2022

Predicting thermal response of the Bio-based bridge in Ritsumasyl using distributed temperature and strain measurements

Heinen, D.A. (David, Student B-CE)

Witteveen + Bos

4-7-2022

Preface

In front of you lies the research "Predicting thermal response of the Bio-based bridge in Ritsumasyl, using temperature distributions and a regression model". It was written to meet the bachelor requirements of the study Bachelor of Civil Engineering at the University of Twente. I was involved in the research from May 2022 to July 2022. The project was undertaken at the company Witteveen + Bos where I did my internship. The research questions were formulated together with my supervisors Roland Kromanis and Chris Jolink. It has been an interesting study that has taught me many new things. I would like to thank my supervisors Roland and Chris for the excellent guidance and support during this process. I would also like to thank the people in my working group at Witteveen + Bos who were always willing to help if needed. It was a pleasure to work together and to gain experience at the Witteveen + Bos office in Deventer.

I hope you enjoy your reading.

David Heinen

Enschede, July 4, 2022

Inhoud

List of figures

List of tables

Summary

This research investigates the application of a regression model, which is built using distributed temperature and response (strain) measurements to predict the thermal response of the bio-based bridge in Ritsumasyl. The aim is to have an accurate regression model that can predict the response of the bridge over diurnal and seasonal temperature changes, which cause large deformations that exceed live loads. The regression-based thermal response prediction (RBTRP) methodology was used to determine the relationship between strain and temperature distribution during a reference period. A multiple linear regression algorithm (MLR) has been applied. By applying data pre-processing and an iterative process to find the smallest prediction error by varying the thermal inertia and number of input measurements for the regression analysis, the thermal response was successfully predicted with a maximal prediction error of 4.5% for the monitoring period. This involved a shift correcting measured data for all strain gauges. This shift was necessary since strain data in the monitoring period deviated significantly from strain data in the reference period at the same temperatures. The first measured value of the monitoring period is set equal to the first predicted strain value of the monitoring period. The difference between these two values is applied to all other measured values. The reason for the deviation is unknown and should be further investigated in a follow-up study. The research also shows that the thermal response can be well predicted using temperature distributions, but the prediction error does differ per strain gauge. Whether this depends on factors like sensor location, applied loads, wind or humidity cannot be concluded from this study. This also requires further research.

1 Introduction

Nowadays, many changes are taking place in the current environment. Climate change is an issue, but CO_2 and nitrogen emissions are also major problems. To become more sustainable, it is important that new and current bridges meet the CO_2 emission standards and the nitrogen limits during construction. The use of new sustainable materials can be beneficial in solving these problems. In addition, structural health monitoring (SHM) is important in extending the life span of a bridge. This can give an adequate indication of the 'fitness for purpose' of a bridge under gradual changes of their state. With this, a lot can be learned from, for example load, but also response mechanisms (Brownjohn, 2006). Bridge thermal response is the dominant response in long-term. Previous studies have shown that diurnal and seasonal temperature variations have a major influence on the structural response. This effect is perhaps even greater than the response for vehicular traffic (Hua et al., 2007). This report looks at the effect of diurnal and seasonal temperature variations on a special bio-composite bridge. It is situated in the small town Ritsumasyl, located in Friesland. The bicycle bridge is made of bio-composite consisting of balsa wood, flax, resin and harder. 80% of the bridge consists of natural materials. Because this material is so new, not much is known about the behaviour of the bridge under certain circumstances. According to Witteveen+Bos, the bridge has a lifespan of 100 years. However, they want to monitor whether the bridge has already shown signs of weakness. Temperature and strain are important indicators for weaknesses in the bridge. Especially because the application is so new, it is interesting to gain more insight into this (Sweco & Witteveen+Bos, 2020).

The structure of the report is the following. First, the context of the problem is made clear. The stakeholders involved, the study area and the temperature strain problem are discussed here. Next, the problem statement and theoretical framework are elaborated, this then leads to the research questions. This is continued with the methodology that will be used to answer the research questions. This is followed by results and a discussion. Finally, a conclusion is given on the research questions and recommendations are given.

1.1 Problem context

1.1.1 Involved parties

The province of Friesland is the initiator of the replacement of the old bridge and the installation of the new bridge near Ritsumasyl. The bridge was built by a special team consisting of contractor Strukton-Spie, composite producer Delft Infra Composites and a combination of Sweco/Witteveen+Bos went through the challenging development process. This also involved close collaboration with Green PAC. These are all parties involved in the construction of the bridge. In addition, of course, there are the users of the bridge. This includes cyclists and walkers, but also water traffic that passes under or along the bridge. It is important to them that the bridge is always passable and does not require frequent maintenance. To ensure that this is not the case, several sensors have been placed on the bridge that provide data that is managed by Sweco, with support from Witteveen+Bos. Because they manage this data, they can also monitor how the bridge is responding and how the bridge is doing. The most important party for predicting the temperature-strain response of the bridge is therefore Witteveen+Bos.

1.1.2 The bio-based bridge

The bridge is located near Ritsumasyl in Friesland. The bridge has been there approximately 2.5 years now. The location is indicated with an arrow in Figure 1. This figure shows provincially and locally where the bridge is located. It can also be seen that the bridge is oriented from North to South.

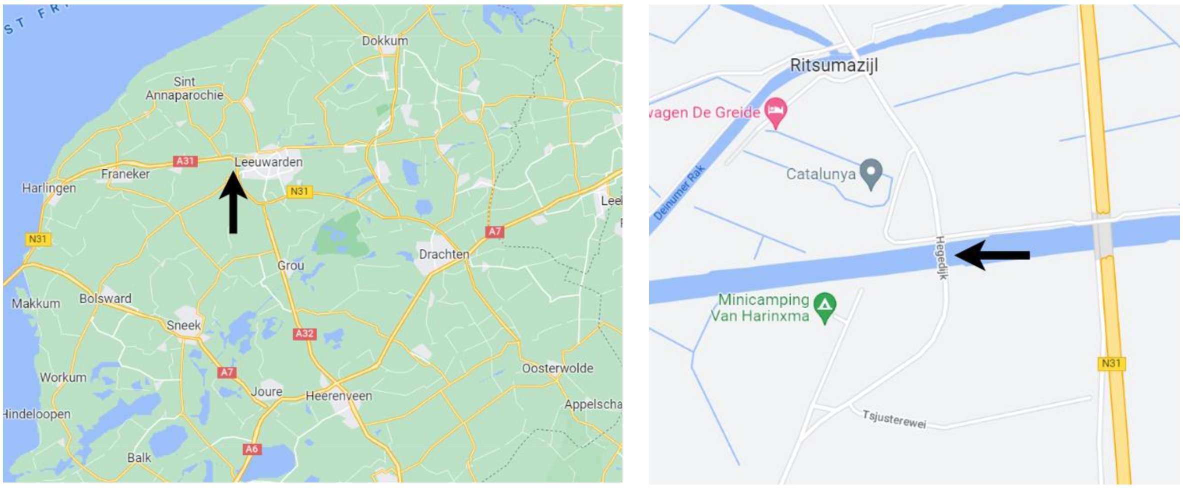

Figure 1: Bio-based bridge location provincially (left) and location locally (right) (Google, (n.d.))

The bridge consists of a fixed part of 34 meters and a rotating part of 32 meters. It opens a few times during the day. In Figure 2, the bridge can be seen when its closed. The left part in the figure is the part that can be turned open. Figure 3 shows the situation when the left part of the bridge is open.


Figure 2: Bio-based bridge (closed)


Figure 3: Bio-based bridge (open)

As mentioned before, it is a bridge made of special material. It consists of 80% natural materials such as balsa wood, flask, harder and resin. The bridge is equipped with sensors that provide a lot of data and can be used for monitoring. The goal of this is to gain insight into the life expectancy and to optimize the maintenance of the bridge. The bridge is equipped with 184 optic fiber sensors. These are strategically placed on the bridge along the bridge deck at the bottom and top. In total, 82 sensor points have been installed on the fixed part. These exist out of 76 strain sensors and 6 temperature sensors. 124 sensor points consisting of 112 strain sensors and 12 temperature sensors are placed on the moving part. With all these sensors, data has been collected from the bridge for more than 2 years.

1.2 Research question

The aim of the research is to generate an accurate regression model that is capable to accurately predict bridge thermal response. This involves looking at temperature and strain. This gives the following main research question and sub-questions.

Main question:

- *Can bridge thermal response be accurately predicted using the knowledge of bridge response and distributed temperature?*

Sub-questions:

- *How can temperature measurements and strain be used to accurately make a regression model to predict the thermal response of the bridge.*
- *Which steps need to be taken for appropriate data pre-processing?*
- *How can the number of input measurements be used to increase the prediction accuracy?*
- *How can thermal inertia be used to increase the prediction accuracy?*
- *Is multiple linear regression an adequate algorithm to predict thermal response?*

2 Theoretical framework

The bridge in Ritsumasyl is made of a special type of material called bio composite. Bio composite often consists of materials from local and renewable sources that offer significant sustainability. Different types of bio composites have already been successfully applied in domestic sectors such as the aerospace industry, circuit boards and automotive applications (Bharath & Basavarajappa, 2016). The properties of this material can influence aspects such as the lifespan of structures. But it can also affect weaknesses in the bridge and where they might arise.

Structural health monitoring (SHM) is very important to identify these aspects. This term is increasingly used to describe various systems that are widely used in civil infrastructure. The aim of SHM is to contribute to and inform about the 'fitness for purpose' of structures under gradual changes in their state in order to learn about load and response mechanisms (Brownjohn, 2006). This allows for forehanded action if structural damage is identified Structural damage is defined as changes to the material and/or geometric properties, including changes to the boundary conditions and system connectivity, which negatively affect the systems performance (Farrar & Worden, 2006).

There are fundamentally two approaches for damage identification. Model-driven methods establish a high-fidelity physical model of the structure often using finite element analysis. This model is then compared with the measured data of the real structure. Data-driven approaches also establish a model, but with using a statistical representation of the system (Worden & Manson, 2006). Statistical learning algorithms are often used for this, which aid in the construction of a data-driven model. The use of a data-driven model instead of a high-fidelity model comes from the idea that capturing all the physics involved becomes increasingly difficult as the complexity of a system increases (Sen & Nagarajaiah, 2018). Several forms of SHM have been used over the past half century, but it has been a few years since computer-based systems have been designed with the aim of assisting operators of outdated infrastructure with timely information about their safe and economic operation. Another problem why SHM has been used less before is that it is very difficult to extract meaningful information from large amounts of data (Kromanis & Kripakaran, 2016).

In research from Kromanis and Kripakaran (2014), a generic approach to evaluate thermal response of bridges from temperature measurements has been conducted as part of structural health monitoring. It elaborates on the challenge of accounting for the thermal response in measurements collected during quasi-static monitoring of bridges. Quasi static is defined as a process that proceeds at a low speed. Quasi-static signals are more commonly used for load identification and damage identification. The quasi-static strain response of the bridge is useful information for a condition assessment, which the SHM of a bridge usually depends on (Lu et al., 2019). These quasi-static reactions of the bridge are often mainly caused by slow temperature changes that follow diurnal and seasonal cycles. Most materials contract or expand with a change in temperature. This therefore also takes place in structures such as bridges. These deform continuously with changes in weather conditions. Previous studies have shown that deformations caused by seasonal temperature effects in large bridges can be up to 10 times greater than when caused by traffic (Catbas et al., 2008). However, it is very difficult to predict what the temperature distribution in a structure will be. It is often assumed by engineers that there are linear temperature gradients to evaluate thermal response. However, this is not appropriate enough for long-term monitoring. The largest response is mainly caused by temperature variations (Kromanis & Kripakaran, 2014). As an example, a plot of the Cleddau bridge in Wales is shown. The figure shows the daily variation in the bearing displacements. In Figure 4 the daily time series of displacement can be seen.

