UNIVERSIDADE FEDERAL DE SANTA CATARINA CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

Felipe Carraro

Inverse problem regularization in Bridge Weigh In Motion systems

Florianópolis

2023

Felipe Carraro

Inverse problem regularization in Bridge Weigh In Motion systems

Tese submetida ao Programa de Pós-Graduação em Engenharia Civil da Universidade Federal de Santa Catarina para a obtenção do título de Doutor em Engenharia Civil.

Orientador: Prof. Rafael Holdorf Lopez, Dr.

Florianópolis

2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Carraro, Felipe
Inverse problem regularization in Bridge Weigh In Motion systems / Felipe Carraro ; orientador, Rafael Holdorf Lopez, 2023.
136 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Civil, Florianópolis, 2023.

Inclui referências.

1. Engenharia Civil. 2. bridge weigth in motion. 3. problemas inversos. 4. regularização. 5. soluções esparsas. I. Lopez, Rafael Holdorf. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Civil. III. Título.

Felipe Carraro

Inverse problem regularization in Bridge Weigh In Motion systems

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. André Teófilo Beck, Dr.
EESC - USP

Prof. Paulo Junges, Dr.
UNILA

Prof. Roberto Caldas de Andrade Pinto, Dr.
UFSC

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Engenharia Civil.

Documento assinado digitalmente
Philippe Jean Paul Gleize
Data: 15/05/2023 10:50:54-0300
CPF: ***.877.589-**
Verifique as assinaturas em https://v.ufsc.br

Coordenação do Programa de
Pós-Graduação

Documento assinado digitalmente
Rafael Holdorf Lopez
Data: 15/05/2023 09:42:44-0300
CPF: ***.262.979-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Rafael Holdorf Lopez, Dr.
Orientador

Florianópolis, 2023.

AGRADECIMENTOS

Agradeço inicialmente aos meus pais por sempre me apoiarem, dando suporte e incentivando os meus estudos.

Agradeço ao Matheus pela parceria de anos, desde o início da graduação até o fim do doutorado, sendo sempre um grande amigo, propondo ótimas discussões e me ajudando a resolver problemas de alta complexidade.

Agradeço à Graziele por ser uma pessoa maravilhosa, pela compreensão e ajuda que ofereceu, me auxiliando como podia ao longo de todo o percurso do doutorado.

Agradeço imensamente ao Professor Rafael, pela ótima forma de orientar, sempre conduzindo os processos de uma forma leve, divertida e amigável, nunca falhando em trazer possíveis soluções para quaisquer questões complicadas que surgiam.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e do Laboratório de Transportes e Logística (LabTrans).

SUMÁRIO

RESUMO

O sistema de pesagem em movimento em pontes (B-WIM) utiliza dados obtidos de sensores sob a estrutura da ponte de modo a avaliar as cargas dos veículos que passam. Ele tem sido aplicado em diversos locais pelo mundo. Os dados obtidos pelo sistema também conseguem prover informações adicionais para a avaliação de segurança da ponte e monitoramento da saúde estrutural.

Existem muitos métodos na literatura aplicados para a pesagem estática B-WIM. Este estudo inicialmente explora os procedimentos de solução existentes provendo uma comparação crítica de um subconjunto de métodos conceitualmente similares, expandindo a discussão em seus aspectos teóricos e conduzindo experimentos numéricos

Uma vez que a saída do sistema, que é a resposta da ponte, é aplicada para estimar as características do tráfego passante, que são as entradas do sistema, o problema pode ser caracterizado como sendo do tipo inverso. A teoria dos problemas inversos é explorada oferecendo um novo ponto de vista para o desenvolvimento de novos procedimentos de solução em sistemas B-WIM.

O estudo atual foca em abordar uma das maiores ausências dos métodos atuais da literatura: a aplicação da informação existente sobre as características do modelo de modo a obter novos procedimentos de solução. O estudo revisa uma interpretação de convolução do problema B-WIM, propondo soluções baseadas na estrutura existente. Além disso, ele remodela o problema da pesagem como um problema de deconvolução esparsa, desenvolvendo um conjunto integrado de métodos para a solução do problema de regularização e relaxando a restrição de detecção de eixos perfeita. Por fim, ele esclarece as relações de complementariedade entre pesagem e calibração propondo o uso de modelos que levem em conta os possíveis erros nas variáveis

As melhorias identificadas são combinadas em novo método de solução proposto que é testado sobre um banco de dados de simulações de sinais desenvolvido além de dados reais de uma ponte instrumentada. Os resultados sugerem que a abordagem é viável e efetiva, melhorando a performance de inferência da pesagem de eixos isolados ao passo que mantém uma performance similar ao método convencional com relação a precisão do Peso Bruto Total.

Palavras-chave: bridge weigth in motion. problemas inversos. regularização. soluções esparsas.

RESUMO EXPANDIDO

Introdução

Pontes são importantes estruturas em termos de deslocamento de recursos e operação de sistemas de transporte ao redor do mundo. No Brasil, muitas das pontes foram construídas na década de 70, sendo que muitas delas sofrem com falta de manutenção. Além disso, pode haver incerteza quanto suas características estruturais visto que os projetos originais podem não estar mais acessíveis. A incerteza quanto as características iniciais das pontes, aliadas ao aumento de tráfego e carga que vem ocorrendo ao longo dos anos e combinado com a falta de manutenção destas estruturas, compõe um cenário onde se torna interessante aplicar mecanismos que nos permitam monitorar a resposta aos esforços causadas nestas estruturas. Este monitoramento visa conhecer melhor as características da ponte e também traçar um histórico da evolução de sua resposta com o passar do tempo. Neste contexto, o sistema de pesagem em movimento em pontes (B-WIM) pode ser visto como uma ferramenta poderosa. Trata-se de um sistema que utiliza sensores sob a ponte e com base nos sinais obtidos fornece uma estimativa do espaçamento e peso de cada eixo de um veículo passando sobre ela.

O sistema B-WIM pode ser visto como um problema do tipo inverso. Em um sistema mecânico qualquer, observam-se as deformações a partir da aplicação de um carregamento. No caso do problema B-WIM, o que se busca é utilizar a leitura das deformações estruturais para se obter uma estimativa da forma e intensidade do carregamento que estava aplicado sobre a estrutura. Uma dificuldade inerente dos problemas do tipo inverso é que múltiplas configurações de forma e intensidade podem resultar em uma mesma deformação na estrutura. Neste sentido técnicas de regularização podem ser necessárias para que se obtenha uma resposta consistente.

De modo geral, o problema B-WIM pode ser visto a partir de ótica de um problema de deconvolução. Está forma distinta de ver o problema, proposta no trabalho de Frøseth *et al.* (2017) permite identificar uma certa estrutura fixa existente. Além da forma estrutural fixa existente na analogia da deconvolução existem outras possibilidades pouco exploradas na literatura para incorporar informações conhecidas para auxiliar na solução do problema. Uma delas é característica esparsa do sinal de um veículo. Um veículo pode ser representado por um vetor de impulsos, onde seu espaçamento e intensidade indicam o espaçamento de eixos e peso por eixo, respectivamente. Nesse contexto, tem-se que métodos de resolução que tentam promover a esparsidade na solução tendem a se aproximar mais da caraterística da resposta esperada. Uma outra característica que pode ser explorada pelos métodos de solução é não-negatividade do vetor que representa o veículo, por conta de o peso dos eixos não poderem ser negativos. Uma vez tratando-se de um problema inverso, torna-se interessante aplicar toda informação disponível de modo a guiar o processo de solução

para a solução desejada. Dentro do contexto da deconvolução, existe um paralelo com a teoria de processamento de sinais e imagens que pode ser utilizado para trazer técnicas de outros campos de pesquisa para a aplicação em engenharia.

Objetivos

Este estudo tem como objetivo propor um método de regularização para a solução do problema inverso associado ao método B-WIM, incorporando informações do sitema físico e características estruturais da modelagem do sistema mecânico.

