From Regression to Learning the Underlying Trajectory: Forecasting the Homelessness

Ishita Khanna

11 October 2023

Walsall Council

MONASH University

Motivation

- Predicting homelessness allow us to focus on prevention rather than reaction.
- With limited resources available for social services and housing programs, it's crucial to allocate resources efficiently.
- Addressing homelessness contributes to the stability of the housing market and reduces the economic burden on healthcare, emergency services, and the criminal justice system.
- Understanding and addressing homelessness can position the UK as a global leader in social policy and serve as a model for other nations facing similar challenges.
- The motivation is to improve the lives of individuals and families, ensuring they have the opportunity for stable housing, better health, education, and employment prospects.

- Refer [1], [2] and [3].

[1] United Nations Commission for Social Development. Affordable housing and social protection systems for all to address homelessness. Report of the Secretary-General. E/CN.5/2020/3. New York: United Nations Commission for Social Development; 2019 Nov 27 [cited 2020 May 10]. Available from: https://undocs.org/en/E/CN.5/2020/3.

[2] Directorate of Employment, Labour and Social Affairs. HC3.1 Homeless population. Paris: Organisation for Economic Co-operation and Development; 2020 [cited 2020 May 3]. Available from: https://www.oecd.org/els/family/HC3-1-Homeless-population.pdf.

[3] **T**sai J. Lifetime and 1-year prevalence of homelessness in the US population: results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. J Public Health (Oxf). 2018;40(1):65–74. pmid:28335013

Preliminaries

Traditional Approaches: Regression-based Analysis

- Given, at time step (i.e. year)
  - Y represents the “Homeless Decisions"
  - X_1 represents the “Year”
  - X_2 represents factors such as “Local Authority Regions”.
  - Then, a simple multi-regression model can be given using
    - $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
    - where,
    - β_0 is the intercept, while β_1 and β_2 are the coefficient parameters, ε is the error terms.

[4] Gutwinski, S., Schreiter, S., Deutscher, K. and Fazel, S., 2021. The prevalence of mental disorders among homeless people in high-income countries: An updated systematic review and meta-regression analysis. *PLoS medicine*, *18*(8), p.e1003750.

Limitations to a Regression-based Analysis

Assumptions about Data- Linearity, Homoscedasticity, Normality & Independence assumption.

Poor Modelling - Regression analysis often ignores the temporal nature of the data while only focusing on $y_t = f(x_{1,t}, \cdots, x_{n,t})$ rather than $y_t = g(y_{t-1}, \cdots, y_0) + f(x_{1,t}, \cdots, x_{n,t})$, where the first term captures the temporal property.

The success of Deep Learning

With the surge in computational capabilities and access to abundant data, neural networks emerged as the state-of-the-art for prediction tasks.

Complicated neural network architectures such as Recurrent Neural Networks (RNNs) were proposed for the time-series predictions.

