Structural health monitoring meets data mining

Miao, S.

Chapter 4

Sensor Dependencies among Multiple Sensor Types

With the rapidly decreasing prices for sensors, data gathering hardware and data storage, monitoring physical systems in the field is becoming a viable option for many domains. In fields such as civil engineering, windmills and aviation, so-called Structural Health Monitoring (SHM) systems are becoming popular to understand the actual workings of the system in situ, as well as to monitor the system for any developing faults. More and more, sensor networks consisting of multiple sensor types are being employed in these environments, and large quantities of data are being collected. New methods are required to deal with the proper analysis and interpretation of such data collections.

When dealing with multiple sensors measuring a physical system, each individual sensor will be sensitive to some aspects of the system, based on the specific characteristics of the type of sensor and on which part of the system the sensor is placed. This is clearly the case for sensors of different types (such as vibration and temperature sensors), but also for identical sensors attached differently to the system. If two sensors are measuring in each other's vicinity, they will likely show some dependency, but in most cases, this dependency will be non-trivial, depending on the location, the orientation and the attachment. As an example, consider an SHM-system employed on an aircraft. In order to measure stresses

on a wing, and potential metal fatigue on the wing attachment, strain gauges are fitted to the wing attachment. During high-g-force manoeuvres, the strain gauges will measure high values of strain on the attachment. Other sensors might be placed at the tip of the wing, to measure vibrations caused by turbulence for example. These vibration sensors however, will not be sensitive to sustained bending of the wing, as the sensor simply moves along with the wing, and is only sensitive to rapid changes in the location of the wing. As such, strain gauges are sensitive to different aspects of the dynamics than vibration sensors, although some overlap exists in the physical phenomena captured by either type.

In this chapter, we provide some examples of modelling the dependencies between (pairs of) sensors on the Hollandse Brug, specifically where multiple sensor types are involved. One of the main challenges here is to understand the specific focus of each sensor type, and to model any relationships across types. Having such a model may help, for instance, to remove certain phenomena measured by one sensor type from the signal of another sensor type. Specifically, we will consider the effect of temperature changes on the strain measurements at various locations on the bridge. As such, we can correct for this temperature effect.

Modelling dependencies between sensors also helps to remove redundancies in the data. Being able to infer the measurements of a particular sensor from the remaining sensor may suggest a smaller, and thus cheaper monitoring set-up. Finally, any modelling over the collection of sensors is beneficial for tracking the health of the bridge over longer periods. Changes in the value of a single sensor will often indicate transient effects, such as traffic or weather, but changes in the *models* of the bridge data indicate structural changes to the actual bridge, warranting further investigation.

A further issue we will be investigating is the effect that location and placement of sensors has on their usefulness within the network. For example, if we wish to understand the effect of temperature on strain measurements, it will be relevant to know where and how these two parameters are being measured. By investigating the dependencies between all pairs of sensors from two types (in this case strain and temperature), we hope to discover practical guidelines for the optimal placement of sensors. In Section 4.4, we use a meta-learning approach based on

