Xin Cai xincai00@gmail.com

Research Proposal

Xin Cai

Title: Using Climatic and Imaging Data to Predict Apple Fruit Ripeness

Keywords: Flowering Time Prediction, Fruit Ripeness Prediction, Multivariate Time Series Analysis, Dynamical Systems, Deep Learning

1. Project Description

1.1 Background

Apple fruit is usually stored for a minimum of six months in a controlled atmosphere before marketing in the UK. It is not unusual for the stored fruit to suffer from 10-15% post-storage losses due to various causes, including physiological disorders and fungal rotting. This leads to not only yield losses but also increased the cost of sorting fruit post-storage. Fruit storability (i.e. post-storage fruit quality) can be affected by many factors, including flowering time, fruit ripeness at picking, fruit surface microflora, and climatic factors. Furthermore, the relationships of fruit quality with these factors are usually non-linear, and the precise causal relationships have yet to be elucidated.

1.2 Research Objectives

There are two specific research objectives: (1) predicting flowering time and (2) predicting fruit ripeness for optimum picking. Predicting the degree of fruit ripeness is critically important since it has been well established that fruit ripeness at picking could significantly affect fruit storage potential.

Predicting flowering time: Historic data collected at XXXX over the last 80 years will be used to study the temporal flowering pattern of several specific cultivars in relation to winter and spring climatic data. Statistical modelling will be carried out to study whether the temporal flowering pattern could be predicted from the winter and spring climatic data alone.

Predicting fruit ripeness: One key research activity is to work with crop physiologist in order to define physiologically/biochemically what is “perfect fruit ripeness” for harvesting. Previous research on predicting fruit ripeness is based on batches of fruit. To understand the variability of fruit ripeness among individual fruit, we propose to follow the development of individual fruit on several popular cultivars via imaging to investigate whether ripeness can be predicted from the imaging information as well as post-blossom temperatures.

2. Related Work & Methodology

Generally, knowledge and skill in multivariate time series (MTS) analysis and dynamical systems will play a vital role in accomplishing those research objectives mentioned above. Besides, the feasibility of incorporating other types of data into prediction frameworks such as hyperspectral images will be explored. More importantly, the research will be focused on developing data-driven algorithms, especially on the extension of deep-learning-based models to their probabilistic counterparts. Hopefully, the proposed algorithms will not only deliver satisfactory results concerning the prediction of temporal flowering pattern and the degree of fruit ripeness but also can be generalized to a broad spectrum of MTS data.

2.1 Multivariate Time Series (MTS) Analysis

2.1.1 Statistical Models

Large collections of time series are ubiquitous and generated in a wide variety of areas, including natural and social sciences, internet of things applications, cloud computing, supply chains and many more. The large-scale time-series data can be leveraged to make better forecasts or to detect anomalies more effectively, which in turn facilitates informed downstream decision making. Therefore, modelling MTS data has long been a subject that has attracted researchers from a diverse range of fields. There is a distinct demarcation in existing MTS analysis approaches, traditional statistical models and deep-learning-based models.

The most well-known model for linear univariate time series forecasting is the autoregressive integrated moving average (ARIMA)[1], which encompasses other autoregressive time series models, including autoregression (AR), moving average (MA), and autoregressive moving average (ARMA). Additionally, linear support vector regression (SVR) [2], treats the forecasting problem as a typical regression problem with time-varying parameters. However, these models are mostly limited to linear univariate time series and do not scale well to MTS. To forecast MTS data, vector autoregression (VAR), a generalization of AR-based models, was proposed. VAR is probably the most well-known model in MTS forecasting. Nevertheless, neither AR-based nor VAR-based models capture non-linearity. Besides, learning generative models of sequences is a long-standing machine learning challenge and historically the domain of dynamic Bayesian networks (DBNs) such as hidden Markov models (HMMs) [3] and Kalman filters [4]. State-space models (SSMs) are particularly well-suited for applications where the structure of the time series is well-understood, as they allow for the incorporation of structural assumptions into the model. However, these SSMs suffer from two major drawbacks: 1) the assumptions are restrictive and are violated in practical applications, and 2) extending linear dynamical systems to their nonlinear counterparts makes learning more difficult and is computationally prohibitive for high-dimensional signals.

