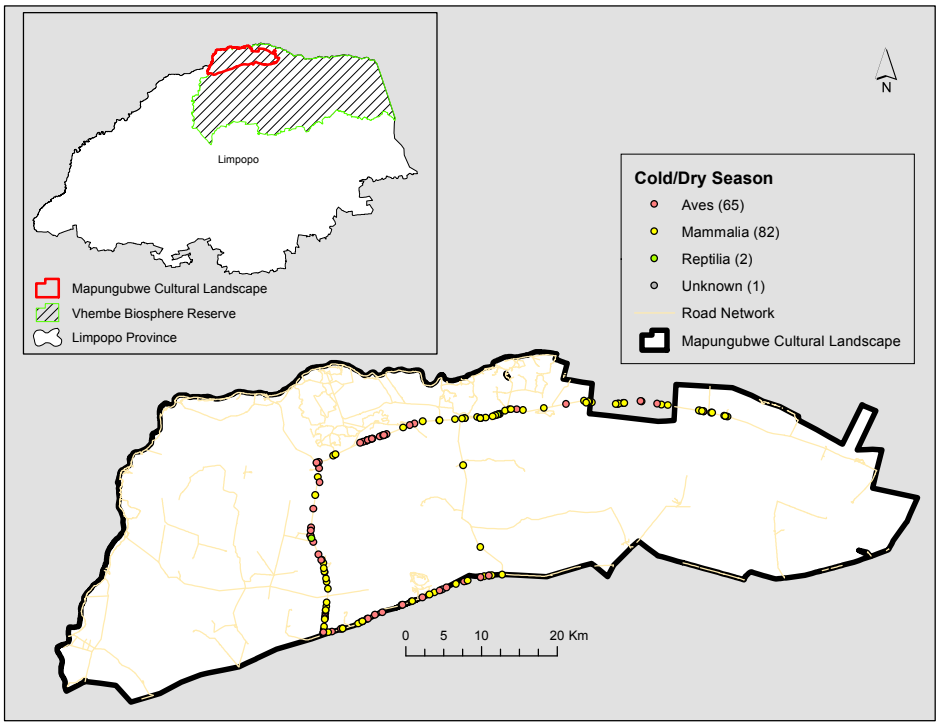


(**a**) Hot–Dry Season

Figure 1. *Cont.*



(**b**) Hot–Wet Season



(**c**) Cold–Dry Season

Figure 1. The dataset is composed of the three ecological seasons of South Africa, which are hot–dry (**a**), hot–wet (**b**), cold–dry (**c**).

These data demonstrate that roads near protected areas increase the probability of death for wild animals through vehicle collisions, and seasons and habitat type play an important role in influencing animal activities and consequently their susceptibility to roadkill (see [6]). These data are in agreement with previous studies [3]. We observed that Mammalia took the lead in the cold/dry season and the concentration of dead animals was observed to be similar on many roads. This is likely due to the different activity patterns of each taxon.

Table 1. Original and Re-sampled Dataset.

Data Collected per Taxa		
Dataset	***Original Samples***	***Re-sampled Daily Data***
Mammalia	300	277
Aves	477	277
Reptilia	230	267
Lepus saxatilis	117	275
Caprimulgus	67	258
Numida meleagris	69	261

4.2. Performance Evaluation Metrics

Evaluation metrics such as the mean absolute error (MAE) and the root mean square error (RMSE) were used in this study [41]. The mean absolute error measures the average absolute difference between the observed and predicted outcomes. In contrast, the RMSE measures the square root of the average squared difference between the observed and predicted outcomes. The lower the MAE and/or the RMSE, the higher the prediction accuracy. In addition, the MAE is less sensitive to outliers compared to the RMSE. Mathematically, the MAE and RMSE are given as

$$MAE = \frac{\sum_{t=1}^{n} |\mathbf{x}_t - \hat{\mathbf{x}}_t|}{n} \tag{21}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\mathbf{x}_t - \hat{\mathbf{x}}_t)^2}, \tag{22}$$

where n is the order of the time-series model. The AIC metric was also considered to identify the most accurate model [42]; the smaller the AIC, the more accurate the model. The AIC is defined mathematically as follows:

$$AIC = \frac{1}{n}(k - ll), \tag{23}$$

where k is the number of parameters in the model, i.e., the number of variables in the model plus the intercept, and ll is the model's log-likelihood calculated on the training data. The log-likelihood measures the best fit; the higher the log-likelihood, the better the fit. In addition, the Pearson correlation coefficient was also used as a metric to evaluate performance [43]. The Pearson correlation coefficient measures the strength and direction of the linear relationship between two variables. This coefficient has a value between +1 and -1, where +1 is a perfect positive linear correlation, 0 is no linear correlation, and -1 is a perfect negative linear correlation. The Pearson correlation coefficient is given as

$$r = \frac{\sum_{t=1}^{n}(\mathbf{x}_t - \bar{\mathbf{x}})(\hat{\mathbf{x}}_t - \bar{\hat{\mathbf{x}}})}{\sqrt{\sum_{t=1}^{n}(\mathbf{x}_t - \bar{\mathbf{x}})^2(\hat{\mathbf{x}}_t - \bar{\hat{\mathbf{x}}})^2}}, \tag{24}$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the actual and predicted values, respectively; $\mathbf{x}_t$ and $\hat{\mathbf{x}}_t$ are the values of $\mathbf{x}$ and $\hat{\mathbf{x}}$ at time step t; $\bar{\mathbf{x}}$ and $\bar{\hat{\mathbf{x}}}$ are the mean values of $\mathbf{x}$ and $\hat{\mathbf{x}}$, respectively. Note that n is

the total sample of $\mathbf{x}$ or $\hat{\mathbf{x}}$. Table 2 shows the rule of thumb for interpreting the size of the correlation coefficient.

Table 2. Guidelines that help interpret the result size of the Pearson correlation coefficient.

Rule of Thumb for Interpreting the Size of a Correlation Coefficient [44]	
Size of Correlation	***Interpretation***
0.90 to 1.00 (−0.90 to −1.00)	Very high positive (negative) correlation
0.70 to 0.90 (−0.70 to −0.90)	High positive (negative) correlation
0.50 to 0.70 (−0.50 to −0.70)	Moderate positive (negative) correlation
0.30 to 0.50 (−0.30 to −0.50)	Low positive (negative) correlation
0.00 to 0.30 (0.00 to −0.30)	negligible correlation

5. Data Analysis and Transformation

LSTM-only is straightforward to apply since it requires non-transformed time-series data as input, with parameter tuning using grid search. The model can minimize exploding gradients by utilizing the power of gates, as described in Section 3. Unlike SARIMAX-only, SARIMAX-LSTM data need to be transformed.

5.1. Data Transformation and Stationarity: Augmented Dickey–Fuller and Autocorrelation Tests

Table 3 shows two kinds of original datasets, i.e., stationary and non-stationary datasets. By observing the non-stationary datasets for Mammalia, Aves, Reptilia, *Lepus saxatilis*, and ecological seasons in Table 3, some weak seasonality and trends can be seen in the data. We used the ADF to test the data stationarity to justify the latter. Given a time-series data $\mathbf{x}_t$, the ADF uses the ordinary least squares estimator $\hat{\rho}$ of ρ to test the null hypothesis; mathematically, this is a regression problem with the form

$$\mathbf{x}_t = \rho \mathbf{x}_{t-1} + \sum_{i=1}^{m} a_{j,m} \Delta \mathbf{x}_{t-i} + \mathbf{e}_{t,m}, \tag{25}$$

where $\Delta \mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$, $\mathbf{e}_{t,m}$ is the white noise, $a_{j,m}$ is a constant, and $m = m(n)$ is the order of the regression that depends on n, the number of sequences in the data. Under the null hypothesis, $\rho = 1$ means the data are non-stationary and $\rho < 1$ means the data are stationary. We used the ρ-value as a probability to confirm the assumption imposed by the null hypothesis. Stronger evidence favoring the hypothesis would require a smaller ρ-value. When we tested the ADF on the original datasets, the results showed that the *Caprimulgus* and *Numida meleagris* datasets were stationary, whereas the rest were non-stationary. The null hypothesis was not rejected for Mammalia, Aves, Reptilia, season, or *Lepus saxalitis* since the data were non-stationary, while the null hypothesis was rejected for *Caprimulgus* and *Numida meleagris* since the data were stationary. We transformed the data by taking the first differencing for Mammalia and *Lepus saxalitis*, as well as transforming Reptilia and Aves, which we differenced with a window size of −4 and 9, respectively. These techniques helped to attain a stationary dataset with a constant mean and variance. A stationary dataset means the data are smooth with no trend, which can improve the models' overall performance. In this study, the transformed data were also tested using ADF. The null hypothesis was rejected since the data were stationary and the ρ-values were smaller than 0.05 and 0.1, which was below the selected threshold and significance level (i.e., 5% and 10%). When performing ADF, the ADF statistic values, as shown in Table 3, were also vital in determining whether a data sample was stationary.

The ADF test statistic values should always be smaller than the critical values at either or all 10% (0.1), 5%, (0.05), or 1% (0.01) in a stationary dataset. For this work, we chose a critical value(s) at 10% (0.1) and 5% (0.05) for the six taxa represented. As such, the more negative the ADF statistic value, the more likely the data were stationary. In Table 3, it

can be seen that the ADF statistic value of the transformed Mammalia data, which was −2.84, was less than the critical value of −2.57 at 10%; the transformed Aves dataset's ADF statistic value of −2.58 was also less than −2.57 at 10%; the ADF statistic value of the Reptilia transformed dataset was −2.98, which was less than the critical values of −2.87 at 5% and −2.57 at 10%; and for the transformed *Lepus saxatilis*, the ADF statistic value was −9.83, which was much lower than the critical values at 10%, 5%, and 1%. In addition, it showed an extremely lower ρ-value. The original datasets for *Caprimulgus* and *Numida meleagris* were stationary with ADF statistic values of −3.00 and −2.92, respectively. Their ADF statistic values were less than their critical values at both 5% and 10%. The results from the stationary test showed that all data from the different taxa were stationary. The ρ-values were equal to 0.05, 0.09, 0.03, 4.88×10^{-17}, 0.03, and 0.04 for Mammalia, Aves, Reptilia, *Lepus saxatilis*, *Caprimulgus*, and *Numida meleagris*, respectively, which were less than the 10% and 5% thresholds.