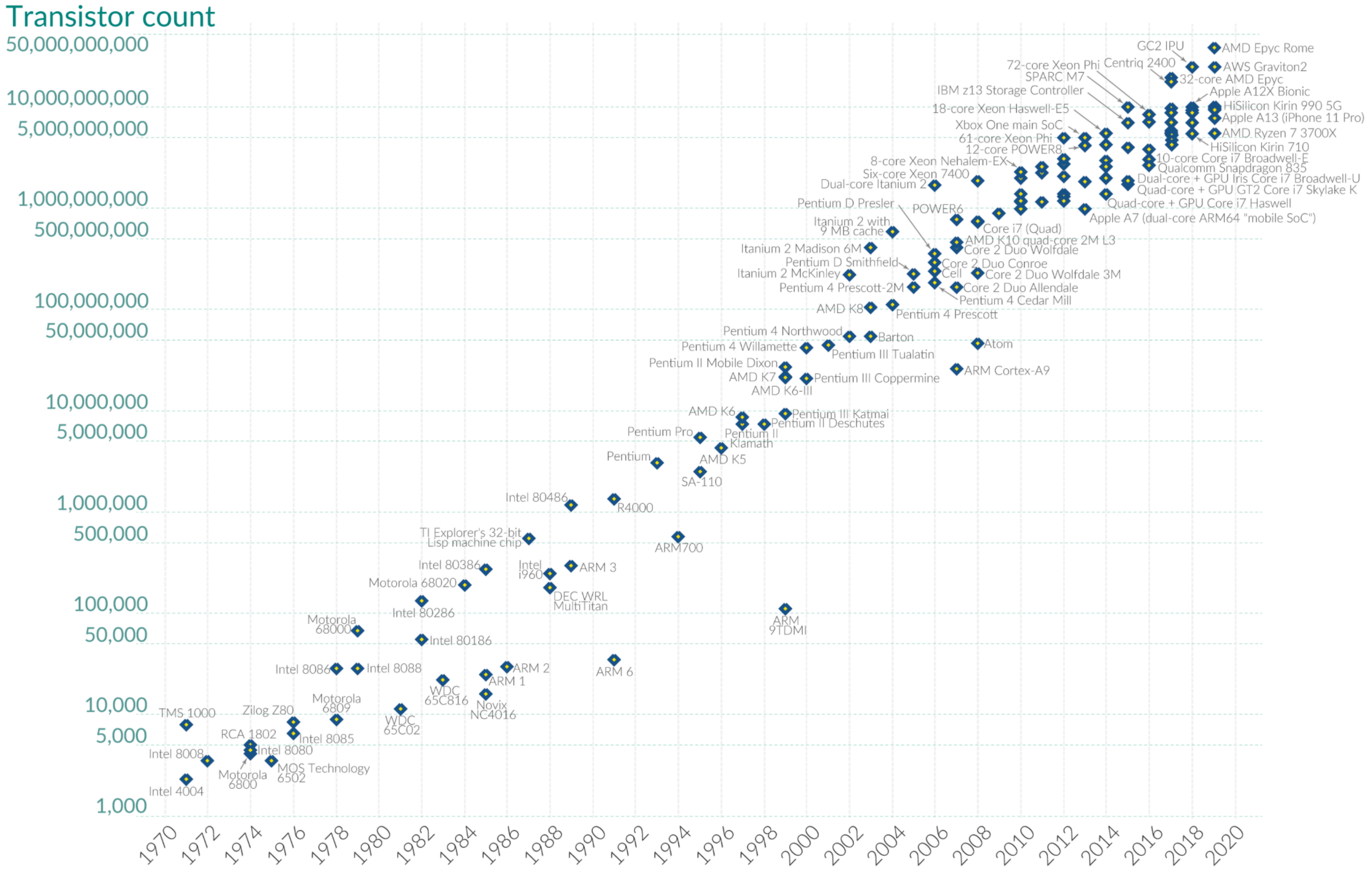


Figure 1. The number of transistors on microchips over years

https://en.wikipedia.org/wiki/Moore's_law

Preliminaries
Traditional Approaches: RNNs for forecasting

The hidden state h_t at time step t is given as

State Update:

$$h_t = \sigma(W_{hh} \cdot h_{t-1} + W_{xh} \cdot x_t + b_h),$$

Where:

$\sigma(\cdot)$is a non-linear activation function

W_{hh} is the weight matrix for the previous hidden state.

W_{xh} is the weight matrix for the current input.

b_h is the bias.

Then, the predictions can be now be given as

Output at time t :

$y_t = W_{hy} \cdot h_t + b_y$Where:

W_{hy} is the weight matrix connecting the hidden state to the output.

b_y is the output bias.