2.1.2 Deep Sequential Learning Models

With the advent of deep learning, it has shown promising results to exploit deep neural networks to perform MTS analysis. Deep sequential learning models have been extensively studied in various research areas, such as natural language processing [5, 6, 7], speech analysis [8, 9], video prediction and generation [10, 11, 12], trajectory forecasting [13, 14, 15] and MTS analysis [16, 17, 18]. As far as the research field of time series analysis is concerned, many approaches have been proposed to address challenges in a wide range of applications, such as traffic forecasting[19, 20], product demand forecasting [21, 22], energy consumption forecasting [23], and disease progression modelling[24, 25]. Existing deep sequential learning models can be subsumed under two broad categories: 1) temporal and 2) spatio-temporal models. Concretely, temporal models only consider temporal dependencies irrespective of forms of input signals. This type of approaches mainly relies on the remarkable capability of deep neural networks, such as recurrent neural networks (RNNs) [5], long short-term memory networks (LSTMs) [26], gated recurrent units (GRUs) [27], temporal convolution networks (TCNs) [28], and transformer architectures [7, 29, 9], capturing nonlinear relationships in historical data for accurate forecasting. To the best of my knowledge, the long- and short-term time-series network (LSTNet) [30] is the first model designed specifically for MTS forecasting with up to hundreds of time series. LSTNet uses TCNs to capture short-term patterns and LSTMs or GRUs for memorizing relatively long-term patterns. Besides, a noticeable characteristic of this type of methods is that many research efforts have been devoted to designing complex neural architectures to enhance expressive capacity. For example, there has been research work [10, 31] attempting to improve the modelling capability for spatio-temporal sequences by incorporating additional memory cells responsible for leveraging spatial dependencies along the depth of stacked RNNs.

Despite the impressive performance achieved by these pure temporal models, it is reasonable to exploit interdependencies between different variables to develop advanced deep sequential learning models for the purpose of increasing predictive accuracy and efficiency simultaneously. Consequently, approaches taking interdependencies between different variables or dimensions of signals into account have been actively studied, referred to as spatio-temporal models [32, 33, 19, 20]. In many real-world applications, such as traffic forecasting and cyber-physical systems, different dimensions of MTS naturally correspond to recordings of sensors deployed in various locations. Nevertheless, spatio-temporal models hereafter will be used to refer to models which employ not merely temporal dynamics but structured information in time-series data. Spatio-temporal models can be further separated by their internal mechanisms of modelling interdependencies between different variables implicitly or explicitly.

Graph convolutional networks (GCNs) are currently dominating solutions for implicitly modelling interdependencies between different variables, especially in those applications where different variables correspond to different physical entities in dynamical systems. For example, different from general correlated time series prediction, research on traffic forecasting pays more attention to spatial correlations among

the traffic series collected from different sources (e.g., sensors deployed in different spatial locations in a road network) except for the temporal correlations. GCNs are a special kind of CNNs generalized for graph-structured data, which have been widely used in node classification, link prediction, and graph classification [34]. There are two mainstreams of GCNs, the spectral- and spatial-based approaches. Spectral-based models [35, 36, 37] smooth a node's input signals using spectral graph convolution filters. Spatial-based models [38, 39, 40, 41] extract a node's high-level representation by aggregating feature representations from its neighbouring nodes. Most of these works focus on graph representation learning, which obtains node embeddings by integrating the features from its local neighbours based on the given graph structure and proposed message-passing mechanisms. For example, GAT [41] learns to weight the features from different neighbouring nodes with attention scores calculated by the multi-head self-attention mechanism. DIFFPOOL [42] enhances GCNs with node clustering/graph coarsening to generate hierarchical graph representations. MTS forecasting can be viewed naturally from a graph perspective. Different dimensions of MTS data can be considered to be nodes in graphs, and they are interlinked through their hidden associations. Therefore modelling MTS data using GCNs has been shown a promising way to preserve temporal trajectories while exploiting the interdependencies among MTS data.

Generally, GCN-based MTS models fall into two broad categories: 1) designing more complicated neural architectures, especially message-passing mechanisms, and 2) constructing dynamical graph structures rather than using pre-defined static graphs. The first type of research work either focuses on the fusion mechanisms for effectively integrating spatial and temporal dependency learning modules or improving the message-passing functions. For example, GA-RNNs [32] proposed to replace multilayer perceptrons (MLPs) in classical RNNs with graph convolutions which take into consideration graph topology, enabling node hidden state updating to be dependent on historical hidden states of its local neighbouring nodes. DCRNN [33] proposed to use diffusion convolutions to model spatial dependency and further integrate it into sequence-to-sequence models. However, GCN-based MTS models relying on predefined static graph structures require substantial domain-specific expertise, and the expressive capacity is sensitive to the quality of prescribed graphs. As a result, there have been research efforts [19, 20] attempting to generate data-dependent adaptive graph structures dynamically. These methods mainly rely on learned node embeddings to infer underlying associations using attention mechanisms with the limitation that the diversity of graph structures that can be modelled may be restricted.

