Machine learning of spatiotemporal processes: investigating livestock changes in Mongolia with LSTM recurrent neural networks

Alexis Comber
University of Leeds
Leeds, UK
a.comber@
leeds.ac.uk

Alison Heppenstall
University of Leeds
Leeds, UK
a.j.heppenstall@
leeds.ac.uk

Taro Takahashi
Rothamsted Research,
North Wyke, UK
taro.takahashi@
rothamsted.ac.uk

Naru Tsutsumida
Kyoto University,
Kyoto, Japan
naru@
kais.kyoto-u.ac.jp

Paul Harris
Rothamsted Research,
North Wyke, UK
paul.harris@
rothamsted.ac.uk

Abstract

This paper explores the application of Long Short-Term Memory (LSTM) recurrent neural networks to a geographical problem. Specifically, it explores the benefits of including spatial (geographic) information in the input parameters to this family of machine learning approaches. The exploratory initial analyses parameterised the LSTM with spatially lagged data derived from a contiguity analysis and temporally lagged under assumptions of both serial and spatial autocorrelation using historical data for 1 to 20 years. These indicated model inferential and predictive gains with the inclusion of spatially and temporally lagged data overall when seeking to predict the spatial distribution of NDVI (as a measure of greenness and livestock pressure). They also indicated that the most salient years to be lags of 1, 2, 17, 18, 19 and 20 years. A number of areas of future work are identified, not least of which is the challenge to circumvent inherent assumptions of serial autocorrelation in many space-time methods and to consider the phase of data relative to the periodicity of the process being considered.

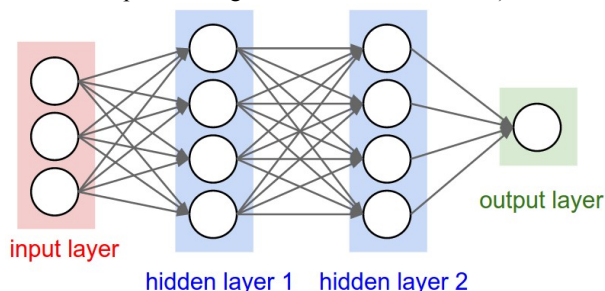*Keywords*: Soum; Space-time; machine learning;

1 Introduction

All data are not only spatial, they are increasingly temporal (Kitchin 2013), 'big' and diverse, with many new forms of data generated in our increasingly digital, connected and GPS-enabled lives (Kitchin and McArdle 2016). Such developments are changing the nature of spatial data analysis (Brunsdon 2016). Space-time relationships are central to GIS research (Yuan 2017) and while process spatial auto-correlation is well-known (Tobler 1970, 2004), the concept of serial auto-correlation is more uncertain for temporal processes. Many processes have periodicities which require consideration of the phase of observation.

There are a number of approaches for analysing space time data and interactions. Evolutionary approaches include cellular automata (e.g. Balzter et al. 1998; Dietzel et al. 2005) and agent-based models (e.g. O'Sullivan and Haklay 2000; Torrens et al. 2011). In these approaches solutions emerge through simulation but can be sensitive to initial parameterisation and difficult to tune. Other methods explicitly model local and heterogenous spatial and temporal interactions (Kyriakidis and Journel 1999; Fotheringham et al. 2015) including ARIMA (autoregressive integrated moving average), STARIMA (space-time autoregressive), panel models and geostatistical approaches (Griffith 2010; Deng et al. 2017). Others approaches such as geographically and temporally weighted regression (Huang et al. 2010; Fotheringham et al. 2015; Liu et al. 2018) explicitly focus on non-stationary spatiotemporal attribute relationships. However, there is an implicit assumption of serial as well as spatial autocorrelation in these approaches. For these reasons, some have identified the need for methods to handle spatiotemporal data (Goodchild 2013; Miller and Goodchild 2015).

In brief a neural network (NN) model consists of a set of adaptive and unidirectionally connected processing elements, typically structured in layers, with input layers, hidden layers and an output layer (see Figure 1). Recurrent neural networks (RNNs) include a looping mechanism to allow information about the hidden state, representing previous inputs, to be passed from one step to the next. Reviews of NNs and RNNs can be found in Tsoi and Back (1997) and Lipton et al (2015). There is a long history of NNs in spatial analysis, starting with Fischer (1996) and recent applications to geographical problems include Li et al (2017), Yu et al (2017) and Lyu et al (2016). On overview can be found in Fischer and Abrahart (2014).