For training, we simply use the loss function as to backpropogate

$$\mathscr{L}_{mse} = \sum_{t=0}^{T} ||y_t - \hat{y}_t||$$

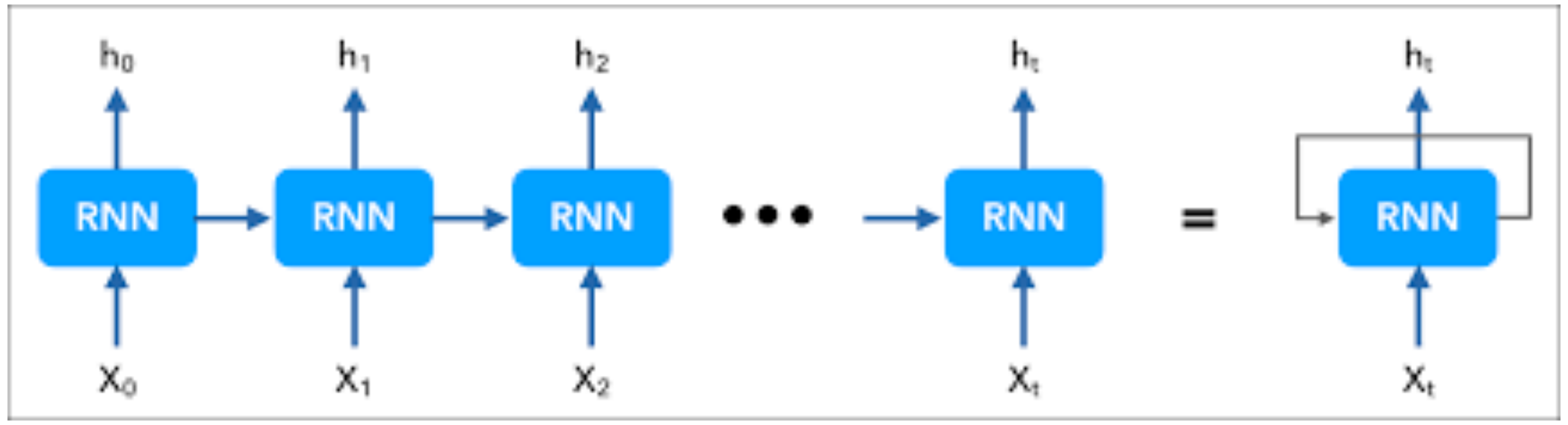


Figure 2. Architecture of the RNN

Downfalls of the traditional approaches

- While traditional methodologies that predict y based on x have shown significant efficacy, they come with inherent challenges in practical applications (where data is collected):

  - **Over-reliance on x Variables**: A heavy dependence on predictor variables can sometimes overshadow the intrinsic properties of the target variable y.

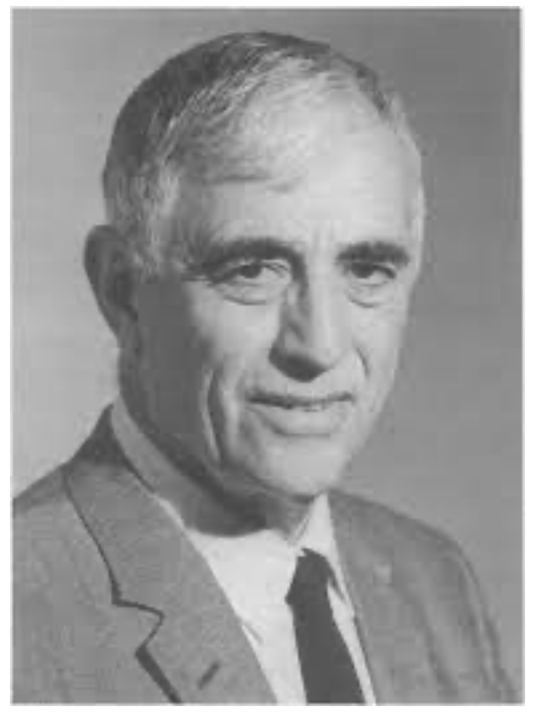
  - **Data Incompleteness**: The presence of missing data can introduce biases and reduce the robustness of the predictions.

  - **Absence of Correlation**: In some instances, the predictors x might not have a meaningful or strong relationship with the target variable y, leading to weak predictive power.