As mentioned previously, GCNs model underlying interactions implicitly by the message-passing function or with the help of an attention mechanism. Consequently, it is difficult to capture different types of interactions between pairs of nodes. Recently, a novel, principled framework called neural relational inference (NRI) [43] has been proposed in order to infer an explicit interaction structure of modelled dynamical systems in an unsupervised fashion. Generally, the NRI model learns the dynamics with a GCN and a latent graph structure where discrete edge types are predicted from observed trajectories using variational inference. It has been demonstrated that NRI models can predict the dynamics many time steps into the future in real motion

capture and sports tracking data with a very small number of edge types.

2.2 Hyperspectral Image Analysis

Apart from numerical data collected either at NIAB EMR or other institutions, such as meteorological stations in the UK, hyperspectral imaging techniques may provide supplementary information for realizing those research objectives outlined in section 1. Indeed, the rising availability of high-quality satellite data (hyperspectral image series) by both state [44] and private [45] sectors opens up numerous high-impact applications for machine learning methods. Among these, satellite image time series (SITS) have been demonstrated to be well-suited for analysing crop phenology [46], crop type classification [47, 48, 49, 50] and crop yield prediction [51, 52, 53, 54]. Despite the achieved success of using traditional machine learning methods, such as random forest (RF) and support vector machine (SVM), it can be seen that the gradual adoption of deep learning methods such as CNNs and RNNs for learning spatial and temporal attributes has brought significant improvements in predictive performance. For example, researchers [52] have proposed to use MODIS reflectance and temperature images to predict crop yield and have shown the transferable potential to crops in different regions. Based on the assumption proposed in [51] that pixels of satellite images may be considered permutation-invariant and spectral bands may be considered uncorrelated, processing SITS data can be reduced to sequential modelling problems where input signals at individual time steps are 2D matrices. Besides, it has been shown promising to integrate heterogeneous data, including surface reflectance, surface temperatures, climatic variables, soil property, and genotypes of crops, into deep sequential learning frameworks to improve crop yield prediction accuracy or agricultural breeding [53, 54]. Recently, a large-scale dataset called EarthNet2021 [55] has been released, containing spatio-temporal Sentinel-2 satellite imagery at 20m resolution, for benchmarking Earth surface reflectance forecasting models. It has been argued that such models would benefit downstream applications, such as crop yield prediction and biodiversity monitoring. The research mentioned above has suggested that incorporating heterogeneous data such as hyperspectral images may be promising for building powerful models that can better achieve the research objectives of this project.

2.3 Challenges & Promising Solutions

This section will be devoted to briefly illustrating challenges characteristic to MTS analysis and corresponding promising solutions, which helps make innovative developments in future research.

2.3.1 Probabilistic Time Series Forecasting

Deep sequential learning models mentioned in section 2 are all deterministic models. The only source of randomness or variability in these models comes from the conditional output probability model, which is assumed to be insufficient to model the

kind of variability observed in highly-structured data. There is recent evidence that when complex sequences such as speech and music are modelled, the performance of RNNs can be significantly improved when uncertainty is included in their hidden states [56, 57, 58]. Time series for dynamic systems have been studied extensively in systems theory. In particular, state-space models (SSMs) [59] have proven to be a powerful tool to analyze and control the dynamics. SSMs can be regarded as a probabilistic extension of RNNs, where the hidden states are assumed to be random variables. Although SSMs have found their widespread applications, their stochasticity has limited their use in the deep learning community as posterior inference for nonlinear models is computationally intractable. Benefiting from recent advances in variational inference, especially stochastic gradient variational Bayes (SGVB) [60, 61], which makes approximate inference of latent variables computationally tractable, there have been research efforts [62, 63, 64, 24] attempting to bridge the gap between SSMs and deep neural networks, giving rise to sequential latent variable models or deep state-space models. For example, researchers proposed to use deep neural networks to enhance classical Kalman filters with arbitrarily complex transition dynamics and emission distributions, resulting in deep Kalman filters [62]. Deep state-space models proposed in [63], a framework for time series forecasting, marries SSMs with RNNs by using RNNs to predict parameters of linear state-space models from which observations are generated. Deep Variational Bayes filters [64], a new method for unsupervised learning and identification of latent Markovian state-space models, make the recognition model predict transition parameters rather than latent states to allow reconstruction errors to be backpropagated into transition models. The recently proposed attentive state-space model [24] for disease progression modelling applies attention to the latent state space, eliminating the restriction caused by the Markovian assumption so that richer distributions can be modelled. All these methods follow similar design principles, i.e., retaining structural assumptions in SSMs while making the stochastic state transitions of SSMs nonlinear. Integrating SSMs with deep neural networks enjoys benefits of both worlds, such as efficient data utilization and the capability to directly process raw time series with considerably less human effort.