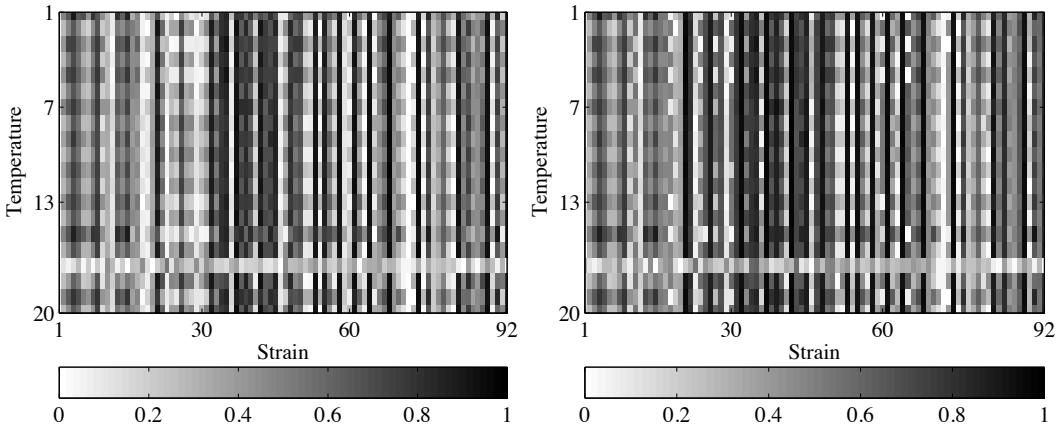

Figure 4.1: Correlation matrices between strain and temperature sensor types. - The numbers on the axes indicate the sensor number; the gray scale indicates the correlation between two sensors, 0 means no dependency, and 1 means strong dependency; St-Te before exponential moving average (left) and after exponential moving average (right).

subgroup discovery to find key characteristics of sensors in terms of their type, location, mode of attachment and orientation.

4.1 The Dependency between Strain and Temperature Sensors

In this section, we study the relationship between two types of sensor: strain and vibration. The sensor network features a total of 91 strain sensors, 44 of which are embedded (14 along X-axis, 28 along Y-axis), and 47 are attached (34 along X-axis, 13 along Y-axis). Of the 20 temperature sensors, one half is embedded in the surface of the deck, and the other half is attached to the underside of the deck.

In Fig. 4.1 on the left, the absolute correlation coefficients between strain and temperature vary from 0 to 0.97. For these sensor pairs with high correlation coefficients, we can simply employ a linear model that assumes the measured strain

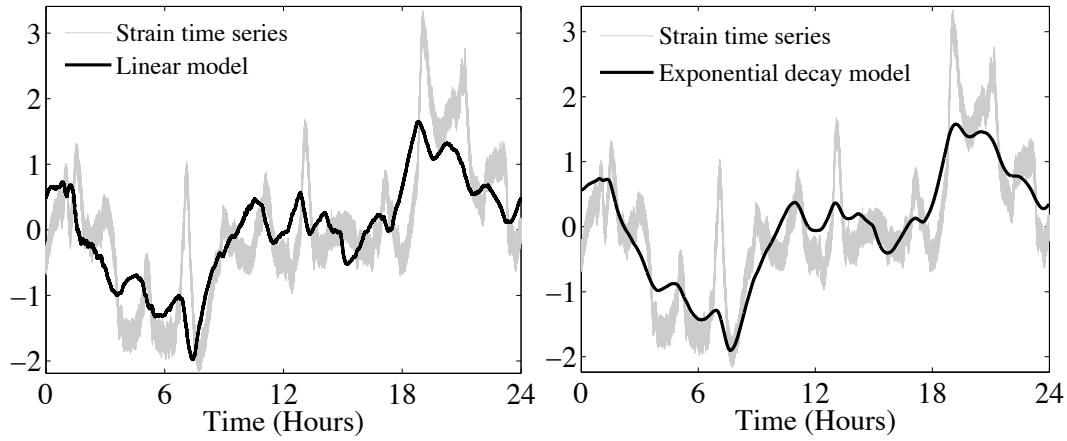

Figure 4.2: Models between strain and temperature time series. - The left picture shows the linear model between strain and temperature time series of 24 hours (100 Hz), in which there are a number of peak delays; The model in the right picture is obtained with an exponential decay model, which improves the delays in the linear model.

is directly influenced by the temperature of one of the temperature sensors:

$$S = a \cdot T + b \tag{4.1}$$

In this model, the coefficients a and b translate between the temperature scale (in Celsius) and the micro-strain scale (in $\mu m/m$). The left graph of Fig. 4.2 shows the effect of this model applied to two time series that are only moderately related, with $a = -3.288$ and $b = 27.547$ obtained through linear regression over a period of 24 hours. The correlation coefficient for this example is $r = 0.776$, which indicates that the selected pair of sensors are moderately correlated. However, when considering the time series in more detail, one can note that there is a dependency of the strain signal on the temperature measurements, but this relation is non-trivial: it involves a degree of delay: the upward and downward movement of the signal appear to be shifted by several hours.