Table 3. Testing for stationarity in the datasets using ADF: results from the original and transformed series. Mammalia transformed data, *Caprimulgus* original data, and *Numida meleagris* original data show all the endogenous or dependent variables that are stationary with a critical value of 5% before or after differencing, Aves transformed data shows the endogenous variable that is stationary with a critical value of 10% even after differencing, *Lepus saxatilis* Transformed data shows the endogenous variable that is stationary with a critical value of 1% even after differencing, and the remaining series shows the non-stationary endogenous and exogenous series. The variables hot/wet, cold/dry, and hot/dry are the ecological seasons. These exogenous variables failed to become stationary even after differencing for all taxa.

Dataset	*ADF Statistic*	*ρ-Value*	*Critical Value (1%)*	*Critical Value (5%)*	*Critical Value (10%)*
Mammalia Original Data	−2.70	0.07	−3.45	−2.87	−2.57
Mammalia Transformed data	−2.84	0.05	−3.45	−2.87	−2.57
Aves Original Data	−2.53	0.10	−3.45	−2.87	−2.57
Aves Transformed data	−2.58	0.09	−3.45	−2.87	−2.57
Reptilia Original Data	−2.45	0.12	−3.45	−2.87	−2.57
Reptilia Transformed data	−2.98	0.03	−3.45	−2.87	−2.57
Lepus saxatilis Original Data	−1.82	0.36	−3.45	−2.87	−2.57
Lepus saxatilis Transformed data	−9.83	4.88×10^{-17}	−3.45	−2.87	-2.57
Caprimulgus Original Data	−3.00	0.03	−3.45	−2.87	−2.57
Numida meleagris Original Data	−2.91	0.04	−3.45	−2.87	−2.57
hot/wet (season)	−1.42	0.57	−3.45	−2.87	−2.57
cold/dry (season)	−0.36	0.91	−3.45	−2.87	−2.57
hot/dry (season)	$t - 1.12$	0.70	−3.45	−2.87	−2.57

Apart from the ADF test, data stationarity was also determined by the autocorrelation function (ACF) and partial autocorrelation function (PACF). The autocorrelation function is a statistical measure that indicates how a time series relates to itself over time. For any

autocorrelation at a given lag for a series, the correlation between the original series and the same original series moved forward in time at a specified number of periods or lags is computed. The partial autocorrelation function is also a statistical measure similar to the ACF that is used to evaluate relationships among the series values. The PACF operates in combination with the ACF to produce autoregressive and moving-average relationship patterns. The PACF is an extension of the ACF, where any intermediate elements are removed. For example, in the PACF, to determine the correlation between the original series and the series time lag at 2, the series time lag at 1 is removed. However, when there are no intermediate elements within the lag, the PACF is equivalent to the ACF. Consider the guidelines in Table 4, where $AR(p)$ is the AR-only model with lag p, $MA(q)$ is the MA-only with lag q, and $ARMA(p,q)$ is used for the ARIMA-only. If the ACF and PACF do not satisfy the criteria of the guidelines in Table 4, there is a high probability that the series are non-stationary, simply because the models and orders cannot be determined. The total samples for each taxon were divided into training and testing sets. All the samples shown in Table 5 were used for training, except for the last 20 series that were used for testing. The total numbers of samples in the original and transformed datasets were not equal because a moving average with different window sizes was applied to each original dataset for each taxon to obtain stationarity.

Table 4. Guidelines for how to choose the $MA(q)$, $AR(p)$, and $ARMA(p,q)$ models and their associated parameters based on the ACF and PACF.

Choosing Model Order			
	AR(p)	***MA(q)***	***ARMA(p, q)***
ACF	Tails off	Cuts off at lag q	Tails off
PACF	Cuts off at lag p	Tails off	Tails off

Note that different samples were discarded before SARIMAX-only during the stationarity filtering process. Alternatively, LSTM-only did not discard any samples during data preparation and training. To obtain a stationary and supervised time series for model prediction, SARIMAX-LSTM also lost different samples for each taxon. We observed and compared the total number of samples for each taxon in the original dataset, which can be seen in Table 1, and the datasets for supervised training, which can be seen in Table 5. The effect of the few discarded samples was negligible and did not affect the prediction.

Table 5. Wildlife roadkill time-series data used by each model; SARIMAX is differenced to obtain a stationary dataset and LSTM is differenced with a shift of one for supervised training.

Wildlife Roadkill Time-Series Datasets per Taxa					
Dataset	***Train***	***Test***	***Total***	***Zero Entries***	***Models Used***
Mammalia Original Data	256	20	276	167	LSTM
Mammalia Transformed Data	255	20	275	167	SARIMAX-only, SARIMAX-LSTM
Aves Original Data	256	20	276	169	LSTM
Aves Transformed Data	247	20	267	169	SARIMAX-only, SARIMAX-LSTM
Reptilia Original Data	246	20	266	191	LSTM
Reptilia Transformed Data	240	20	260	191	SARIMAX-only, SARIMAX-LSTM
Lepus Saxatilis Original Data	254	20	274	199	LSTM

Table 5. *Cont.*

Wildlife Roadkill Time-Series Datasets per Taxa					
Dataset	***Train***	***Test***	***Total***	***Zero Entries***	***Models Used***
Lepus Saxatilis Transformed Data	253	20	273	199	SARIMAX-only, SARIMAX-LSTM
Caprimulgus Original Data	237	20	257	218	LSTM
Caprimulgus Original Data	237	20	257	218	SARIMAX-only, SARIMAX-LSTM
Numida meleagris Original Data	240	20	260	219	LSTM
Numida meleagris Original Data	240	20	260	219	SARIMAX-only, SARIMAX-LSTM

5.2. SARIMAX-Only and Parameter Selection

With regard to Table 4, the stated standard guidelines helped us determine the kind of ARIMA linear models that best suited a specific dataset. The SARIMAX-only contained the non-seasonal $SARIMAX(p,d,q)$ and seasonal $SARIMAX(P,D,Q)S$ parameters, where P,D,Q are the seasonal lags and S is the seasonal period. In addition, this model was determined with the help of the ACF and PACF. The ACF tailed off smoothly for the $AR(p)$ model, and the PACF cut off at lag p for the seasonal $AR(P)$ model. For the $MA(q)$ model, the ACF cut off at lag q or seasonal lag Q, and the PACF tailed off smoothly. A $SARIMAX(p,d,q) \times (P,D,Q)S$ model was determined when the ACF and PACF tailed off smoothly. The non-seasonal parameters were p, the AR component in ARIMA-only; d, the same as the I component in ARIMA, which denoted the number of times the successive differencing operation was performed to make the series stationary; and q, which was the same as the MA component in ARIMA. The seasonal parameters were S (seasonal period), which denoted the number of periods in the series after which a seasonal pattern was repeated, and P (seasonal AR component), which was a parameter that captured the effect of past lags on the current observations. Unlike p, the past lags were separated from the current lag by multiples of S lags, e.g., if $S = 3$ and $P = 2$, then the current observation was estimated using observations from the third and sixth lags. D (seasonal I component) was the number of times seasonal differences needed to be calculated to make the series stationary. Unlike d, where the differencing was determined using successive observations (first lagged series), the differencing was determined between the series and the Sth lagged series. Q (seasonal MA component) was a parameter that captured the effect of past residuals on the current observations. Unlike q, the past lags were separated from the current lag by multiples of S lags. For example, if $S = 3$ and $Q = 2$, then the current observation was estimated using the residuals from the third and sixth lags. It is worth noting that for an $AR(p)$, seasonal $AR(P)$, $MA(q)$, and seasonal $MA(Q)$, the PACF at lag p and the seasonal P, as well as the ACF at lag q and the seasonal Q, should be non-zero. As a result, each lag in an ACF and PACF represents the parameters p, seasonal P, q, and seasonal Q associated with the seasonal and non-seasonal parameters. In addition, we selected model orders automatically using the algorithm described in [45], which combines the unit root tests, minimization of the Akaike information criterion (AIC), and maximum likelihood estimation to obtain $SARIMAX(p,d,q) \times (P,D,Q)S$. The algorithm performed a grid search through all combinations of parameters p,d, and q, as well as P,D,Q, and S, at a given interval and then selected a model with the minimum AIC value. The results of the model and parameter selection in this study are shown in Table 6.

Table 6. Selecting models with the least *AIC*. Positive values mean much information is lost and negative values mean less information is lost. The smaller the *AIC* value, the lesser the information lost, and the higher the *AIC* value, the bigger the information lost.

Evaluating Models in Wildlife Roadkill Time-Series Datasets per Taxa			
Dataset	***Best Model Order***	***Total Fit Time (Seconds)***	***AIC***
Mammalia Transformed Data	ARIMA(1,1,1)(1,1,1)[3]	7.485	929.443
Aves Transformed Data	ARIMA(2,1,2)(1,1,1)[3]	41.393	1031.378
Reptilia Transformed Data	ARIMA(2,1,0)(2,1,0)[3]	5.128	913.220
Lepus Saxatilis Transformed Data	ARIMA(2,1,0)(2,1,0)[3]	7.336	809.445
Caprimulgus Original Data	ARIMA(0,1,1)(0,1,1)[3]	4.630	476.820
Numida meleagris Original Data	ARIMA(2,1,0)(2,1,0)[3]	8.185	523.346

5.3. Application of the Error Correction Framework

The transformed data were first passed into the SARIMAX-only model; the LSTM-only used the prediction errors from the SARIMAX-only to correct or calibrate the predicted output of the SARIMAX-only. In other words, the SARIMAX-only output the roadkill prediction with some error residuals; these error residuals were the inputs of the LSTM-only. The latter predicted the error residual, which was used further to correct the prediction of the SARIMAX-only. The flowchart in Figure 2 shows the sequential order in which the dataset traveled through the different machine learning models to generate the error correction model. The raw patterns of the wildlife roadkill data were transformed, as discussed in Section 5.1.