De forma específica, o estudo visa sintetizar as contribuiçãoes de diferentes métodos B-WIM, implementando os métodos mais promissores e avaliando seu desempenho numérico. Além disso, busca-se expor uma análise unificada dos variados métodos relacionando suas propriedades e hipóteses numéricas.

Com relação as técnicas de solução do problema B-WIM, objetiva-se estudar os métodos de regularização para solução de problemas inversos. Adicionalmente, é dado foco para recuperação do vetor de impulso de forma completa, que nada mais é do que uma forma de representação do veículo que atravessa a ponte. Dentre os métodos de regularização buscas-se implementar o método de regularização ℓ_1 em função de sua propriedade de promoção da esparsidade que se adequa a forma do problema em questão. Além disso, busca-se incorporar informações adicionais disponíveis sobre o sistema tornanodo-o mais robusto a partir da modelagem de erro nas variáveis.

Metodologia

A partir de um estudo inicial, diversos métodos de solução de pesagem estática associados ao problema B-WIM são implementados. Os métodos são comparados dentro de uma abordagem unificadora, estabelencendo a relação teórica entre os diferentes métodos, suas semelhanças e hipóteses básicas. Além disso, os métodos são também comparados numericamente. Inicialmente, são aplicados para eventos simulados a partir de um modelo numérico, considerando a interação dinâmica entre a ponte e o veículo por meio de um sistema composto de massa, mola e amortecedor. Os parâmetros da simulação são escolhidos a partir de recomendações da literatura e caraterísticas usuais dos veículos reais que trafegam em pontes brasileiras. A resolução é feita por um precediemento iterativo de integração numérica. Adicionalmente, a comparação numérica se dá a partir de dados reais, obtidos de uma ponte intrumentada.

Posteriormente a análise inicial dos métodos é implementado uma caixa de ferramentas criada para a solução do problema de regularização ℓ_1. Este problema tem aplicação prática para B-WIM uma vez que fornece soluções contendo modelos esparsos.

Visando solucionar da melhor forma possível o problema inverso, um novo método é

proposto combinando diversas ideias da literatura com foco na incorporação de informações previamente conhecidas do problema. O método combina técnicas de regularização, a forma estrutural obtida da deconvolução e a consideração de erros nas variáveis em um processo de solução de duas fases. De modo a verificar o desempenho do método proposto, testes comparativos são conduzidos. Os testes compreendem tanto os dados numéricos simulados quanto os dados reais da ponte.

Resultados e Discussão

A comparação numérica direta entre métodos da literatura por si só pode ser vista como uma contribuição deste estudo. Além da comparação numérica, métodos de pesagem estática promissores foram analisados do ponto de vista teórico, o que rendeu a proposição de melhorias para alguns métodos como o pBWIM e considerações teóricas relevantes, como a de o Método 2D Modificado recair no método matricial no caso da correta estimação dos fatores de distribuição de carga. Em termos gerais a comparação numérica apresentou resultados similares, havendo pequena vantagem para o método da máxima verossimilhança (MLE).

Observar a interpretação da deconvolução com mais atenção permitiu que este estudo focasse com mais ênfase na recuperação do vetor de impulso como um todo, ao invés de se obter resultados dos eixos isoladamente, como usualmente é feito na literatura. Essa mudança de paradigma trouxe o problema B-WIM para o contexto de técnicas de regularização e problemas inversos. Dentro das técnicas de regularização, a regularização ℓ_1 foi vista como a mais aplicável, em função de promover a esparsidade da resposta. Ainda assim, existem diversas técnicas diferentes para a solução do problema ℓ_1. Esta falta de consenso sobre qual a melhor alternativa de solução propiciou outra contribuição relevante, a criação de uma caixa de ferramentas com múltiplos métodos para solução do problema de regularização ell_1. Explorando-se o paralelo com a literatura de processamento de sinais, foi possível buscar técnicas que incorporavam as características estruturais do problema, como a forma matricial da matriz de dados do problema B-WIM, caracterizada por possuir uma estrutura Toeplitz. Além disso, sabendo-se da complementaridade existente entre os processos de calibração e pesagem foi possível incorporar na proposta de solução a existência de incertezas nas variáveis de ambos os processos.

Foi proposto novo método de solução que faz uso das informações conhecidas do problema como esparsidade, estrutura e erros nas variáveis. O método apresentou desempenho satisfatório, com respostas semelhantes aos métodos atuais em termos de estimativa de Peso Bruto Total, porém com melhora do desempenho em termos da resposta por eixo. O método conta com uma fundamentação teórica mais robusta e pode permitir a estimativa da posição dos eixos diretamente a partir da recuperação do vetor de impulso completo, sem a necessidade de se instalar sensores específicos na ponte para este fim.

Considerações finais

O presente trabalho ofereceu um novo enfoque para o problema B-WIM, indicando novos caminhos de solução e possíveis considerações que podem ser incorporadas aos métodos de solução que até então não haviam sido profundamente exploradas. Foi possível contribuir de forma relevante para literatura especialmente considerando a proposição de um novo método de solução que visa oferecer mais robustez para as soluções, maiores garantias teóricas da forma da solução além de possibilitar a redução do número de sensores adotado.

Palavras-chave: bridge weigth in motion. problemas inversos. regularização. soluções esparsas.

ABSTRACT

Bridge weigh in motion (B-WIM) systems employ data gathered by sensors underneath the bridge structure in order to evaluate the loads of passing vehicles. It has been applied in multiple locations around the world. The data gathered by the system may also be able to provide additional information for accessing bridge safety and structural health monitoring.

Multiple methods exist in literature applied for B-WIM static weighing. This study initially explores existing solution procedures providing a critical comparison on a subset of conceptually similar B-WIM methods, further extending the discussion on their theoretical assumptions and conducting numerical experiments.

Since the output of the system, which is the bridge response, is applied for estimating the characteristics of the passing traffic, which are the input of the system, the problem can be characterized as being of the inverse type. The inverse problem theory is explored providing a new point of view for the development of novel solution procedures in B-WIM systems.

The present study focuses on addressing one the main absences of current methods in literature: the application of known information about the model characteristics in order to devise new solution approaches. The study reviews a convolution interpretation of the B-WIM problem, proposing solutions based on the existing underlying structure. Moreover, it recasts the weighing problem as a sparse deconvolution, develops a toolbox for the solution of the regularization problem and relaxes the constraint on perfect axle detection. Lastly, it enlighten the complimentary characteristics of both weighing and calibration proposing the use of models that take into account errors in variables.

The improvements identified are combined in a novel proposed solution method which is tested against a developed public dataset of simulation signals as well as real-world data from an instrumented bridge. The results suggest the approach is viable and effective, improving on single-axle inference performance while retaining similar Gross Vehicle Weight accuracy.

Keywords: inverse problems. bridge weigh in motion. regularization. sparse solutions.

LIST OF ABBREVIATIONS AND ACRONYMS

ADMM	Alternating Direction Method of Multipliers
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BP	Block Pivoting
B-WIM	Bridge Weigh in Motion
CD	Coordinate Descent
EBIC	Extended Bayesian Information Criterion
FAD	Free-of-axle detector
FFT	Fast Fourier Transform
FISTA	Fast Iterative Soft-Thresholding Algorithm
GVW	Gross Vehicle Weight
ISTA	Iterative Soft-Thresholding Algorithm
Lasso	Least absolute shrinkage and selection operator
MLE	Maximum Likelihood Estimate
OESOM	Orthantwise Enriched Second Order Method
SVD	Singular Value Decomposition
TLS	Total Least Squares

1 INTRODUCTION

Infrastructure constructions like bridges are usually designed to last decades. Much of the current bridges utilized in Brazil were constructed in the 70s and several of them suffer from lack of maintenance. Even more concerning, the original design plans may not be available. Therefore, given the natural degradation from constant use, the possible lack of corrective procedures and knowledge about its structural components, it becomes important to have a way to assess and monitor the bridge's behavior.