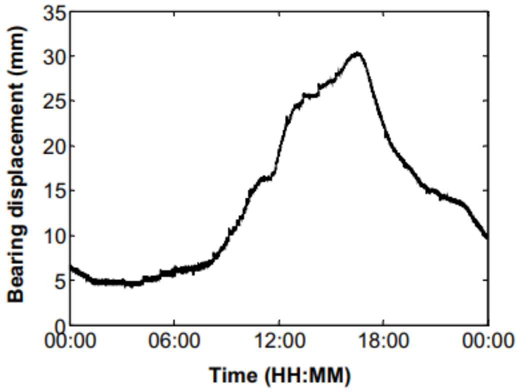

Figure 4: Displacement measurements of the Cleddau bridge collected at a bearing over 1 day (Kromanis & Kripakaran, 2014)

Figure 4 shows very clearly that the displacement is greatest in the middle of the day. It is also clearly visible that the displacement of the bridge decreases again at sunrise and sunset. This effect becomes even more apparent when looking at different sections of the bridge. The Cleddau bridge is oriented from north to south. This means that the right side of the bridge mainly gets sun in the morning and the left side has the sun on it in the afternoon. This can be clearly seen in Figure 5.

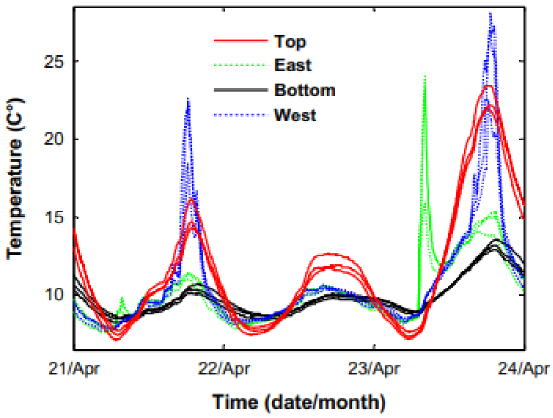

Figure 5: Temperature variations Cleddau bridge (Kromanis et al., 2015)

In Figure 5 large differences in temperature in the bridge can be seen. It can also be seen that bridge sections heat up differently at different times of the day. Due to this difference in temperature at different parts of the bridge, there is also a difference in strain over these different parts. An increase in temperature causes the material to expand. If one side gets warmer than the other, it can cause, for example, stretching of the east side of the bridge and compressing of the west side at a simultaneous moment. This can cause very large stresses in the bridge that must be considered in connection to SHM. It is therefore clear that temperature effects must be included in the measurement interpretation process. For this, a strategy has been developed that supports a bridge management paradigm. The paradigm can be seen in Figure 6.

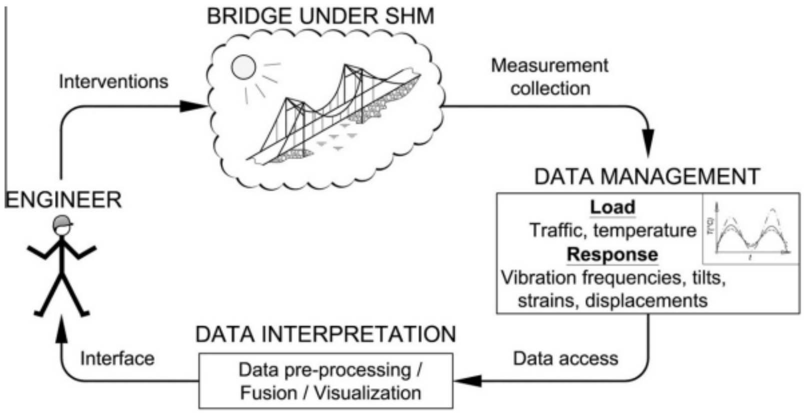

Figure 6: Bridge management paradigm (Kromanis & Kripakaran, 2014)

The research of Kromanis and Kripakaran (2014) looked at the development of measurement interpretation strategies to support this part in the bridge management paradigm. The focus is on developing approaches that ensure that the thermal response of the measurements is isolated. A data-driven strategy is proposed for building models that can reliably predict the thermal response with a given reference set of measurements. How the bridge reacts with a change in temperature depends on the temperature distribution across the bridge. However, it is impossible to accurately measure the temperature of every exact point of a bridge or structure. However, this can be estimated with measurements of distributed sensors. The research of Kromanis and Kripakaran (2014) proposes to apply distributed temperature measurements to predict the thermal response of a bridge. This way of applying is shown in a flowchart. This can be seen in Figure 7.

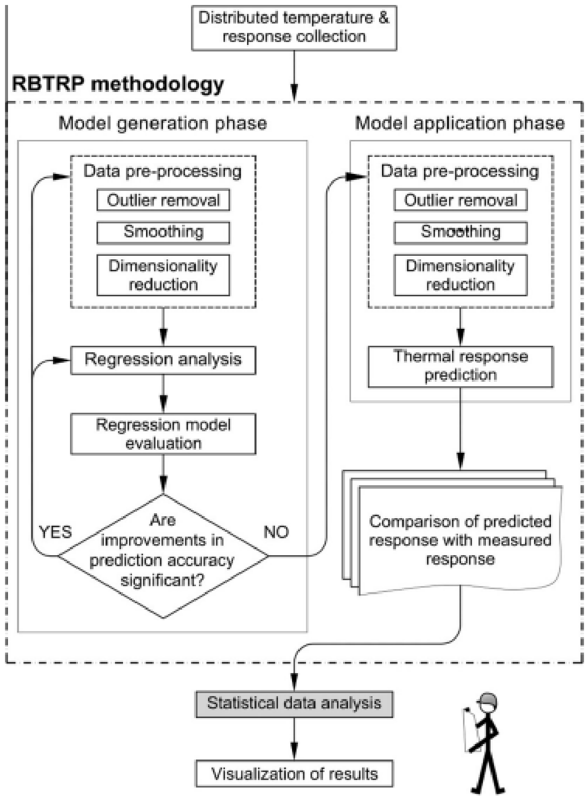

Figure 7: Flowchart of methodology to predict thermal response of a bridge (Kromanis & Kripakaran, 2014)

The flowchart shows the two phases with all associated steps that are required for the process. A statistical algorithm can be used for the regression analysis. Most algorithms have significant reduced capabilities in damage detection for a small number of incorrect measurement values (Posenato et al., 2010). It is therefore very important that outliers are removed from the data. There are several methodologies for finding outliers and replacing them with appropriate values. Posenato et al. (2010) describes 3 different algorithms: Three – σ analysis, auto regressive analysis, and interquartile range analysis (IQRA). These are all described in the report by (Posenato et al., 2010). Of these three algorithms, IQRA has the best features. IQRA is based on a robust analysis, which means that it can also be used when there are multiple outliers. IQRA uses a moving window with certain thresholds. If it finds an outlier by comparing to the other values within this window, the outlier will be replaced by the median of this moving window (Posenato et al., 2010). These create regression models that examine the relationship between temperature distribution and structural response of collected data from a reference period.

There are several algorithms that can be used to make statistical models showing the relationship between temperature and structural response. Kromanis and Kripakaran (2014) elaborated four algorithms in their report that already have been used in the field of SHM. These are Multiple linear regression, Robust regression, artificial neural networks, and support vector regression. For application to the biobased bridge in Ritsumasyl, the focus is mainly on multiple linear regression (MLR). In a simple regression, measurements of a variable (y) and an explanatory variable (x) are used to create a function. With such a function it is possible to predict values of (y) using (x). However, it is often the case that not one, but more explanatory variables must be used to correctly predict the values of (y). MLR can be used for this (Kromanis & Kripakaran, 2014). MLR has several advantages and disadvantages. MLR can be used to determine the relative influence of one or more predictor variables to the criterion variable. Another advantage is that MLR can identify outliers. A disadvantage is that MLR is sensitive to erroneous data (Weedmark, 2019). It is also possible that the assumption of linearity between two variables is not applicable. Overfitting and overtraining are also problems that can occur. Overfitting is a condition in which a statistical model describes a random error in the data instead of the relationship between variables (Frost, 2021). Overtraining is when a model can predict training examples with very high accuracy but cannot properly predict new data. This often occurs when using too little data or data that is very homogenous (2021). To determine whether the regression model produces the desired result, the error must be examined. To find the error of the prediction made by an algorithm, Equation 1 can be used.

Equation 1: $e = \frac{1}{n}\sum_{i=n}^{n} |Y_{p,i} - Y_{r,i}|$

The average prediction error (e) can be determined with the predicted (Yp), the measured response (Yr) and the number of measurements (n) (Kromanis & Kripakaran, 2014).

For the different regression algorithms, it is necessary to vary the parameters and to find the highest accuracy in the prediction. The parameters that fit the highest possible accuracy are then used in the algorithm to predict the structural response based on temperature distributions. This process is called the regression-based thermal response prediction (RBTRP) (Kromanis & Kripakaran, 2014). This will ultimately help to predict how the bridge will respond to changes in temperature in the future. It can also provide insight into the current state of the bridge. A parameter that can be used to optimize the accuracy of the prediction error is the thermal inertia. Thermal inertia indicates how long it takes for a material to reach the same temperature as its surroundings (Sala Lizarraga & Picallo-Perez, 2020). Materials with a high thermal mass and low thermal conductivity may have internal temperatures that are significantly behind the ambient temperatures.

The data sets used for generating regression models are often very large due to the high frequency of measurements. For highly complex regression models, this can cause the computation time to increase significantly with the size of the data set. To speed up this process, the dimensionality of the data set can be reduced. Principle component analysis (PCA) is often applied for this. This is a statistical technique that reduces the dimensionality of the data but at the same time preserves the accuracy and variation in the data (Ringnér, 2008). This is done by identifying directions, called the PC's, along which the variation is maximal. By using fewer components, each sample can consist of fewer numbers instead of thousands of variables. Another way to speed up the process is to reduce the frequency of measurement collection. Omitting measurements increases the speed of calculation with little loss in prediction accuracy. In this way only the size of the training set is changed. This means that the data set still presents full variability in the measurements because only the number of input measurements is changed. This RBTRP method will lead to regression models that can accurately predict the thermal response of a bridge using distributed temperature measurements (Kromanis & Kripakaran, 2014).

3 Methodology

To predict the thermal response of the Bio-based bridge, using temperature distributions, the RBTRP methodology is applied. The RBTRP method consists of two phases (Figure 8): the model generation phase and model application phase (Kromanis & Kripakaran, 2017).

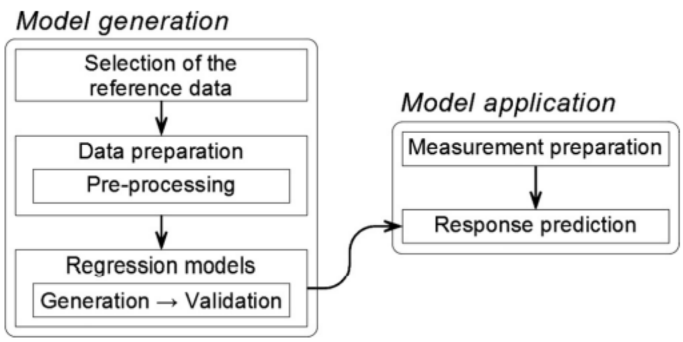

Figure 8: Simplified RBTRP methodology (Kromanis & Kripakaran, 2017)

The model generation phase generates a regression model that uses temperature distributions of the bridge as input to predict the thermal response. It involves several steps that need to be taken. First, a reference set is chosen. This reference set must meet a few requirements. During this reference period it must be assumed that the bridge behaves normally. It should also be considered which temperature and strain sensors are suitable as input for the regression analysis in the reference period. If a sensor collects incorrect measurements, it is not suitable to use as input for the regression analysis (Weedmark, 2019). Incorrect values are values that are far from the expected values for no visible reason. The collected measurements in this period must be split into a training and test set. This is needed to train and evaluate the performance of the model. For this it is important that the training set and test set are approximately the same size. They should also contain a varied collected measurements data set from different seasons so that the model does not get overtrained (2021).