Learning the underlying trajectory

Koopman Theory

- The discrete mapping between the target variable y can be represented as:
  - $y_{t+1} = F(y_t)$,
    - where $F(\cdot)$ is a non-linear function mapping which is usually unknown.
  - Lifting operation to represent in linear space using an observable function $\varphi(\cdot)$, which gives
    - $\mathbf{z}_t = \varphi(y_t) \in \mathbb{R}^k$, where k can be infinite dimensional long.
    - The mapping, now can be given using a linear operation,
      - $\mathbf{z}_{t+1} = \mathbf{K}\mathbf{z}_t$, where $\mathbf{K}$ is the Koopman Operator.
    - Finally, we have
      - $\hat{y}_{t+1} = \varphi^{-1}(\mathbf{K}\varphi(y_t))$ which represents a one step prediction.
    - We can exploit the property of the linearity to make T time step predictions by sampling multiplying the $\mathbf{K}$ matrix T times.
      - $$\begin{aligned} \hat{y}_{t+1} &= \varphi^{-1}(\mathbf{K}\varphi(y_t)) \\ \hat{y}_{t+2} &= \varphi^{-1}(\mathbf{K}^2\varphi(y_t)) \\ &\vdots \\ \hat{y}_{t+T} &= \varphi^{-1}(\mathbf{K}^T\varphi(y_t)) \end{aligned}$$



B.O. Koopman (1931)

"It is possible to represent a nonlinear dynamical system in terms of an infinite-dimensional linear operator"

Approximating the Koopman Operator

Neural Network approach

- We do not discuss the previously proposed approaches such as Dynamic Mode Decomposition (DMD) or Extended DMD (EDMD) and rather focus on the state-of-the-art approach for learning the Koopman Operator.

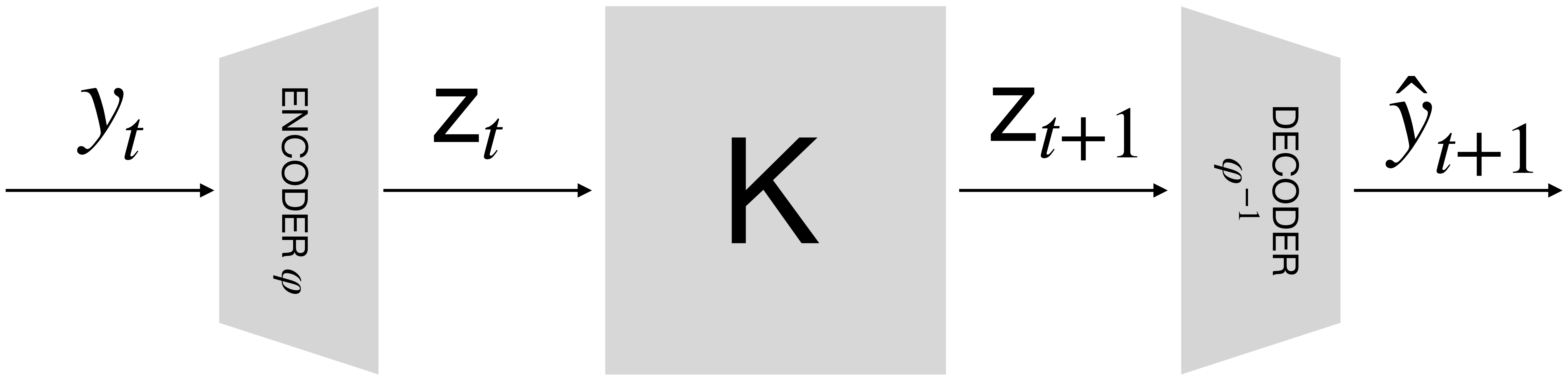


Figure 3. The Koopman Autoencoder (KAE).

- Encoder φ and Decoder φ^{-1} are blocks of fully connected layers.

- K is a fully connected layer without bias or non-linear activations.

Lusch, B., Kutz, J.N. and Brunton, S.L., 2018. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications*, 9(1), p.4950.