2.3.2 Multimodal Nature

One of the greatest challenges in pedestrian trajectory forecasting is the multimodal distribution of human behaviour. Instead of predicting a single mode of human behaviour with high variance, it has been shown beneficial to allow models to generate multiple plausible or socially acceptable trajectories. For example, Social-GAN [13] is the pioneering work where generative adversarial network (GAN) [65] is used to distinguish real trajectories from synthesized ones instead of the commonly used L2 loss that tends to make models learn the ”average behaviour”. In Social-BiGAT [14], researchers proposed a graph-based GAN to generate realistic, multimodal trajectory predictions and employed a reversible mapping similar to that used in BicycleGAN [66] to improve the diversity of generated samples. It has been observed that simply incorporating random noises as additional inputs to conditional GAN does not guarantee increased variations of the generated outputs as mode collapse may still easily

occur. Therefore, increasing the diversity of generated samples of GAN is nontrivial. Consequently, it is still an open problem and has been actively studied, especially in the research field of image-to-image translation [66, 67, 68, 69]. The promising solutions include enforcing bijective consistency [66, 67] between the latent code space and output space and/or disentangled representation learning [68, 69] in deep generative models. It is reasonable to expect that methods developed for tackling this particular challenge would also inspire model design for MTS and the possibility of providing a better alternative for traditional evaluation metrics adopted in MTS, such as mean absolute percentage error (MAPE) and mean absolute scaled error (MASE).

2.3.3 Time Series Representations

It has been acknowledged that the great success achieved by deep learning models can be attributed to the representation learning capability. Therefore, there have been research efforts attempting to explore more effective time series representations for deep learning models. For example, in [70] memory cells in traditional LSTMs have been superseded by state frequency memory which resembles Discrete Fourier Transform (DFT), allowing models to capture multi-frequency patterns. The recently proposed EvoNet [71] transforms raw time series into lower-dimensional temporal dynamics by constructing evolutionary state graph sequences where nodes in graphs represent representative temporal patterns so that anomalous events can be detected when certain kinds of state transitions occur. Similarly, Time2Graph [23], a novel representation learning algorithm for time series modelling, transforms time series to shapelet [72] evolution graphs. N-BEATS [18], a novel architecture for univariate time series forecasting, uses MLPs to predict basis function expansions and their coefficients, and then combine them to obtain predicted values in the specified forecast horizon, showing impressive performance on heterogeneous time-series datasets. These methods share the similarity of decomposing raw time series into basic building units and model dynamics existing in these units, introducing hierarchical representation learning into time series modelling.

3. Expected Outcomes & Relevance to Research in CSEE

3.1 Expected Outcomes

- Building a standard preprocessing pipeline for data collected at XXXX and possibly from other sources to create curated and analysis-ready datasets;
- Conducting extensive experiments and comprehensive comparisons using statistical and deep-learning-based models to realize two main objectives of this research project, i.e., predicting flowering time and the degree of fruit ripeness;
- Based on established baseline models, making further improvements following

the potential research directions outlined in 2.3 and summarizing research outputs for making publications on international conferences and journals;

- Possibly further testing the generalization ability of proposed algorithms, especially for other similar applications in agriculture and horticulture and making a toolkit publicly available in the hope of facilitating quick experimentation and prototype development for researchers in the community.

3.2 Relevance to Research in CSEE

This research project is characterized by being interdisciplinary, requiring expertise in biology, plant science, data science, artificial intelligence, etc., and having significant practical values. Therefore, I believe it strongly aligns with the research interests and vision of the CSEE department.

References

[1] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[2] L.-J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Transactions on neural networks*, vol. 14, no. 6, pp. 1506–1518, 2003.

[3] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[4] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," 1961.