Figure 1: Example of a feed forward neural network (from http://cs231n.github.io/neural-networks-1).
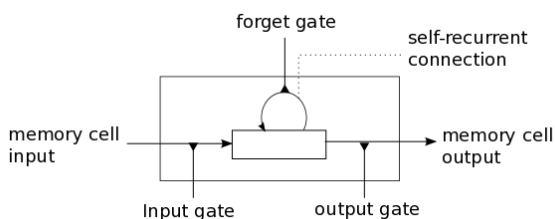

Fischer (1996) comments that NNs are used in machine learning because of their computational adaptivity, their *speed of computation* through their parallel distributed architectures

support real-time applications, their inherent nonlineariy which supports complex tasks such as pattern and speech recognition and their ability to cope with noisy, messy data. An excellent and informative application of RNNs and LSTMs with respect to land cover can be found in Minh et al (2018).

This paper examines the application to spatiotemporal geographical problems of Long Short-Term Memory (LSTM) recurrent neural networks (RNNs). These were specifically developed to model temporal sequences and their long-range dependencies more reliably than NNs and RNNs. Specifically, RNNs may in some circumstances be unable to apply relevant information to the current state of the solution if that information was identified much earlier in the "learning" process: they can become incapable of handling such "long-term dependencies" (i.e. unable to connect the distant information). LSTMs (Hochreiter and Schmidhuber, 1997) were developed to avoid the long-term dependency problem and to be able to apply information learnt much early in their process. An accessible and informative overview of LSTMs is provided in Olah (2015).

One of the problems with propagation approaches like NNs is that information (signals) can be multiplied by weights a large number of times, depending on the number of timesteps (iterations or epochs). If the weights are small, the problem of vanishing signals, making the tasks of deep learning difficult. Equally if the weights are large the machine learning process can explode. LSTMs deal with this by having a different structure to their cells. Each is composed of four elements: an input gate, a neuron with a connection to itself (a self-recurrent connection), a forget gate and an output gate. The gates serve to dampen or modulate the memory cell interactions. The input gate can allow signal in to alter the memory state or block it and similarly the output gate can block or not. The forget gate modulates the cell's self-recurrent connection, allowing the cell to remember or forget its previous state, as needed.

Figure 2: Example of an LSTM cell
(http://deeplearning.net/tutorial/lstm.html).

The aim of this paper is to explore the application of LSTMs to explicitly spatiotemporal geographical problems. The analysis evaluates the degree to which model outputs (predictions) improve when they are parameterised with *spatially* lagged variables as well as the usual temporally lagged ones, and the discussion considers a number of future areas of research in the general area of identifying patterns (and therefore processes) within spatiotemporal data by using a variety of different methods. The R code and data used in this analysis are available at https://github.com/xyz.

2   Methods

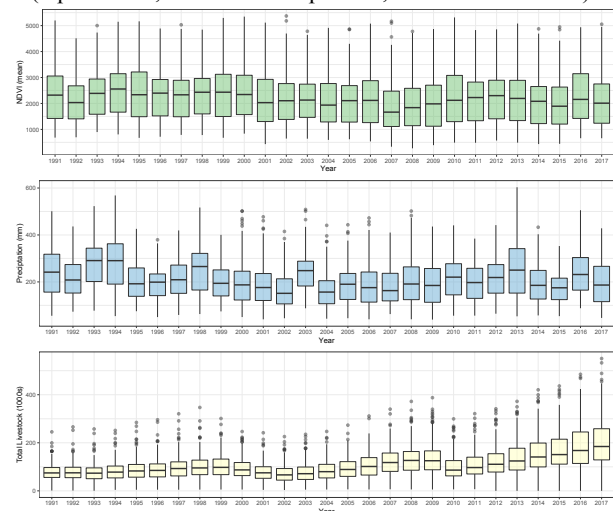### 2.1   Case Study: historical Mongolian livestock

Nomadic pastoralism in Mongolia has been a sustainable model for thousands of years. In the communist period livestock production was centrally planned and managed, with nomadic herders raising state-owned livestock and encouraged to organize collectives locally. Collectives self-regulated their seasonal travel and where their livestock grazed, resulting in good pasture maintenance. However, since the early 1990s, Mongolia has been transformed into free-market economy which has resulted in a rapid increase of livestock and herder populations (Fernandez-Gimenez 2006) and pastures no longer being collectively managed. This has led to serious sustainability and land management concerns (Togtokh 2008) but as yet little research has been undertaken to quantify the overgrazing problem, and the threat to fragile grassland environments (Liu et al, 2013).