Training the KAE

- Training the KAE considers the following loss functions
  - $\mathscr{L}_{rec} = \sum_{t=1}^{T} ||y_t - \hat{y}_t||_2^2$
  - $\mathscr{L}_{linear} = \sum_{t=1}^{\tau} ||\mathrm{z}_t - \mathbf{K}^t \mathrm{z}_1||_2^2$, where we enforce linearity over τ time steps.
  - $\mathscr{L}_{pred} = \sum_{t=1}^{S_p} ||\varphi^{-1}(\mathbf{z}_{t+S_p}) - \varphi^{-1}(\mathbf{K}^{S_p}\mathbf{z}_t)||_2^2$, S_p determines the prediction window.
  - The total loss function is written as a weighted sum of the
  - $\mathscr{L}_{KAE} = \alpha_1 \mathscr{L}_{rec} + \alpha_2 \mathscr{L}_{pred} + \alpha_3 \mathscr{L}_{linear}$

Interesting use-cases

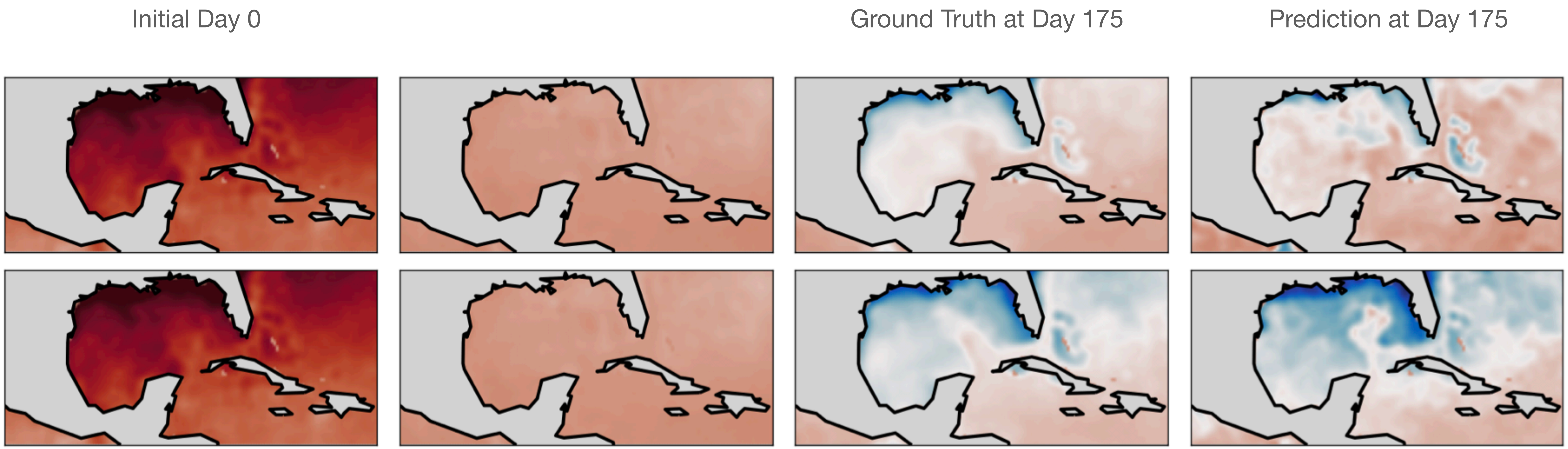


Figure 4. Predicting SEA surface levels.

Predictions were made using Consistent KAE (CKAE), where modifications were made to the original KAE.

Azencot, O., Erichson, N.B., Lin, V. and Mahoney, M., 2020, November. Forecasting sequential data using consistent Koopman autoencoders. In *International Conference on Machine Learning* (pp. 475-485). PMLR.

