Dynex

Validation Report

September 2025

Validation Report	
Project	Proof of Concept CIMATEC-Dynex - Air Pollution Forecasting
Document	[Título] Technical Report on the Proof of Concept on the implementation of an air pollution prediction model using the Dynex SDK
Company	Dynex
Written by	André Saimon Santos Sousa e Otto Menegasso Pires
Revised by	
Approved by	João Marcelo Silva Souza

Revisões	
Data	***Descrição da Revisão***

Table of Contents

1. Introduction

The accurate prediction of air pollutant concentrations is a critical and complex challenge within environmental science and public health. As urban populations grow and industrial activities expand, understanding and forecasting air quality has become essential for informed public policy, urban planning, and regulatory compliance with global standards set by organizations such as the World Health Organization (WHO). Traditional methods for predicting air quality often struggle to effectively process the vast, high-dimensional, and non-linear datasets involved. This has created a demand for more sophisticated and computationally efficient solutions to manage a problem with significant broader impacts on respiratory health, ecosystem stability, and climate change mitigation efforts.

Existing classical machine learning models, while powerful, face several key limitations when applied to the dynamic problem of air pollutant prediction. These approaches are often constrained by the sheer scale and complexity of the data, requiring significant computational resources that result in prolonged simulation times and high operational costs. Furthermore, the inherent non-linear dependencies and subtle correlations within environmental data can be challenging for classical algorithms to model accurately and in a timely manner. These technical and operational bottlenecks led SENAI CIMATEC to seek an alternative computational paradigm, which led us to explore the potential of Quantum Machine Learning (QML) and the Dynex platform as a promising new approach to overcome these constraints.

The purpose of this validation study is to rigorously evaluate the performance and efficiency of the Dynex environment for a specific QML application: predicting air pollutant concentrations. Our primary objective was to test and validate four distinct QML models—varying in complexity, qubit count, and the use of a Wavelet transform preprocessing—to assess their predictive capacity. The scope of this evaluation includes a detailed analysis of each model's performance using established metrics like Mean Squared Error (MSE) and Pearson's R correlation. We also sought to investigate the impact of "lookback" on prediction accuracy, determine if the Wavelet transform improves results, and assess whether increased qubit utilization leads to more accurate predictions. In addition, this report will provide critical feedback to Dynex by documenting the usage and development proccess of the Quantum Bridge Framework.

2. Methods

2.1. Database construction and preparations

The data was collected from the iSCAPE (Improving the Smart Control of Air Pollution in Europe) project, accessed from their public platform. Specifically, we include Living Lab Station data from Sutherland Memorial Park, Guildford (DS_TS_099), collected from June to October 2019, and from Stoke Park, Guildford (DS_TS_103), collected from February to September 2019. Both sites are in Guildford, UK, a high-density area with significant vehicle-induced air pollution. The datasets, collected hourly under open-road conditions, encompass key air pollutants such as PM2.5, Carbon Monoxide (CO), Nitrogen Dioxide (NO2), and Ozone (O3), alongside atmospheric variables like air temperature, humidity, and pressure.

To ensure consistency with previous research and to leverage an established pre-processing technique for temporal data, a five-level wavelet transform was applied to each feature, following the approach of GALVÃO et al. (2022). For each feature, we chose the most suitable wavelet family through the process proposed by ZUCATELLI et al. (2021), which minimizes the Root Mean Square Error (RMSE) between the original signal and the reconstructed approximation signal. As a result, five reconstructed detail and approximation signals were obtained for each feature.

Bibliography:

GALVãO, S. L. J. et al. Particulate matter forecasting using different deep neural network topologies and wavelets for feature augmentation. Atmosphere, v. 13, n. 9, 2022. ISSN 2073-4433. Available at: <https://www.mdpi.com/2073-4433/13/9/1451>.

ZUCATELLI, P. J.; NASCIMENTO, E. G. S.; SANTOS, A. A. B.; MOREIRA, D. M. Nowcasting prediction of wind speed using computational intelligence and wavelet in brazil. International Journal for Computational Methods in Engineering Science and Mechanics, Taylor & Francis, v. 21, n. 6, p. 343–369, 2020. Available at: <https://doi.org/10.1080/15502287.2020.1841335>.

2.2. Quantum Machine Learning Approaches

A comprehensive visual representation of our proposed QNN model's pipeline is provided in Figure 2.1, detailing all stages : Data Preparation, Ansatz, and Model Output.

The Data Preparation stage involves encoding classical data into a quantum state. This is done in two steps: first, a Superposition Layer applies Hadamard gates to all N qubits, one for each feature, to create a superposition state. Second, an Angle Encoding Layer applies a single-qubit Y-rotation gate, $R_y(\theta_j)$, to each qubit, where the rotation angle θ is directly determined by the normalized input feature.