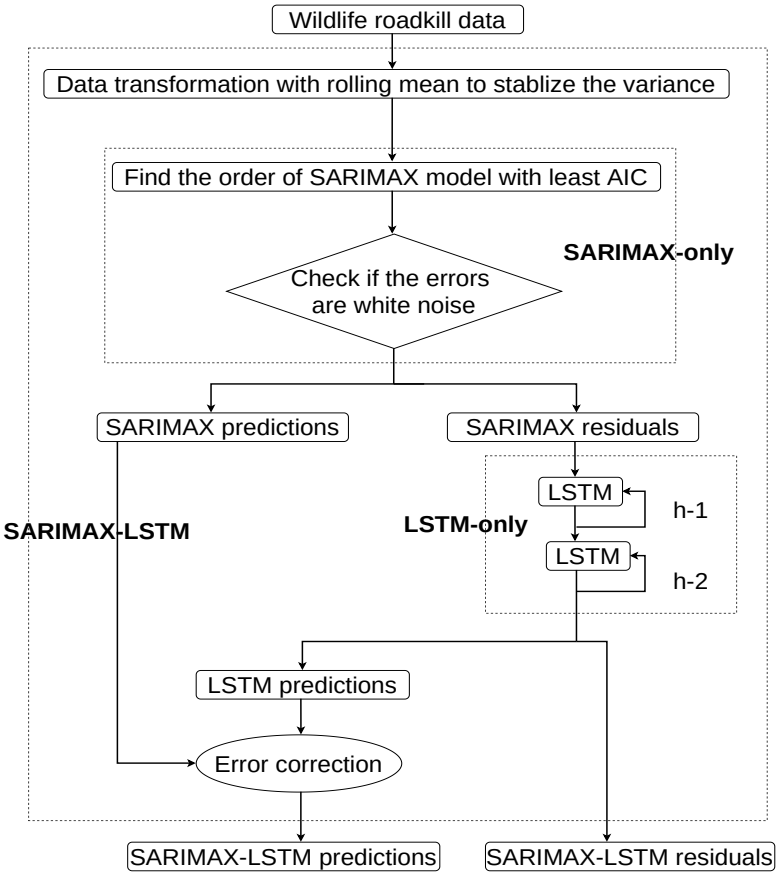


Figure 2. Flow chart showing the process for forecasting roadkill patterns using the SARIMAX-LSTM model.

The transformed data were used as input to the SARIMAX-only model. Next, the order of the SARIMAX-only model and its associated model parameters were determined, as discussed in Section 5.2. As such, SARIMAX-only predictions and residual errors were generated as model outputs. The SARIMAX-only model order that generated the residual errors, which were white noise, was considered. Otherwise, a new model order was determined. The SARIMAX-only residual errors became the input to the LSTM-only, and the LSTM-only output the LSTM-only predictions and LSTM-only residuals. Finally, the LSTM-only predictions were used to correct the SARIMAX-only predictions to generate the hybrid SARIMAX-LSTM model predictions and residuals. In this flowchart, the SARIMAX-only can be substituted by ARIMA-only, AR-only, ARMA-only, or MA-only to predict the ARIMA-LSTM, AR-LSTM, ARMA-LSTM, or MA-LSTM, respectively. This study ran experiments on only the hybrid SARIMAX-LSTM model. We selected the SARIMAX-only model because it considered external variables that ecologically impacted WVC.

The error correction model was optimized via the gradient descent algorithm, and cross-validation was used during model training. All the models had LSTM layers and one dense layer. The regularization used the sigmoid activation function to improve the performance of the LSTM-only and SARIMAX-LSTM models. All the layers for all models were associated with the number of units, as shown in Table 7. Fine-tuning was applied to the different models and executed; the models and hyperparameters seen in Table 7 were considered.

Table 7. Model architectures for LSTM-only and SARIMAX-LSTM for each model and associated taxa as well as the datasets. Each dataset used a learning rate scheduler with an initial learning rate of 1×10^{-2}, a dense output layer of 1 unit, an Adam optimizer, and a sigmoid activation function. The models show the number of layers used and the hyperparameters.

Taxa	Model	Layers	Hyper Parameters
(1) Aves (2) Mammalia (3) *Numida meleagris*	(1) LSTM-only	LSTM	64 neurons
	(2) SARIMAX-LSTM	LSTM	32 neurons
		LSTM	16 neurons
		Dropout	1
(1) Reptilia	(1) LSTM-only	LSTM	35 neurons
	(2) SARIMAX-LSTM	LSTM	15 neurons
		LSTM	5 neurons
		Dropout	0.01
(1) *Caprimulgus*	(1) LSTM-only	LSTM	30 neurons
	(2) SARIMAX-LSTM		
		LSTM	5 neurons
(1) *Lepus saxatilis*	(1) LSTM-only	LSTM	60 neurons
	(2) SARIMAX-LSTM		
		LSTM	40 neurons
		LSTM	30 neurons

6. Results

6.1. Model Identification and Estimation

In this work, we visualized the ACF and PACF for lag 40 on both the original and the transformed dependent variables of daily wildlife roadkill, as well as the exogenous variables, as shown in Figures 3–6. To obtain the p values and seasonal P values for the AR and seasonal AR for the SARIMAX model, we referred to the PACF plots and looked at the ACF plot to obtain the q value and seasonal Q.

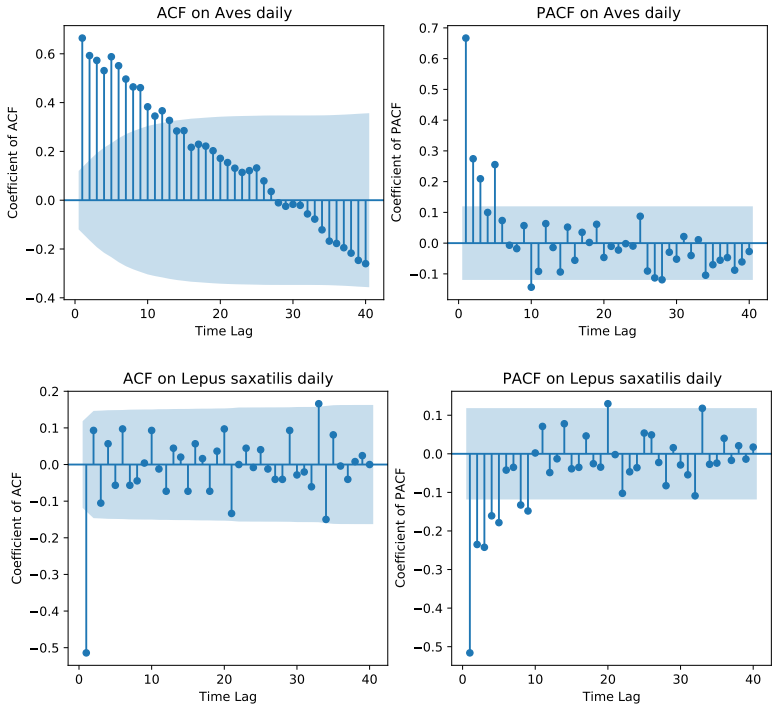


Figure 3. ACF (top-left panel) and PACF (top-right panel) on the Aves data showing stationarity. ACF (bottom-left panel) and PACF (bottom-right panel) on the *Lepus saxatilis* data showing non-stationarity on endogenous variables and over differenced *Lepus saxatilis* series with a lot of negative values.

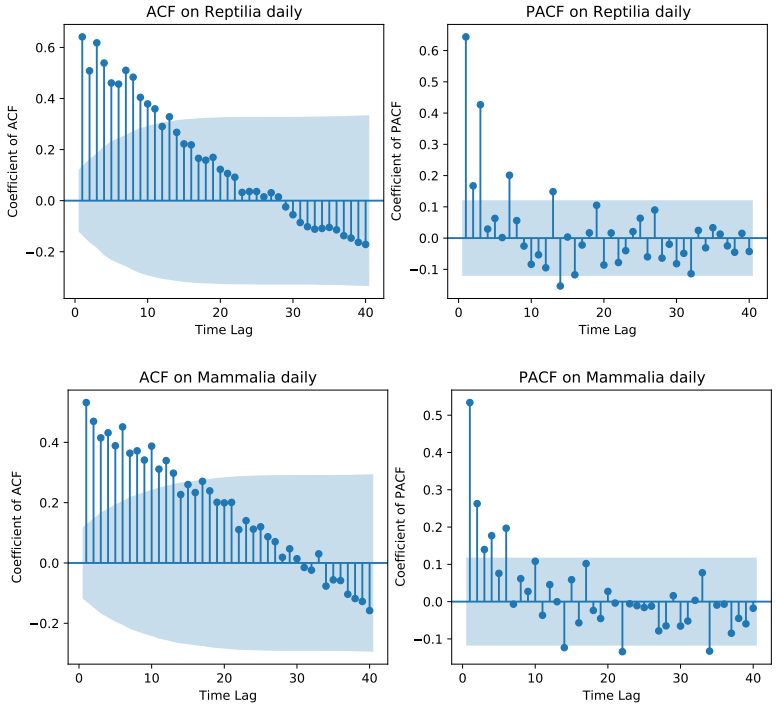


Figure 4. ACF (top-left panel) and PACF (top-right panel) on the Reptilia data showing stationarity. ACF (bottom-left panel) and PACF (bottom-right panel) on the Mammalia data showing stationarity on endogenous variables.