Simulation using a circular slow-manifold dataset

$$x(t) = r(t)\cos\left(2\pi\frac{t \mod 50}{50}\right)$$

$$y(t) = r(t)\sin\left(2\pi\frac{t \mod 50}{50}\right), \text{where}$$

$$r(t) = 1 + 0.5\sin\left(\frac{2\pi t}{T}\right)$$

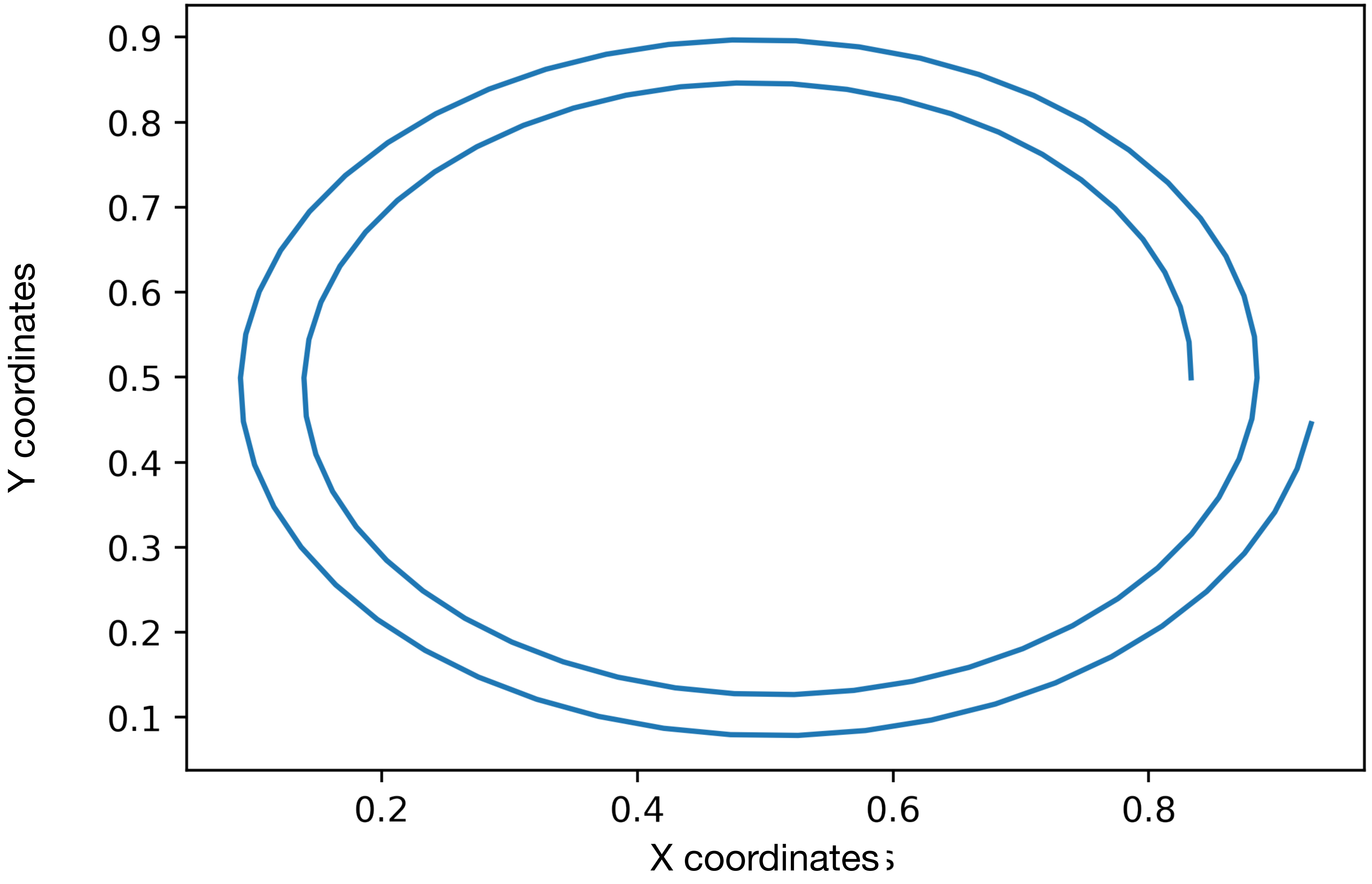


Figure 5. Trajectory depicted over 50 time steps.