The linear model fails to capture the complete effect of temperature on the strain, because the temperature sensor does not actually measure the bridge temperature, but rather the outside temperature. The temperature of the bridge is of

course mostly influenced by the outside temperature, but this influence is spread over time, and the bridge temperature will follow changes of outside temperature with a delay. The amount of delay depends on the size and material of the structure, with larger structures (such as the bridge in question) being less sensitive to sudden changes of outside temperature. In other words, a large concrete bridge has a large capacity to store heat, which is mirrored in a slow response of the strain signal.

In the systems analysis field, systems with a capacity are often modelled as a *Linear Time-Invariant* (LTI) system [22]. *Time-invariant* indicates that the response of the system does not change over time, which is a reasonable assumption for a bridge, if subtle deterioration of the structure is ignored. LTI systems are *linear* because their 'output' is a linear combination of the 'inputs'. In terms of the bridge, the temperature of the bridge is modelled as a linear combination of the outside temperature over a certain period of time (typically the recent temperature history):

$$T_{bridge}(t) = \sum_{m=0}^{\infty} h(m)T(t-m) \tag{4.2}$$

where $T_{bridge}(t)$ is the internal temperature and h is an impulse response (to be defined below). Note that this is a special case of convolution, which is defined in Chapter 2. Of the many impulse response functions h, which include for example the well-known moving average operation, we decide to model the delayed effect of the outside temperature using the exponential decay function $h_e(m) = e^{-\lambda m}$ (for $m \geq 0$). In this function, λ is the decay factor, which determines how quickly the effect of past values reduces with time. Note that the resulting equation

$$S = a \cdot h_e * T + b \tag{4.3}$$

where $h_e(m) = e^{-\lambda m}$, is the solution to a linear differential equation that is known as *Newton's law of cooling*, which states that the change in temperature of the bridge is proportional to the difference between the temperature of the bridge and its environment:

$$\frac{dT_{bridge}}{dt} = -r \cdot (T_{bridge}(t) - T(t)) \tag{4.4}$$

This is a somewhat simplified representation of reality, in that it assumes that the systems consists of two 'lumps', the bridge and the environment, and that within each lump the distribution of heat is instantaneous. Although in reality this is clearly not the case, it turns out that this model performs fairly well.

For a given pair of sensors and the associated data, we will have to choose optimal values for a, b and λ. It turns out that λ behaves very decently, with only a single optimum for given a and b, such that simple optimisation with a hill-climber will produce the desired result. For Equation 4.3, we obtain a fitted model for the selected sensor pair as shown in Fig. 4.2 on the right, which clearly demonstrates that the exponential decay model has removed the apparent delay in the data. The fitted coefficients were $a = -12.147, b = 30.463$, and $\lambda = 3 \cdot 10^{-5}$, with a correlation coefficient $r = 0.867$. Considering every possible pair of sensors from St and Te, we find that the correlation coefficients of 47.4% of sensor pairs are improved by the exponential decay model. Indeed, the successful modelling of the dependency for a given pair of sensors still depends on the location and placement of either sensor. In Section 4.4, we look into the question of finding suitable pairs of sensors in more detail, when we apply meta-learning to the modelling of St-Te sensor pairs.

4.2 The Dependency between Strain and Vibration Sensors

Our sensor network contains 34 vibration sensors, 15 of which are attached to the bridge deck, while the remaining 19 sensors are attached to the bridge girders. As mentioned in Chapter 3, both vibration and strain sensors are used to measure the dynamic stresses acting on the bridge. In theory, there should thus be some degree of correlation. However, we failed to detect a strong linear dependency between any pair. As illustrated in Fig. 4.3 (left), the correlations between most sensor pairs are quite weak, the highest one for this data being 0.1557. To demonstrate what types of modelling can be done for these two types of sensors, we selected one pair of sensors with a moderate correlation coefficient, as shown in the time