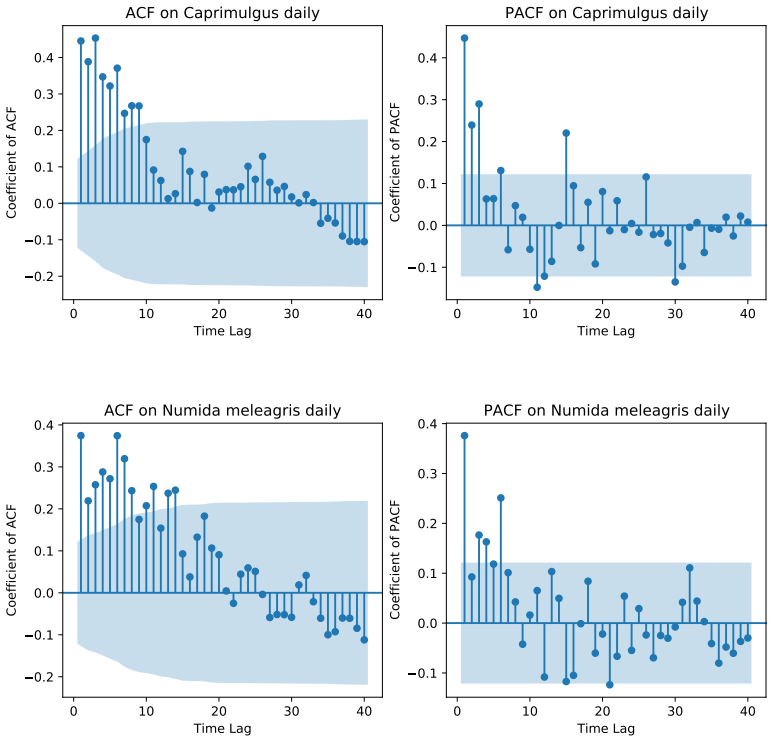


Figure 5. ACF (top-left panel) and PACF (top-right panel) on the *Caprimulgus* data showing stationarity. ACF (bottom-left panel) and PACF (bottom-right panel) on the *Numida meleagris* data showing stationarity on endogenous variables

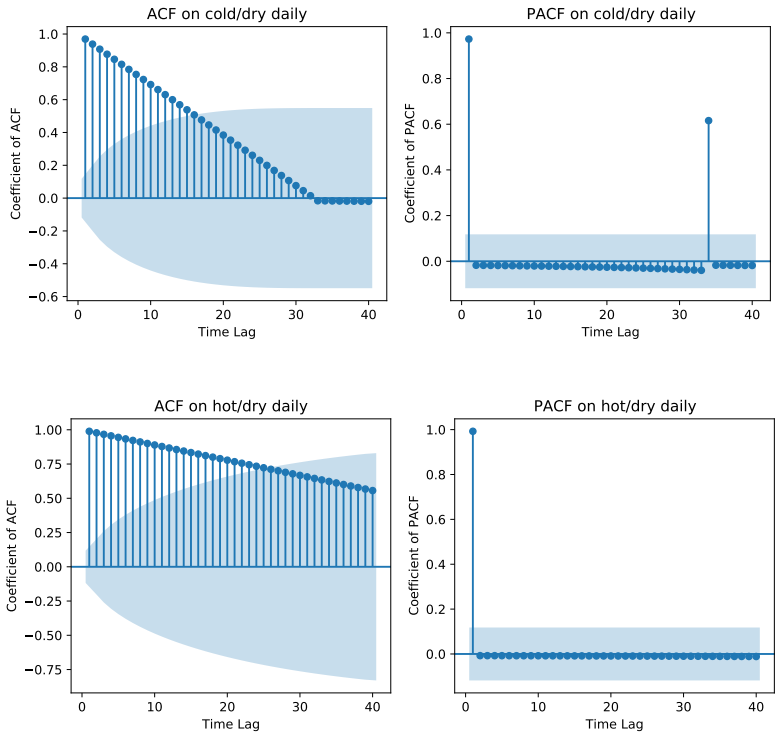


Figure 6. *Cont.*

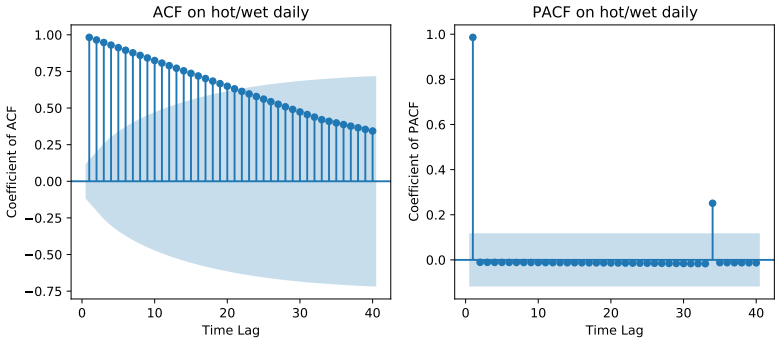


Figure 6. ACF (top-left panel) and PACF (top-right panel) on the cold/dry data showing non-stationarity. ACF (mid-left panel) and PACF (mid-right panel) on the hot/dry data showing non-stationarity. ACF (bottom-left panel) and PACF (bottom-right panel) on the hot/wet data showing non-stationarity on exogenous variables.

The ACF lags of the Aves data (Figure 3, top-left panel) strikes vary from 1 to 12, which means that we can compute MA(1), MA(2), MA(3), MA(4), MA(5), MA(6), MA(7), MA(8), MA(9), MA(10), MA(11), and MA(12), and the seasonal spikes are at lags 3, 6, 9, and 12 for MA(1), MA(2), MA(3), and MA(4), whereas the PACF lags (Figure 3, top-right panel) spike at 1, 2, 3, and 5, thus allowing us to compute AR(1), AR(2), AR(3), and AR(5), as well as the seasonal AR(1) at lag 3. The *Lepus saxatilis* data shows that the ACF lags (Figure 3, bottom-left panel) spike at 1, implying an MA(1) with no strike at the seasonal MA(0), whereas the PACF lags (Figure 3, bottom-right panel) spike at 1, 2, 3, 4, 5, 8, and 9, thus AR(1), AR(2), AR(3), AR(4), AR(5), AR(8), and AR(9), with the seasonal AR at 3, 9 for AR(1) for lag 3, and AR(2) for lag 9.

In Figure 4, the Reptilia data show stationarity with the ACF and PACF lags; the ACF lags (top-left panel) spike at 1 to 11, which means that we can compute MA(1), MA(2), MA(3), MA(4), MA(5), MA(6), MA(7), MA(8), MA(9), MA(10), and MA(11); and the seasonal spikes are at lags 3, 6, and 9 for MA(1), MA(2), and MA(3), whereas the PACF lags (top-right panel) spike at 1, 2, 3, 7, 14, and 15; thus, AR(1), AR(2), AR(3), AR(7), AR(14), and AR(15), as well as the seasonal AR(1) at lag 3, are computed. The Mammalia data show the ACF (Figure 4, bottom-left panel) spiking at 1 to 12, thus MA(1), MA(2), MA(3), MA(4), MA(5), MA(6), MA(7), MA(8), MA(9), MA(10), MA(11), and MA(12), and the seasonal spikes are at lags 3, 6, 9, 12 for MA(1), MA(2), MA(3), MA(4), whereas the PACF lags (Figure 4, bottom-right panel) spike at 1, 2, 3, and 7 and thus AR(1), AR(2), AR(3), AR(7), as well as the seasonal AR(1) at lag 3.

In Figure 5, the *Caprimulgus* data show the ACF lags (top-left panel) spiking from 1 to 7, thus MA(1), MA(2), MA(3), MA(4), MA(5), MA(6), MA(7), and MA(8), and the seasonal spikes are at lags 3, 6, 9 for MA(1), MA(2), and MA(3), whereas the PACF lags (top-right panel) spike at 1, 2, 3, 7, and 15, thus AR(1), AR(2), AR(3), AR(7), and AR(15), as well as the seasonal AR(1) at lag 3. The *Numida meleagris* ACF (bottom-left panel) spikes at 1 to 8, thus MA(1), MA(2), MA(3), MA(4), MA(5), MA(6), MA(7), and MA(8), and the seasonal spikes are at lags 3 and 6 for MA(1) and MA(2), whereas the PACF lags (bottom-right panel) spike at 1, 3, 5, and 6, thus AR(1), AR(3), AR(5), and AR(6), as well as the seasonal AR(1) at lag 3 and AR(2) at lag 6.

Generally, using ACF and PACF plots to identify model order is not straightforward, especially for the kind of data used in this case, which were composed of many nonlinear components due to the different animal behaviors witnessed on highways. To obtain a correct model order, one ensures no autocorrelation is left in the residuals by visualizing the residual correlogram until a model removes the autocorrelation. We witnessed more residual lags exceeding the confidence bands when we used manual model orders from the ACF and PACF plots discussed. Such behaviors show that our time series have characteristics of long memory, e.g., power-like decay of autocorrelation plots, which overly complicate the

ARIMA model orders. To minimize residual lags exceeding the confidence bands and overcome long memory, we used the auto_arima function in Python (https://www.python.org/ accessed on 11 July 2022), which returned the model order according to AIC, a Bayesian information criterion [46]. The function searched over possible models within the order constraints provided, and the residual lag spikes beyond the confidence bands were minimal, as shown in Figure 7.

After visualizing the ACF and PACF plots for each taxa, we fed the data to the auto_arima to fit the best SARIMAX. In this study, we considered model orders detected by the auto_arima as proposed by [45] because they have the smallest AIC, minimizing residual lags from exceeding confidence lags and minimizing long memory problems [47]. Table 6 shows the auto_arima-predicted model orders as the best model orders. The orders are ARIMA(1,1,1)(1,1,1)[3] for Mammalia, ARIMA(2,1,2)(1,1,1)[3] for Aves, ARIMA(0,1,1)(0,1,1)[3] for *Caprimulgus*, and ARIMA(2,1,0)(2,1,0)[3] for Reptilia, *Lepus saxatilis*, and *Numida meleagris*.

Even though the exogenous variables were non-stationary, as shown in Table 3, with ρ-values greater than the critical values at 1%, 5%, and 10%, we observed some spikes on the cold/dry, hot/dry, and hot/wet data, as seen in Table 6. This shows a weak AR process. Together with each endogenous variable, these variables were passed into the SARIMAX model to help in the model fitting and prediction of roadkill. The four-time series finally considered for modeling and prediction were Mammalia, Aves, *Caprimulgus*, and *Numida meleagris*; these taxa were associated with the ACF and PACF for each taxon, respectively, and are shown in Figure 8.

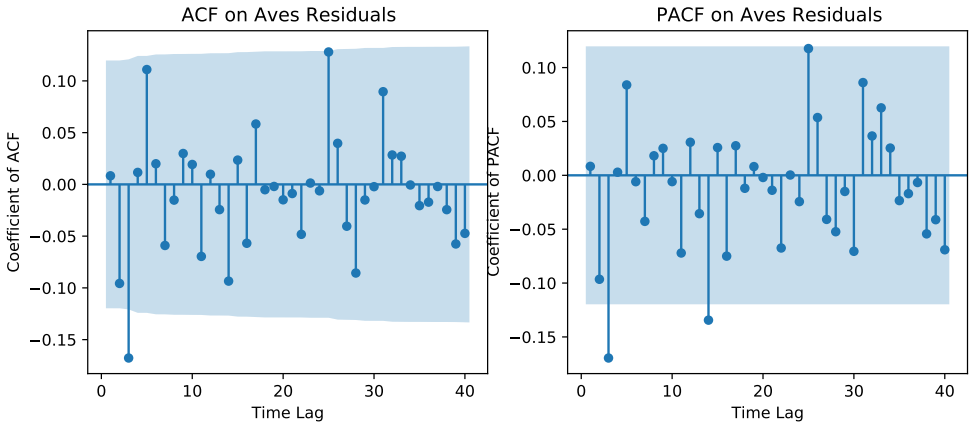


Figure 7. ACF and PACF on Aves residuals are white noise with 95% of the forecast errors below the error criteria. The Aves residuals from SARIMAX-only are the inputs of the LSTM-only.

