POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Mitigating the influence of environmental variability by using machine learning for Structural Health Monitoring applications

TESI MAGISTRALE IN MECHANICAL ENGINEERING – INGEGNERIA MECCANICA

AUTHOR: ANIKET PANDE

ADVISOR: Prof. SIMONE CINQUEMANI

ACADEMIC YEAR: 2022-2023

1. Introduction

Structural Health Monitoring (SHM) uses sensing and data analytics to evaluate structural health and detect damage. However, environmental and operational variability poses challenges. Temperature, wind, humidity, and solar radiation cause fluctuating dynamic response. Traffic congestion and special events also alter stresses. This variability can mask damage or cause false alarms, limiting SHM effectiveness.[1]

This thesis aims to mitigate variability's influence on damage detection. The objectives are to:

1) Review methods like data normalization that reduce variability effects.
2) Implement suitable techniques on the Livenza Railway Bridge using monitoring data.
3) Develop general recommendations for enhancing SHM robustness.

Although applied to a specific bridge, the research will focus on broadly useful techniques. One possible solution by developing hybrid machine learning techniques to isolate indicators of structural damage from environmental impacts in SHM data is proposed, and demonstrated on a case study of a Livenza railway bridge monitoring system. A local outlier removing function is used to mitigate the operational variability and Machine learning models are used for removing the environmental variability.

2. Thesis Breakdown

- The primary goal of the thesis is to develop a model that can remove the operational and environmental variability.
- The literature review and state of the art was conducted to find the current area of research.
- Data-preproccessing
- Hybrid Model building and training on livenza bridge data.
- Introduction of virtual damage to the dataset to access the performance of hybrid model.

3. Model Development Phase

This stage primarily involves a diverse range of tasks, spanning from data preprocessing to the construction of machine learning models.

3.1. Variability due to operations and environment

Time series data from bridge response sensor E35 and environmental sensor T39 over a 1-week period, shown in Figure 1, demonstrate that fluctuations in E35 closely match variations in temperature T39. This is statistically validated by the high correlation coefficient between E35 and T39 evident in the correlation matrix, indicating that environmental factors are a primary source of response variability.

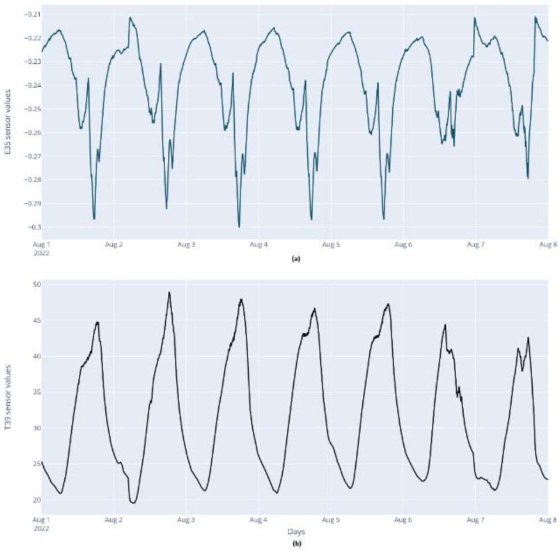

Figure 1: Time series of (a) Bridge response sensor E35; (b) Environmental sensor T39

Additionally, as shown in Figure 4, some isolated spikes visible in the I15 time series are likely caused by operational variability from traffic loads. These outliers are addressed through the local outlier removal technique.

3.2. Data cleaning

Data cleaning is the process of finding, correcting, and removing errors or inconsistencies from a data set. In the thesis, Outlier Removal, Exclusion, Scaling was widely used to clean up the data.

