Towards Sustainable WWTP Operations: Forecasting Energy Consumption with Explainable Disentangled Graph Convolutional Networks

Louis Allena, Joan Cordinera,\*

aThe Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom

\*j.cordiner@sheffield.ac.uk

Abstract

Understanding and predicting factors influencing energy consumption in wastewater treatment plants (WWTP) is critically important. Such insights offer opportunities for significant cost savings and can facilitate the operation of more environmentally sustainable processes. Machine learning (ML) has shown promise in modeling the complex non-linear relationships inherent to WWTPs, however many models suffer from interpretability issues limiting their real-world application. Furthermore, conventional ML methods often struggle with long-term forecasting in the face of variable input data, a common challenge in WWTP operations.

Our study proposes a novel approach that combines operating sensor data with core chemical engineering knowledge to address these challenges. We propose a disentangled graph convolutional network tailored for time series forecasting, drawing upon sensor data from a WWTP in Melbourne. This method learns disentangled representations of latent factors within the dataset which can help explain consumption trends, offering insight that can help reduce overall energy use. When benchmarked against traditional ML models, our approach exhibited superior prediction accuracy and model reliability. Our findings underscore the potential of integrating domain-specific knowledge into data-driven methodologies, especially in intricate manufacturing settings. The standout performance of disentangled graph convolutional networks, as evidenced in our study, offers a promising avenue for delivering transparent and actionable energy consumption forecasts, empowering WWTP operators to make decisions that enhance process sustainability.

**Keywords**: Energy Forecasting, Explainable Machine Learning, Wastewater Treatment

* 1. Introduction

Wastewater treatment is a critical challenge for urban communities, vital for health and various activities (Ahmad and Chen, 2018). As urban populations grow exponentially, the demand for clean water intensifies, posing global challenges like climate change and increased energy consumption in wastewater treatment, constituting 40% of urban energy usage (Bagherzadeh et al., 2021). Addressing this issue is crucial for sustainable global goals.

Understanding the drivers of high energy consumption is pivotal for reducing overall energy usage in wastewater treatment plants (WWTPs), and long-term energy consumption forecasts are key for informed decision-making (Ahmad and Chen, 2018; Bagherzadeh et al., 2021). However, forecasting energy consumption in WWTPs is intricate due to the nonlinear nature of the biological reaction process and diverse influencing factors (Li et al., 2019; Jiang et al., 2014).

Machine learning (ML) has shown promise in accurately capturing consumption trends, but concerns persist about model robustness to noise (Picos-Benítez et al., 2020; Bagherzadeh et al., 2021; Alali et al., 2023). Graph convolutional networks (GCNs) offer a novel approach for complex time series forecasting, explicitly leveraging the relational structure of data to capture complex dependencies and patterns (Hu et al., 2022). Combining GCNs with long short-term memory (LSTM) models enhances accuracy in time series forecasting, as demonstrated in financial time series forecasting (Valaskova et al., 2023). However, this graph-based approach has not been applied to energy consumption (EC) forecasting in WWTPs, possibly due to GCN architectures neglecting latent factors, limiting robustness and explainability (Ma et al., 2019).

To address these challenges, this paper introduces DisenGC-LSTM, a novel approach combining disentangled graph convolutional networks with LSTM architectures for accurate and explainable time series forecasting in WWTPs (Ma et al., 2019). DisenGC-LSTM, aims to overcome the limitations of traditional GCNs in neglecting latent factors and compromising model robustness and explainability (Ma et al., 2019).

* 1. Methodology
		1. DisenConv Layer

Let be an undirected graph where is a finite set of nodes such that corresponds to the number of variables in a dataset , where each node has a feature vector **.**  is a set of edges such that shows the existence of an edge between nodes and .

Figure 1 shows the disentangled convolutional layer which outputs a disentangled representation of a node where describes the aspect of node by iteratively searching each of the neighbors of node and assigning a probability that the nodes are connected due independent factor . This is done by projecting the feature vector into different subspaces:

|  |  |
| --- | --- |
|  | (1) |

where and are the parameters of channel and is a ReLU activation function, is the size of the output of the DisenConv layer. Since feature vector may be incomplete for a real world system, we cannot directly map to . Instead, Ma et al. propose a neighborhood routing mechanism to iteratively mine information from the surrounding neighbors of node to construct . We only want to use information pertinent to factor and therefore should look to only collect information from neighbors of node connected *because* of factor . To do this, let represent the probability that node and are connected due to factor , where and .