Following data preparation, the Ansatz of our model consists of L repeated layers. Each layer contains a Variational Operator, U_{var}, which applies a sequence of single-qubit $R_z(\vartheta)R_y(\varphi)R_z(\omega)$ gates with trainable angles, and an Entanglement Operator, U_{ent}, which uses a circular entanglement scheme with CNOT gates applied between adjacent qubits. The number of layers L can be adjusted to influence the model's learning capacity.

Finally, the Model Output stage generates predictions. This involves performing a computational basis measurement on all N qubits to obtain the expectation value for each qubit. These expectation values are then fed into a classical post-processing layer consisting of fully connected neurons with a ReLU activation function. This classical layer performs a linear transformation on the quantum output to produce the final predictions for a specific forecast horizon, which is equal to the number of nodes in the layer.

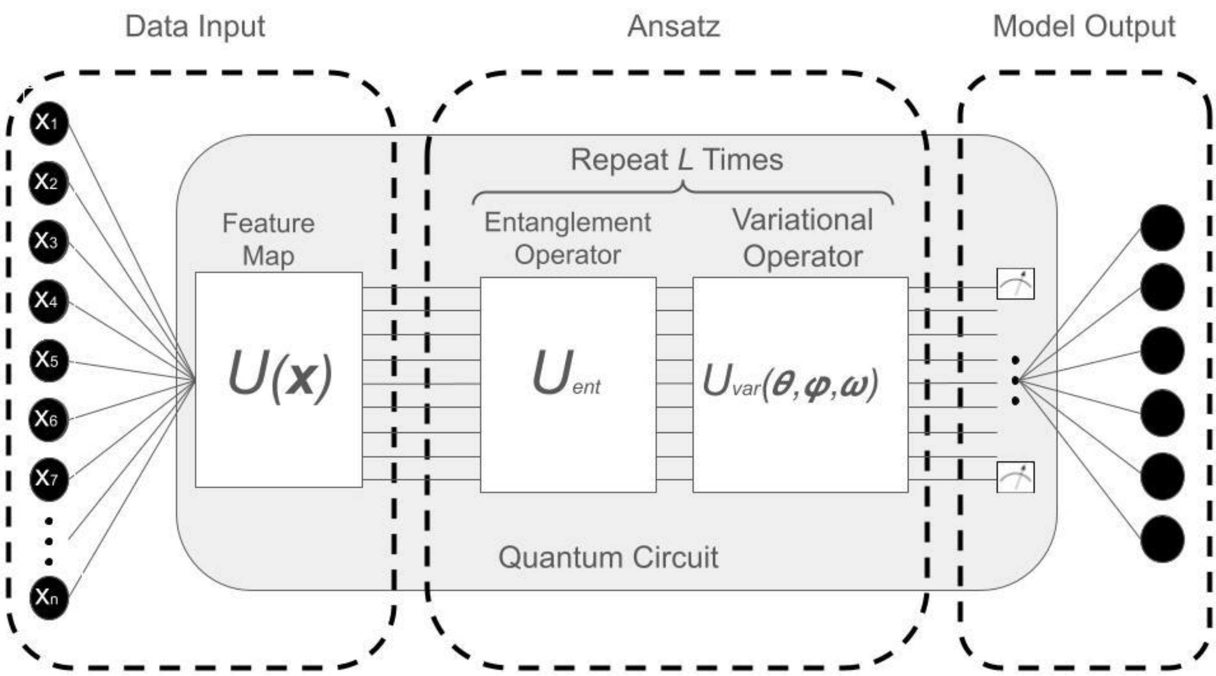

Figure 2.1: Visual representation of the proposed QNN model

2.3. Dynex Quantum Bridge Framework

Dynex has recently introduced a new feature to its SDK, enabling support for QNNs. This update is a significant step toward making quantum machine learning more accessible and practical for both researchers and industry professionals.

The new functionality is built around integrations with the libraries PennyLane and TensorFlow. A core component of this update is the TensorFlow/Keras Integration, which allows users to integrate quantum circuits directly into deep learning models using the well-known TensorFlow and Keras APIs. Furthermore, the SDK includes an Adaptive Quantum-Classical Architecture, an intelligent system for dynamically balancing computational resources between quantum and classical processing. The update also supports Advanced Embedding Techniques, including Angle, Amplitude, and Rotation-based methods for encoding classical data into a quantum state. Finally, the new functionality provides Uncertainty Quantification, offering reliable confidence metrics for quantum predictions, which is vital for building trustworthy and robust models.