The collected measurements in the reference period that will be used also needs to be prepared. This step is called data pre-processing. Any outliers of the collected measurements must be removed and replaced using IQRA. The data also needs to be smoothened using a moving average filter to ensures less noise in the data. With a large dataset, the training of the model can be computationally demanding. In this case, the measurements can be down sampled to a suitable frequency so that the computation time decreases. Finally, the dimensionality of the temperature measurements of the data set can be reduced. Principal component analysis (PCA) can be used for this. This will also help reduce the computational time. A regression model needs a strain value and temperature value at the same point in time as input to be able to make a regression analysis. A period in which collected temperature measurements are available but no collected strain measurements are available, and vice-versa, is unusable for a regression analysis. The last step of data pre-processing is to filter these periods from the data sets.

When the collected measured values are correctly per-processed, the generation and validation of a regression model is the next step. The regression model is trained using the training data set. The performance of the model is then tested on the test data set. Training and testing can be done with different regression models such as support vector regression and multiple linear regression. To get the lowest possible prediction error, the models are generated iteratively by varying parameters. Parameters that can be used for this are the thermal inertia, number of input measurements and number of principle components. The combination of parameter values for the lowest prediction error are saved and used for the next phase. This is the model application phase. Here, the regression models are used with the highest possible accuracy to predict real-time thermal response using measured temperature distributions. For this, the temperature and strain measurements must also be pre-processed to be used as input for the regression model. By looking at the measured values, it can be checked whether the model can correctly predict the thermal response. A prediction with an error below 5% is considered an acceptable prediction (Swanson, 2015).

4 Results

4.1 Sensor identification

The first step in the RBTRP methodology is to select sensors that have suitable collected measurements. Suitable sensors are sensors that collect little incorrect data and are representative for the behaviour of the bridge. The bridge consists of a moving and a fixed part. For consistency it is decided to choose one of these parts for the analysis. The moving part has more temperature sensors thus seems more suitable. The moving bridge part can be seen in Figure 9 and Figure 10.

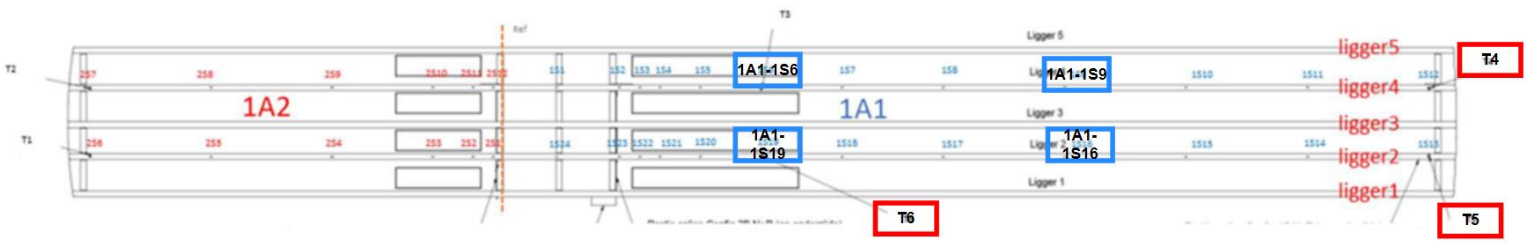

Figure 9: Sensor locations top, moving part

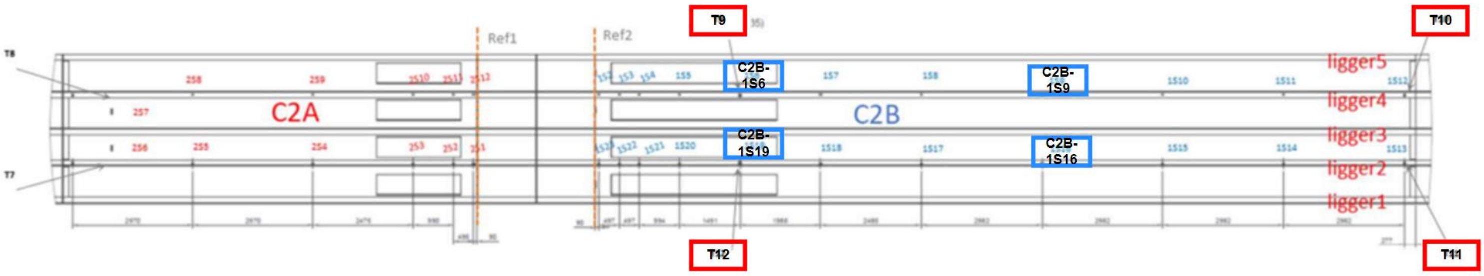

Figure 10: Sensor locations bottom, moving part

The moving bridge part consists of a short (1A2, C2A) and a long part (1A1, C2B). Small blue and red numbers and letters can also be seen in Figure 9 and Figure 10. These are all strain sensors placed on the bridge. The colour indicates whether the sensors are on the short or long part of the moving bridge part. The temperature sensors are indicated with small black letters and numbers and with an arrow pointing to the location of the sensor on the bridge. The long part has eight temperature sensors and is therefore preferred over the short part where two temperature sensors are placed. Sensor T3 is malfunctioning so this one will not be used. Eight strain sensors have been chosen on this same long section of the moving bridge section. These are strategically chosen near the temperature sensors and right in the middle between the 8 temperature sensors. The selected temperature and strain sensors are indicated with red (temperature) and blue (strain) squares in Figure 9 and Figure 10. The strain sensors and temperature sensors used are numbered and are shown in Table 1.

Table 1: Strain gauges (SGs) and temperature sensors (T)

Nr	Strain gauge (SG)	Temperature (T)
1	1A1-1S6	T4
2	1A1-1S9	T5
3	1A1-1S16	T6
4	1A1-1S19	T9
5	C2B-1S6	T10
6	C2B-1S9	T11
7	C2B-1S16	T12
8	C2B-1S19	

4.2 Data pre-processing

The next step in the RBTRP methodology is to pre-process the data so that it can be used for the regression analysis. First, it must be considered which measurements will be used for the reference period and monitoring period for training and testing the regression model. Strain and temperature data is available from the bridge from two years of measurements. The data is collected with a measurement interval of approximately 10 seconds. Because this is only two years of data, it is decided to use the first year as a reference period and the second year for monitoring. In this case, the reference period is used to train and test the model to find the smallest possible prediction error. The beginning of the measurements started at the end of November 2019. This is therefore used as the start of the reference period. Because in the year 2020 very large amounts of incorrect data were collected from mid-August, it is decided not to use the entire year, but only the data collected up to and including August 1. The reference period therefore runs from November 27, 2019, to August 1, 2020. For the monitoring period it was decided to choose approximately the same period a year later. The monitoring period therefore runs from 1 January 2021 to 31 August 2021. To represent these periods schematically, Figure 11 and Figure 12 indicate all collected strain and temperature measurements used for the reference period and measurements used for monitoring.

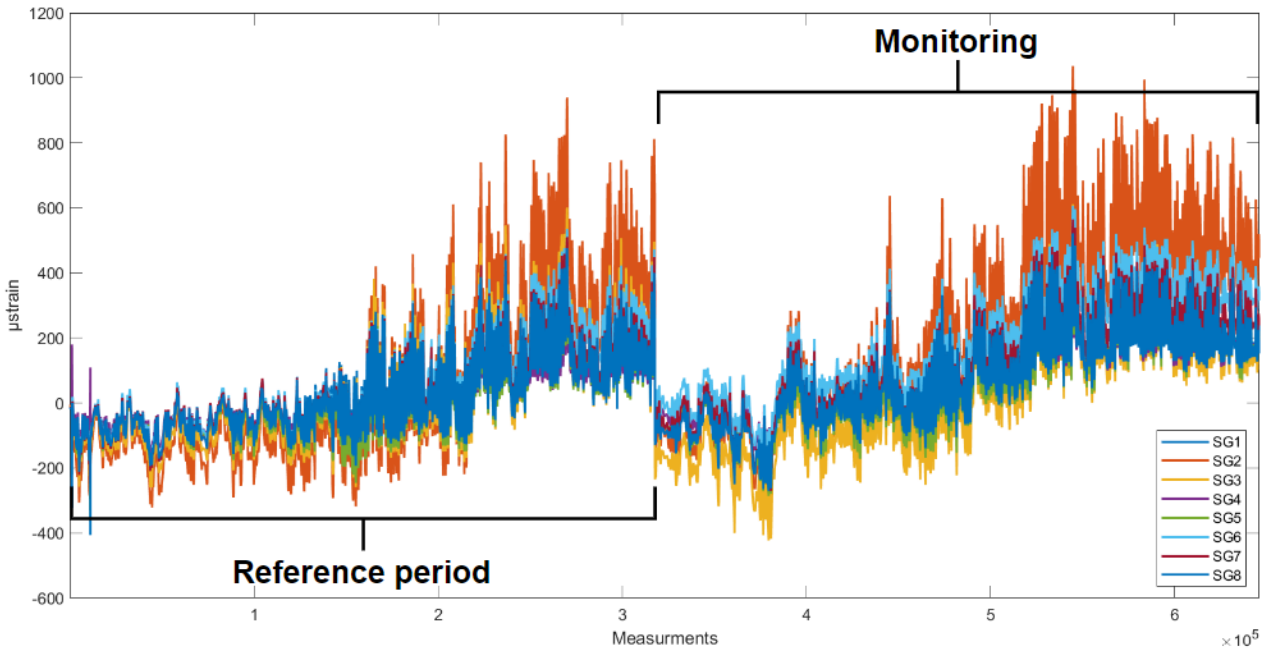

Figure 11: Strain measurements used for reference period and monitoring

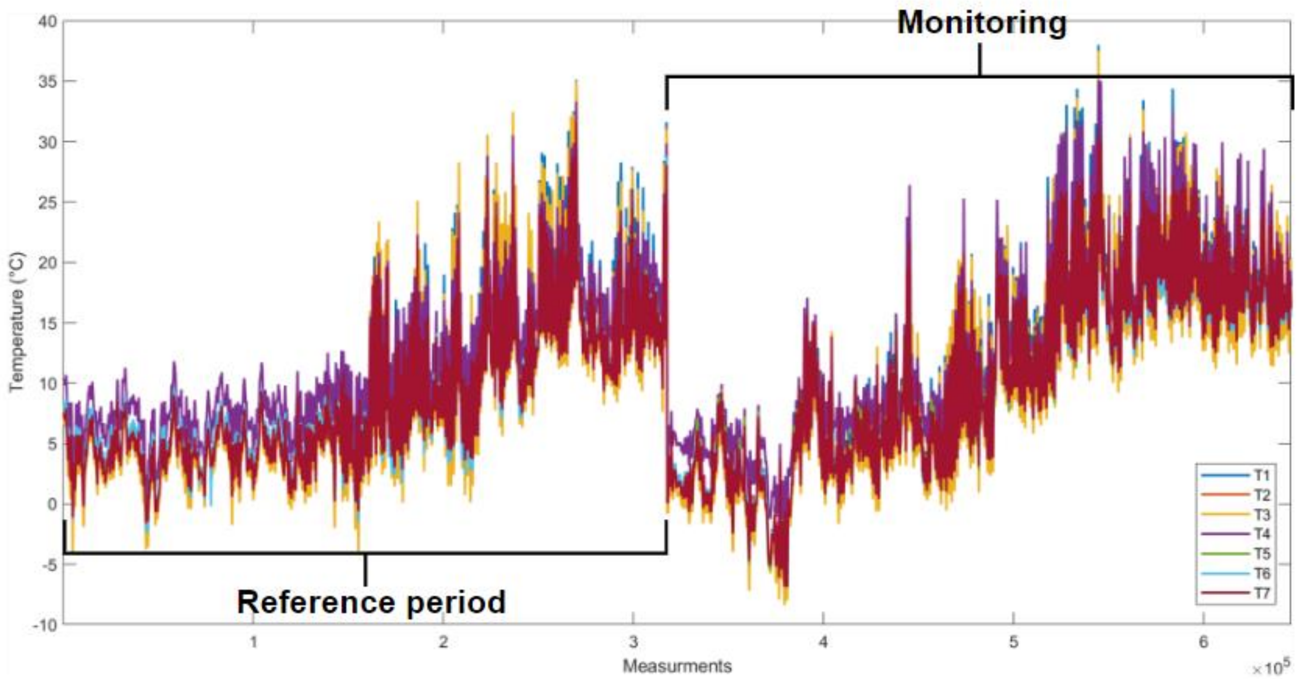

Figure 12: Temperature measurments used for reference period and monitoring

In Figure 11 and Figure 12 a drop can be seen in the measured values between the reference period and monitoring phase. This is caused by no data being used from August 2, 2020, to January 1, 2021. The data selected for the reference period and monitoring was further pre-processed before it is suitable for the regression analysis. This process is shown in Figure 13.