Data of annual livestock populations (sheep, goat, horse, cattle and camel) for 334 *soums* in Mongolia for the period 1991 to 2017 were downloaded from the Mongolian Statistical Information Service (http://www.1212.mn). Soums are second level administration units. The soums and the spatial adjacency are shown in Figure 3. Annual data of mean NDVI (vegetation greenness) and total precipitation were also collated for each soum (see Figure 4). The aim was to develop a predictive model of NDVI.

Figure 3: The soums in Mongolia and their contiguity.

Figure 4: The distributions of the annual data over the soums (top = NDVI, middle = Precipitation, bottom = Livestock).

The 27 years of data for each soum were pre-processed in the following way. The individual livestock types were combined to create a livestock total for each sou. Then the data arranged in long format for the 3 variables (i.e. with a record for each soum and for each year for each of the 3 variables). They were ordered by soum and by year and then grouped by soum and the precipitation and livestock data were lagged for 1 to 20 years. Finally, spatially lagged measures of livestock were calculated for each soum. These were derived from the neighbouring soums, using a bi-square distance decay model with a 180km limit. This was used to determine a distance weighted mean of neighbouring livestock counts for each soum. All the variables were individually normalised to the range [0,1]. The input data for the analysis thus included NDVI for 2011 to 2017 as the target variable, and lagged data years -1 to -20 for precipitation and livestock number, and spatially lagged livestock data for -1 to -20 years as the predictor variables.

2.2 Analysis

A LSTM model was constructed after lengthy investigations of different numbers of hidden layers, different number of cells in each layer and different types of layer. The "best" LSTM structure will depend on a number of factors including the size of the data (number of records) and the data variability. There are no formal guidelines on how to determine the number of layers or the number of memory cells in a LSTM and so this process was guided by the following heuristics / rules of thumb:

```
1.Start with a single hidden layer with a small
number of cells
2.Increase the number of cells
3.If this does not work try adding another layer
and repeat
```

The aim here is to identify a model that avoids under- and over-fitting. Investigations determined that a model with 2 hidden layers performed well across different analysis: a single LSTM layer with 32 cells and a *tanh* activation, and a dense matrix vector multiplication layer with 4 cells. The latter applies a rotation, scaling, translation transform and whose values are the trainable parameters which get updated during back propagation. The investigations were undertaken using different lengths of lagged data (from 1 to 20 years) and different training validation splits, using a mean absolute error metric under an Adam optimiser.

3 Results

The final model specification was determined by running multiple LSTM models with different lengths of spatially and temporally lagged data with a training and validation split of 2014. This allowed long and short lags of data to be evaluated and used to predict NDVI in each soum for 2015, 2016 and 2017. The R^2 fits for these runs are shown in Figure 4. These show an improvement with the inclusion of spatially lagged data and suggest that models for NDVI prediction in Mongolia should include lagged data from years -1, -2, and -17 to 20

Lagged data from years -1, -2, -17, -18, -19 and -20 were used to parameterise the LSTM model inputs to predict NDVI at soum level. The fits after different epochs are shown in Figure 5 and indicate a close-fitting model that is neither being under-

or over-fitted. The results of predicted minus observed NDVI for each predicted year are shown in Figure 6. These show the spatial distribution of the differences. In 2015 and 2017 there is some spatial structure but there is considerably more in 2016 with a distinct region of underestimation running east west in the northern part of the study area. Future work will seek to understand the drivers of these variations.