The underlying engineering of this implementation focuses on practicality and scalability. Key highlights of the new system include dynamically adaptive resource allocation based on training progress, such as the number of epochs and batch sizes. The SDK is also designed for architecture optimization, adapting the model to the specific complexity of the problem being solved. Additionally, the system can recognize patterns in quantum solutions to enhance future results. Advanced gradient estimation techniques are also implemented, which are crucial for training QNNs effectively.

2.4. Evaluation Metrics

The evaluation of the proposed QML model involves a comprehensive analysis of relevant performance metrics. The following metrics are proposed:

$$\text{NMSE} = \frac{\frac{1}{n}\sum^{n} i = 1(x_i - x'i)^2}{\sigma^2(x)}$$

$$\rho(r) = \frac{\sum^{i} \quad = 1^{n}(x_i - \overline{x_i})\left(x_i' - \overline{x'i}\right)}{\sqrt{\sigma^2(x)\sigma^2(x')}}$$

where n is the number of samples, NMSE is the normalized mean squared error and ρ(r) is the Pearson correlation coefficient, x and x' are the observed and predicted datasets, respectively, x_i and x'_i are the i-th observed and predicted samples, respectively, and σ^2 is the variance of a set of input variables. These metrics provide quantitative insights into the accuracy and correlation of the predictions.

3. Results

3.1. Machine Configuration

Initial model validation and preliminary analyses were primarily performed using a quantum simulator. Specifically, we leveraged SENAI-CIMATEC's Kuatomu, a high-performance computing cluster equipped with 192 processing cores, 3 TB of RAM, and 4 NVidia V100 GPU accelerator cards (32 GB each).

3.2. Quality Metrics

Table 3.1 presents a comparative overview of the best results obtained by the proposed QNN model with a baseline MLP and LSTM.

Lookback 1		
Ansatz	NMSE	R (Pearson)
QNN depth 1	0.784	0.551
QNN depth 2	**0.696**	0.589
QNN depth 3	0.973	0.570
QNN depth 4	1.082	**0.599**
MLP	1.236	0.303
LSTM	1.299	0.288

Lookback 2		
Ansatz	NMSE	R (Pearson)
QNN depth 1	1.005	0.485
QNN depth 2	1.188	0.513
QNN depth 3	1.068	0.539
QNN depth 4	**0.907**	**0.550**
MLP	1.641	0.294
LSTM	1.322	0.310

Lookback 1		
Ansatz	NMSE	R (Pearson)
Wavelets depth 1	0.506	0.748
Wavelets depth 2	0.636	0.690
Wavelets depth 3	**0.496**	**0.760**
Wavelets depth 4	0.545	0.708
MLP	**0.496**	0.718
LSTM	0.597	0.662

Lookback 2		
Ansatz	NMSE	R (Pearson)
Wavelets depth 1	0.644	0.688
Wavelets depth 2	0.546	0.708
Wavelets depth 3	0.754	0.581
Wavelets depth 4	**0.487**	**0.729**
MLP	0.589	0.695
LSTM	0.563	0.703

Figure 3.1: Comparison of the statistical analyses (NMSE and Pearson's R) for the best results of the particulate matter forecasting by the QNN model and the classical benchmark models. Best values for each metric are highlighted in bold.

4. Discussion

Our validation study demonstrated that the Dynex platform, utilizing our QNN model, is a highly effective alternative for predicting air pollutant concentrations. The most significant finding is the superior performance of the QNN models over classical machine learning baselines (MLP and LSTM) in nearly all tested configurations. Notably, the QNN models with a *lookback = 1* consistently outperformed the baselines in both normalized mean squared error (NMSE) and Pearson's R correlation, indicating a more accurate and robust predictive capacity. The use of a five-level wavelet transform as a pre-processing step further enhanced the performance of the QNN models, with the *Wavelets depth 3* model achieving the lowest NMSE (0.496) and highest correlation (0.760) among all QNN variants. This also matched the best-performing MLP model with wavelets, highlighting the competitive, and often superior, predictive power of the QNN approach. While increasing the lookback to two did not uniformly improve results, the best-performing QNN model in this category still demonstrated a significant advantage over its classical counterparts. These results indicate that Dynex's QML solution can effectively model the complex, non-linear dependencies within environmental data.

The validation study revealed a significant advantage of the Dynex solution: its ability to execute complex quantum circuits with a high number of algorithmic qubits, far exceeding the capabilities of current noisy intermediate-scale quantum (NISQ) hardware. Simulating a 172-qubit circuit for the air pollutant prediction model would be unfeasible on physical quantum computers due to their limited size. Dynex's QML framework bypasses these physical limitations, allowing for the practical application of large-scale quantum algorithms today. This is a critical advantage for complex, real-world problems where classical baselines may struggle to find optimal solutions. However, a key limitation was that the models with the highest qubit counts did not consistently yield the best performance, indicating a non-linear relationship between complexity and accuracy. Furthermore, in one instance, a classical MLP model with wavelet pre-processing achieved a NMSE equal to the best-performing QNN, demonstrating that while the Dynex solution offers a powerful new tool, it is not a silver bullet and its advantages are most pronounced where classical methods fail to perform adequately.