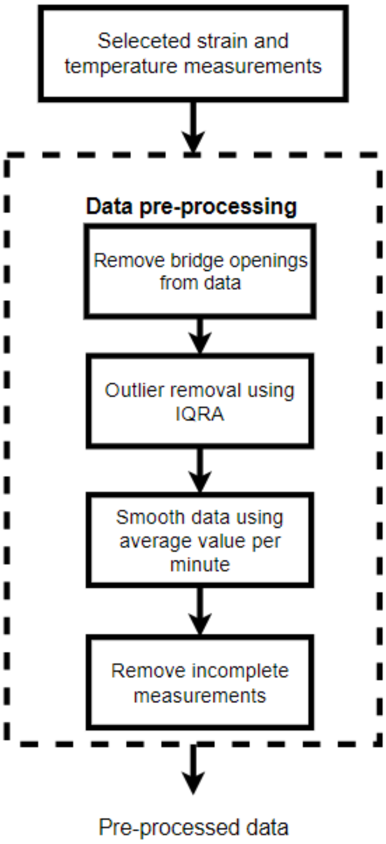

Figure 13: Flowchart for data pre-processing

After applying all steps from Figure 13 the data is suitable for regression analysis. In Figure 14 the result of pre-processing can be seen.

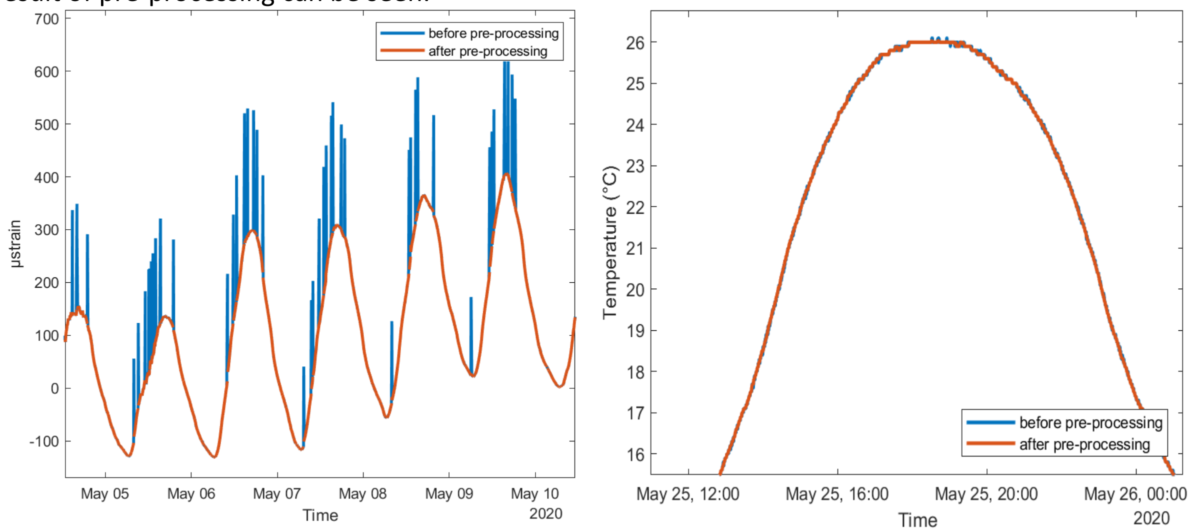

Figure 14: Strain (SG1) and temperature (T1) measurements before and after pre-processing

In Figure 14 a situation for the strain and temperature data is taken showing the effect of pre-processing. For the strain it is clear that the peaks caused by the bridge openings have been removed. With the temperature data there were already few errors with the raw data. The effect of smoothing is only slightly visible. The processed data can now be used for the regression analysis. According to Kromanis and Kripakaran (2014) the use of an MLR is sufficient for a correct prediction. To determine the error of this regression, the mean is taken of the absolute difference between predicted and measured values, divided by the total range of the measured values. In this way the error in percentage can be obtained.

4.3 Minimizing the prediction error

To optimize the regression model, an iterative process was used to find the smallest prediction error. The parameters used for this are the number of input measurements and the thermal inertia. The model was trained on the reference period, from the end of 2019 to August 2020. The model is also tested on this period. The model has been optimized by looking with different parameter values to see what causes the lowest error between predicted and measured values. This has been applied to all eight strain sensors using all temperature sensors. In Figure 15 the prediction error is set against the thermal inertia (Thermal I) and number of input measurements (No. of inpM). The number of input measurements refers to that around a certain number of measurements a point is taken that is used as input for the regression. With a number of input measurements of 100, therefore, every 100th point of the total data set is used for the regression as input. The higher the number, the less data is used for the regression. In Figure 15 the prediction error can be seen from all strain gauges.

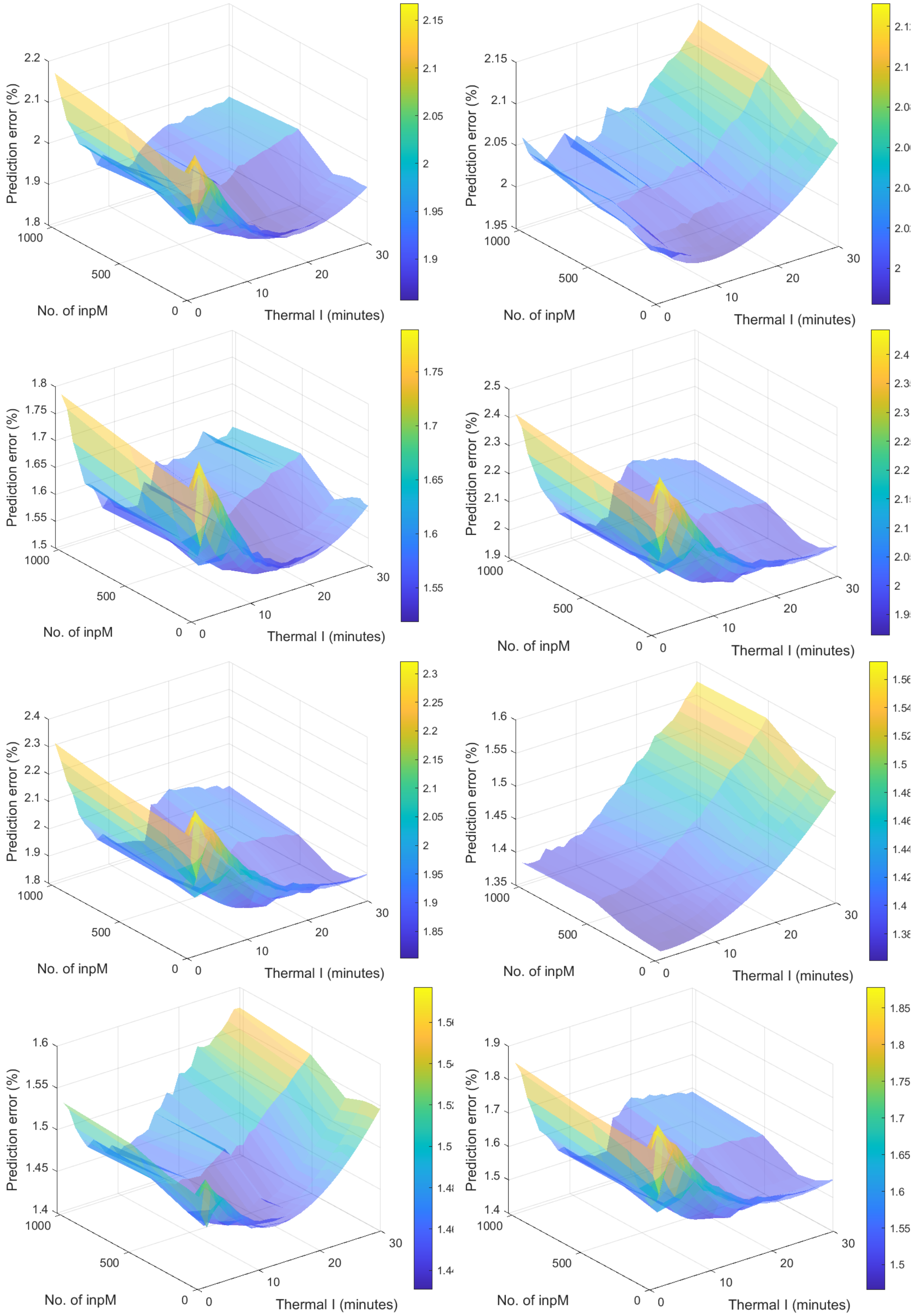

Figure 15: Prediction error versus thermal inertia and number of input measurements from SG1, SG3, SG5, SG7 (Top left to bottom left) and SG2, SG4, SG6, SG8 (Top right to bottom right)

The number of input measurements has a range from 1 to 1000 in Figure 15. The thermal inertia has a range of 0 to 30 minutes. These ranges were chosen because the prediction error within this range in the case of number of input measurements flattens out after about 500 and the thermal inertia outside this range increases the prediction error substantially. So, it has been decided not to increase the range as no improved prediction errors are expected and this would cost a lot of extra computing time. The optimal value for both parameters is therefore taken within these ranges. What stands out for the number of input measurements is that for some SGs with a low No. of inpM value (1 for example) the prediction error gets higher. This while it is expected that the prediction error would become smaller when using a low value. In Figure 15 all SGs have a specific value for the number of input measurements combined with the thermal inertia that ensures that the prediction error is minimal. These specific values are saved and used as input to train the regression model. The values that provide the highest accuracy for the prediction error are shown in Table 2.

Table 2: Optimal values for number of input measurements and thermal inertia for best accuracy prediction error

SG	No. of inpM	Thermal I (minutes)
1	256	18
2	8	9
3	256	13
4	256	16
5	256	16
6	2	2
7	256	11
8	256	16

4.4 Prediction error

With the optimal parameters as input, the regression model can predict the thermal response of the bridge using the temperature distribution. This has been applied for all SGs. The model has been applied on the reference and monitoring period. In Figure 16 the measured and predicted values of both periods can be seen. The reference period includes all data up to the approximately the 300000th measurement and the monitoring period includes everything after approximately the 300000th measurement. The results of SG1 are used as an example. Figure 17 shows a zoomed in view on the prediction of SG1 in the monitoring period.