3.2.1. Exclusion

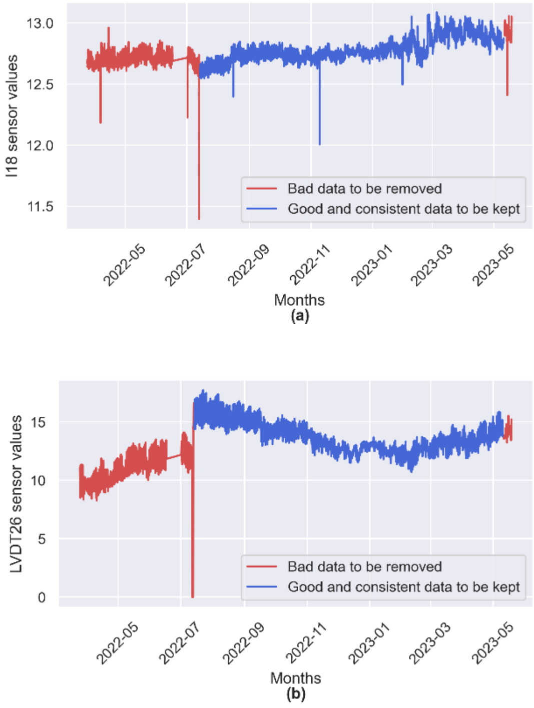

Figure 2 Removing data by exclusion

Imputing too many values can result in a number of potential issues like bias in data or overfitting of models. So, it is better to remove this large inconsistent data before applying any machine learning method. This method of removing the data is referred as Exclusion.

Data selected for the study was 14th July 2022 to 10th May 2023.

Only good performing sensors were selected, and faulty sensors were dropped.

Also, the sensors which are not useful in calculation of residuals can also be dropped, i.e., the sensors which have constant values over the period.

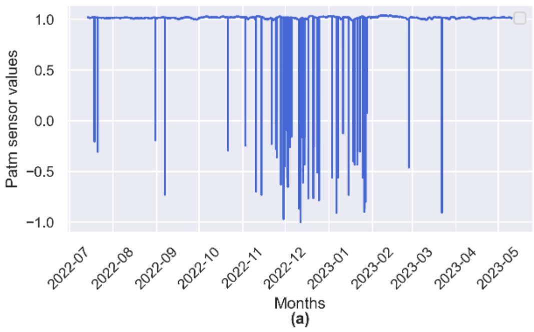

Figure 3 Sensor with almost constant value over period

3.2.2. Outliers and missing values

In the present case, the outliers are also said to be due to the operation of the bridge itself. This can be due to passing of a train which in turn changes the dynamic response of the bridge.
A local outlier removal algorithm is developed to remove the operational variability.

Algorithm 1: Imputing local outliers/ operational data function
Input: Unfiltered dataframe
Output: Filtered dataframe
Defining Function: Replace Outliers (Input arguments – time series, number of neighbors, threshold)
for (*i* from number of neighbors to (size of time series − number of neighbors − 1) **do**
Slicing timeseries
if (absolute value of (current value−average of sliced timeseries)>threshold) **then**
Replace the value with the mean value of the window
end if
end for
return modified timeseries

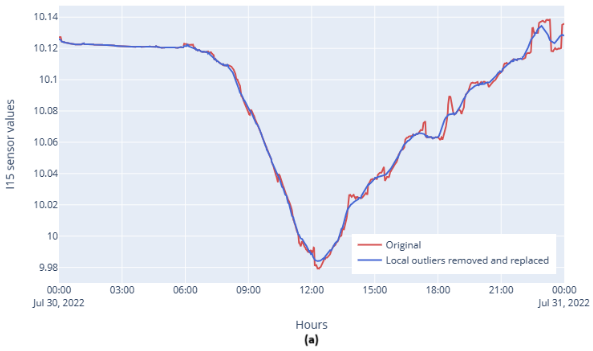

Figure 4 Local outlier removal

Figure 4 shows successful removal of outliers.

3.2.3. Handling missing values

For imputing missing data, the second-degree polynomial interpolation approach has a number of benefits. It can identify intricately curved patterns in the data and offer reasonably precise estimates for values that are absent.