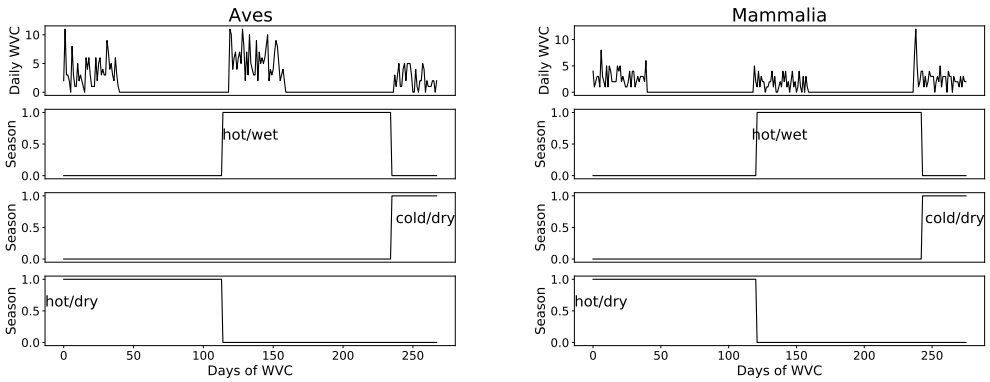


Figure 8. *Cont*.

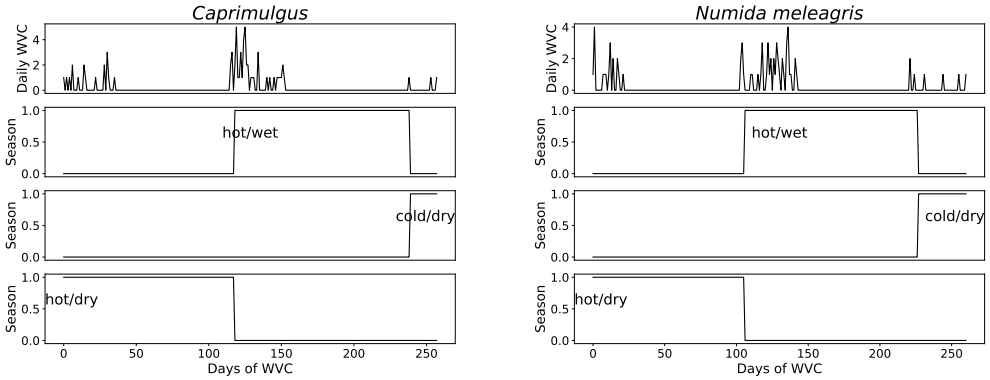


Figure 8. Taxonomic wildlife roadkill daily data showing the endogenous and exogenous variables on the Aves (top-left panel), Mammalia (top-right panel), *Caprimulgus* (bottom-left panel), and *Numida meleagris* (bottom-right panel) datasets.

6.2. Model Diagnostic Check

Diagnostic checks, such as residual and parameter diagnostics, helped to determine the model's validity. Whenever the model validity failed, alternative model (s) were suggested. For example, in each loop, different model parameters were estimated and applied until a model with the fewest errors was determined. Next, the ACF and PACF were visualized to examine if the residuals from SARIMAX were white noise. Figure 7 shows no ACF and PACF among the residuals.

6.3. Model Prediction Results

In this work, daily data on WVC on South African highways were collected from EWT. A stationary dataset was attained by transforming the original wildlife roadkill data. Next, the LSTM-only and SARIMAX-only models were used to model and predict wildlife roadkill patterns and the results were compared with our proposed error correction. For each model result in Table 8, we display the MAE, RMSE, and Pearson correlation coefficient metric in Table 2. We selected models for each taxon with very high positive (negative) correlations on the actual and predicted data as the best and focused majorly on the generalization performance of the models. The SARIMAX-only model generally performed poorly due to the high nonlinear temporal tendencies in the data compared to the linear components. The LSTM-only model obtained the best performance in three out of six experiments. The three time series were *Lepus Saxatilis*, *Numida meleagris*, and Reptilia. Although SARIMAX-LSTM was expected to outperform single models, it failed to overcome LSTM-only while predicting some taxa. The LSTM-only's best performance can be attributed to (1) The ability of LSTMs to capture long-term dependencies using the availability of the recurrent memory gates. The LSTM gates considered previous training series in their prediction process, which created a better understanding of the past data and a more robust prediction process. (2) The availability of fewer categories of exogenous variables, which limited the capability of LSTMs to learn long-term dependencies from the SARIMAX residuals. (3) The absence of adequate extreme cases in the residual training set, which caused bias in the training process toward the prediction of more regular cases. (4) The lower total number of equal distribution of samples in the cold/dry, hot/dry, and hot/wet seasons. The best-selected taxa are visualized in Table 8 in red and bold text. These are also represented in Figure 9 for SARIMAX-only and Figure 10 for LSTM-only.

The residual error of SARIMAX-only was input to an LSTM model for prediction; the results are shown in Figure 11. Notably, these residual errors were not zero everywhere, with a maximum peak of 10 wildlife roadkills for Mammalia and Aves. Figure 12 shows the results for SARIMAX-LSTM from the correction using the predicted error. However, it failed to generalize cases whose test samples had fewer extreme cases. By visualizing

these results, one can see that the predictions of the error correction methods surpassed the single-model SARIMAX-only and LSTM-only for some cases. Table 8 confirms the visual inspection with the results of the MAE, RMSE, and correlation coefficient for the different models' performances for each individual taxa. All taxa achieved very high positive (negative) correlation coefficients for either the LSTM-only or SARIMAX-LSTM models. We also recorded correlation coefficients for moderate positive (negative), low positive (negative), and negligible cases. These results are motivated by the total samples for each taxon used for training the models shown in Table 1. Table 8, models in red and bold black were considered because they showed the lowest MAE and RMSE.

Table 8. Prediction performance of models showing error values of SARIMAX-only, SARIMAX-LSTM, and LSTM-only for each taxon. For each dataset, the best values of the metrics and model are highlighted in bold and their generalization correlation coefficient is a very high positive correlation, meaning the actual and predicted data are strongly correlated, as shown in Table 2.

Taxa	Model	Dataset Sample	MAE	RMSE	Correlation Coefficient
Aves	LSTM-only	Full	0.381521	0.450312	0.986556
		Train	0.385948	0.453470	0.986903
		Test	0.324627	0.407560	0.980473
	SARIMAX-only	Full	1.224962	2.112995	0.660432
		Train	1.204622	2.141005	0.669376
		Test	1.477176	1.728364	0.263314
	SARIMAX-LSTM	Full	0.080921	0.127924	0.998877
		Train	0.075987	0.126021	0.999015
		Test	0.142109	0.149522	0.999780
Mammalia	LSTM-only	Full	0.125066	0.173092	0.996019
		Train	0.124645	0.174720	0.995873
		Test	0.130476	0.150607	0.995824
	SARIMAX-only	Full	0.849971	1.464794	0.573942
		Train	0.829077	1.465029	0.569484
		Test	1.117425	1.461778	−0.360313
	SARIMAX-LSTM	Full	0.062675	0.116218	0.997964
		Train	0.065726	0.120334	0.997817
		Test	0.023624	0.032274	0.999861
Reptalia	**LSTM-only**	Full	0.088569	0.136402	0.997526
		Train	0.072142	0.090745	0.999191
		Test	0.291438	0.382994	1.0
	SARIMAX-only	Full	0.861354	1.778986	0.518128
		Train	0.907954	1.846819	0.510820
		Test	0.297493	0.434515	−0.003334
	SARIMAX-LSTM	Full	0.962019	1.892603	0.430450
		Train	0.914437	1.916095	0.450477
		Test	1.537755	1.580922	−0.078020

Table 8. *Cont.*

Taxa	Model	Dataset Sample	MAE	RMSE	Correlation Coefficient
Caprimulgus	LSTM-only	Full	0.221570	0.278943	0.944529
		Train	0.190457	0.231701	0.963569
		Test	0.591814	0.604054	0.832552
	SARIMAX-only	Full	0.361386	0.688052	0.522025
		Train	0.374627	0.707808	0.525577
		Test	0.203815	0.381116	0.071954
	SARIMAX-LSTM	Full	0.107907	0.208612	0.963236
		Train	0.066180	0.122092	0.988346
		Test	0.604448	0.619685	0.973613
Lepus Saxatilis	**LSTM-only**	Full	0.018158	0.027782	0.999617
		Train	0.017084	0.026798	0.999591
		Test	0.031855	0.038159	0.999711
	SARIMAX-only	Full	0.679132	1.249148	0.021462
		Train	0.631545	1.220008	0.011280
		Test	1.283485	1.572944	0.124298
	SARIMAX-LSTM	Full	0.767327	1.576533	−0.217262
		Train	0.700520	1.532019	−0.356469
		Test	1.615778	2.059804	−0.798428
Numida meleagris	**LSTM-only**	Full	0.018468	0.029443	0.999550
		Train	0.018857	0.030349	0.999541
		Test	0.013784	0.014638	0.999999
	SARIMAX-only	Full	0.381317	0.761554	0.411592
		Train	0.386667	0.781032	0.417769
		Test	0.316855	0.466821	−0.145816
	SARIMAX-LSTM	Full	0.029861	0.053740	0.997248
		Train	0.027875	0.053512	0.997531
		Test	0.053786	0.056411	0.999203