Bridge Weigh in Motion (B-WIM) consists in a system that transforms a bridge in a weighing mechanism. The implementation of such a system enables gathering data for several activities such as traffic studies, pavement design, overweight law enforcement, structural health monitoring and issuing Special Traffic Authorizations. The system offers advantages with respect to other WIM systems. It is portable and the sensors can be reallocated to other bridges. Moreover, the system does not disrupt traffic, since sensors are hidden from conductors, being located beneath the bridge slab. Figure 1 illustrates the strain signal obtained from the system.

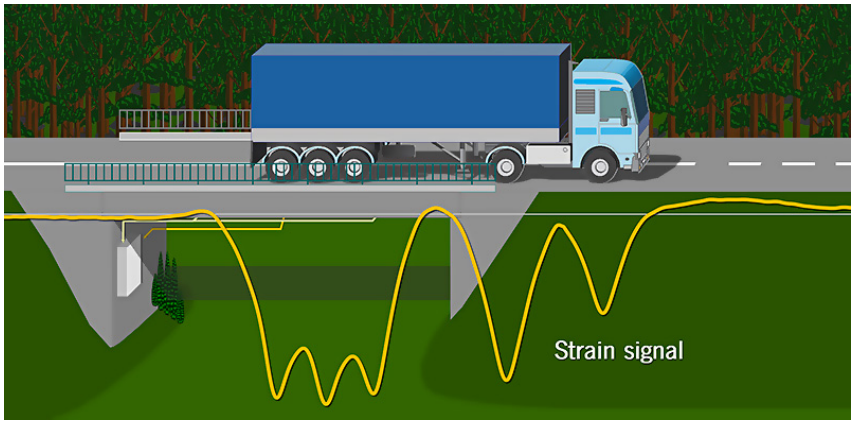


Figure 1 – B-WIM system signal acquisition

Source: (ISWIM..., 2019)

The B-WIM system synergizes well with several activities, allowing acquisition of information about the pavement, the passing vehicles, traffic characteristics and the structure itself in which it is installed. Considering how different studies and decision-making may be highly dependent on the data generated from the system, it becomes important to ensure an acceptable weight estimation performance. Measurements may suffer from systematic forms of error related to dynamic effects, temperature conditions, incorrect estimation of axle spacing, etc. Therefore, despite usually having plenty of data available, the system needs to account for the inherent uncertainties of measurement acquisition and the model employed to represent the system.

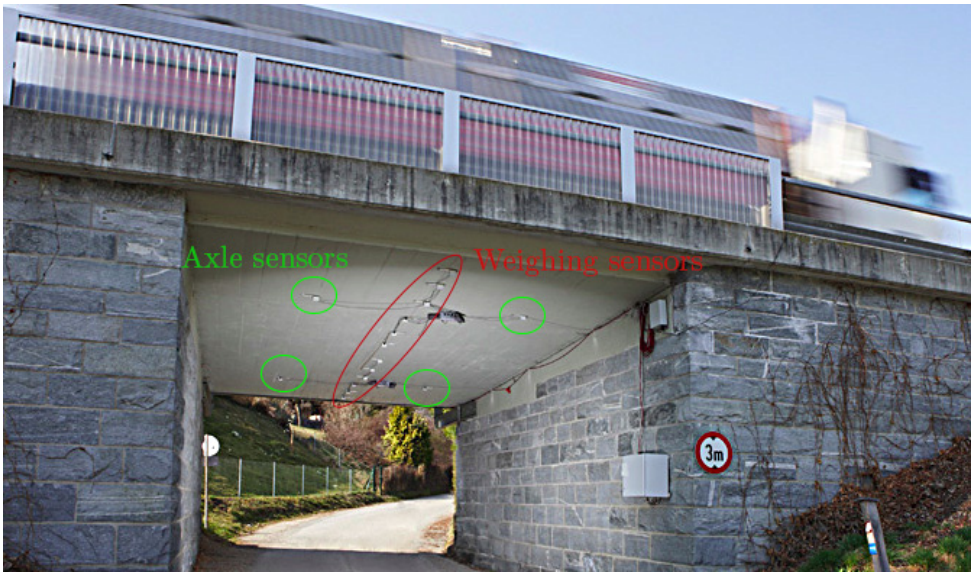


Figure 2 – B-WIM sensors

Source: Adapted from Road Traffic Technology (2019)

The use of the B-WIM system can be seen as an inverse problem. What one seeks are the inputs that generate a given process output. The goal is to infer the weight of a vehicle passing over the bridge given the measured strains and axle spacing detected by sensors during its traversal. Figure 2 illustrates the sensors installed beneath a bridge. An issue that may arise in this type of problem is the instability of solutions. There may exist multiple configurations of axle spacing and weights that can generate the same measured strains. Thus, it becomes important having ways for enforcing solutions forms that match what is expected from a real vehicle passing over a bridge, *e.g.* positive axle weight.

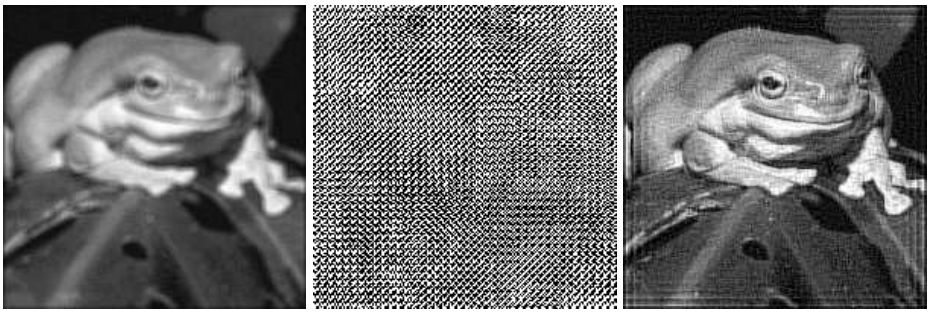
Different studies on the B-WIM estimation problem have been proposed in recent years. However, there is a lack of comparison between the novel approaches in literature. In addition to methods comparison from numerical experiments, an investigation from the theoretical point of view is missing. Considering that difference between some methods consists of relaxing a constraint or introducing a different model hypothesis or probabilistic/statistic assumption, a theoretical comparison can offer a more precise evaluation of the generalization capabilities of each method.

Regarding the computational point of view, a recent study has proposed another view on the B-WIM problem, establishing the relation between input and output of the system with a convolution operation. As a result of this interpretation, the direct/inverse problem can now be viewed as a convolution/deconvolution problem. This perspective may offer important insights on obtaining more efficient solutions to the problem as well as enabling the use of techniques often employed in the field of signal and image processing.

One of the techniques that shall be incorporated in this study is regularization. In order to overcome the issue of multiple valid inputs that generate the same output,

regularization allows for setting additional constraints to the problem. These additional constraints limit the possible solution forms in order to achieve more reasonable solutions given a certain metric.

With the convolution interpretation of the problem, regularization becomes even more important since it allows for additional constraints to be exploited. Moreover, it offers an interesting parallel with signal and image processing literature to be explored. Images are usually represented by matrices. Convolving the matrix representation with a particular kernel (matrix) results in a blurred image, *i.e.* the forward problem. However, it may be of interest the inverse problem, where one has a blurred image, which may also be corrupted from additional noise sources, and wants to recover the original image. As exemplified in Figure 3, simply performing the deconvolution on the matrix may result in an indistinguishable image. On the other hand, the application of regularization may largely aid the accuracy of the recovery.



blurred direct inverse regularized inverse

Figure 3 – Deconvolution of blurred image by direct inverse and regularization approaches

Source: Adapted from Fox *et al.* (2010)

Both the B-WIM weight estimation and the image restoration problem in Figure 3 can be viewed as a discrete inverse problem where the shift between input and output is aided by the convolution operator. Therefore, this originates the opportunity of bringing the tools already employed in other fields to help improving the solution of this practical engineering problem. Apart from the objective similarity, the employment of the convolution operator enables the solution of the problem with greater efficiency by exploiting the possibility of solving the problem in frequency domain.