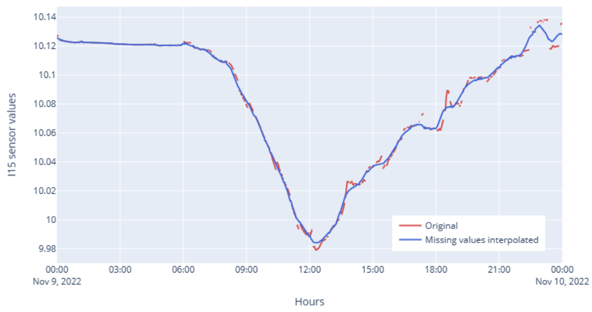

Figure 5: Interpolation of missing data

3.2.4. Feature Transformation

The sensor data, initially in electrical voltages, underwent an electrical-to-physical transformation to convert them into their respective physical units. Additionally, to handle a year-long dataset, the observation frequency was adjusted from 1 minute to 1 hour by averaging the data over this time interval.

4. Hybrid model training phase

The processed dataset considering only bridge response sensors are scaled and then a PCA transform is applied to it, which transforms original data from original subspace to Principal subspace.[2]
The first principal component should explain the maximum variation due to environmental effects in the dataset, which can be proved by plotting a correlation matrix of all principal components and environmental data.
Thus, this principal component was extracted and a dynamic regression model having a 24-hour time lag was trained on this PCA transformed data.[3]
The dynamic regression model was chosen carefully after considering other machine learning model also. All the models were trained using a python library called DARTS, library for time series forecasting that offers a variety of models,

from classics such as ARIMA to state-of-the-art deep neural networks. [4]
Figure 6 shows comparison of various models considered during training phase.

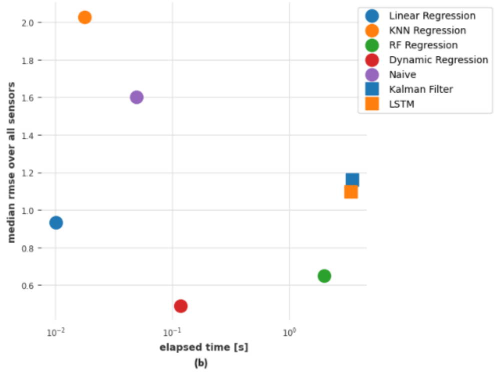

Figure 6: Comparison of Model Performance and Efficiency.

Above figure shows performance of each model.

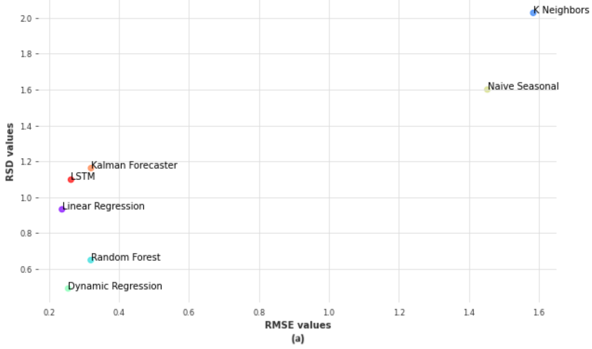

Figure 7: RMSE vs median RSD

Figure 7 shows efficacy of all models in terms of successfully removing the environmental variability.
After selecting the dynamic regression model, residuals are calculated by subtracting the predicted values from the actual values of the chosen principal component (PC). These residuals represent the bridge response data that cannot be solely attributed to environmental changes, mainly reflecting fluctuations related to structural degradation. To bring these residuals back to the original subspace, an inverse PCA transform is applied, reconstructing them into the original feature space. These refined residuals offer valuable insights for damage detection and analysis.
Statistical analysis techniques and visualization methods can be applied to these refined residuals. They enable the identification and interpretation of damage indicators within the bridge monitoring data. By isolating the variations associated with structural damage and removing the influence of temperature, this PCA-based method enhances the accuracy and reliability of damage detection in bridge monitoring systems.