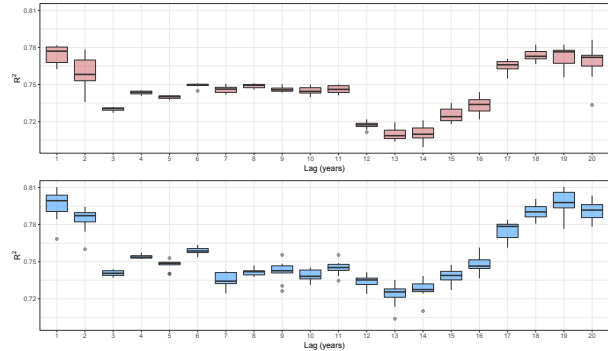Figure 4: The R^2 fit values for lagged data (top = temporal lagged data, bottom = spatial and temporal lagged data).
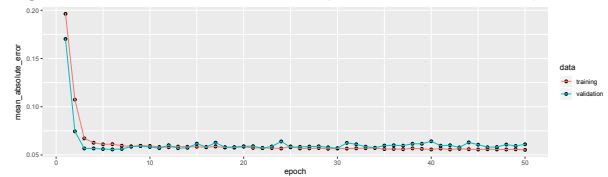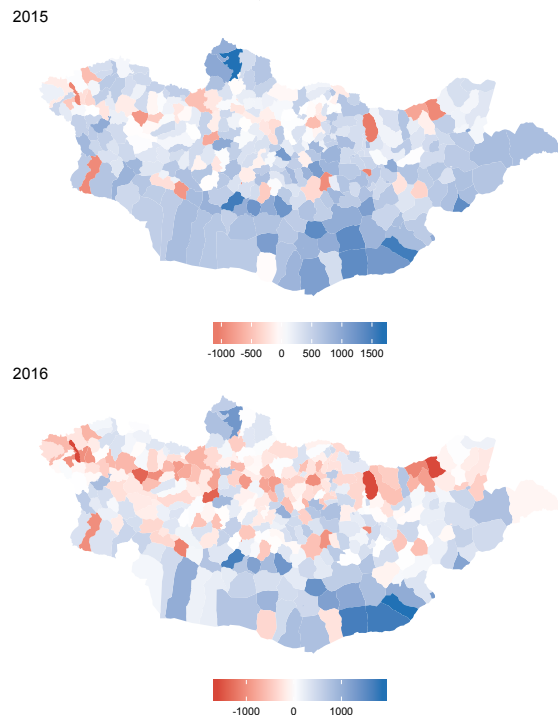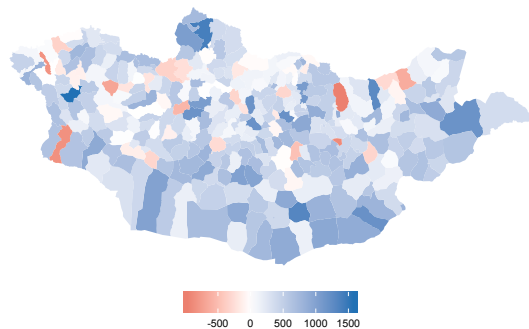

Figure 5: The LSTM model fits (Mean Absolute Error).

Figure 6: The distributions of predicted minus observed NDVI for 2015, 2016 and 2017.

2017

4 Discussion points

This paper reports initial work that is seeking to understand some of the landscape scale processes in Mongolia and to develop machine learning methods that are sensitive to geographical and spatial processes.

Here spatial information was incorporated into the LSTM model by attribute spatial lagging using contiguity and distance decay measures. The incorporation of spatial information in the form of spatially lags improved the predictive model fits greatly in comparison with model prediction based simply on temporally lagged data. This suggests that the inclusion of explicitly spatial attributes in machine learning approaches could help to solve geographical problems and that further development of neural approaches able to reason with spatial interactions between observations could be of benefit. Future work will develop these ideas further in order to make machine learning approaches more geographically aware. Specifically, we will seek to include methods for accounting for spatial structure (and not just spatially lagged variables as here) within LSTM models.

A number of other areas will be also addressed in future work, using variables excluded from this analysis describing socio-economic activity (income, local government income, employment rates), demographics (population changes) and livestock related activity (deaths by diseases, by climate events, households engaged directly in livestock management). These will allow a number of other questions to be investigated, particularly if data such population changes can be used to infer flows and thus be used to evaluate urbanisation pressures in Mongolia using gravity and spatial interaction models.

References

Balzter, H., Braun, P.W. and Köhler, W., 1998. Cellular automata models for vegetation dynamics. *Ecological Modelling*, *107*(2-3), pp.113-125.

Brunsdon, C. 2016. Quantitative methods I: Reproducible research and quantitative geography. *Progress in Human Geography*, *40*(5), 687-696.