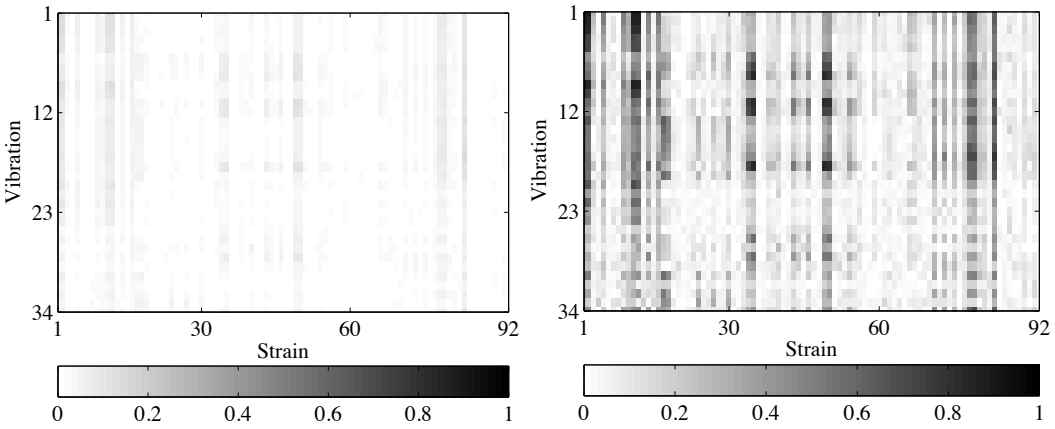

Figure 4.3: Correlation matrices between strain and vibration sensor types. - The correlation matrix in the left picture is obtained with raw strain and vibration time series of 5 minutes (100 Hz), in which the correlations are fairly weak; the correlation matrix in the right picture is obtained by applying a bandpass filter to the vibration time series, in which the correlations between several strain-vibration pairs are improved significantly.

domain in Fig. 4.4 (left top). The graphs show that the vibration sensor is a symmetric signal, while the strain sensor time series is not. However, the peaks in both occur consistently, which indicates that they are related. Using a simple correlation, this effect is hidden by the symmetric nature of the vibration signal.

In order to extract the amplitude of the vibration signal, which should correspond to the magnitude of the strain on the bridge, we apply an *envelope* operation. In the simplified situation where a signal consists of a single frequency s_f, modulated by another signal e as $s = s_f \cdot e$, we can simply obtain this envelope e by dividing s and s_f. However, in the presence of complex signals and noise, the envelope will have to be approximated by detecting peaks and interpolating between them. In our method, we define a peak as the maximum between two consecutive zero-crossings of the signal. To suppress the influence of noise, we first smooth the vibration time series with a small Gaussian kernel, and then interpolate adjacent peaks with piece-wise linear approximation. After being processed with the envelope operation, the vibration signal, shown in Fig. 4.4 (left bottom), shows