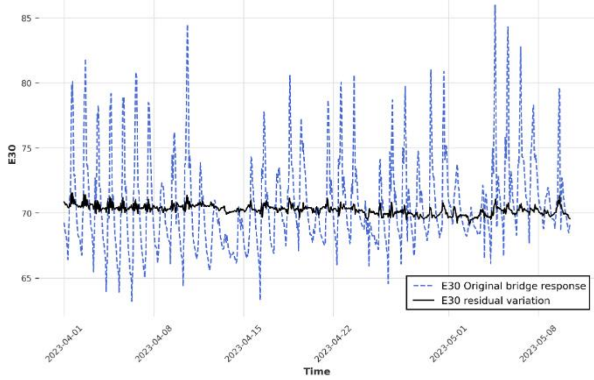

Figure 8: Variation of residuals and original bridge response for sensor LVDT25

The trained hybrid model when used on the sensor data from another railway bridge "Piave bridge" was also able to remove the environmental effects, this can be shown in figure below.

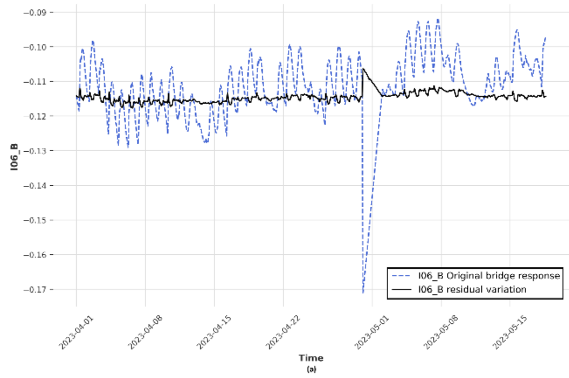

Figure 9: Residuals obtains for "Piave bridge" after applying trained hybrid model.

This proves the generality of the trained hybrid model.

5. Application on Simulated damage scenario.

The structural health monitoring system's design and tuning for damage detection involve simulating the bridge's response to various forms of structural damage under operating conditions. This includes investigating notional damage scenarios with different magnitudes and extents of damage to understand the impact on sensor-measured indicators like displacements, rotations, natural frequencies, and modal eigenvectors. To assess the effectiveness of Principal Component Analysis (PCA) and regression modeling in removing environmental effects, virtual damage is induced in mechanical sensors. This virtual damage manipulation includes altering strain values, sensor readings, or data relationships to mimic structural damage. One possible scenario is shown below.

Livenza Bridge Response Sensors			Intact Structure	SCENARIO #14: Bottom, Top, Diagonal damage (δ=10%)	
Measurements	Sensor ID	Model ID	Value (δ=0)	Value	Δ
Rotation[deg]	I15	1	-4.67×10^{-2}	-4.81×10^{-2}	-1.41×10^{-3}
	I18	56	4.67×10^{-2}	4.81×10^{-2}	-1.41×10^{-3}
Longitudinal Displacement [mm]	LVDT25	28	4.64	4.80	0.16
	LVDT26	56	4.64	4.80	0.16
Strain (Delta in microepsilon)	E29	1	-609.38	-581.94	9.79
	E31	1	-609.38	-581.94	9.79
	E30	41	-609.28	-581.84	9.79
	E32	41	-609.28	-581.84	9.79
	E33	33	974.16	918.67	-10.12
	E35	33	974.16	918.67	-10.12
	E34	73	974.15	918.65	-10.12
	E36	73	974.15	918.65	-10.12

The above hybrid trained model was applied to this new dataset.