The neighborhood routing mechanism iteratively infers a value for before constructing , starting by initializing where is a hyper-parameter that controls the ‘hardness’ of the assignment (Ma et al., 2019). The mechanism searches for the largest cluster in each subspace, with the constraint that a neighbor can only belong to a single subspace. Since each channel receives different subsets of neighbors we can maintain that each channel represents an independent factor. The following are completed on every iteration:

*Figure 1: Disentangled convolutional layer (DisenConv) which takes in a graph and produces a disentangled representation of the graph for each node shown here for node u.*

|  |  |
| --- | --- |
|  | (2) |
|  | (3) |

for each iteration where . The iterations output where is the center of the subspace. This DisenConv layer can be integrated into a recurrent neural network architecture to enhance the accuracy and explainability of time-series forecasting models.

* + 1. Time-series Forecasting with DisenConv Layers

Given is an undirected graph representation of a WWTP, is a dataset such that is the number of features in the dataset and is the number of time steps in the dataset, we aim to forecast the energy consumption of the plant. To do this we integrate the DisenConv layer described above with a long-short-term memory (LSTM) architecture (DisenGC-LSTM) architecture in which latent factors are disentangled through the convolution, and temporal information is learned by the LSTM. This is an adaptation of the GC-LSTM model demonstrated for link prediction by (Chen et al., 2021). The impact of the DisenConv layer can therefore be directly observed by comparing the proposed DisenGC-LSTM model with a GC-LSTM model where the architecture is kept identical apart from the convolution method. The models will be assessed quantitatively using the root-mean-squared error (RMSE) and R2 value.

* 1. Case Study

The Melbourne Easter Treatment Plant (ETP) treats wastewater from around 2.5 million homes, approximately half of Melbourne’s total sewage (Melbourne Water, 2005). This equates to around 330 million liters of sewage per day. The treatment process begins with physical cleaning of the raw wastewater, such as settling large particles, before moving on to biological treatment.

* + 1. Data Description

The dataset comprises of 4 wastewater quality attributes collected from sampled sensors - Ammonia (NH4-N), total nitrogen (TN), chemical oxygen demand (COD) and biological oxygen demand (BOD), along with 2 hydraulic parameters in the inlet flowrate of sewage and outlet flowrate from the plant. Since climate effects have a significant impact on the treatment of wastewater data from the Melbourne airport weather station was joined to give an augmented dataset (Bagherzadeh et al., 2021). This offers an additional 9 features including average temperature (Tavg), maximum temperature (Tmax), minimum temperature (Tmin), atmospheric pressure (AP), average humidity (H), total precipitation (Pr), average visibility (VIS), average wind speed (WSavg), and maximum wind speed (WSmax). Finally, total electricity consumption (EC) is taken via revenue quality meters. This gives a dataset of 16 features with approximately 1000 samples taken across a period of 5 years.

* + 1. Graph Data

We format the ETP dataset into a graph that can be read by both the DisenGC-LSTM and GCLSTM models along with the data. While there are many methods of graph construction, from leveraging mechanistic relationships to establish graph edges, to data-driven causal inference methods, the goal of this work is to establish if the DisenConv layer can group important variables into distinct channels that represent latent factors. Therefore, input to the model is a fully connected, undirected graph where since each node represents a dataset feature, and . Data for the energy consumption is not included in the graph. Instead, this data is supplied as targets for the prediction model to train against.

*Figure 2: Results from the plotting of the real energy consumption data (MWh/day) against the forecast from the DisenGC-LSTM model (green) and the GC-LSTM model (red), demonstrating the performance of the DisenGC-LSTM model.*

* + 1. Model Training

The problem is formulated as a supervised machine learning problem in which each of the models is given a set of training data where is the number of features of the dataset that each correspond to a node in the graph, and is the length of the training dataset. The models are trained to predict a the energy consumption of the plant based on collected training samples. Prior to training, both the samples and the targets are scaled in the region of [0, 1]. The models were trained on 80% of the available data, with the remaining data split into 3 test sets of 10% each. This is so the results can be validated across 3 testing sets giving a more reliable performance estimate. The models were trained over 100 epochs in which the data was divided into mini-batches of 32 samples. The Adam optimizer was used with a learning rate of 0.001 (Kingma and Ba, 2014). The loss function used was the mean squared error (MSE) between the predicted samples and the training samples. Hyper-parameters for each of the models were tuned using the grid search method.