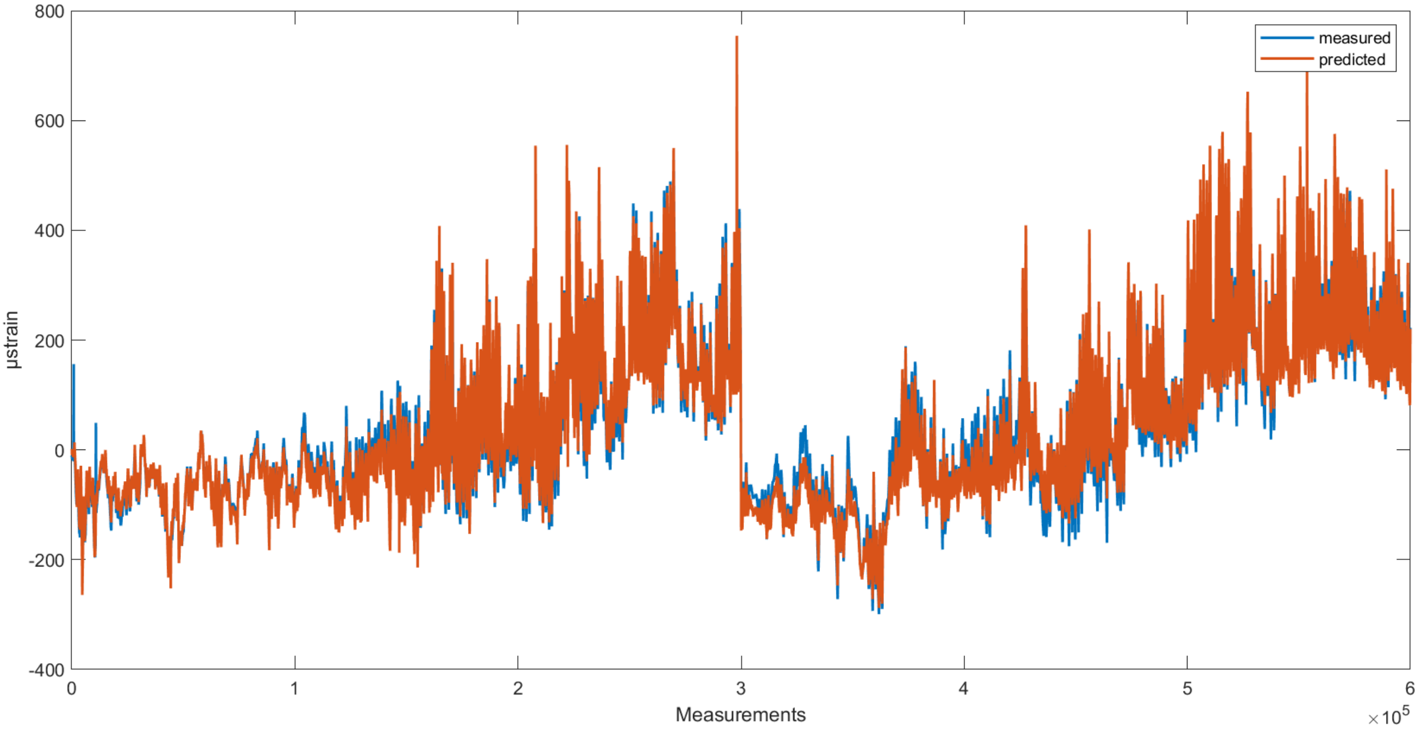

Figure 16: Measured and predicted strain values reference and monitoring period (SG1)

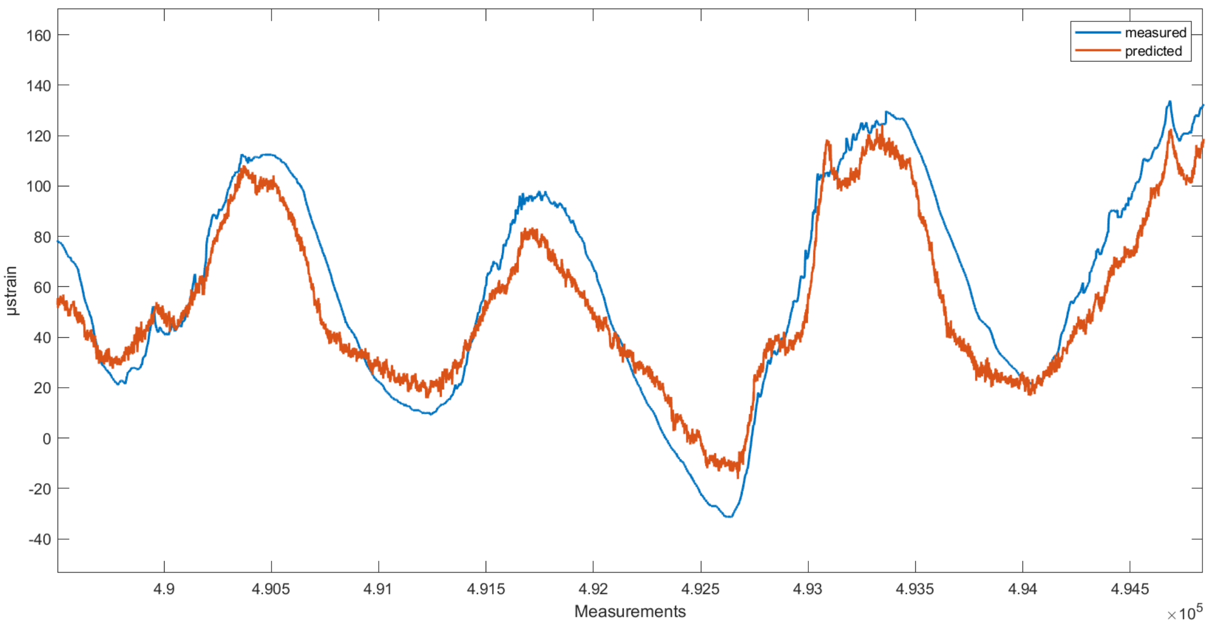

Figure 17: Measured and predicted strain values zoomed in monitoring period (SG1)

In Figure 16 and Figure 17 it can clearly be seen that the measured and predicted values do not always correspond well. The prediction error can be used to see how accurate the predicted values are. The prediction error can be determined by looking at the difference between the measured and predicted values. The prediction error has been determined for the reference and monitoring period. IQRA has also been applied to the prediction error to replace most of the outliers. In Figure 18 the prediction error can be seen before and after applying IQRA. In the figures, the prediction error is indicated in micro strain.

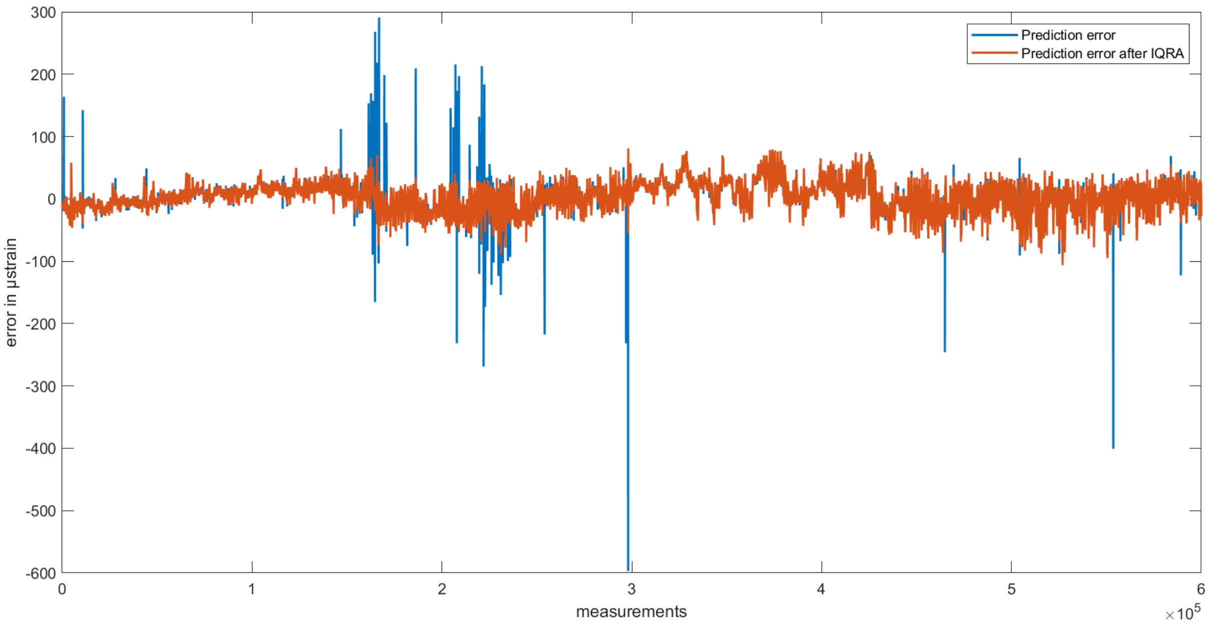

Figure 18: Prediction error before and after aplying IQRA (SG1)

In Figure 18 the prediction error is determined by subtracting the predicted values from the measured values. The more the prediction error deviates from 0, the more inaccurate the prediction is. This has been applied for all SGs. For this, the prediction errors have been converted into percentages in relation to the total range. A prediction with an error below 5% is considered an acceptable prediction (Swanson, 2015). In Table 3 all errors are shown for all SGs for the reference and monitoring period.

Table 3: Prediction errors in percentages for all SG's

	SG1	SG2	SG3	SG4	SG5	SG6	SG7	SG8
Prediction error reference period (%)	1.86	1.98	1.51	1.91	1.95	1.36	1.43	1.47
Prediction error monitoring period (%)	2.65	7.16	2.12	8.49	3.48	9.70	5.87	3.90

Table 3 shows that a few SGs have a significantly higher error in the monitoring period than in the reference period. When looking at the strain data from the beginning of the reference period (2019) and the beginning of the monitoring period (2021), something should be noted. When the measurements started, the strain data from all strain gauges were still close to each other. This can be seen in Figure 19.

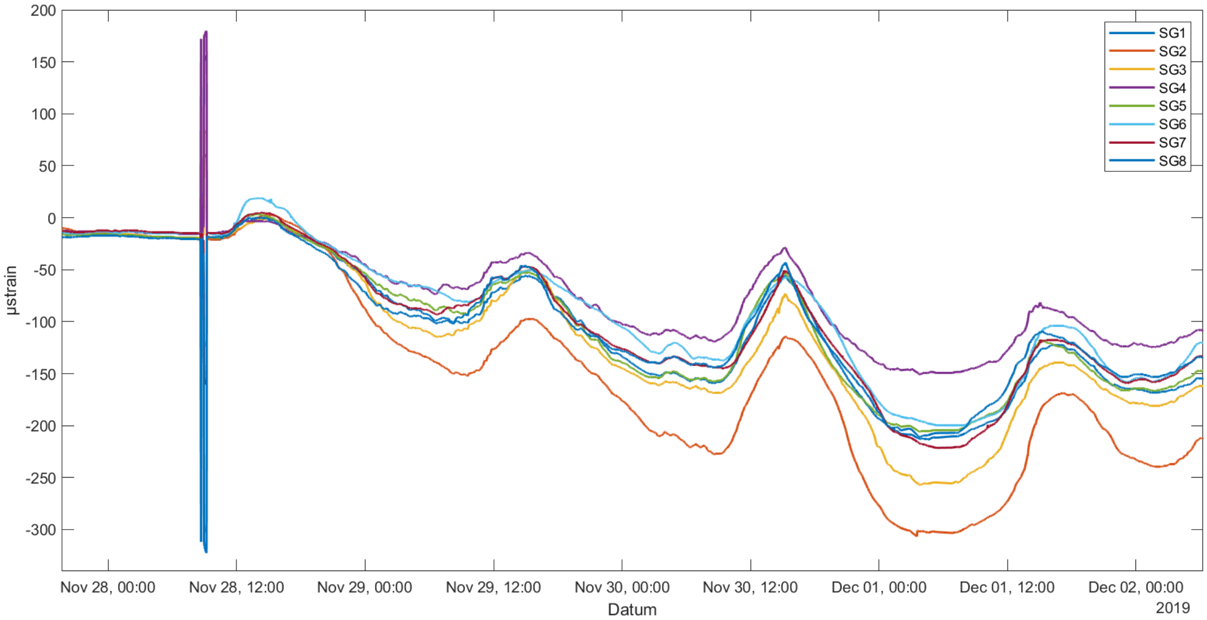

Figure 19: Strain data start of reference period (2019)

However, the strain data is more diverging at the beginning of the monitoring period, visible in Figure 20, compared to the reference period in Figure 19. The temperature in both periods was similar.