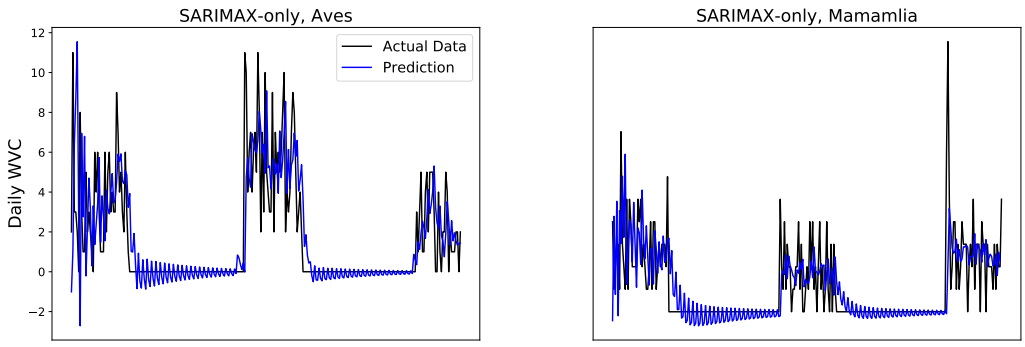


Figure 9. *Cont.*

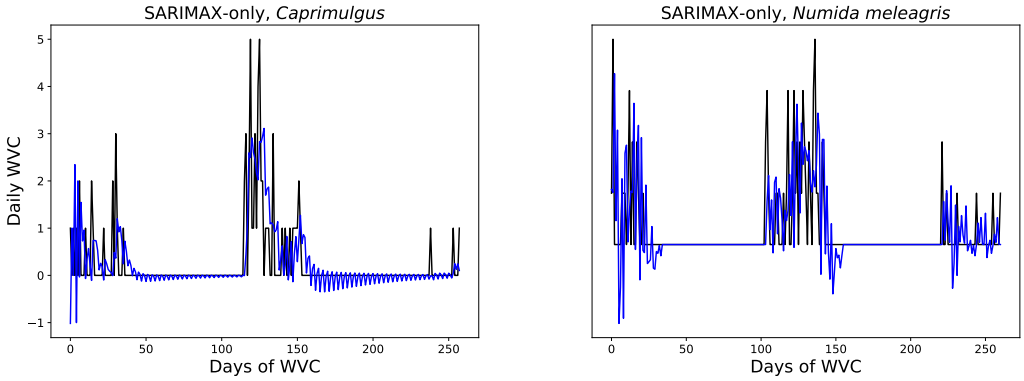


Figure 9. A time series of daily wildlife roadkill showing actual data and predicted data using SARIMAX-only on the Aves (top-left panel), Mammalia (top-right panel), *Caprimulgus* (bottom-left panel), and *Numida meleagris* (bottom-right panel) datasets.

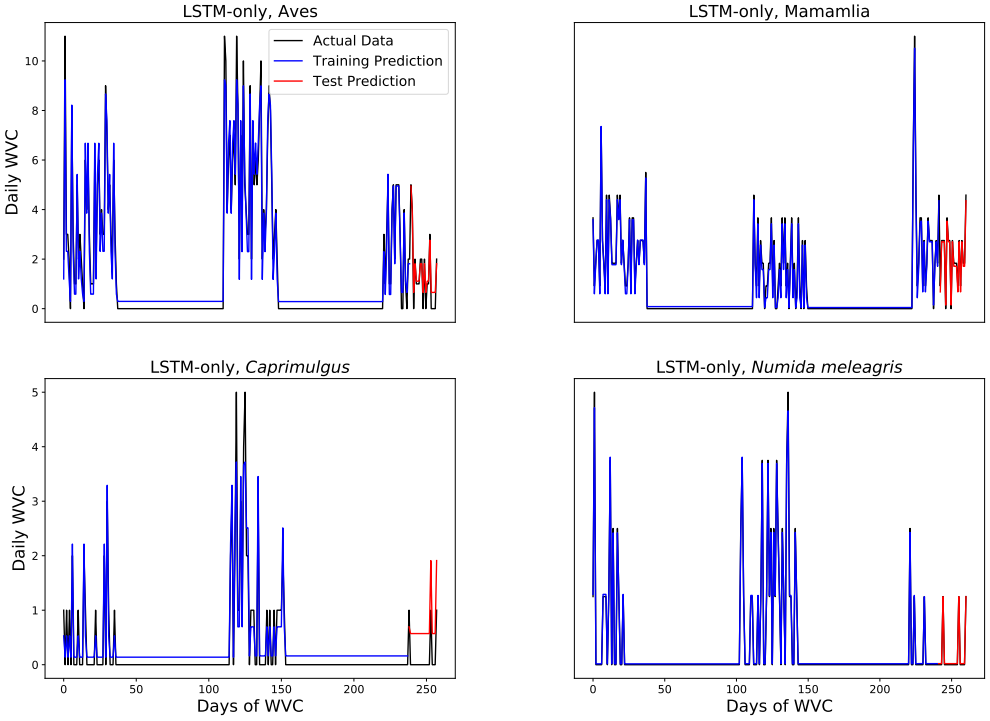


Figure 10. A time series of daily wildlife roadkill showing actual data and predicted data using LSTM-only on the Aves (top-left panel), Mammalia (top-right panel), *Caprimulgus* (bottom-left panel), and *Numida meleagris* (bottom-right panel) datasets.

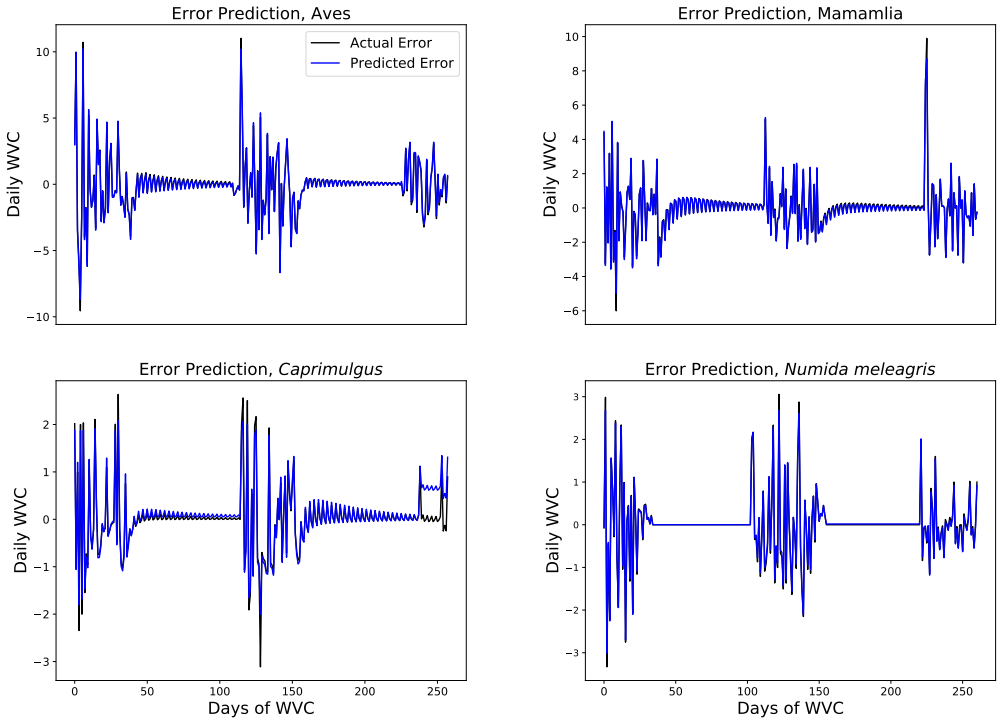


Figure 11. Daily actual residuals and predicted residuals of Aves taxa from the SARIMAX-only model.

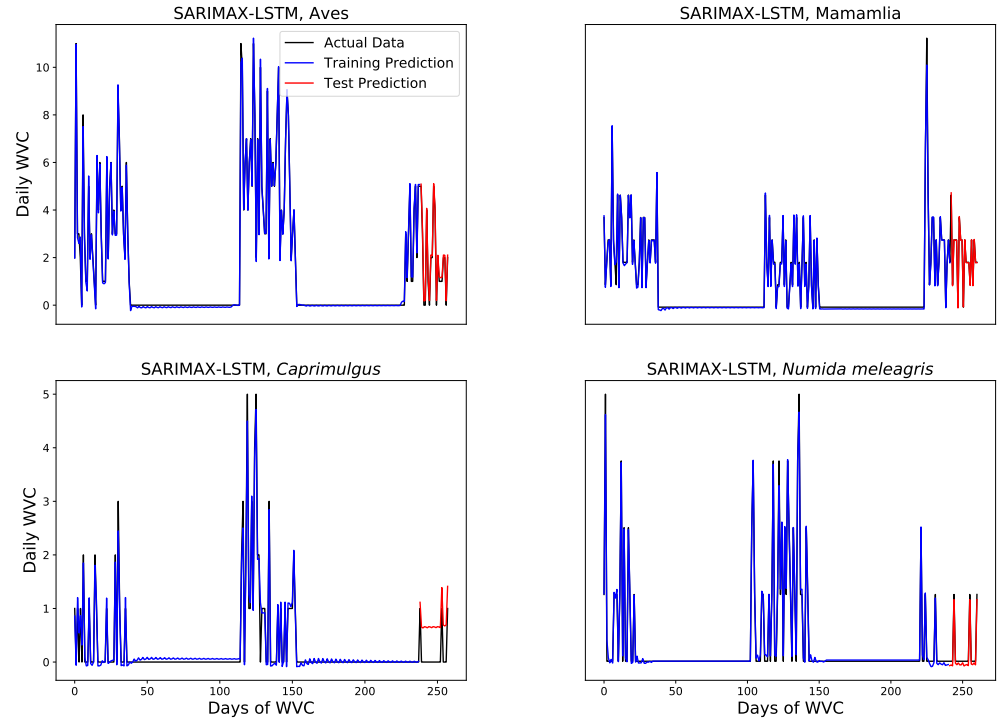


Figure 12. A time series of daily wildlife roadkill showing actual data and predicted data using SARIMAX-LSTM on the Aves (top-left panel), Mammalia (top-right panel), *Caprimulgus* (bottom-left panel), and *Numida meleagris* (bottom-right panel) datasets.

7. Discussion

Researchers have developed several prediction methods. To date, the SDM is extensively applied to predict the future wildlife roadkill risk. We have compared the

current results in the literature that use the SDM, such as logistic regression, GLMM, LCM, and PCA (see Appendix A). These models reveal that roadkill positively correlates to the road density, proportion of forests within a 1000 m buffer, proportion of pasture, route type, and distance. In addition, these models determine and forecast roadkill risk maps, vehicle-strike hotspots, and mortality rates. However, SDM models cannot be generalized using limited data samples characterized by sparse distributions. On the other hand, linear models, such as AR-only, ARIMA-only, and SARIMAX-only, execute excellent results even when using small data samples by sparse distributions. These models can capture complex linear relationships and rely on regressing a variable on past values and/or forecast errors. The SARIMAX-only model uses input features, such as endogenous variables or both endogenous and exogenous variables, whereas the ARIMA-only, AR-only, ARMA-only, and MA-only input only the endogenous variables. In addition, the LSTM-only model easily captures nonlinear elements on big datasets and supports endogenous and exogenous variables as input features.