Simulating the trajectory using the KAE

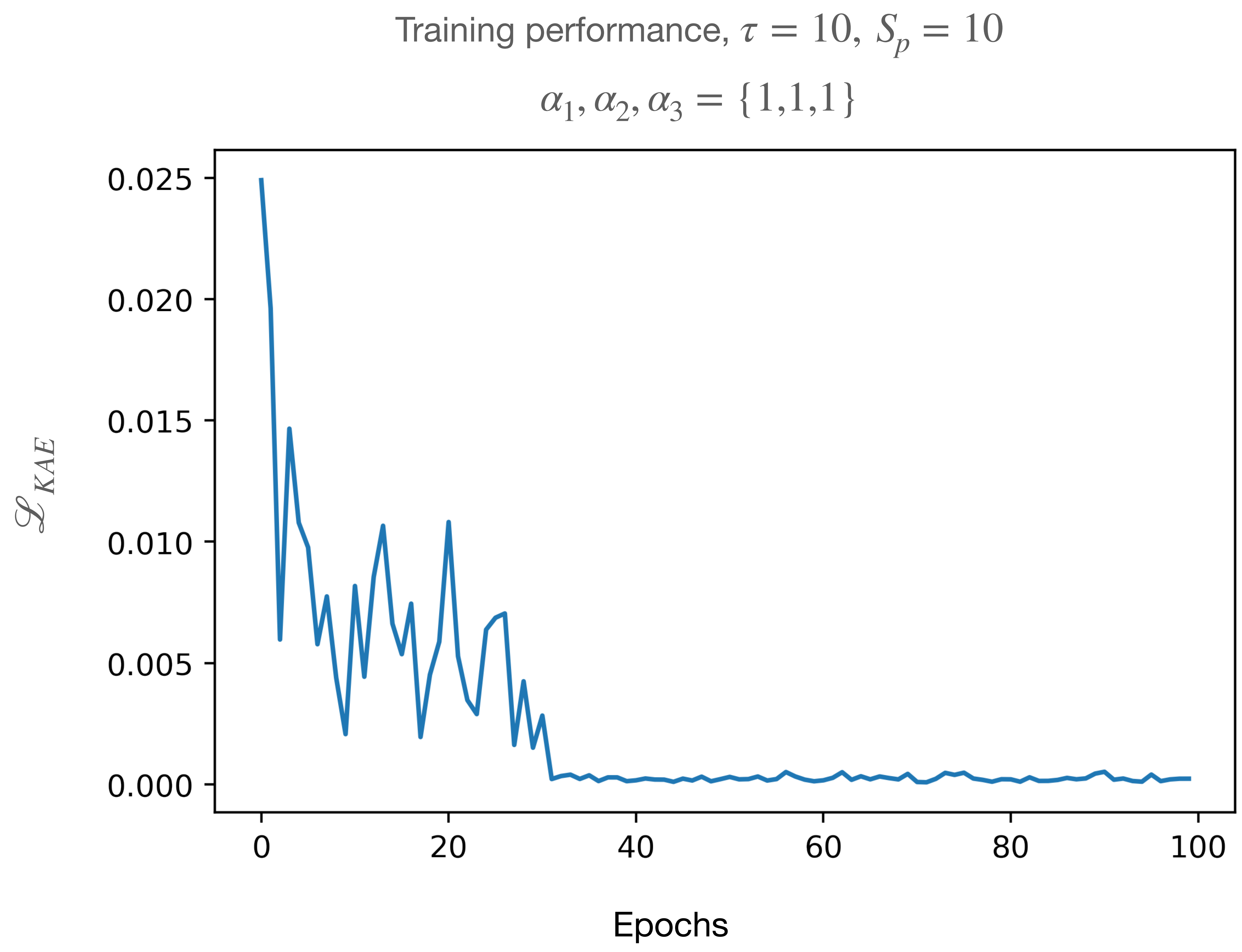


Figure 6. Training performance