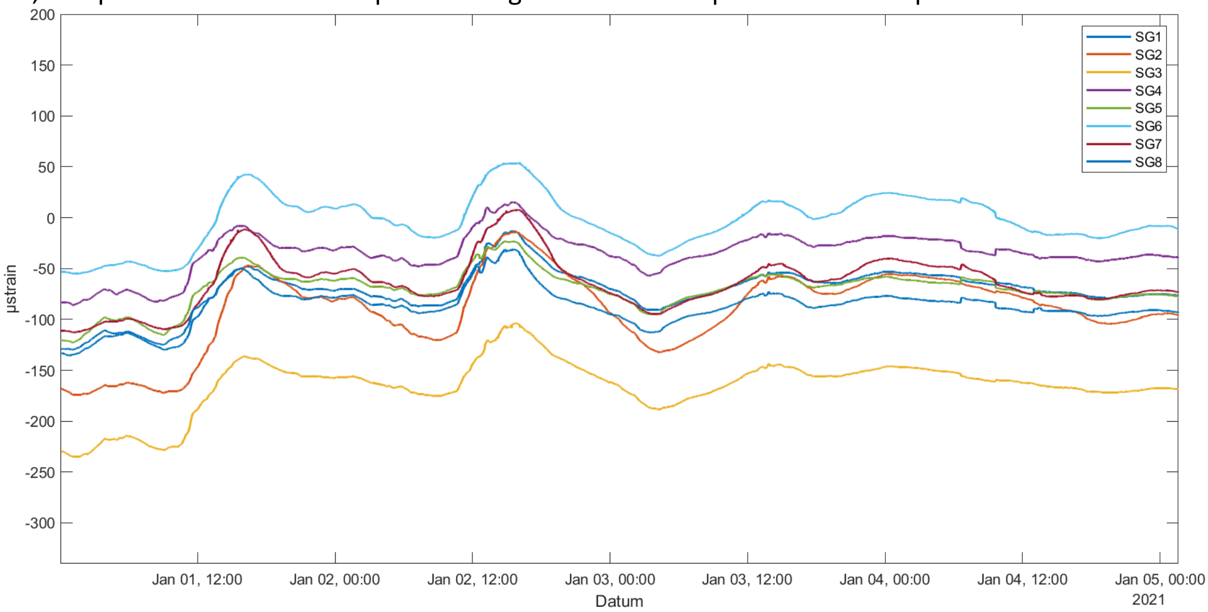

Figure 20: Strain data start of monitoring period (2021)

This is one possible reason why the prediction error for monitoring is much larger than for the reference period. It is clear that something has happened to the SG's data. It is possible that at the beginning of the reference period the bridge still had to be placed. It could be that because the bridge has only just been placed, the bridge still had to sag a bit. That would mean that the strain values at the beginning of the reference period may not have been representative yet. However, this cannot be said with certainty. For a correct model it is therefore important that the data is corrected in an appropriate way. Assuming that the predicted values for monitoring are correct, the measured data can be corrected with this. This assumption is based on the prediction error in the reference period that gives consistently low values. The shift is done by comparing the first measured value of 2021 with the first predicted value of 2021. The difference between these two values is then subtracted from all data points of the measured data in 2021. In Figure 21 it can be seen how this is applied. The size of the shifts is shown in Table 4.

Table 4: Shifts applied to measured data monitoring period

	SG1	SG2	SG3	SG4	SG5	SG6	SG7	SG8
Shift in µstrain	-10.06	-122.76	42.86	-21.41	-35.67	-82.92	-27.37	-9.13

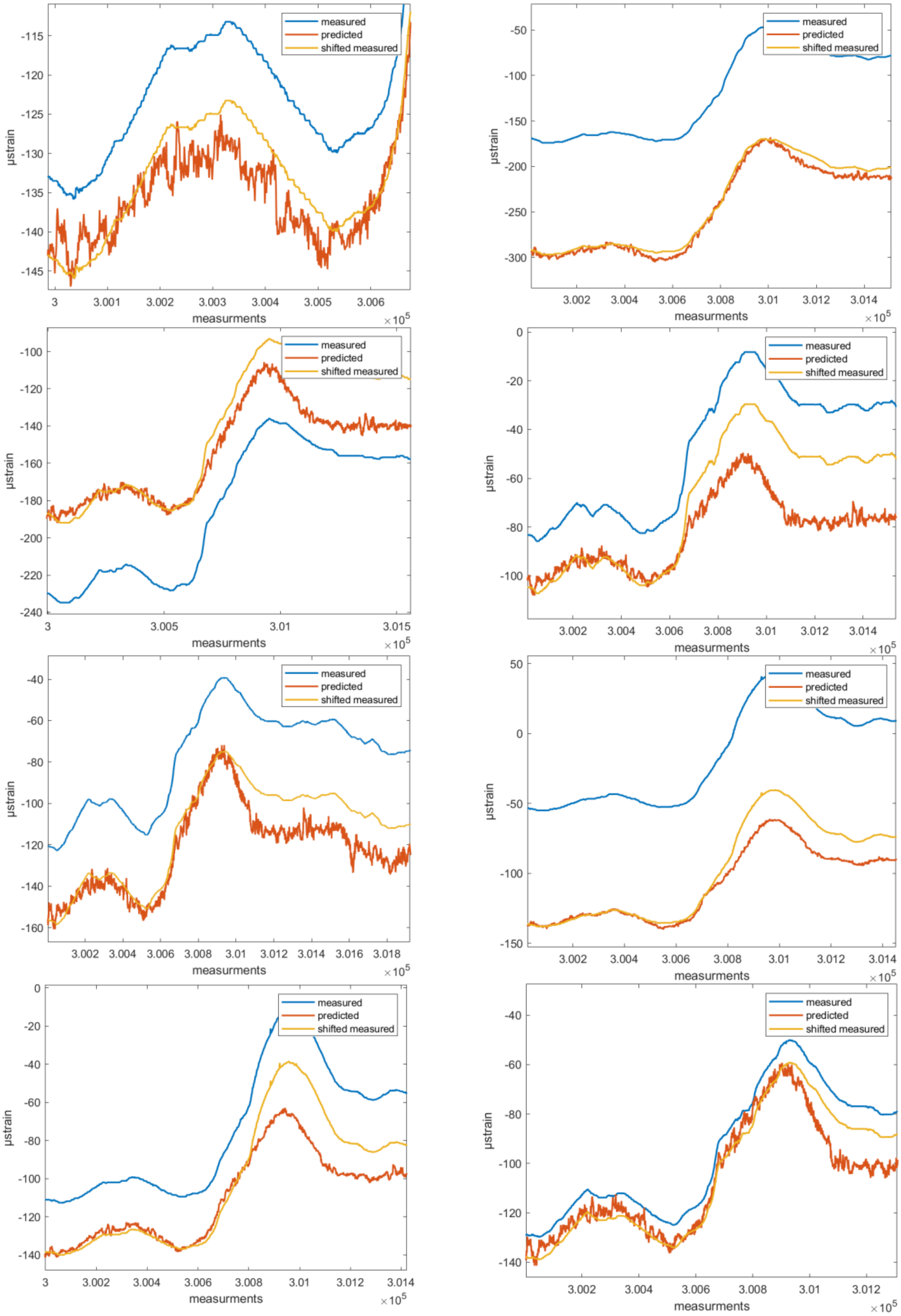

Figure 21: Shifted measurements monitoring period from SG1, SG3, SG5, SG7 (Top left to bottom left) and SG2, SG4, SG6, SG8 (Top right to bottom right)

In Figure 21 it can be seen that the shifted measurement fits better with the predicted values than the original measurement for each SG. The new prediction errors can be seen in Table 5.

Table 5: Prediction errors in percentages for all SG's including error shifted measurements 2021

	SG1	SG2	SG3	SG4	SG5	SG6	SG7	SG8
Prediction error reference period (%)	1.86	1.98	1.51	1.91	1.95	1.36	1.43	1.47
Prediction error monitoring period (%)	2.65	7.16	2.12	8.49	3.48	9.70	5.87	3.90
Prediction error monitoring period shift (%)	2.15	2.91	2.18	4.54	2.39	2.68	2.93	2.55

Table 5 shows that the shifted measurement values have ensured that the prediction error for the monitoring period has significantly improved. All prediction values are now within the 5% range which means that the predictions are appropriate.

5 Discussion

5.1 Data pre-processing

To make a successful and accurate regression model, pre-processing is very important. The first step is to choose a reference and monitoring period so that the model can be trained and tested. The training period includes winter and summer, to include varying temperatures for the training of the regression model. In the winter of the monitoring period the temperature was at some point lower than the temperature has been in the winter of the reference period. The model was trained with collected measurements from the reference period. Because the model is not trained at these lower temperatures from the monitoring period, the prediction for the monitoring period may be less accurate at colder temperatures. For an improved regression model, data for the reference period could have been taken up to and including 2021. The first data set ends at the end of August 2020. By also including data from after August 2020, it might have been better to see why the strain data at the beginning of 2021 is so much further apart than a year earlier. This is namely not yet the case in the first half of 2020. After choosing a suitable reference set, it is important to remove outliers from the data and make the data smoother. Incorrect data has a major effect on the outcome of MLR. This is also apparent from the prediction error of the monitoring period that can be seen in Table 3. Data is pre-processed by using IQRA and removing the data when the bridge is open to remove as much incorrect data as possible. However, there are a few incorrect values that have not been removed by IQRA. This is not a common occurrence but may have influenced the regression analysis. The incorrect collected measurement data of the monitoring period can have various causes. The sensors may have had an error. For example, due to a reset of the sensors that has taken place. It is also possible that the characteristics of the bridge have changed. However, the chance of this seems small because the bridge has only been there for 2.5 years. Another reason could be that the data may not have been representative for the bridge yet. It is possible that the bridge needed some time to sink after installation. A shift had to take place of the measured data in the monitoring period because it deviated from the reference period. The fact that the bridge had yet to sink could be a reason for this, but this should be further investigated.

5.2 Optimal thermal inertia and No. of input measurements

For an appropriate regression model, it is important that the optimal thermal inertia and number of input measurements are used. The bridge has only seven working temperature sensors that are used for the regression analysis. The application of principle component analysis is therefore not of great importance in this situation and has therefore been omitted. With larger datasets it is useful to apply this because it both improves accuracy and calculation speed. It is important that the optimal values per parameter are found because it can have a lot of influence on the prediction error. For each SG an optimal value has been found for the number of input measurements in the range of taking every measurement and taking every 1000st measurement. This range has been chosen because it can be clearly seen that after a value of 500 for number of input measurements, the prediction error flattens out and does not change much anymore. It can also be seen that for some SGs (see Figure 21) the prediction error increases again with a low value for number of input measurements, while the expectation is that with a low value the prediction error would be the most accurate. There is no direct reason for this and could therefore be further investigated. Thermal inertia is important to give the predicted values and measured data the same phase. This will decrease the prediction error. An optimal value has also been found within the range for the thermal inertia for the lowest possible prediction error. A range of 0 to 30 minutes has been chosen because within this range is already visible that the prediction error becomes a lot higher at certain values. It was therefore decided not to increase the range because no better prediction errors are expected. The results of the iterative process and thus the optimal No. of input measurements and thermal inertia can be seen in Table 2. It can be clearly seen that most SGs have a No. of input measurements of every 256th point and a thermal inertia of approximately 15 minutes. However, there are 2 exceptions that require a lower No. of input measurements and thermal inertia. These are SG2 and SG6. The similarity between these 2 sensors is that they are located in the middle of the bridge on the east side (see Figure 9 and Figure 10). There can be several reasons why these sensors have a different thermal inertia compared to the other sensors. One reason could be that these sensors are affected more quickly by the sun than the other sensors. However, this cannot be stated with certainty and further research is needed.