These single models have their drawbacks and advantages, as discussed. To benefit from the advantage of the SARIMAX-only and LSTM-only, we introduce SARIMAX-LSTM, an error correction model that improved predictions in univariate and multivariate time series compared to either SARIMAX-only or LSTM-only.

This model is vital in this study as it can predict wildlife roadkill data collected from different seasons of the year. In addition, the data contain different species of animals with different animal behaviors and traits. These behaviors and traits lead to different linear and nonlinear variations in the wildlife roadkill patterns during the different ecological seasons in South Africa. The error correction model can help policymakers to be aware of expected future wildlife roadkill incidences in the coming week(s), month(s), and year(s). If there is an increase in wildlife roadkill patterns, policymakers can help by making decisions for the better implementation of mitigation measures for conserving biodiversity. Mitigation measures widely used include crossing structures, fencing, road verges, vegetation cover, and animal detection systems. If there is a significant drop in wildlife roadkill cases, policymakers can use diverse resources to increase revenue and decrease costs. Generally, mitigation measures affect the likelihood of increasing or decreasing WVCs. The proposed error correction model has the following advantages: (1) Even though the single-model SARIMAX-only and LSTM-only can predict wildlife roadkill, using the error correction model increases the prediction performance and the correlation coefficient since jointly linear and nonlinear components are easily captured in big or small data regimes. (2) The experiments show that the error correction model can compensate for the shortcomings of the single models and improve the prediction accuracy. (3) In addition, the error correction model results can facilitate lawmakers' decision making with good assurance, as it provides accurate predictions of wildlife roadkill incidences. However, there are still some limitations to the proposed model, especially in practical applications and model deployment and when the nonlinear components in the series are more than the linear ones. The proposed model may experience slightly lower performance results when tested for out-of-distribution shifts during model deployment. With in-distribution performance outperforming out-of-distribution performance, a huge problem is experienced in the artificial intelligence field [48]. In addition, computational requirements become an issue in the big data regime while the error correction model is in action because LSTM and hybrid systems based on the LSTM model are more costly regarding computational effort. More dataset samples are required for SARIMAX-LSTM to achieve a very high positive (negative) correlation or extremely minimal MAE and RMSE. A diverse sample of exogenous features helps the model to predict accurately.

8. Conclusions

Roadkill prediction and forecasting are usually widely implemented by SDMs. These models fail to accurately predict if the data are small and sparsely distributed. In addition, the SDM struggles to make forecasts when the dataset has fewer exogenous variables. We adopted the power of SARIMAX linear architecture and LSTM nonlinear architecture to

attempt this challenge. This work implemented an error correction SARIMAX-LSTM model applied to wildlife roadkill time-series data. Several models were also implemented for different taxa to compare with the proposed error correction model. A transformed stationary dataset's features were fed as inputs to the SARIMAX linear models. Later, the model performance was evaluated for all models in the study. The *AIC*, MAE, and RMSE values were calculated, and the lowest values showed high accuracy for the model, and the highest values showed a high correlation coefficient between the actual and predicted values.

The results showed that the error correction framework outperformed SARIMAX-only when the linear components were more than the nonlinear components. In addition, LSTM-only outperformed SARIMAX-LSTM when the nonlinear components were more than the linear components, as shown in Table 8. We attribute this to the insufficient datasets collected and the inability of SARIMAX-LSTM to accurately detect seasonality on residuals from SARIMAX-only and minimize the SARIMAX-LSTM errors. We concluded that this method could be a promising technique if applied to a big dataset to improve prediction performance, especially for time-series data in ecology. This work opens the door to future investigations; we suggest collecting more data for each taxon, which would improve the model's ability to predict. A hybrid linear SDM with LSTM is also a possible future study since some work has shown promising results with hybrid PCA-LSTM for time-series prediction. We further suggest using meta-heuristic algorithms, which have been successful in other domains [49,50].

Author Contributions: Conceptualization, I.N. and M.A.; Data curation, I.N. and N.M.; Formal analysis, I.N., M.A., N.M., and F.T.; Funding acquisition, M.A.; Investigation, I.N.; Methodology, I.N., M.A., S.T.S., P.O., R.R., T.A.-N., J.L.E.K.F., and F.T.; Supervision, M.A. and P.O.; Validation, I.N., M.A., and P.O.; Visualization, I.N., M.A., N.M., S.T.S., R.R., T.A.-N., J.L.E.K.F., and F.T.; Writing—original draft, I.N., N.M., S.T.S., P.O., R.R., T.A.-N., J.L.E.K.F., and F.T.; Writing—review and editing, I.N., M.A., N.M., S.T.S., P.O., R.R., T.A.-N., J.L.E.K.F., and F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out with financial support from Rhodes University.

Informed Consent Statement: Not applicable.

Institutional Review Board Statement: Ethical review and approval were waived for this study by the Animal Research Ethics Committee of Rhodes University, South Africa because the study did not involve the use of live animals and therefore did not require ethical clearance.

Data Availability Statement: Daily WVCs data were collected from EWT; previously, the same dataset was published partially as roadkill cases from 2011 to 2012 in South Africa [4]. For full access to data, contact EWT.

Acknowledgments: This paper is dedicated to the memory of Marco Andrea Garuti, who passed away suddenly while this paper was still in preparation. IN, MA, and PO acknowledge Rhodes University for the support. The authors convey a special thanks to EWT in South Africa for providing the dataset for this study. We further express sincere gratitude to Gabi Teren and Wendy Collinson, who worked with EWT, for the expert advice and valuable discussions. We would like to thank the anonymous reviewers, for their comments that improved the paper.

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

Appendix A. Discussion and Comparative Studies of Some Works That Use SDM to Model and Forecast Wildlife Roadkill Patterns Showing Their Performances, Strengths, and Weaknesses

Table A1. Traditional methods used to forecast wildlife roadkill patterns in terms of their performances, strengths, and weaknesses.

Reference	Title	Data	Model	Results	Strengths	Weaknesses
[33]	Life in the fast lane: roadkill risk along an urban–rural gradient	• Citizen science database of wildlife roadkill • Data from Cardiff University Otter Project (source: https://www.cardiff.ac.uk/otter-project accessed on 15 September 2020) • Open access data from national biodiversity network Atlas (source: https://registry.nbnatlas.org/public/show/dp205 accessed on 15 September 2022)	• SDM • PCA • MaxEnt model	• Assessed how roadkill risk varied along an urban–rural gradient and what abiotic factors underlie this risk • Determined if risk was a function of a species' distribution or not.	• Roadkill risk was positively associated with road density	• Implementing road mitigation strategies in dense urban road networks may be difficult and expensive due to increased roadkill
[13]	Snakes on a lane: Road type and edge habitat predict hotspots of snake road mortality	• Road mortality data for snakes for over 15 years, i.e., 2003 to 2018	• Logistic regression	• September was the month with the highest average number of roadkill per survey • April was the month with the lowest average number of roadkill encountered per survey • 2011 had the highest average number of roadkill encountered per survey • 2003 had the lowest average number of roadkill encountered per survey • High numbers of roadkill were recorded on state routes with the least numbers recorded on township routes	• Study showed that the proportion of pasture, the route type, and the distance to Athens were important factors in predicting road mortality • The work was able to associate high-risk areas with high-traffic roads throughout the study area, including in regions not well covered by the surveys • The study can inform management and conservation about species occurrence and also about potential risks to population persistence	• Sparse data for some species were observed and this contributed to the poor performance of the model.
[14]	A forecasting map of avian roadkill-risk in Europe: A tool to identify potential hotspots	• Bird casualty data recorded in roadkill events • Data on road density	• GLMM	• The study standardized the relative frequency of roadkill from 0 (species never recorded in bird casualty reports) to 1 (species with the higher number of roadkill) • A cumulative risk of roadkill for each bird assemblage was calculated by considering the sum of the values estimated for each bird species in Europe identified in the previous step • Calculations of the road density in each spatial unit • A forecasting map was developed of the potential avian roadkill risk by combining the data on road density and cumulative roadkill risk	• There was a pairwise relationship between road density and cumulative risk of roadkill in avian species assemblages for the whole of the European continent	• Size bias in roadkill surveys through over-representation of larger birds compared to smaller birds • Alternatively, larger birds may be underestimated in roadkill foot surveys due to the bias of them being easily detected and removed by road maintenance crews, drivers, or scavengers

Table A1. *Cont.*

Reference	Title	Data	Model	Results	Strength	Weaknesses
[15]	Roadkill and space use data predicted vehicle-strike hotspots and mortality rates in a recovering bobcat (*Lynx rufus*) population	• Bobcat roadkill were recorded between 1978 and 2017 • Land cover data were from the 2011 national land cover database • Road trait variables were from the Ohio Department of Transportation spatial database for 2017	• Logistic regression • Spherical semivariogram • Additive	• The study showed that bobcat roadkill were positively related to the proportion of forests • Calculations for the probability of mortality at each roadkill location	• Predicted mortality of wildlife roadkill at each location	• This work showed roadways with higher traffic and four lanes contributing to causing higher bobcat mortality
[16]	Habitat suitability vs. landscape connectivity determining roadkill risk at a regional scale: a case study on European badger (*Meles meles*)	• Data of both living and road killed species	• HSM • LCM • Roadkill risk model	• Landscape connectivity was more important than habitat suitability in determining roadkill risk for the Eurasian badger. • The density of regional roads was the most important variable in determining roadkill risk. • Higher values of risk were located in hillside areas and valleys with large urban surfaces.	• Both HSM and LCM were used as predictors in combination with road characteristics to fit a roadkill risk model	• A bad kill risk exhibited poor or no relationship with the distance from urban areas and the density of highways.