[5] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[6] M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi, "Bidirectional attention flow for machine comprehension," *arXiv preprint arXiv:1611.01603*, 2016.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[8] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International conference on machine learning*. PMLR, 2014, pp. 1764–1772.

[9] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu *et al.*, "Conformer: Convolution-augmented transformer for speech recognition," *arXiv preprint arXiv:2005.08100*, 2020.

[10] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 879–888.

[11] Y. Ye, M. Singh, A. Gupta, and S. Tulsiani, "Compositional video prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 10 353–10 362.

[12] S. Tulyakov, M.-Y. Liu, X. Yang, and J. Kautz, "Mocogan: Decomposing motion and content for video generation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1526–1535.

[13] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2255–2264.

[14] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, S. H. Rezatofighi, and S. Savarese, "Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks," *arXiv preprint arXiv:1907.03395*, 2019.

[15] B. Ivanovic and M. Pavone, "The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2375–2384.

[16] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," *arXiv preprint arXiv:1704.02971*, 2017.

[17] S.-Y. Shih, F.-K. Sun, and H.-y. Lee, "Temporal pattern attention for multivariate time series forecasting," *Machine Learning*, vol. 108, no. 8, pp. 1421–1441, 2019.

[18] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-beats: Neural basis expansion analysis for interpretable time series forecasting," *arXiv preprint arXiv:1905.10437*, 2019.

[19] L. Bai, L. Yao, C. Li, X. Wang, and C. Wang, "Adaptive graph convolutional recurrent network for traffic forecasting," *arXiv preprint arXiv:2007.02842*, 2020.

[20] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph wavenet for deep spatial-temporal graph modeling," *arXiv preprint arXiv:1906.00121*, 2019.

[21] Y. Liu, X. Shi, L. Pierce, and X. Ren, "Characterizing and forecasting user engagement with in-app action graph: A case study of snapchat," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2023–2031.

[22] L. Bai, L. Yao, S. Kanhere, X. Wang, Q. Sheng *et al.*, "Stg2seq: Spatial-temporal graph to sequence model for multi-step passenger demand forecasting," *arXiv preprint arXiv:1905.10069*, 2019.

[23] Z. Cheng, Y. Yang, W. Wang, W. Hu, Y. Zhuang, and G. Song, "Time2graph: Revisiting time series modeling with dynamic shapelets," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 3617–3624.

[24] A. Alaa and M. van der Schaar, "Attentive state-space modeling of disease progression," 2019.

[25] B. Lim and M. van der Schaar, "Disease-atlas: Navigating disease trajectories using deep learning," in *Machine Learning for Healthcare Conference*. PMLR, 2018, pp. 137–160.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[27] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[28] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.

[29] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," *arXiv preprint arXiv:2012.07436*, 2020.

[30] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long-and short-term temporal patterns with deep neural networks," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 95–104.

[31] Y. Wang, L. Jiang, M.-H. Yang, L.-J. Li, M. Long, and L. Fei-Fei, "Eidetic 3d lstm: A model for video prediction and beyond," in *International conference on learning representations*, 2018.

[32] R.-G. Cirstea, C. Guo, and B. Yang, "Graph attention recurrent neural networks for correlated time series forecasting–full version," *arXiv preprint arXiv:2103.10760*, 2021.

[33] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *arXiv preprint arXiv:1707.01926*, 2017.

[34] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, 2020.

[35] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *arXiv preprint arXiv:1312.6203*, 2013.

[36] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *arXiv preprint arXiv:1606.09375*, 2016.

[37] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[38] J. Atwood and D. Towsley, "Diffusion-convolutional neural networks," in *Advances in neural information processing systems*, 2016, pp. 1993–2001.

[39] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International Conference on Machine Learning*. PMLR, 2017, pp. 1263–1272.

[40] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *arXiv preprint arXiv:1706.02216*, 2017.

[41] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[42] R. Ying, J. You, C. Morris, X. Ren, W. L. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," *arXiv preprint arXiv:1806.08804*, 2018.

[43] T. Kipf, E. Fetaya, K.-C. Wang, M. Welling, and R. Zemel, "Neural relational inference for interacting systems," in *International Conference on Machine Learning*. PMLR, 2018, pp. 2688–2697.

[44] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort *et al.*, "Sentinel-2: Esa's optical high-resolution mission for gmes operational services," *Remote sensing of Environment*, vol. 120, pp. 25–36, 2012.