Deng, M., Yang, W. and Liu, Q. 2017. Geographically Weighted Extreme Learning Machine: A Method for Space–Time Prediction. *Geographical Analysis*, *49*(4), 433-450.

Dietzel, C., Herold, M., Hemphill, J.J. and Clarke, K.C., 2005. Spatio-temporal dynamics in California's Central Valley: Empirical links to urban theory. *International Journal of Geographical Information Science*, *19*(2), pp.175-195.

Fernandez-Gimenez, M. E. 2006. Land Use and Land Tenure in Mongolia: A Brief History and Current Issues. In Rangelands of Central Asia: *Proceedings of the Conference on Transformations, Issues, and Future Challenges*, eds. D. J. Bedunah, E. D. McArthur, and M. E. Fernandez-Gimenez, 30–36. Salt Lake City: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Fischer, M.M. and Abrahart, R.J., 2014, June. 13 Neurocomputing for GeoComputation. Pp 307-328 in Robert J. Abrahart and Linda M. See (eds) *GeoComputation* . CRC Press.

Fotheringham, A.S., Crespo, R. and Yao, J. 2015. Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, *47*(4), 431-452.

Goodchild, M.F. 2013. Prospects for a space–time GIS: Space–time integration in geography and GIScience. *Annals of the Association of American Geographers*, *103*(5), 1072-1077.

Griffith, D.A. 2010. Modeling spatio-temporal relationships: retrospect and prospect. *Journal of Geographical Systems*, *12*(2), 111-123.

Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, *9*(8), pp.1735-1780.

Huang, B., Wu, B. and Barry, M. 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, *24*(3), 383-401.

Kitchin, R. and McArdle, G. 2016. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), doi:10.1177/2053951716631130

Kitchin, R. 2013. Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, *3*(3), 262-267.

Kyriakidis, P.C. and Journel, A.G. 1999. Geostatistical space–time models: a review. *Mathematical Geology*, 31(6), 651-684.

Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C. and Chi, T., 2017. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, *231*, pp.997-1004.

Lipton, Z.C., Berkowitz, J. and Elkan, C., 2015. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.

Liu, Y., Lam, K.F., Wu, J.T. and Lam, T.T.Y. 2018. Geographically weighted temporally correlated logistic regression model. *Scientific Reports*, *8*(1), p.1417.

Liu, Y.Y., Evans, J.P., McCabe, M.F., De Jeu, R.A., van Dijk, A.I., Dolman, A.J. and Saizen, I., 2013. Changing climate and overgrazing are decimating Mongolian steppes. *PloS one*, *8*(2), p.e57599.

Lyu, H., Lu, H. and Mou, L., 2016. Learning a transferable change rule from a recurrent neural network for land cover change detection. *Remote Sensing*, *8*(6), p.506.

Miller, H.J. and Goodchild, M.F. 2015. Data-driven geography. *GeoJournal*, *80*(4), 449-461.

Minh, D.H.T., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F. and Maurel, P., 2018. Deep recurrent neural networks for winter vegetation quality mapping via multitemporal SAR Sentinel-1. *IEEE Geoscience and Remote Sensing Letters*, *15*(3), pp.464-468.

O'Sullivan, D. and Haklay, M., 2000. Agent-based models and individualism: is the world agent-based?. *Environment and Planning A*, *32*(8), pp.1409-1425.

Olah C (2015). *Understanding LSTM Networks.* http://colah.github.io/posts/2015-08-Understanding-LSTMs/ Tobler 1970, 2004

Togtokh, C. 2008. Climate change adaptation strategies for pastoral communities of Mongolia's central mountainous region. *IHDP Update*, 2:53–58.

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(sup1), 234-240.

Tobler, W. 2004. On the first law of geography: A reply. *Annals of the Association of American Geographers*, *94*(2), 304-310.

Torrens, P., Li, X. and Griffin, W.A., 2011. Building agent-based walking models by machine-learning on diverse databases of space-time trajectory samples. *Transactions in GIS*, *15*, pp.67-94.

Tsoi, A.C. and Back, A., 1997. Discrete time recurrent neural network architectures: A unifying review. *Neurocomputing*, *15* (3-4), pp.183-223.

Yu, H., Wu, Z., Wang, S., Wang, Y. and Ma, X., 2017. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors*, *17*(7), p.1501.