* 1. Results & Discussion

Figure 2 shows the forecasts for both the GC-LSTM model and the DisenGC-LSTM model with the proposed disentanglement mechanism. We see that the DisenGC-LSTM model provides a superior forecast. The DisenGC-LSTM forecast (green) fits very closely to the real data (gray) and is clearly able to predict the pattern of the data well, capturing both the trend of the data and largely able to replicate and predict the noise that is seen by the real energy consumption data. For example, between timestamps 1233 and 1263 we see that the DisenGC-LSTM prediction accurately captures the shape of the real data, despite there being variations in the data. This is in contrast to the GC-LSTM model which does not predict the rise in energy consumption across these timestamps. In fact, the GC-LSTM model predicts poorly and, despite capturing rough trends, tends to miss almost all important characteristics of the energy consumption data. Proof of this can be seen when comparing the GC-LSTM prediction to a straight-line forecast taken from the mean average of the training data, as shown in Table 1. We can see that the average forecast actually predicts better, with an RMSE of 20.36 MWh/day compared to 28.37 MWh/day for the GC-LSTM model. Comparatively, the DisenGC-LSTM model achieves an RMSE of just 10 MWh/day, outperforming the GC-LSTM model considerably and showing its effectiveness as a model compared to using an average forecast. The closeness of this fit is also reflected in the R2 scores for both models, where we see the DisenGC-LSTM model achieving an R2 score of 0.66, where the GC-LSTM model has a negative R2 score showing that a simple averaging model performs better.

Table 1: Summary of results comparing the two run models against a base line average prediction.

**Model**

Average

**RMSE (MWh/day)**

20.36

GC-LSTM 28.37

DisenGC-LSTM 10.91

**R**2

0.00

-0.15

0.66

* 1. Conclusion

In this paper we introduced a novel approach by combining graph convolutional networks (GCNs) with disentangled graph convolutional layers for time series forecasting of energy consumption in a WWTP. The DisenConv layer aimed to disentangle latent factors within the graph, providing enhanced accuracy and explainability compared to traditional GCN models. Our methodology was applied to the Melbourne Eastern Treatment Plant (ETP) dataset, incorporating various wastewater quality attributes, hydraulic parameters, and climate data. The results demonstrated the superiority of the proposed DisenGC-LSTM model over the standard GC-LSTM model. The DisenGC-LSTM accurately captured the complex dynamics of energy consumption, outperforming the traditional graph convolutional model which have historically been shown to exhibit high accuracy. The DisenGC-LSTM model’s ability to disentangle latent factors within the graph led to improved forecasting accuracy and a more nuanced understanding of the underlying processes.

This work contributes to the advancement of machine learning applications in the domain of wastewater treatment, offering a promising avenue for optimizing energy consumption in WWTPs. The DisenGC-LSTM model’s success opens the door to further exploration of disentangled graph- based approaches for addressing the challenges posed by the intricate and nonlinear dynamics of wastewater treatment processes.

References

T. Ahmad, H. Chen, 2018. Utility companies strategy for short-term energy demand forecasting using machine learning based models. Sustainable Cities and Society401 39, 401–417.

URL https://doi.org/10.1016/j.scs.2018.03.002

Y. Alali, F. Harrou, Y. Sun, 7 2023. Unlocking the potential of wastewater treatment: Machine learning based energy consumption prediction. Water (Switzerland) 15.

F. Bagherzadeh, A. S. Nouri, M. J. Mehrani, S. Thennadil, 10 2021. Prediction of energy consumption and evaluationof affecting factors in a full-scale wwtp using a machine learning approach. Process Safety and Environmental Protection 154, 458–466

J. Chen, X. Wang, X. Xu, J. Chen, X. Wang, X. Xu, 2021. Gc-lstm: Graph convolution embedded lstm for dynamic network link prediction.

Y. Hu, X. Cheng, S. Wang, J. Chen, T. Zhao, E. Dai, 2 2022. Times series forecasting for urban building energy consumption based on graph convolutional network. Applied Energy 307, 118231.

Y. Jiang, W. Fu, L. Mao, F. Ren, L. Yang, J. Xiang, R. Liang, H. Hao, Z. Wang, 2 2014. Influence factors analysis of urban sewage treatment plant on energy consumption. Beijing Jiaotong Daxue Xuebao/Journal of Beijing Jiaotong University 38, 33–37.

D. P. Kingma, J. Ba, 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Z. Li, Z. Zou, L. Wang, 2019. Analysis and forecasting of the energy consumption in wastewater treatment plant. Mathematical Problems in Engineering 2019.

J. Ma, P. Cui, K. Kuang, X. Wang, W. Zhu, 2019. Disentangled graph convolutional networks. pp. 4212–4221.

W. Melbourne Water, 2005. Melbourne water. Social and Environment Data 2006.

A. R. Picos-Benítez, B. L. Martínez-Vargas, S. M. Duron-Torres, E. Brillas, J. M. Peralta-Hernández, 11 2020. The use of artificial intelligence models in the prediction of optimum operational conditions for the treatment of dye wastewaters with similar structural characteristics. Process Safety and Environmental Protection 143, 36–44.

K. Valaskova, A. Lazcano, P. J. Herrera, M. Monge, 2023. Combined model based on recurrent neural networks and graph convolutional networks for financial time series forecasting. Mathematics224 11. URL https://doi.org/10.3390/math11010224