1.1 OBJECTIVES

1.1.1 General Objective

Propose a regularization procedure for the solution of B-WIM inverse problem, exploiting the structure and prior (available) information of the system under analysis.

1.1.2 Specific Objectives

- Synthesize contributions from different B-WIM methods. Implement promising methods, assess their numerical performance and provide a unifying analysis of their theoretical assumptions and properties.
- Identify potential approaches exploiting the problem structure and prior knowledge based on the comparative study;
- Study the use of regularization techniques for the solution of the B-WIM inverse problem;
- Focus on the recovery of the full impulse vector which represents a vehicle over the bridge;
- Implement ℓ_1 regularization as an alternative to promote sparseness of the solution;
- Verify how to incorporate additional information to the model, such as accounting for errors in variables.

1.2 THESIS STRUCTURE

Chapter 2 concisely describes the relevant theoretical background that is employed for describing the problem and its possible solution procedures.

In Chapter 3 the problem being addressed is described and detailed. In addition, the scope of the study is delimited, with the discussion of possible solution alternatives for the problem.

In Chapter 4 the main B-WIM solution methods are presented under a common knowledge base. The methods are discussed highlighting their conceptual similarities and differences.

In Chapter 5 numerical experiments are conducted in order the evaluate the presented B-WIM solution procedures and access the validity of the theoretical findings discussed in the previous chapter.

Chapter 6 introduces the main ideas that arose from analyzing and evaluating the proposed subset of B-WIM solution methods. In essence, the ideas aim to better exploit the available knowledge of the problem.

Chapter 7 introduces the novel proposed method employed for accounting for the available information from the B-WIM model. Moreover, numerical experiments are presented evaluating the method in both simulated and real-world data.

Chapter 8 draws concluding remarks about the topic studied and the analyses presented.

Chapter 9 lists potential research lines further developing the novel ideas presented in this study.

In Appendix A is attached the research paper that resulted from initial part of the conducted studies.

2 PRELIMINARY THEORY

In order to better understand how different approaches can be employed in the B-WIM problem discussed in chapter 1, some initial theoretical remarks should be presented first. This chapter presents the theoretical preliminaries and important definitions that allow further discussion of the proposed study in the context of the B-WIM system in chapter 3.

2.1 INVERSE PROBLEMS THEORY

Inverse problems may arise when trying to infer a cause from an effect. In a generic evaluation procedure, one has an input that is directed through a process to obtain an output. The inverse problem occurs when one seeks the input that generates the observed output given a certain process, as illustrated in Figure 4.

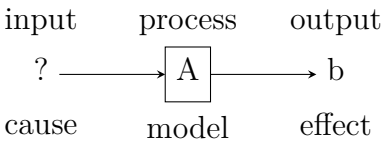


Figure 4 – Inverse problem

In general, inverse problems are more complex than their direct counterpart, since they are usually ill-posed. A well-posed problem as defined by Hadamard (1902) satisfies the following conditions:

- Existence: a solution for the problem exists for every input;
- Uniqueness: the solution is unique;
- Stability: the solution depends continuously on the data.

If any of the above conditions are not satisfied, the problem is said to be ill-posed.

Given the possible ill-posedness of inverse problems, distinct solution techniques from their direct counterpart should be applied. Figure 5 illustrates the general frameworks employed in solving inverse problems. Techniques usually follow two main paths, employing either regularization approaches or making use of Bayesian reasoning. In both cases, one tries to provide additional problem information to guide the solution procedure.

Following the regularization framework, methods could be divided into two main categories: Penalization and Subspace projection. In the Penalization approach, additional constraints are added to solution form in order to penalize solutions that do not conform

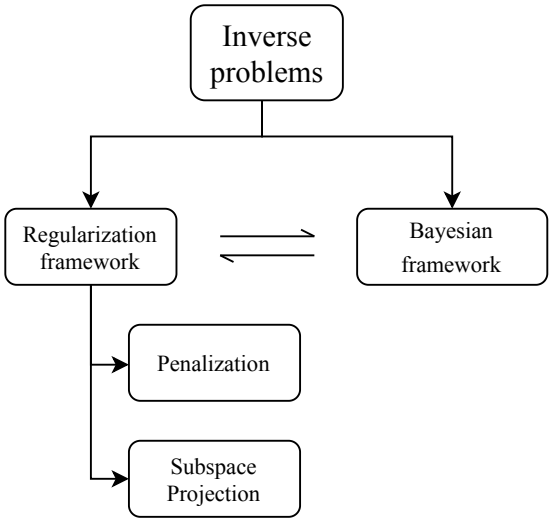


Figure 5 – Alternatives for inverse problem solution

with what is expected (THEODORIDIS, 2015). This branch comprises regularization methods such as Tikhonov, Lasso, Elastic Net and Regularized Least Absolute Deviations (WANG *et al.*, 2006). The other main branch in regularization approaches is that of Subspace Projection (JENSEN, 2006). It consists of finding solutions by projection on lower-dimensional subspaces, offering desirable properties. This achieves a regularizing effect which may be sufficient for many applications (KILMER; O'LEARY, 2001). Such methods, usually offer dimensionality reduction that helps when dealing with high-dimensional problems. Instances of such methods include Truncated Singular Value Decomposition and the family of Krylov subspace methods, which originates many iterative algorithms such as Conjugate Gradient (HESTENES; STIEFEL, 1952), LSQR (PAIGE; SAUNDERS, 1982) and Golub-Kahan iterative bidiagonalization (GOLUB; KAHAN, 1965).

It is also possible to apply Bayesian reasoning for solving inverse problems. In fact, there is much interplay between Regularization and Bayesian approaches, where several Regularization procedures have equivalent interpretations under a Bayesian reasoning. For example, a L1 regularization can be viewed as the maximum *aposteriori* estimate given an Laplacian prior (HASTIE *et al.*, 2015). The advantages of the Bayesian approach are the possibility of better incorporating the uncertainties on the model and the possibility of recovering not only a single value but a related probability distribution of the results (GONÇALVES *et al.*, 2022). The drawback is often the computational cost associated with the estimation procedure. As examples of studies employing this approach one could cite (TITTERINGTON, 1985; ARCHER; TITTERINGTON, 1995; FLORENS; SIMONI, 2012).

2.2 CONVEXITY

Convexity is an important characteristic of an optimization problem. Knowing beforehand that a function is convex enables the use of specialized methods that exploit this property. Thus, the search becomes faster if compared to more general methods and the solution is guaranteed to be a global optimum when a local search method such as gradient descent is employed (WRIGHT *et al.*, 1999).

In order to define a convex function, the definition of a convex set is required. A set $\mathbb{S}$ is said to be convex if given any two points p_1 and p_2 in $\mathbb{S}$, the line segment $\overline{p_1 p_2}$ is also in $\mathbb{S}$. This concept can be extended for the n-dimensional space. Mathematically, a parametric representation of a line segment between points $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$ can be formulated as follows:

$$\mathbf{d} = \alpha \mathbf{d}^{(2)} + (1 - \alpha)\mathbf{d}^{(1)};\ 0 \leq \alpha \leq 1. \tag{2.1}$$

If the entire line segment is in $\mathbb{S}$, then it is a convex set (ARORA, 2004).

Considering now a function $f(\mathbf{d})$ defined on a convex set $\mathbb{S}$, this function is said convex if it satisfies:

$$f(\alpha \mathbf{d}^{(2)} + (1 - \alpha)\mathbf{d}^{(1)}) \leq \alpha f(\mathbf{d}^{(2)}) + (1 - \alpha) f(\mathbf{d}^{(1)});\ 0 \leq \alpha \leq 1. \tag{2.2}$$

This condition is necessary and sufficient and applies to n-dimensional functions. An illustrative example of a single variable convex function is shown in Figure 6.