[45] M. Technologies., "Helping facebook connect the world with deep learning, accessed nov. 2019." 2016. [Online]. Available: http://blog.digitalglobe.com/news/helping-facebookconnect-the-world-with-deep-learning/

[46] A. Vrieling, M. Meroni, R. Darvishzadeh, A. K. Skidmore, T. Wang, R. Zurita-Milla, K. Oosterbeek, B. O'Connor, and M. Paganini, "Vegetation phenology from sentinel-2 and field cameras for a dutch barrier island," *Remote sensing of environment*, vol. 215, pp. 517–529, 2018.

[47] M. Rußwurm and M. Korner, "Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 11–19.

[48] M. Rußwurm and M. Körner, "Convolutional lstms for cloud-robust segmentation of remote sensing imagery," *arXiv preprint arXiv:1811.02471*, 2018.

[49] V. S. F. Garnot, L. Landrieu, S. Giordano, and N. Chehata, "Time-space tradeoff in deep learning models for crop classification on satellite multi-spectral image time series," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 6247–6250.

[50] ——, "Satellite image time series classification with pixel-set encoders and temporal self-attention," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 325–12 334.

[51] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep gaussian process for crop yield prediction based on remote sensing data," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[52] A. X. Wang, C. Tran, N. Desai, D. Lobell, and S. Ermon, "Deep transfer learning for crop yield prediction with remote sensing data," in *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, 2018, pp. 1–5.

[53] J. Sun, Z. Lai, L. Di, Z. Sun, J. Tao, and Y. Shen, "Multilevel deep learning network for county-level corn yield estimation in the us corn belt," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5048–5060, 2020.

[54] J. Shook, T. Gangopadhyay, L. Wu, B. Ganapathysubramanian, S. Sarkar, and A. K. Singh, "Crop yield prediction integrating genotype and weather variables using deep learning," *arXiv preprint arXiv:2006.13847*, 2020.

[55] C. Requena-Mesa, V. Benson, J. Denzler, J. Runge, and M. Reichstein, "Earthnet2021: A novel large-scale dataset and challenge for forecasting localized climate impacts," *arXiv preprint arXiv:2012.06246*, 2020.

[56] O. Fabius and J. R. Van Amersfoort, "Variational recurrent auto-encoders," *arXiv preprint arXiv:1412.6581*, 2014.

[57] J. Chung, K. Kastner, L. Dinh, K. Goel, A. Courville, and Y. Bengio, "A recurrent latent variable model for sequential data," *arXiv preprint arXiv:1506.02216*, 2015.

[58] S. Gu, Z. Ghahramani, and R. E. Turner, "Neural adaptive sequential monte carlo," *arXiv preprint arXiv:1506.03338*, 2015.

[59] J. Durbin and S. J. Koopman, *Time series analysis by state space methods*. Oxford university press, 2012.

[60] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[61] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *International conference on machine learning*. PMLR, 2014, pp. 1278–1286.

[62] R. G. Krishnan, U. Shalit, and D. Sontag, "Deep kalman filters," *arXiv preprint arXiv:1511.05121*, 2015.

[63] S. S. Rangapuram, M. W. Seeger, J. Gasthaus, L. Stella, Y. Wang, and T. Januschowski, "Deep state space models for time series forecasting," *Advances in neural information processing systems*, vol. 31, pp. 7785–7794, 2018.

[64] M. Karl, M. Soelch, J. Bayer, and P. Van der Smagt, "Deep variational bayes filters: Unsupervised learning of state space models from raw data," *arXiv preprint arXiv:1605.06432*, 2016.

[65] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv preprint arXiv:1406.2661*, 2014.

[66] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," *arXiv preprint arXiv:1711.11586*, 2017.

[67] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[68] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, "Diverse image-to-image translation via disentangled representations," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 35–51.

[69] A. Almahairi, S. Rajeshwar, A. Sordoni, P. Bachman, and A. Courville, "Augmented cyclegan: Learning many-to-many mappings from unpaired data," in *International Conference on Machine Learning*. PMLR, 2018, pp. 195–204.

[70] L. Zhang, C. Aggarwal, and G.-J. Qi, "Stock price prediction via discovering multi-frequency trading patterns," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 2141–2149.

[71] W. Hu, Y. Yang, Z. Cheng, C. Yang, and X. Ren, "Time-series event prediction with evolutionary state graph," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, pp. 580–588.

[72] L. Ye and E. Keogh, "Time series shapelets: a novel technique that allows accurate, interpretable and fast classification," *Data mining and knowledge discovery*, vol. 22, no. 1, pp. 149–182, 2011.