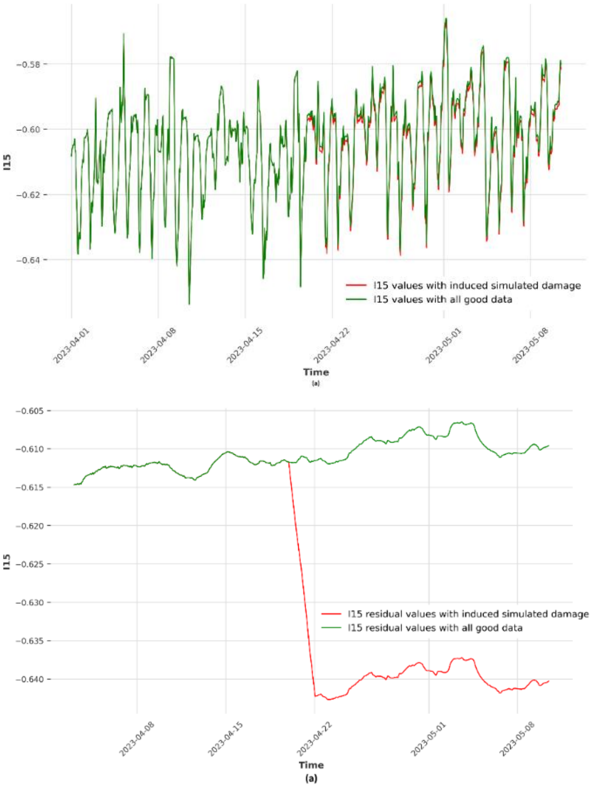

Figure 10: Variability before and after application of hybrid model

It can be seen from Figure 10 that after processing the data through the PCA+Regression model described in previous section, the damage induced can be clearly identified.

The ability of the refined residuals to primarily capture the variations due to induced damage while effectively removing temperature variations would demonstrate the success of this work.

6. Conclusion

- The methodology of using machine learning models to reduce environmental and operational variability in structural health monitoring (SHM) was successfully developed and validated on the Livenza Railway Bridge case study.
- Visualization and statistical techniques confirmed the presence of temperature and train loading effects on the bridge response data, highlighting the need for data normalization.
- Regression models were effective in predicting and removing temperature variations. Dynamic regression with a 24-hour lag performed the best.
- Combining regression with PCA further filtered out unmeasured variability not captured by regression. This integrated approach reliably isolated damage indicators.
- Simulated damage scenarios validated the methodology's ability to extract damage patterns from raw data affected by environmental impacts.
- While demonstrated on a bridge, the machine learning techniques have potential for broader SHM applications susceptible to environmental or operational influences.
- The computational framework enables distinguishing between external variability and underlying structural degradation for more accurate SHM analysis.

7. Future Prospects

- Optimize LSTM model hyperparameters like layers, nodes, regularization to further improve performance.
- Apply advanced system identification techniques like SINDy to derive enhanced physics-based model tailored to bridge dynamics.
- Develop hybrid models integrating machine learning with physics-based principles to better isolate environmental effects.
- Explore unsupervised learning methods like autoencoders for detecting damage from normalized data.

1. Bibliography

[1] H. Zhang, J. Guo, X. Xie, R. Bie, and Y. Sun, "Environmental effect removal based structural health monitoring in the internet of things," *Proceedings - 7th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS 2013*, pp. 512–517, 2013, doi: 10.1109/IMIS.2013.91.

[2] A. M. Yan, G. Kerschen, P. De Boe, and J. C. Golinval, "Structural damage diagnosis under varying environmental conditions - Part I: A linear analysis," *Mech Syst Signal Process*, vol. 19, no. 4, pp. 847–864, Jul. 2005, doi: 10.1016/j.ymssp.2004.12.002.

[3] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 3rd ed. Melbourne, Australia, 2021. Accessed: Sep. 03, 2023. [Online]. Available: OTexts.com/fpp3

[4] J. Herzen *et al.*, "Darts: User-Friendly Modern Machine Learning for Time Series," *Journal of Machine Learning Research*, vol. 23, no. 124, pp. 1–6, 2022, [Online]. Available: http://jmlr.org/papers/v23/21-1177.html