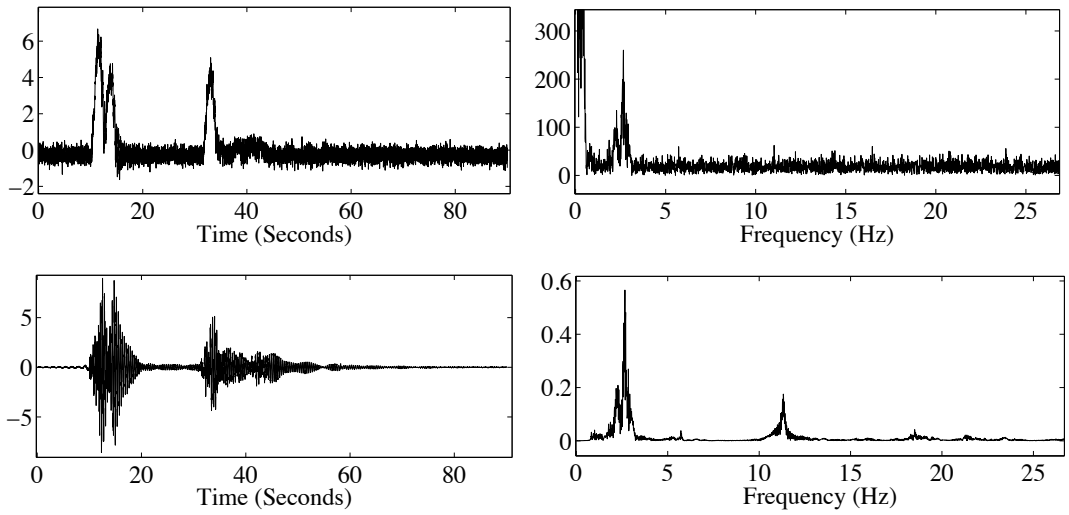

Figure 4.4: Strain and vibration time series in the time and frequency domain. - The left top picture is a strain time series of 90 seconds in the time domain; the right top picture is the spectrum of the strain time series; the left bottom picture is a vibration time series in the time domain; the right bottom picture is the spectrum of the vibration time series.

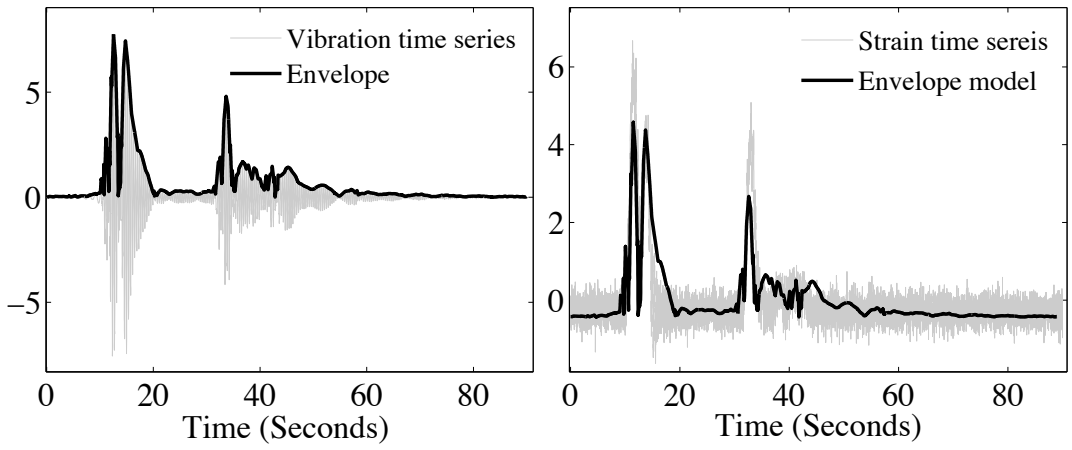

Figure 4.5: Envelope model - The envelope in the left picture is obtained by detecting peaks from vibration time series and interpolating them; the envelope model is obtained by shifting the envelope in the left picture by 105 data points, and then modelling it with the strain time series.

a better correlation with strain signal, improved from −0.16 to 0.44, as demonstrated in the left picture of Fig 4.5. By aligning the envelope, derived from the vibration time series, with the strain time series, we can obtain an improved envelope model, with a correlation of 0.80. We can also detect the dependencies between strain and vibration time series in the frequency domain.

Fig. 4.4 (right), which features the spectra obtained for the two signals, shows that despite a lack of a direct relation in the time domain, the signals are actually fairly similar in parts of the spectrum, notably where frequencies above 1 Hz are concerned. Note the big peak around 2.8 Hz in both spectra. In fact, what is missing in the vibration spectrum are the lower frequencies, which correspond to slower bridge movements. In other words, the vibration sensors are not sensitive to gradual changes in the deflection of the bridge, as the sensors themselves simply move along with the bridge. The strain gauges, on the other hand, *are* sensitive even to the slowest changes in bridge deflection. However, both sensors measure shaking of the bridge (frequencies above 1 Hz) in a similar fashion.