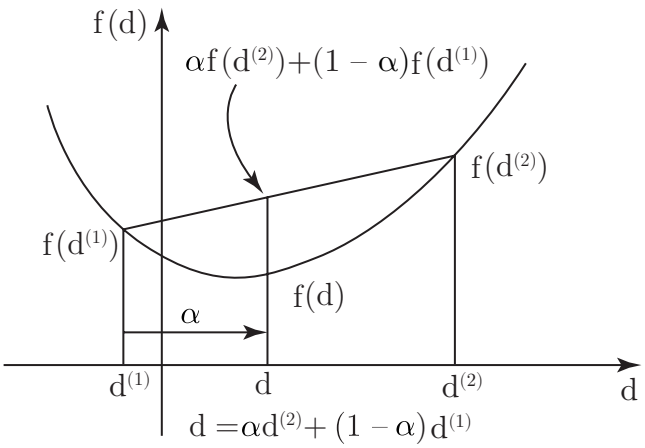


Figure 6 – Convexity check on unidimensional function

Source: Adapted from Arora (2004)

In practice, applying Equation 2.2 is difficult, as an infinite number of pair of points must be checked. Checking the Hessian of the function is a simpler alternative. A function

is said to be convex when its Hessian $\nabla^2 f(\mathbf{d})$, defined as:

$$\nabla^2 f(\mathbf{d}) := \begin{bmatrix} \dfrac{\partial^2 f(\mathbf{d})}{\partial d_1^2} & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_1 \partial d_2} & \cdots & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_1 \partial d_n} \\ \dfrac{\partial^2 f(\mathbf{d})}{\partial d_2 \partial d_1} & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_2^2} & \cdots & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_2 \partial d_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(\mathbf{d})}{\partial d_n \partial d_1} & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_n \partial d_2} & \cdots & \dfrac{\partial^2 f(\mathbf{d})}{\partial d_n^2} \end{bmatrix}, \tag{2.3}$$

is at least positive semidefinite everywhere, that is, has non-negative eigenvalues for all points in $\mathbb{S}$.

2.3 ORDINARY LEAST SQUARES

Data points from input-output systems may be corrupted by imperfect measurements that include random error. A linear least squares problem is usually the starting point for many modeled systems in which the presence of data errors is admitted. The general formulation can be written as:

$$\arg\min_{\boldsymbol{x}} \boldsymbol{r}^T \boldsymbol{r}; \quad \boldsymbol{r} = A\boldsymbol{x} - \boldsymbol{b} \tag{2.4}$$

where A is the design matrix, $\boldsymbol{r}$ is the vector of residuals, $\boldsymbol{b}$ is the vector of observations and $\boldsymbol{x}$ is the vector of parameters. Since the sum of squared residuals term $\boldsymbol{r}^T\boldsymbol{r}$ is usually different from zero, one seeks the solution $\boldsymbol{x}$ which minimizes this metric. The solution can be found analytically, resulting in the well known normal equations (HANSEN *et al.*, 2013):

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b}, \tag{2.5}$$

where the value of $\boldsymbol{x}$ that minimizes Equation 2.4 is found to be:

$$\boldsymbol{x} = (A^T A\boldsymbol{x})^{-1} A^T\boldsymbol{b}. \tag{2.6}$$

2.4 MAXIMUM LIKELIHOOD ESTIMATION

Given any machine learning context where predictions must be made, maximum likelihood estimation (MLE) becomes an important tool for the development of solution procedures. It consists of a general method that can be applied to any estimation problem where a joint probability density function can be assigned to the observations (ASTER *et al.*, 2018).

The likelihood function associated with a certain joint probability distribution can be interpreted as a function of the distribution parameters given the observations. Therefore, when employing MLE, one wants to find from all parameters that could define a model distribution, those which are most likely to have generated what has been observed. This concept is relevant for this study since it allows for different formulations for equating input and output. Moreover, assuming distinct forms for the distribution of errors between observation and expected results enables the development of more general methods.

When trying to derive an estimator for the system given these data, one may seek the solution which is best from a statistical point of view. In this regard, maximum likelihood is one of the most popular techniques for deriving estimators (CASELLA; BERGER, 2002).

The error function in Equation 2.4 defines a least square problem. From a statistical point of view, such an approach is the maximum likelihood estimator when errors are independent and normally distributed random variables (ASTER *et al.*, 2018). These errors are related to each measured ordinate and at the same event of calibration. However, these underlying error assumptions may not be met when considering practical cases. For example, missing information on formulation could be seen as correlated errors (WASHINGTON *et al.*, 2010). Thus, if the model description in Equation 2.4 is not accurate enough, a least squares estimate would not provide the best result. However, useful information can still be extracted from the application of the method, although there are minor violations of the assumptions (CHATTERJEE; HADI, 2015).

2.5 ℓ_p NORMS

Different vector norms are employed in this study. Their concept can be generalized with the introduction of the ℓ_p norm, defined as:

$$\|\boldsymbol{x}\|_p = \sum_{i=1}^{\infty} |x_i^p|^{\frac{1}{p}}. \tag{2.7}$$

Figure 7 illustrates the different contours generated by employing different p parameters in the ℓ_p norm.

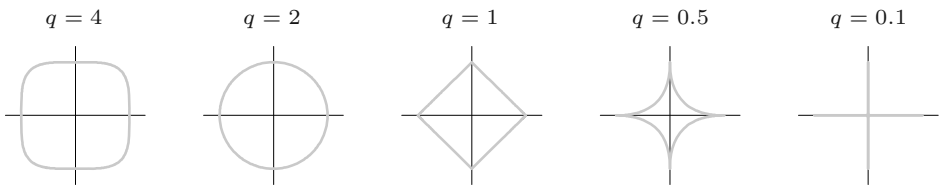


Figure 7 – Contours of different ℓ_p norms

Source: (HASTIE *et al.*, 2009)

This measure is relevant for this study since both of most usual forms of regularization procedures employ the ℓ_p norms in their formulation. Tikhonov regularization, as shall be discussed, is defined according to the ℓ_2 norm, while Lasso regularization is defined according to the ℓ_1 norm.

Considering the B-WIM context, another relevant norm, named the infinity norm (ℓ_∞), is also employed as a manner for comparing different distances between influence line curves. It can be defined as an asymptotic case of the ℓ_p norm:

$$\|\boldsymbol{x}\|_\infty = \lim_{p\to\infty} \|\boldsymbol{x}\|_p = \max(\boldsymbol{x}) \tag{2.8}$$

Additionally, when studying systems involving matrices, the Frobenious norm is often considered. It can be seen as the equivalent of the ℓ_2 vector norm for the space of matrices. It is defined as the square root of the sum of the squares of all matrix entries. For a m by n real matrix A that can be written as:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} = \sqrt{\text{trace}(A^T A)}. \tag{2.9}$$

The relation with the ℓ_2 norm becomes more evident after defining the Frobenius inner product:

$$A : B = \text{trace}(A^T B). \tag{2.10}$$

This operation works on the space of matrices and follows all requirements to be considered an inner product such as linearity, symmetry and positive definiteness. Therefore, the two expressions:

$$\|\boldsymbol{x}\|_2^2 = \boldsymbol{x} \cdot \boldsymbol{x}, \tag{2.11}$$

$$\|X\|_F^2 = \boldsymbol{x} : \boldsymbol{x}, \tag{2.12}$$

are the norms induced by the inner products of the vector space, known as the dot product and the inner product of the matrix space, respectively.