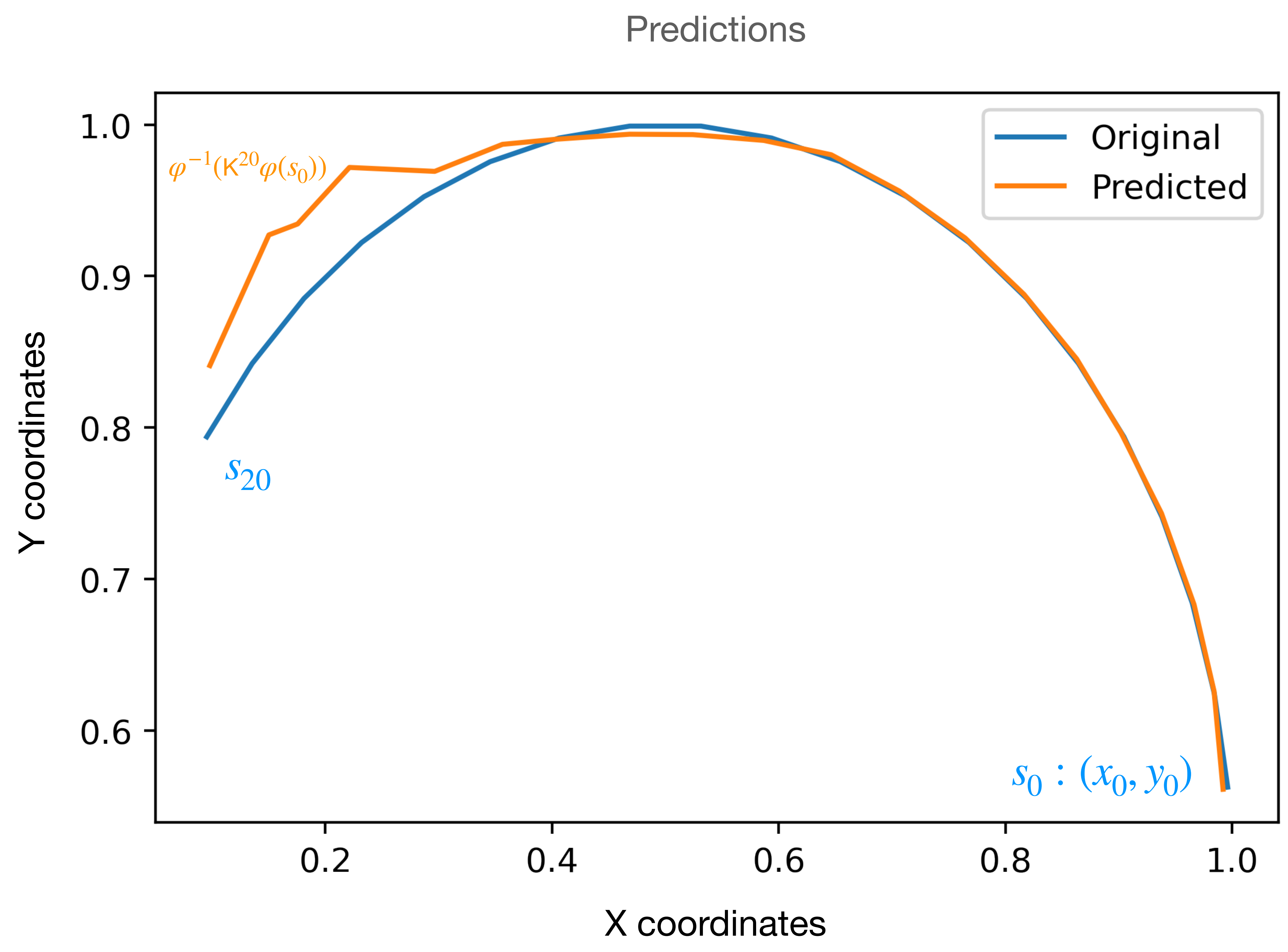


Figure 7. Ground truth vs Predictions over 20 time steps