Based on these observations, an obvious way to relate St to Vi is to focus on a fairly specific range of frequencies. To obtain the boundaries of each well-matched spectrum range, we first transfer the spectrum of strain and vibration time series into one spectrum ratio signal by diving the strain spectrum with the vibration spectrum. Shown as the left picture in Figure 4.6, there are some "flat" ranges, which indicate that the spectral components of the strain and the vibration time series are similar during these ranges. We propose an approach to detect these flat spectrum ranges, based on sliding windows and the standard deviation. The standard deviation is used to measure the amount of variation from the average. A low standard deviation indicates a flat spectrum range. Given a time series $T = (t_1, t_2, \ldots, t_n)$, the standard deviation σ can be represented as:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - \mu)}, \text{ where } \mu = \frac{1}{n}\sum_{i=1}^{n} t_i$$

We employ a sliding window of a small length, which is 0.1 Hz (9 data points) in this work. Each sliding window can be featured with two parameters: the

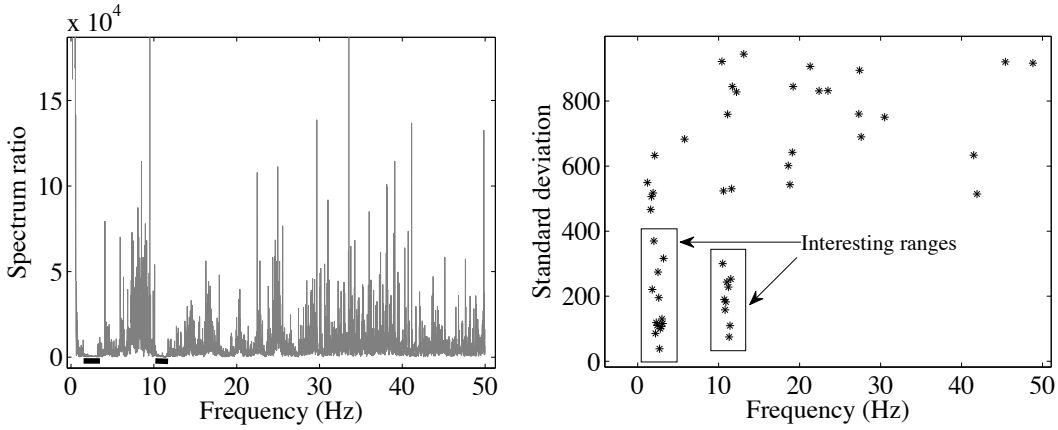

Figure 4.6: Interesting spectrum ranges detection - The left picture shows the spectrum ratio between the strain and the vibration time series. The right picture illustrates the spectrum ratio distribution, which is obtained by calculating the standard deviation of a sliding widow of 0.1 Hz (9 data points).

mean frequency and standard deviation. Given a vibration time series of 9,000 data points, we obtain 1,000 pairs of features. The scatter plot of these features is shown as the right picture in Fig. 4.6, in which there are two interesting spectrum ranges, whose standard deviations are below 400. The first spectrum range is between 2.0 Hz and 3.2 Hz, and the second one is between 10.5 Hz and 11.5 Hz.

In our experiments, we employ a *bandpass filter* to remove all spectral components outside the interesting ranges. The linear model between the strain and vibration time series then becomes:

$$\sum_{i=1}^{n} BPF_i(S) = a \cdot \sum_{i=1}^{n} BPF_i(V) + b \tag{4.5}$$

in which BPF_i stands for the band-pass filter operation on the ith frequency range, and n is the total number of interesting spectral ranges. After applying the band-pass filter operation to both St and Vi, the correlation coefficient improves from -0.16 to -0.90, as is shown in Fig. 4.7.

The model we achieved through the band-pass filter operation works well for a small selection of sensor pairs. In Fig. 4.3 on the right, information is displayed