2.6 ℓ_2 REGULARIZATION

In ℓ_2 regularization, additionally to the error term, the ℓ_2-norm of the solution is also minimized by incorporating a parameter λ. It can be viewed, from an optimization perspective, as an additive penalty term to the residual function. This inclusion imposes a new constraint, where not only the ℓ_2-norm of the residual should be minimized but also the ℓ_2-norm of the solution. The trade-off between the optimization of these two quantities is controlled by the parameter λ, which can be viewed as a Lagrange multiplier. (ARORA, 2004).

Tikhonov regularization is perhaps one of the most widely used techniques for regularizing discrete ill-posed problems (ASTER *et al.*, 2018). One the main difficulties when employing this type of procedure is how to reliably estimate the optimal λ parameter. The usual formulation reads:

$$\hat{\boldsymbol{x}}_{\ell_2} = \arg\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2. \tag{2.13}$$

From Equation 2.13 one can derive the argument that minimizes this function as a function of λ by setting the gradient to zero. Taking the differential of the previous expression, one can identify the gradient as in:

$$f = \|A\boldsymbol{x}_\lambda - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{x}_\lambda\|_2^2 = \boldsymbol{r} : \boldsymbol{r} + \lambda\boldsymbol{x}_\lambda : \boldsymbol{x}_\lambda \tag{2.14}$$

$$df = 2\boldsymbol{r} : d\boldsymbol{r} + 2\lambda\boldsymbol{x}_\lambda : d\boldsymbol{x}_\lambda \tag{2.15}$$

$$= 2(A\boldsymbol{x}_\lambda - \boldsymbol{b}) : Ad\boldsymbol{x}_\lambda + 2\lambda\boldsymbol{x}_\lambda : d\boldsymbol{x}_\lambda \tag{2.16}$$

$$= 2(A^T(A\boldsymbol{x}_\lambda - \boldsymbol{b}) + \lambda\boldsymbol{x}_\lambda) : d\boldsymbol{x}_\lambda \tag{2.17}$$

$$\frac{\partial f}{\partial \boldsymbol{x}_\lambda} = 2(A^T(A\boldsymbol{x}_\lambda - \boldsymbol{b}) + \lambda\boldsymbol{x}_\lambda). \tag{2.18}$$

where in the above derivation, the colon denotes the Frobenius inner product, *i.e.* $A : B = \text{trace}(A^T B)$, with the following properties derived from the underlying trace function:

$$\begin{aligned} A : BC &= B^T A : C \\ &= AC^T : B \\ &= A^T : (BC)^T \\ &= BC : A. \end{aligned} \tag{2.19}$$

Setting the gradient in Equation 2.18 to zero and solving for $\boldsymbol{x}_\lambda$ one obtains:

$$0 = 2(A^T(A\boldsymbol{x}_\lambda - \boldsymbol{b}) + \lambda \boldsymbol{x}_\lambda \tag{2.20}$$

$$A^T\boldsymbol{b} = (A^TA + \lambda I)\boldsymbol{x}_\lambda \tag{2.21}$$

$$\boldsymbol{x}_\lambda = (A^TA + \lambda I)^{-1}A^T\boldsymbol{b}. \tag{2.22}$$

It can be noted that the main difference from the usual normal equations that arise from the application of least squares is the added regularization term in the diagonal of the A^TA matrix.

2.7 SUBGRADIENTS AND SUBDIFFERENTIAL

When applying optimization procedures on convex nonsmooth functions the concept of subgradient becomes important. When gradients are available, the gradient descent method converges quadratically to a local solution, which is also the global solution for convex functions. However, for nonsmooth functions the gradient may not exist everywhere. To solve the issue subgradients can be seen as a generalization of gradients applicable to nonsmooth functions. The subgradient $\boldsymbol{s}$ is defined as a vector that satisfies the relation:

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \mathbf{s}^T(\boldsymbol{y} - \boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom} f(\boldsymbol{x}). \tag{2.23}$$

Then, the set of all subgradients of a function f at a point $\mathbf{x}$ is the subdifferential, denoted $\partial f(\mathbf{x})$ and written as:

$$\partial f(\boldsymbol{x}) = \{\boldsymbol{s} | f(\boldsymbol{y}) \geq f(\boldsymbol{x})\boldsymbol{s}^T(\boldsymbol{y} - \boldsymbol{x}), \forall y \in \mathrm{dom} f\}. \tag{2.24}$$

An important property for optimization is that $\boldsymbol{x}$ is a minimum of the convex function f if and only if f is subdifferentiable at $\boldsymbol{x}$ and:

$$0 \in \partial f(\boldsymbol{x}). \tag{2.25}$$

2.8 PROCEDURES FOR FINDING REGULARIZATION PARAMETERS

2.8.1 L-curve

A trade-off exists between point-wise fidelity and the generalization introduced by smoothing the solution given a certain regularization parameter. The L-curve method tries to locate the optimal parameter employing a graphic procedure which relates these two situations. It plots on a log-log scale the least-squares gap, *i.e.* the norm of the regularized solution versus the norm of the corresponding residual. Calling the residual norm as E_{norm} and the regularized solution norm as F_{norm}, the quantities plotted in the L-curve are:

$$E_{norm} = \|A\boldsymbol{x}_\lambda - \boldsymbol{b}\|_2 \tag{2.26}$$

$$F_{norm} = \|\boldsymbol{x}_\lambda\|_2. \tag{2.27}$$

When a small regularization is introduced, the total error is dominated by the noise error. This situation is called under-smoothing and corresponds to the vertical part of the L-curve. On the other hand, when the amount of regularization introduced is large, the total error becomes more influenced by the smoothing error. This originates the horizontal part of the L-curve (DOICU *et al.*, 2010). Therefore, in the L-curve method one seeks the best trade-off between smoothing and noise error, which corresponds to the corner of the L-curve.

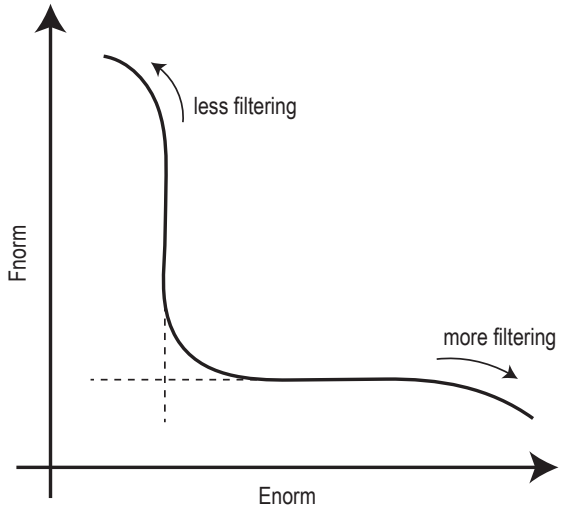


Figure 8 – The generic form of the L-curve plotted in double-logarithmic scale, adapted from: (HANSEN, 1998)

The main issue of depending on the L-curve for the assessment of the optimal regularization parameter is the lack of a standard procedure for obtaining the corner of the curve. One possible systematic way of evaluating this corner is by minimizing the Euclidean distance from the origin of the graph. Another similar technique, proposed by Hansen e O'Leary (1993), seeks the point that minimizes the L-curve curvature. Nevertheless, when dealing with the discrete case, more robust methods are needed since corners may appear. In this regard Hansen *et al.* (2007) proposes a technique for addressing this issue with an adaptive pruning procedure.

2.8.2 Cross-validation

The basic assumption behind cross-validation is that a solution computed on a reduced set of data points should give a good estimate of missing points. Thus in general,

one aims to verify the suitability of a parameter by employing it on different partitions of the same training set. Different techniques arise from different forms of partitioning the training set. Two usual forms are Leave One Out, where each partition contains all but one element and K-Fold Validation, where K represents the partition size (HASTIE *et al.*, 2009).