5.3 Multiple linear regression model

Multiple linear regression (MLR) was used to predict the thermal response of the bridge. This is one of several algorithms that can be used. In a follow-up study, it could be examined whether the prediction can be made better by using another algorithm. MLR has a few disadvantages. MLR is sensitive to incorrect data and MLR assumes linearity when this may not be the case. However, research by Kromanis and Kripakaran (2014) showed that MLR was sufficient. With the MLR it is possible to predict the thermal response using the distribution of the bridge. However, there has been an adjustment to the monitoring data to improve the prediction. The reason for this is that the strain data at the beginning of 2021 were much further apart than in 2019. This is noticeable in colder periods. It is possible that this deviation can already be seen in autumn and winter months in 2020, but these periods are not used for the regression analysis. Therefore, it only becomes noticeable at the start of the monitoring period in the beginning of 2021. It is clear, that something has happened to the sensors or data. This can have various causes. For example, the sensors may have been reset once because an error occurred. In September 2020 there was a lot of incorrect data that could not be used. In Figure 22 the raw data from SG1 is given as an example that from approximately the 2100000th measurement incorrect data is collected. This happens approximately from mid-August. This happens with all SGs.

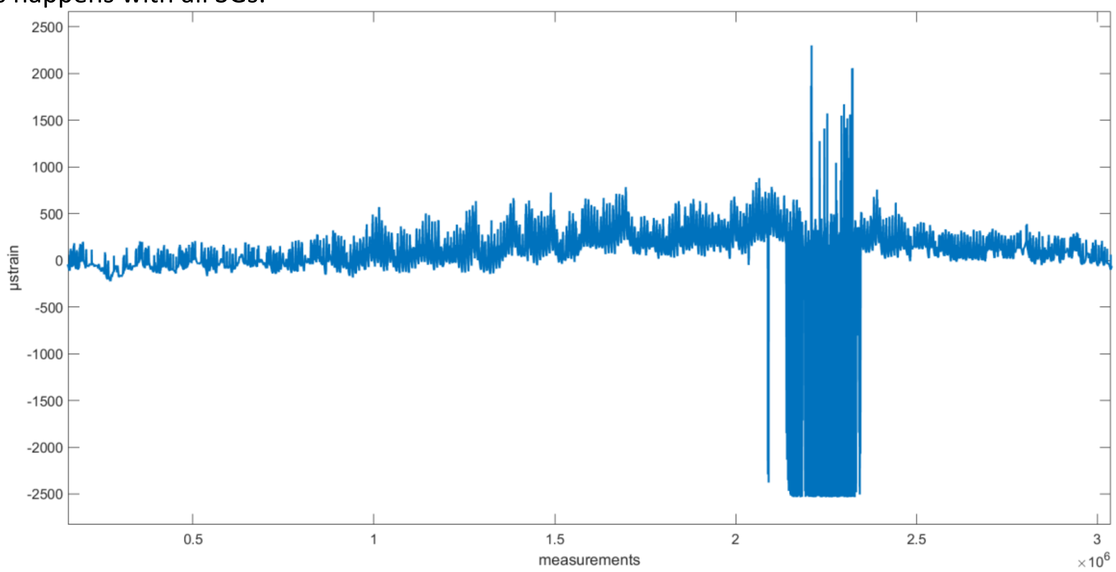

Figure 22: Raw strain data SG1

Another reason could be that the material has started to react differently to external effects over time. It could also be that because the bridge still had to be placed at the beginning of the reference period, the bridge needed some time to sink. As a result, it is possible that the strain data in the reference period was not yet truly representative. Subsequent research should then examine whether training the model with a different reference period yields more accurate results. However, a real reason cannot be concluded with the information from this study. Further research could be done for this. After the shift in the measured data, the prediction error for all SGs in the monitoring period is within 5% (see Table 5). However, SG4 does have a prediction error that is close to 5% with 4.5%. If the prediction of SG4 is compared with, for example, SG1, which has the most accurate prediction, a few things can be seen. In Figure 23 and Figure 24 the prediction of the strain can be seen where the temperature is between 3 and 11 degrees. At these temperatures, the strain is predicted differently by the regression model for SG1 and SG4. The shifted measured values are used in the figures.

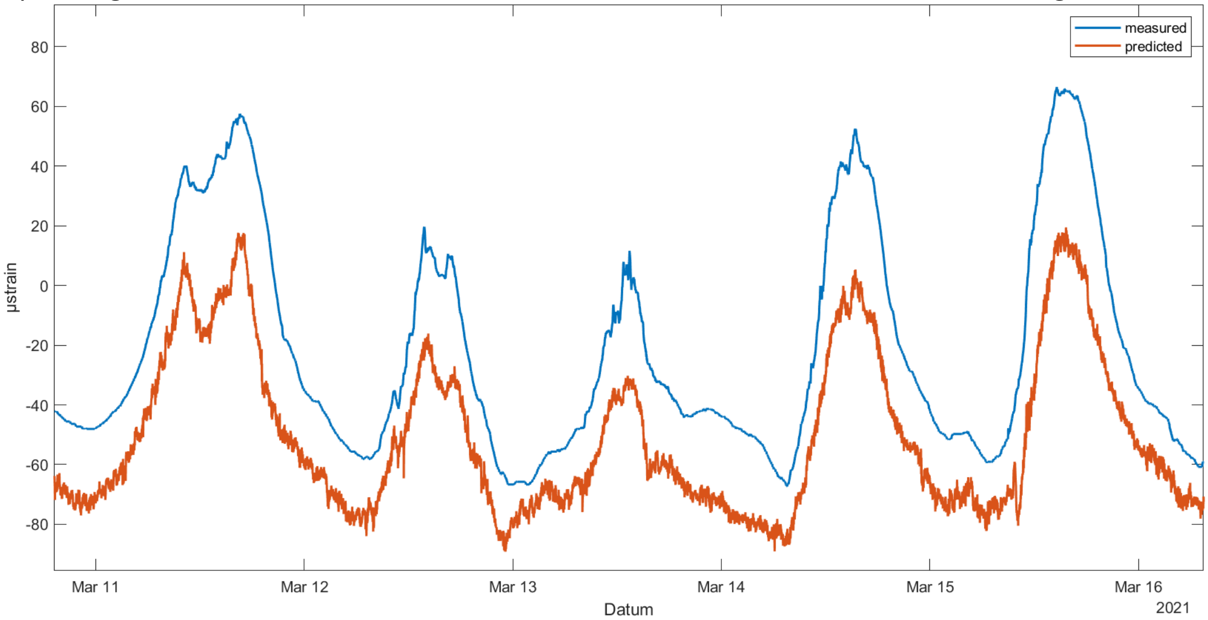

Figure 23: Prediction SG1 March 11-March16

In Figure 23 the predicted values are constantly estimated to be lower than the measured values. In Figure 24 this is also the case, but it is clear to see that the troughs are estimated even lower than in SG1 compared to the measured values. Despite this difference, the two predictions are quite similar in behaviour.

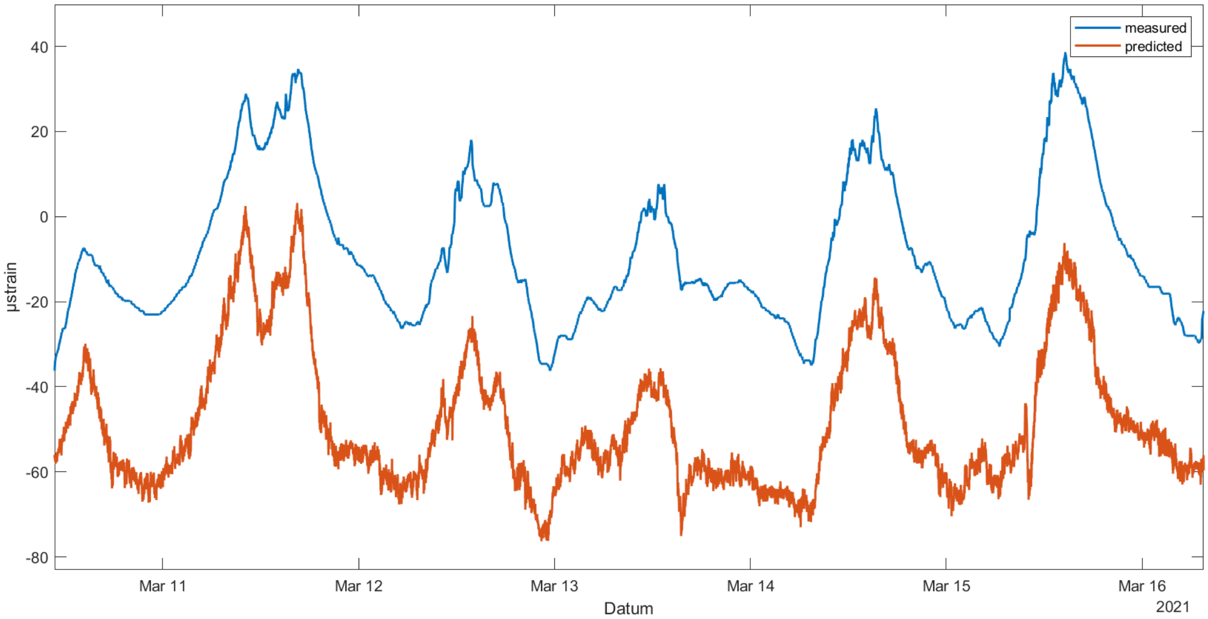

Figure 24: Prediction SG4 March 11-March 16

At higher temperatures, a different pattern can be seen. In Figure 25 can be seen that at higher temperatures, between 13 and 30 degrees, the peaks are predicted to be higher than they actually are for SG1. The troughs in the prediction of SG1 are reasonably at the same values as the measured values.

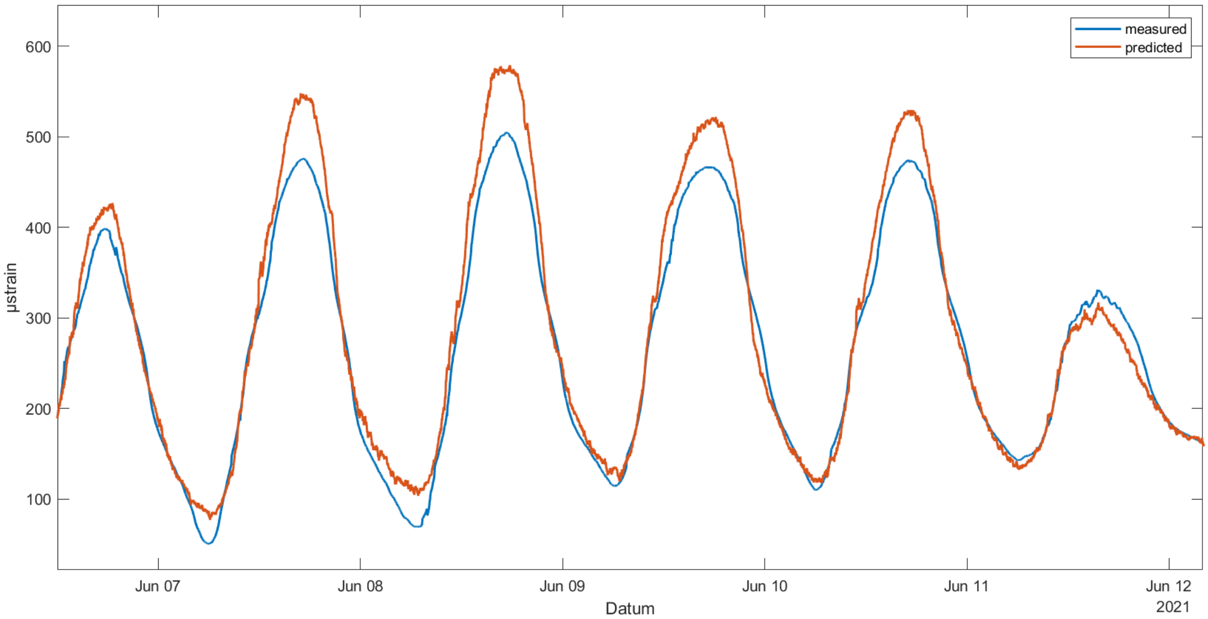

Figure 25: Prediction SG1 June 7-June 12

In Figure 26 can be seen that the predicted strain for SG4 at higher temperatures estimates the peaks reasonably well but predicts the troughs a slightly lower.