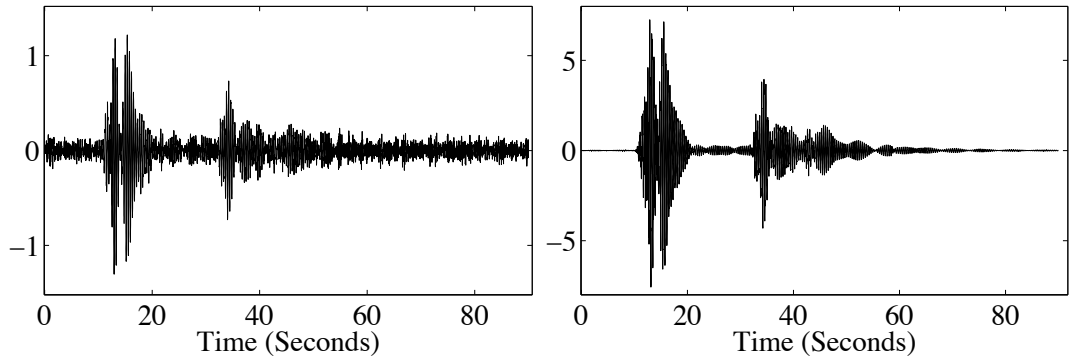

Figure 4.7: Bandpass filter model. - The left picture is the bandpass filter model derived from the strain signal. The right picture is the bandpass filter derived from the vibration signal.

on which sensor pairs specifically gain from this operation. Note that some strain gauges correspond well to most of the vibration sensors (dark columns in the matrix). These sensors are primarily located on the right-hand side of the bridge. The few exceptions are located on the girder entirely on the other side of the bridge. We look into such observations in more detail in the coming meta-learning section (Section 4.4).

4.3 The Dependency between Vibration and Temperature Sensors

As mentioned in the previous section, the vibration spectrum shows little activity in the range below 1 Hz, which happens to be where all of the temperature changes occur (for example due to the daily difference between day and night). For this reason, we do not expect significant dependencies between the sensors from Vi and Te. However, the vibration of the bridge does depend on the temperature. It is well known that bridges tend to oscillate at specific frequencies, and that these frequencies are determined by the stiffness of the structure, which in turn is influenced by changes in the temperature of the material. In a simplified model of a span of the bridge [38], the *natural frequency* f_n of the span is computed as

follows:

$$f_n = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{4.6}$$

In this equation, m refers to the mass of the bridge (including the possible load on the bridge), and k is a stiffness coefficient that depends on several factors such as material, humidity, corrosion, etc., but also on temperature. Note that an increasing temperature leads to a decreasing stiffness k, and hence a decrease in frequency, such that we expect a negative relationship between Vi and Te sensors.

The effect of temperature on natural frequencies is widely studied [39, 40]. After external excitation, for example traffic or wind, a bridge can vibrate in different *modes* [41]. Each mode stands for one way of vibration, which can be vertical, horizontal, torsional or more complicated combinations thereof, and there is one natural frequency corresponding to each. We will introduce a number of modal parameter extraction methods in Chapter 7, and look into the dependencies between temperature derived from temperature time series and frequencies derived from vibration time series.

4.4 Meta-learning

As mentioned at the ends of Section 4.1 and Section 4.2, we can accurately model some of the strain sensor signals using temperature sensor signals, and correlate some vibration sensors with strain sensors. However, the models we obtained are not universal for every pair of sensors. To further look into why some sensor pairs work well and others not, we analysed them in a meta-learning setting, which takes various sensor properties such as location and orientation into account.

Table 4.1: Example of the data that was used in the meta-learning experiment.

Strain							Temperature						corr.
sensor	x	y	embed.	orient.	lane	layer	struct.	sensor	x	y	embed.	lane	layer	struct.	
St1	14	0	attach	X-axis	right	girder	girder	Te1	13	7	embed	right	top	deck	0.139
St1	14	0	attach	X-axis	right	girder	girder	Te2	13	5	attach	right	bottom	deck	0.024
St1	14	0	attach	X-axis	right	girder	girder	Te3	9	7	embed	middle	top	deck	0.068
...															
St2	14	2	attach	X-axis	right	girder	girder	Te1	13	7	embed	right	top	deck	0.277
St2	14	2	attach	X-axis	right	girder	girder	Te2	13	5	attach	right	bottom	deck	0.472
...															

Table 4.2: The $d \leq 2$ results for the St-Te models ($\mu_0 = 0.533$).