The validation results have significant practical implications for the deployment of advanced air quality forecasting systems. The demonstrated superior performance of the QNN models on the Dynex platform suggests a new paradigm for environmental monitoring and management. This technology could be integrated into urban planning tools and public health initiatives, providing more accurate, real-time forecasts to inform policy and warn the public. While the study did not directly measure computational speed, the successful execution of these complex models on the Dynex platform signals potential for faster simulation times compared to classical high-performance computing, which is a critical factor for operational costs and timely data delivery. Future work should focus on optimizing the QNN architecture and exploring the scalability of the platform for larger datasets, as well as evaluating the cost-effectiveness and ease of integration into existing environmental data pipelines. The findings provide a strong foundation for moving beyond the Proof of Concept and towards a full-scale, real-world deployment.

4.1. Learning Proccess

The initial proposal was that the POC commenced on **July 1st, 2025**, and concluded on **July 31st, 2025**. However, due to the experimental nature of the new Quantum Bridge Framework, CIMATEC needed an extension of the previous deadline because of the waiting time for the corrections of errors found by the developing team. The experiments ended on August 28th.

Early in the process, the CIMATEC development team encountered several technical roadblocks, including an error with the StronglyEntanglingLayers template and the absence of a QKerasLayer implementation in the Dynex SDK. To overcome these initial hurdles, the Dynex team provided access to a dedicated Jupyter Lab environment. This environment offered a solution to the local configuration issues, allowing the CIMATEC team to proceed with model development and training with minimal latency.

Once the official project commenced, new and more complex challenges emerged. The QNN model initially failed to learn, with its loss function stagnating at zero. This issue, along with subsequent inconsistent learning behavior, highlighted the intricacies of the interface between the quantum and classical components of the hybrid model. Observations showed that the Val_loss was unusually volatile, displaying signs of both overfitting and underfitting in different runs. These issues were a direct result of how the quantum bridge handled circuit parameters and the initial gradient calculation methods. In response, the Dynex team implemented critical updates to the SDK, adding a new Quantum Solution Validator and improving the gradient calculation function.

With these key updates, the model began to learn correctly. The final experiments confirmed the QNN's superior performance, demonstrating better NMSE and Pearson's R values than the classical baseline models.

5. Conclusions

Based on our rigorous validation study, the Dynex platform presents a powerful and viable solution for advancing air pollutant prediction through quantum computing. The results not only confirm the platform's technical capabilities but also provide a clear pathway for its future application and development.

5.1. Key Findings

- **Superior Predictive Performance:** The QNN models running on the Dynex platform consistently outperformed classical MLP and LSTM baselines. In some configurations, the QNN models achieved a lower NMSE and a higher Pearson's R correlation, indicating a more accurate and reliable prediction of air pollutant concentrations.
- **Impact of Wavelet Pre-processing:** Applying a five-level wavelet transform significantly improved the predictive accuracy of the QNN models. The best-performing QNN with this pre-processing step achieved a remarkable NMSE of 0.496, matching the classical model and demonstrating the value of this technique for complex, non-linear time-series data.
- **Scalability of Qubit Simulation:** The Dynex framework successfully simulated quantum circuits with up to 172 qubits. This is a critical finding, as it proves the platform can handle a scale currently unfeasible for physical NISQ quantum computers.
- **Non-linear Relationship with Complexity:** We observed that increasing the model complexity (e.g., higher qubit count or lookback) did not always lead to a proportional improvement in performance. The optimal results were often found with a specific combination of parameters, suggesting that a "bigger is better" approach does not apply here.

5.2. Future Directions

Based on these findings, we offer the following recommendations for Dynex and future implementations of this technology:

- **Investigation of Simulation Time and Cost:** While the study validated performance, future work should focus on a detailed analysis of the time and cost associated with running these models on the Dynex platform. Quantifying the operational efficiency will be vital for justifying large-scale, commercial deployment to public health and urban planning organizations.
- **Expansion to New Datasets:** We recommend testing the Dynex solution on a broader range of environmental datasets with different complexities and time granularities to further validate its versatility and robustness across various geographical and climatic conditions.
- **Portability with PyTorch:** The successful execution of this POC relied upon the existing integration of the TensorFlow SDK with PennyLane. However, we recommend the

development of a native PyTorch plugin. This recommendation is predicated on the planned deprecation of PennyLane's support for TensorFlow. Implementing a PyTorch plugin would proactively address this potential compatibility issue, mitigating future disruption to ongoing research and development efforts.