For the application on ℓ_2 regularization, Golub *et al.* (1979) proposes a form of generalized cross-validation (GCV). This method requires the minimization of the GCV function:

$$V(\lambda) = \frac{\|(I - A_\lambda^\#)\boldsymbol{b}\|_2^2}{(\mathrm{tr}(I - A_\lambda^\#))^2}, \tag{2.28}$$

where A_λ for Tikhonov regularization is:

$$A_\lambda^\# = (A^T A + \lambda I)^{-1} A^T. \tag{2.29}$$

2.8.3 Information-based methods

Another approach for choosing the optimal regularization parameter is a class of methods based on information theory. The advantage of these methods is that they are much cheaper computationally when compared to data-driven methods such as cross-validation.

The Akaike Information Criterion (AIC) is a model selection criterion derived by Akaike (1998) which aims to find the best approximate model to the unknown true data generating process. The metric for evaluating the approximation quality is the Kullback–Leibler (KL) divergence. Since the true model is unknown, the derivation relies on some asymptotic identities (WIT *et al.*, 2012). The derivation results in the following criterion when applied to the lasso problem:

$$\lambda_{AIC} = \arg\min_\lambda (n \log(SSE_\lambda) + 2s) \tag{2.30}$$

where SSE_λ is the sum of squared errors associated to the considered λ, namely $\|b - Ax_\lambda\|_2^2$, and s in the cardinality of the support, *i.e.*, the number of nonzero elements in the solution.

Another popular criterion in called the Bayesian Information Criterion (BIC), proposed by Schwarz (1978). In the Bayesian approach of model selection one aims to maximize the posterior probability of a model given the data. Assuming all models are equally likely maximizing the posterior probability of a model given the data is the same as maximizing the marginal likelihood:

$$P(b_1, ..., b_n | M_i) = \int_{x_i} L(x_i | b_1, ..., b_n) pdf(x_i) dx_i \tag{2.31}$$

where L is the likelihood function, x_i are the model parameters associated with model M_i and $pdf(x_i)$ is the probability density function of the model parameters. BIC uses Laplace method to approximate the integral and further assumptions of Independent and identically distributed errors as well as big n in order to simplify the expression into the following criterion (BHAT; KUMAR, 2010):

$$\lambda_{BIC} = \arg\min_{\lambda}(n\log(SSE_\lambda) + s\log n) \tag{2.32}$$

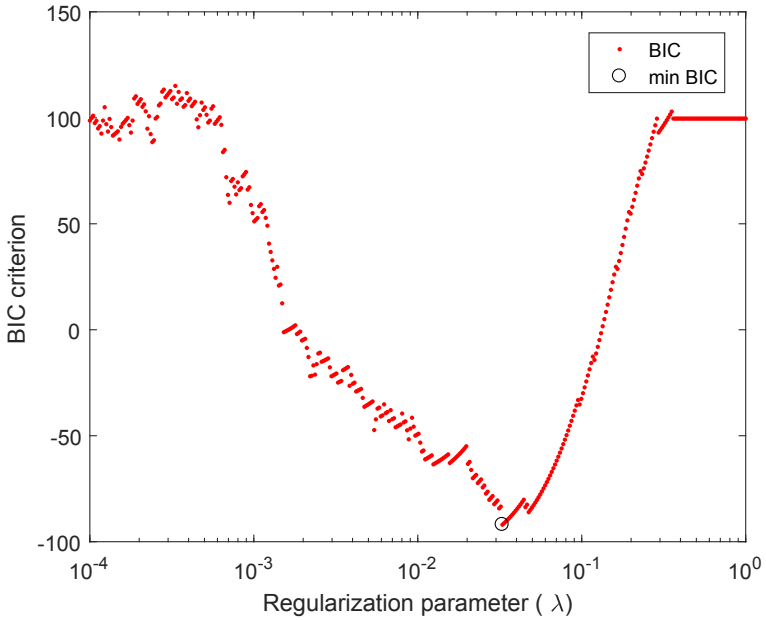


Figure 9 – Example BIC curve and its minimum

There is also the EBIC method proposed by Chen e Chen (2008) which is a extension of the classic BIC. It is aimed at correcting a tendency of exceeding number variables entering the model in high-dimensional problems. This increased sparseness is achieved by by adding an additional penalty term on the BIC for the growing number of parameters in the model. It is defined as:

$$\lambda_{EBIC} = \arg\min_{\lambda}(n\log(SSE(\lambda)) + s(2\log p + \log n)). \tag{2.33}$$

2.9 ℓ_1 REGULARIZATION

Another important regularization procedure is the LASSO (least absolute shrinkage and selection operator). This procure is also known as ℓ_1 regularization since it employs the ℓ_1 norm in the formulation.

The major aspect of this regularization procedure regarding this study is the characteristic sparseness promoting feature of the method. This means that solutions

obtained from this regularization approach have a smaller number of non-zero entries. This sparseness promoting feature is readily visualized when comparing the 2D case between ℓ_1 and ℓ_2 regularization, presented in Figure 10. In this figure, when employing ℓ_1 the regularized solution induced w_2 to become zero, which does not happen in the ℓ_2 norm case.

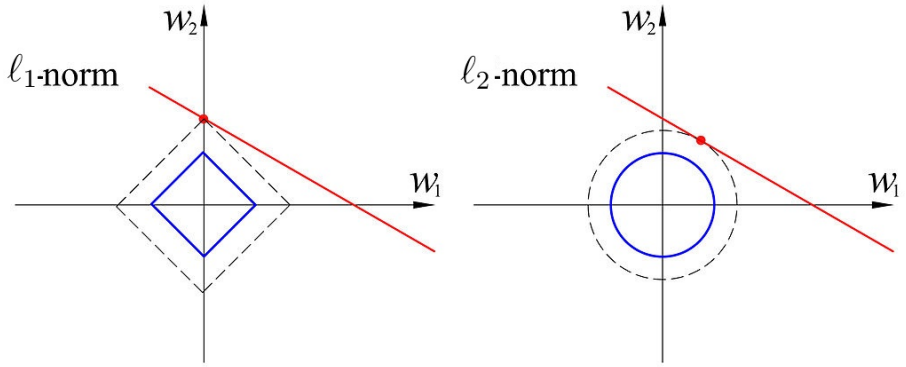


Figure 10 – Difference between ℓ_1 and ℓ_2 norms in regularization

Source: adapted from Hastie *et al.* (2009)

The sparseness-promoting feature can also be inferred from the formulation, since the minimization of the absolute value of components will penalize small components more severely than the large counterparts. The formulation reads, for a linear system of the form $A\boldsymbol{x} = \boldsymbol{b}$:

$$\hat{\boldsymbol{x}}_{\ell_1} = \arg\min \|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \tag{2.34}$$

where λ is again the regularization parameter. The solution of the ℓ_1 regularization problem is not as straightforward as its ℓ_2 counterpart. This is mainly due to the fact that the regularization requires absolute values of parameters, and the absolute value function is non-differentiable at zero. Nevertheless, different algorithms exist implementing solution procedures for this type of regularization, such as LARS (EFRON *et al.*, 2004), coordinate descent (WU *et al.*, 2008), iterative shrinkage thresholding (BECK; TEBOULLE, 2009) and proximal gradient descent (COMBETTES; PESQUET, 2011).

Similarly as ℓ_2 regularization, ℓ_1 regularization can be interpreted from a Bayesian perspective, where the coefficients have a Laplace prior distribution. The Laplace distribution has a sharp peak at zero, due to a discontinuity in its first derivative. Therefore, an important mass of probability becomes concentrated closer to zero than what is observed with the Gaussian distribution. This constitutes another explanation for the sparseness promoting characteristic of this type of regularization.