Rules	Coverage%	Quality	Average
St vertical = inside deck & St horizontal ≤ 7	11.0	18.2	0.89
St vertical = inside deck & St orientation = Y-axis	9.9	17.8	0.90
St vertical = inside deck	16.5	16.1	0.79
St vertical = inside deck & Te horizontal ≤ 9	13.2	15.9	0.82
St vertical = inside deck & Te horizontal ≥ 5	13.2	14.1	0.79
St vertical = inside deck & Te embedding = attach	8.5	12.2	0.81
St vertical = inside deck & Te horizontal ≤ 5	6.6	11.2	0.82
St vertical = inside deck & Te embedding = embed	8.2	10.6	0.77
St embedding = embed	47.3	10.3	0.63

The software we used to conduct the meta-learning is called Cortana [42], which is a generic toolbox for Subgroup Discovery tasks, including the regression setting that is required here.

Table 4.1 shows the structure of our data in the meta-learning procedure. We represent each sensor pair and their properties, including the correlation of the best model, in one row. In the St-Te model, we have $91 \cdot 20 = 1,820$ rows, and $91 \cdot 34 = 3,094$ rows for the St-Vi model. The sensor locations are represented using x and y coordinates, but we also introduced several intervals in both dimensions to group sensors based on the part of the bridge they are placed.

St-Te models In meta-learning for the strain and temperature sensors, we take the absolute correlation value of each sensor pair as the primary target, and employ *z-score* as quality measure. The first 9 subgroups (sets of pairs of St-Te sensors) with search depth $d \leq 2$ are shown in Table 4.2. The average correlation over the entire set of pairs is $\mu_0 = 0.533$.

This table shows 2 subgroups of depth one and 7 subgroups of depth two. The depth-one subgroups indicate that the interesting vertical position for strain sensors is inside the deck, and that embedded strain sensors are influenced more than attached ones. The depth-two subgroups show more detailed information. For the strain sensors inside the deck, their interesting horizontal position is the

Table 4.3: The $d \leq 2$ results for the St-Vi models ($\mu_0 = 0.139$).

Rules	Coverage %	Quality	Average
St vertical = girder	17.4	31.3	0.36
Vi vertical = girder & St vertical= girder	10.2	28.0	0.40
St vertical = girder & St horizontal = right	6.5	24.8	0.43
St embedding = attach & St orientation = X-axis	38.0	17.7	0.21
St vertical = girder & Vi vertical = under deck	7.2	15.3	0.31
sensor = St1 & Vi vertical = girder	0.6	12.8	0.62
St vertical = girder & St horizontal = left	6.5	12.2	0.28
sensor = St83 & Vi vertical = girder	0.6	11.4	0.56
sensor = St11 & Vi horizontal = right	0.4	11.4	0.68

left side of the bridge, and the orientation is along the Y-axis. The attached temperature sensors perform slightly better than those embedded in the bridge. The horizontal position of the temperature sensors is also important, which indicates that the strain sensors inside the deck correspond well with temperature sensors located on the middle and left side of the bridge.

St-Vi models In meta-learning for the strain and vibration models, our target is the improvement of correlation after band-pass filtering. Table 4.3 presents the top-9 subgroups. From these results, it is clear that the interesting vertical locations for both strain and vibration sensors are the girders. For vibration sensors, placement under the deck is also interesting. The interesting horizontal locations for strain sensors are either side of the bridge, rather than the middle lanes. The strain sensors on the right-hand side measure more variation than those on the left. This is explained by checking the video stream: there was more traffic on the right lanes during this period of time. We also identified several specific strain sensors located on both sides of the bridge, which are consistent with the other subgroups in Table 4.3. Note that these selected sensors correspond to the three darkest columns in the correlation matrix in Fig. 4.3 (right).

4.5 Conclusion

We have demonstrated the use of a number of key data mining and signal processing techniques to model dependencies among multiple sensor types. We have built a linear model to correlate strain and temperature readings, and improved this model through convolution with an exponential response function. In the frequency domain, we used band-pass filters to detect the correlated spectra between strain and vibration sensor time series. Finally, we conducted meta-learning on the models obtained in Section 4.1 and Section 4.2, and extracted subgroups to explain the effects of sensor placement. The extracted rules can be used as guidelines for designing more (cost-)effective networks on future Structural Health Monitoring installations.