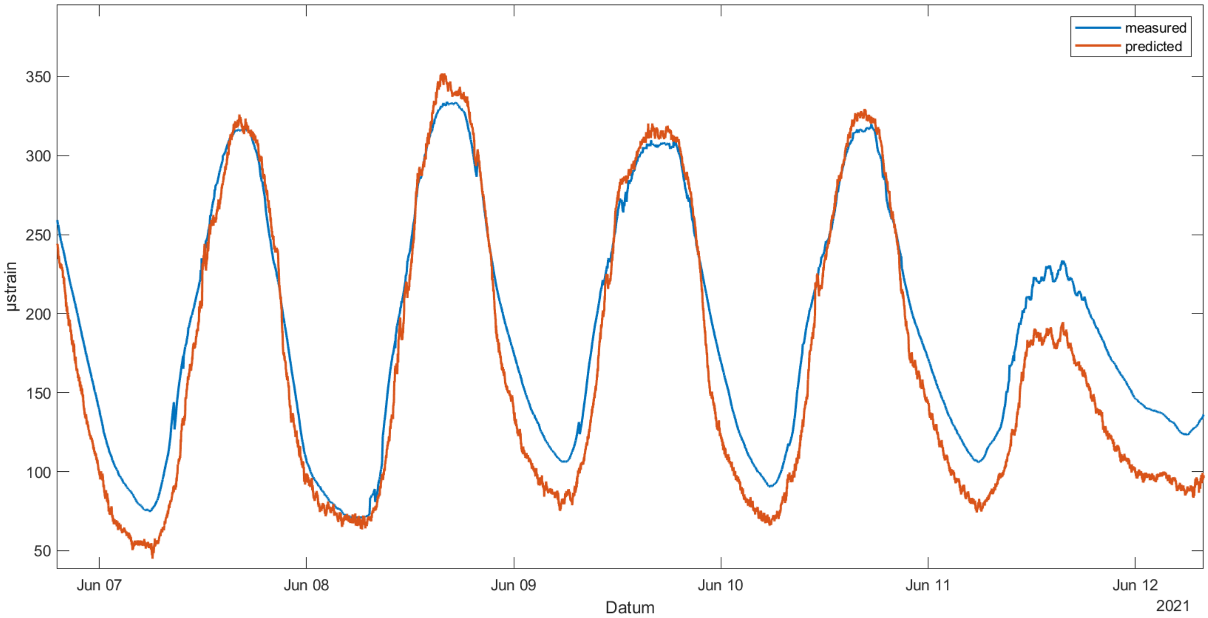

Figure 26: Prediction SG4 June 7-June 12

It can be deduced from the figures that the prediction error at higher temperatures is caused by other reasons than at lower temperatures. However, much is still unclear. The prediction shown at SG1 corresponds to the other SGs. Why SG4 is an exception to this is not clear. This may be due to the placement of the sensor or, for example, an error in the sensor. All temperature sensors were used for all predictions. It may also be the case that this is not representative for SG4. It is possible that only a specific combination of temperature sensors is representative for the prediction of SG4. In Table 6 all regression coefficients are shown per strain sensor and temperature sensor. The higher the value in absolute terms, the more dependent the strain is on the temperature measured at the specific temperature sensor.

Table 6: Regression coefficients of all strain and temperature sensors

	SG1	SG2	SG3	SG4	SG5	SG6	SG7	SG8
T1	9.81	22.77	4.70	0.63	5.85	6.08	4.87	2.48
T2	-36.59	-63.69	-40.85	-20.57	-33.18	-32.87	-28.98	-28.11
T3	33.58	56.32	46.07	17.77	26.25	23.63	23.07	25.47
T4	-28.40	-46.49	-47.53	-19.40	-31.39	-19.11	-28.85	-27.65
T5	-21.23	-4.23	-37.30	-1.26	-17.99	12.78	-5.94	-15.33
T6	20.82	16.37	47.94	-0.48	16.68	-1.71	14.28	11.51
T7	38.43	49.03	45.78	35.39	48.98	31.09	39.47	46.50

In Table 6 can be seen that SG4 has a stronger relationship with certain temperature sensors than with others. However, based on the location of the temperature sensors, nothing can be concluded. This also applies to other SGs. It is not the case that if a temperature sensor is further away from a strain sensor, the relationship is always less significant. It is also not clear why the prediction is underestimated at lower temperatures and the prediction is too high at higher temperatures. Only temperature is included in the prediction in this study. But there are many more factors such as humidity, wind and loads on the bridge that influence the strain of the bridge. So, more research is needed to determine the cause of multiple factors.

6 Conclusion and further research

This research uses the RBTRP method to predict thermal response in the biobased bridge in Ritsumasyl using temperature distributions. A multiple linear regression model has been applied. The research has led to several findings. The thermal response of the biobased bridge can be accurately predicted using temperature distribution using MLR. The largest prediction error found is 4.5%, which falls within the 5% range. However, this required a shift in the measured data since the measured strain data in the monitoring deviates from the measured strain data in the reference period at the same temperature. The first measured strain values in the monitoring period are set equal to the first predicted strain values in the monitoring period. The difference between these has been applied to all strain data per SG. This raises doubts about the accuracy of the model. The reason why the strain deviates cannot be concluded from this study.

Strain and temperature data can be used as input for the regression model when a reference period is chosen where the bridge behaves 'normally', and pre-processing is applied correctly.

The steps needed for appropriate pre-processing are choosing a reference period, remove data when the bridge is open, apply IQRA to detect and replace outliers, smooth the data, and remove incomplete data from the data set.

The optimal number of input measurements is different for each SG. It varies between taking every 2 measurement and every 256 measurements. The optimal number of input measurements within the range of taking each measurement and taking every 1000th measurement was examined. It can be concluded that with a number of input measurements greater than 500 for all SGs, the prediction error does not change much. The thermal inertia can also be used to improve the prediction error by means of an iterative process. The thermal inertia varies per SG and has an optimal value between 2 and 18 minutes.

Applying multiple linear regression results in appropriate prediction error results of the thermal response. Hence, MLR is a suitable algorithm to predict thermal response. However, it is important that no incorrect values are used with MLR. Data pre-processing must therefore be performed correctly. MLR can only be used when there is linearity. It should be considered carefully if this is the case.

Many questions remain after this research so there is plenty of room for further research. The model could be applied in the future once more data has been collected. Then it can be checked whether the prediction is still correct and whether changes have taken place in the behaviour of the bridge. The model can also be trained on another year to see if this reduces the prediction error. It is recommended to investigate the cause of the difference in measured strain data in the reference period and the monitoring period. It is also recommended to investigate why the regression model over or underestimates predicted strain values for some SGs. It is also possible to investigate by which factors SG4 is influenced, so that the prediction error found for the sensor is higher than other strain sensors. Research into the influence of factors such as temperature location, wind, humidity, and loads can be done to make the prediction error even more accurate and to better understand it.

7 References

Bharath, K. N., & Basavarajappa, S. (2016). Applications of biocomposite materials based on natural fibers from renewable resources: a review. Science and Engineering of Composite Materials, 23(2), 123–133. https://doi.org/10.1515/secm-2014-0088

Brownjohn, J. (2006). Structural health monitoring of civil infrastructure. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *365*(1851), 589–622. https://doi.org/10.1098/rsta.2006.1925

Catbas, F. N., Susoy, M., & Frangopol, D. M. (2008). Structural health monitoring and reliability estimation: Long span truss bridge application with environmental monitoring data. *Engineering Structures*, *30*(9), 2347–2359. https://doi.org/10.1016/j.engstruct.2008.01.013

Farrar, C. R., & Worden, K. (2006). An introduction to structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1851), 303–315. https://doi.org/10.1098/rsta.2006.1928

Frost, J. (2021, April 5). Overfitting Regression Models: Problems, Detection, and Avoidance. Statistics By Jim. https://statisticsbyjim.com/regression/overfitting-regression-models/

Google. (n.d.). [Google maps location of the bio-based bridge in Ritsumasyl]. Retrieved July 1, 2022, from https://www.google.nl/maps/@53.1023062,5.806599,9.5z?hl=nl

Hua, X. G., Ni, Y. Q., Ko, J. M., & Wong, K. Y. (2007). Modeling of Temperature–Frequency Correlation Using Combined Principal Component Analysis and Support Vector Regression Technique. *Journal of Computing in Civil Engineering*, *21*(2), 122–135. https://doi.org/10.1061/(asce)0887-3801(2007)21:2(122)

Kromanis, R., & Kripakaran, P. (2014). Predicting thermal response of bridges using regression models derived from measurement histories. *Computers & Structures*, *136*, 64–77. https://doi.org/10.1016/j.compstruc.2014.01.026

Kromanis, R., Kripakaran, P., & Harvey, B. (2015). Long-term structural health monitoring of the Cleddau bridge: evaluation of quasi-static temperature effects on bearing movements. *Structure and Infrastructure Engineering*, *12*(10), 1342–1355. https://doi.org/10.1080/15732479.2015.1117113

Kromanis, R., & Kripakaran, P. (2016). SHM of bridges: characterising thermal response and detecting anomaly events using a temperature-based measurement interpretation approach. Journal of Civil Structural Health Monitoring, 6(2), 237–254. https://doi.org/10.1007/s13349-016-0161-z

Lu, H., Gao, Z., Wu, B., & Zhou, Z. (2019). Dynamic and quasi-static signal separation method for bridges under moving loads based on long-gauge FBG strain monitoring. Journal of Low Frequency Noise, Vibration and Active Control, 38(2), 388–402. https://doi.org/10.1177/1461348418822375

M. (2021, October 1). 3 ways to make your Machine Learning projects successful. Reality AI. https://reality.ai/machine-learning-projects-3-ways-to-make-it-successful/

Posenato, D., Kripakaran, P., Inaudi, D., & Smith, I. F. (2010). Methodologies for model-free data interpretation of civil engineering structures. Computers & Structures, 88(7–8), 467–482. https://doi.org/10.1016/j.compstruc.2010.01.001

Ringnér, M. (2008). What is principal component analysis? Nature Biotechnology, 26(3), 303–304. https://doi.org/10.1038/nbt0308-303

Sala Lizarraga, J. M. P., & Picallo-Perez, A. (2020). Exergy analysis of heat transfer in buildings. Exergy Analysis and Thermoeconomics of Buildings, 263–343. https://doi.org/10.1016/b978-0-12-817611-5.00004-7

Sen, D., & Nagarajaiah, S. (2018). Data-Driven Approach to Structural Health Monitoring Using Statistical Learning Algorithms. Intelligent Systems, Control and Automation: Science and Engineering, 295–305. https://doi.org/10.1007/978-3-319-68646-2_13

Swanson, D. A. (2015). On the relationship among values of the same summary measure of error when used across multiple characteristics at the same point in time: an examination of MALPE and MAPE. Review of Economics and Finance, 5(1).

Sweco & Wittveen+Bos. (2020, August). *Biobased composiet fietsbrug Ritsumasyl*. Sweco Nederland B.v. and Witteveen+Bos Raadgevende Ingenieurs B.V.

Weedmark, D. (2019, March 2). The Advantages & Disadvantages of a Multiple Regression Model. Sciencing. https://sciencing.com/advantages-disadvantages-multiple-regression-model-12070171.html

Worden, K., & Manson, G. (2006). The application of machine learning to structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1851), 515–537. https://doi.org/10.1098/rsta.2006.1938