References

1. Grilo, C.; Koroleva, E.; Andrášik, R.; Bíl, M.; González-Suárez, M. Roadkill risk and population vulnerability in European birds and mammals. *Front. Ecol. Environ.* **2020**, *18*, 323–328. [CrossRef]
2. Medrano-Vizcaíno, P.; Espinosa, S. Geography of roadkills within the Tropical Andes Biodiversity Hotspot: Poorly known vertebrates are part of the toll. *Biotropica* **2021**, *53*, 820–830. [CrossRef]
3. Kioko, J.; Kiffner, C.; Jenkins, N.; Collinson, W. Wildlife roadkill patterns on a major highway in Northern Tanzania. *Afr. Zool.* **2015**, *50*, 1–6. [CrossRef]
4. Collinson, W.J.; Parker, D.M.; Bernard, R.T.; Reilly, B.K.; Davies-Mostert, H.T. An Inventory of Vertebrate Roadkill in the Greater Mapungubwe Transfrontier Conservation Area, South Africa. *Afr. J. Wildl. Res.* **2015**, *45*, 301–311. [CrossRef]
5. Périquet, S.; Roxburgh, L.; le Roux, A.; Collinson, W.J. Testing the Value of Citizen Science for Roadkill Studies: A Case Study from South Africa. *Front. Ecol. Evol.* **2018**, *6*, 15. [CrossRef]
6. Collinson, W.J.; Parker, D.M.; Bernard, R.T.F.; Reilly, B.K.; Davies-Mostert, H.T. Factors influencing the spatial patterns of vertebrate roadkill in South Africa: The Greater Mapungubwe Transfrontier Conservation Area as a case study. *Afr. J. Ecol.* **2019**, *57*, 552–564. [CrossRef]
7. Lester, D. Effective wildlife roadkill mitigation. *J. Traffic Transp. Eng.* **2015**, *3*, 42–51.
8. Rytwinski, T.; Soanes, K.; Jaeger, J.A.; Fahrig, L.; Findlay, C.S.; Houlahan, J.; Van Der Ree, R.; van der Grift, E.A. How effective is road mitigation at reducing road-kill? A meta-analysis. *PLoS ONE* **2016**, *11*, e0166941. [CrossRef]
9. Bager, A.; Alves da Rosa, C. Priority ranking of road sites for mitigating wildlife roadkill. *Biota Neotrop.* **2010**, *10*, 6. [CrossRef]
10. Meena, S.D.; Loganathan, A. Intelligent animal detection system using sparse multi discriminative-neural network (SMD-NN) to mitigate animal-vehicle collision. *Environ. Sci. Pollut. Res.* **2020**, *27*, 39619–39634. [CrossRef]
11. Nandutu, I.; Atemkeng, M.; Okouma, P. Intelligent Systems Using Sensors and/or Machine Learning to Mitigate Wildlife–Vehicle Collisions: A Review, Challenges, and New Perspectives. *Sensors* **2022**, *22*, 2478. [CrossRef]
12. Zhou, D. Real-Time Animal Detection System for Intelligent Vehicles. Ph.D. Thesis, Université d'Ottawa/University of Ottawa, Ottawa, ON, Canada, 2014.
13. Ryan, B.W.; Carl, R.B.; Viorel, D.P. Snakes on a lane: Road type and edge habitat predict hotspots of snake road mortality. *J. Nat. Conserv.* **2021**, *61*, 125978.
14. Morelli, F.; Benedetti, Y.; Delgado, J. A forecasting map of avian roadkill-risk in Europe: A tool to identify potential hotspots. *Biol. Conserv.* **2020**, *249*, 108729. [CrossRef]
15. Bencin, H.; Prange, S.; Rose, C.; Popescu, D. Roadkill and space use data predict vehicle-strike hotspots and mortality rates in a recovering bobcat (*Lynx rufus*) population. *Sci. Rep.* **2019**, *9*, 2045–2322. [CrossRef] [PubMed]
16. Fabrizio, M.; Di Febbraro, M.; D'Amico, M.; Frate, L.; Roscioni, F.; Loy, A. Habitat suitability vs landscape connectivity determining roadkill risk at a regional scale: A case study on European badger (*Meles meles*). *Eur. J. Wildl. Res.* **2019**, *65*, 1439–0574. [CrossRef]
17. Mondal, P.; Shit, L.; Goswami, S. Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. *Int. J. Comput. Sci. Eng. Appl.* **2014**, *4*, 13–29. [CrossRef]
18. Al-Douri, Y.; AL-Chalabi, H.; Lundberg, J. Time Series Forecasting using Genetic Algorithm A case study of maintenance cost data for tunnel fans. In Proceedings of the ADVCOMP 2018: The Twelfth International Conference on Advanced Engineering Computing and Applications in Sciences, Athens, Greece, 18–22 November 2018.
19. Elamin, N.; Fukushige, M. Modeling and forecasting hourly electricity demand by SARIMAX with interactions. *Energy* **2018**, *165*, 257–268. [CrossRef]
20. Arunraj, N.S.; Ahrens, D.; Fernandes, M. Application of SARIMAX model to forecast daily sales in food retail industry. *Int. J. Oper. Res. Inf. Syst. (IJORIS)* **2016**, *7*, 1–21. [CrossRef]
21. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef]
22. Gjika, E.; Ferrja, A.; Kamberi, A. A Study on the Efficiency of Hybrid Models in Forecasting Precipitations and Water Inflow Albania Case Study. *Adv. Sci. Technol. Eng. Syst. J.* **2019**, *4*, 302–310. [CrossRef]
23. Hyeong, K. Stock Price Correlation Coefficient Prediction with ARIMA-LSTM Hybrid Model. *arXiv* **2018**, arXiv:1808.01560.
24. Khandelwal, I.; Adhikari, R.; Verma, G. Time Series Forecasting using Hybrid ARIMA and ANN Models based on DWT Decomposition. *Procedia Comput. Sci.* **2015**, *48*, 173–179. [CrossRef]
25. McDonald, S.; Coleman, S.; McGinnity, T.M.; Li, Y. A hybrid forecasting approach using ARIMA models and self-organising fuzzy neural networks for capital markets. In Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 4–9 August 2013. [CrossRef]
26. Rathod, S.; Naveena, K.; Singh, S.; Singh, A. Hybrid ARIMA-ANN Modelling for Forecasting the Price of Robusta Coffee in India. *Int. J. Curr. Microbiol. Appl. Sci.* **2017**, *6*, 1721–1726.
27. Belmahdi, B.; Louzazni, M.; Bouardi, A. A hybrid ARIMA–ANN method to forecast daily global solar radiation in three different cities in Morocco. *Eur. Phys. J. Plus* **2020**, *135*, 2190–5444. [CrossRef]
28. Deng, Y.; Fan, H.; Wu, S. A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits. *J. Ambient. Intell. Humaniz. Comput.* **2020**, 1868–5145. [CrossRef]

29. Dave, E.; Leonardo, A.; Jeanice, M.; Hanafiah, N. Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM. *Procedia Comput. Sci.* **2021**, *179*, 480–487. [CrossRef]
30. Trancart, T.; Acou, A.; De Oliveira, E.; Feunteun, E. Forecasting animal migration using SARIMAX: An efficient means of reducing silver eel mortality caused by turbines. *Endanger. Species Res.* **2013**, *21*, 181–190. [CrossRef]
31. Yulian, W.; Peiguang, L.; Xiushan, N. Research of Stock Price Prediction Based on PCA-LSTM Model. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *790*, 012109.
32. Choi, S.; Kim, B. Applying PCA to Deep Learning Forecasting Models for Predicting PM2.5. *Sustainability* **2021**, *13*, 3726. [CrossRef]
33. Kent, E.; Schwartz, A.; Perkins, S. Life in the fast lane: Roadkill risk along an urban–rural gradient. *J. Urban Ecol.* **2021**, *7*, juaa039. [CrossRef]
34. McElroy, T. Matrix formulas for nonstationary ARIMA signal extraction. *Econom. Theory* **2008**, *24*, 988–1009. [CrossRef]
35. Hua, Y.; Zhao, Z.; Li, R.; Chen, X.; Liu, Z.; Zhang, H. Deep Learning with Long Short-Term Memory for Time Series Prediction. *arXiv* **2018**, arXiv:1810.10161.
36. Léon, B. Stochastic gradient descent tricks. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 421–436.
37. Langen, T.A.; Ogden, K.M.; Schwarting, L.L. Predicting hot spots of herpetofauna road mortality along highway networks. *J. Wildl. Manag.* **2009**, *73*, 104–114. [CrossRef]
38. Guiden, P.W.; Orrock, J.L. Seasonal shifts in activity timing reduce heat loss of small mammals during winter. *Anim. Behav.* **2020**, *164*, 181–192. [CrossRef]
39. Wone, B.; Beauchamp, B. Movement, home range, and activity patterns of the horned lizard, Phrynosoma mcallii. *J. Herpetol.* **2003**, *37*, 679–686. [CrossRef]
40. Robbins, C.S. Effect of time of day on bird activity. *Stud. Avian Biol.* **1981**, *6*, 275–286.
41. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model. Dev.* **2014**, *7*, 1247–1250. [CrossRef]
42. Arnold, T.W. Uninformative parameters and model selection using Akaike's Information Criterion. *J. Wildl. Manag.* **2010**, *74*, 1175–1178. [CrossRef]
43. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
44. Mukaka, M.M. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* **2012**, *24*, 69–71.
45. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *27*, 1–22. [CrossRef]
46. Smith, T.G. Tips to Using Auto_arima. Copyright 2017–2018. 2022. Available online: https://alkaline-ml.com/pmdarima/1.1.1/tips_and_tricks.html (accessed on 16 September 2022).
47. Bhansali, R.; Kokoszka, P. Prediction of Long-Memory Time Series: A Tutorial Review. In *Processes with Long-Range Correlations: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 621, pp. 3–21.
48. Koh, P.W.; Sagawa, S.; Xie, S.M.; Zhang, M.; Balsubramani, A.; Hu, W.; Yasunaga, M.; Phillips, R.L.; Gao, I.; Lee, T.; et al. Wilds: A benchmark of in-the-wild distribution shifts. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 5637–5664.
49. Ahmed, A.N.; Van Lam, T.; Hung, N.D.; Van Thieu, N.; Kisi, O.; El-Shafie, A. A comprehensive comparison of recent developed meta-heuristic algorithms for streamflow time series forecasting problem. *Appl. Soft Comput.* **2021**, *105*, 107282. [CrossRef]
50. Ngo, N.T.; Truong, T.T.H.; Truong, N.S.; Pham, A.D.; Huynh, N.T.; Pham, T.M.; Pham, V.H.S. Proposing a hybrid metaheuristic optimization algorithm and machine learning model for energy use forecast in non-residential buildings. *Sci. Rep.* **2022**, *12*, 1065. [CrossRef] [PubMed]