2.10 TRUNCATED SVD

Another possible regularization involving linear systems arises from applying a Linear Algebra matrix decomposition. Suppose that a linear system of the form $A\boldsymbol{x} = \boldsymbol{b}$ must be solved by the least-squares method. Using the Singular Value Decomposition (SVD) of the matrix A, which has the form:

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T = U\Sigma V^T \tag{2.35}$$

the solution vector x can be written as:

$$\boldsymbol{x}_{lsq} = \sum_{i=1}^{n} \frac{u_i^T \boldsymbol{b}}{\sigma_i} v_i \tag{2.36}$$

where σ_i are the singular values of A and the vectors u_i and v_i are left and right singular vectors of A, respectively.

The methods of regularization often try to filter out the contributions to the solution corresponding to the small singular values. Therefore, they produce solutions of the form (GUERRA; HERNANDEZ, 2001):

$$\boldsymbol{x}_{reg} = \sum_{i=1}^{n} f_i \frac{u_i^t \boldsymbol{b}}{\sigma_i} v_i \tag{2.37}$$

where f_i are called filter factors.

Similarly to the ℓ_2 regularization, the TSVD also depends on a parameter, namely, the truncation parameter (VOGEL, 1986). It corresponds to the number of singular value terms summed in the solution. In fact, following the filtering factors scheme in Equation 2.37, both TSVD and Tikhonov can be seen as the application of the same regularization framework where the distinction between methods lies in the choice of filter factors. The following filter factors are employed (GUERRA; HERNANDEZ, 2001):

$$f_i = \frac{\sigma_i^2}{\lambda^2 + \sigma_i^2} \qquad \text{Tikhonov}(\lambda > 0), \tag{2.38}$$

$$f_i = \begin{cases} 1 & \text{if } i \leq n - \lambda \\ 0 & \text{if } i > n - \lambda \end{cases} \qquad \text{TSVD}(\lambda = 0, 1, \ldots, n-1). \tag{2.39}$$

Such methods based on filter parameters are known in literature as spectral filtering methods (HANSEN *et al.*, 2006).

2.11 ELASTIC NET REGULARIZATION

Elastic net is related to a recent hybrid method proposed by Zou e Hastie (2005). The method can be viewed as a middle term between Tikhonov (ℓ_2) and Lasso (ℓ_1) regularization. It aims at solving stability issues in the application of Lasso, where solutions may become too dependent on the data.

2.12 CONVOLUTION

A major point of this study relies on the concept of convolution between two signals. Convolution is a binary operation on functions, *i.e.*, it takes as input two functions and outputs another function. Recalling the definition, the convolution between two functions x and h can be written as:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau, \tag{2.40}$$

where τ is a delay between signals and $*$ is the convolution operator.

The discrete version of the above convolution, which encompasses an increased relevance for this study, can be written as:

$$y(t) = x(t) * h(t) = \sum_{m\in\mathbb{Z}} x(m)h(n-m). \tag{2.41}$$

The convolution theorem is also useful for enabling the determination of solutions in the frequency domain. It states that:

$$\mathcal{F}\{x * h\} = \mathcal{F}\{x\} \cdot \mathcal{F}\{h\}, \tag{2.42}$$

where

$$\mathcal{F}(x(t)) = \mathcal{X}(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt. \tag{2.43}$$

Therefore, the convolution can be performed by multiplying both signals in frequency domain and reverting the transformation for the time domain as in:

$$x * h = \mathcal{F}^{-1}\Big(\mathcal{F}(x) \cdot \mathcal{F}(h)\Big), \tag{2.44}$$

where:

$$\mathcal{F}^{-1}(\mathcal{X}(\omega)) = x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{X}(\omega)e^{i\omega t}d\omega. \tag{2.45}$$

This same equation holds for the discrete case. Thus, in case of limiting computational resources, any convolution procedure can be accelerated by solving it in frequency domain, where it enables the use of the efficient Fast Fourier Transform (FFT) algorithm. Thus, the convolution between two signals (vectors) can be written as:

$$\boldsymbol{x} * \boldsymbol{h} = \text{ifft}(\ \text{fft}(\boldsymbol{x}) \odot \text{fft}(\boldsymbol{h})), \tag{2.46}$$

where $\odot$ is the element-wise (Hadamard) product and fft and ifft are the fast Fourier transform and its inverse counterpart, respectively.

2.12.1 Convolution as matrix multiplication

Restricting the analysis to the discrete domain, it is possible to evaluate the convolution between two signals, represented by finite-length vectors as a matrix-vector multiplication. The matrix being multiplied, which is a linear mapping between the input vector and the output vector, always follows a special form known as Toeplitz (GRENANDER; SZEGÖ, 1958). This type of matrix is constructed by taking every row equal a zero-padded right-shifted version of an initial vector. This effectively results in a matrix where diagonals are always constant. In general, when a computing the convolution between a vector $\boldsymbol{a} \in \mathbb{R}^n$ with a vector $\boldsymbol{b} \in \mathbb{R}^m$ the Toeplitz matrix has the form $T \in \mathbb{R}^{(n \cdot m-1) \times n}$.

As an example, the following vectors $\boldsymbol{t} \in \mathbb{R}^4$, $\boldsymbol{a} \in \mathbb{R}^5$ are convolved:

$$\boldsymbol{t} = \{t_1, t_2, t_3, t_4\}^T; \boldsymbol{a} = \{a_1, a_2, a_3, a_4, a_5\}^T. \tag{2.47}$$

Thus, the following Toeplitz matrix $T \in \mathbb{R}^{8 \times 5}$ can be created for performing the convolution. Toeplitz matrices are used in section 3.3 in a convolution interpretation of the use of influence lines for obtaining axle weights.

$$T = \begin{bmatrix} t_1 & 0 & 0 & 0 & 0 \\ t_2 & t_1 & 0 & 0 & 0 \\ t_3 & t_2 & t_1 & 0 & 0 \\ t_4 & t_3 & t_2 & t_1 & 0 \\ 0 & t_4 & t_3 & t_2 & t_1 \\ 0 & 0 & t_4 & t_3 & t_2 \\ 0 & 0 & 0 & t_4 & t_3 \\ 0 & 0 & 0 & 0 & t_4 \end{bmatrix} \tag{2.48}$$

Operating the convolution as a matrix multiplication results in:

$$\boldsymbol{y} = \boldsymbol{t} * \boldsymbol{a} = T\boldsymbol{a} = \begin{bmatrix} a_1 t_1 \\ a_1 t_2 + a_2 t_1 \\ a_1 t_3 + a_2 t_2 + a_3 t_1 \\ a_1 t_4 + a_2 t_3 + a_3 t_2 + a_4 t_1 \\ a_2 t_4 + a_3 t_3 + a_4 t_2 + a_5 t_1 \\ a_3 t_4 + a_4 t_3 + a_5 t_2 \\ a_4 t_4 + a_5 t_3 \\ a_5 t_4 \end{bmatrix}. \tag{2.49}$$

Since the convolution operation is commutative the same procedure can be reapplied considering a new convolution Toeplitz matrix A with respect to the vector a. That is:

$$\boldsymbol{y} = \boldsymbol{t} * \boldsymbol{a} = \boldsymbol{a} * \boldsymbol{t} = A\boldsymbol{t} = \begin{bmatrix} a_1 & 0 & 0 & 0 \\ a_2 & a_1 & 0 & 0 \\ a_3 & a_2 & a_1 & 0 \\ a_4 & a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 & a_2 \\ 0 & a_5 & a_4 & a_3 \\ 0 & 0 & a_5 & a_4 \\ 0 & 0 & 0 & a_5 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix} = \begin{bmatrix} a_1 t_1 \\ a_1 t_2 + a_2 t_1 \\ a_1 t_3 + a_2 t_2 + a_3 t_1 \\ a_1 t_4 + a_2 t_3 + a_3 t_2 + a_4 t_1 \\ a_2 t_4 + a_3 t_3 + a_4 t_2 + a_5 t_1 \\ a_3 t_4 + a_4 t_3 + a_5 t_2 \\ a_4 t_4 + a_5 t_3 \\ a_5 t_4 \end{bmatrix}. \tag{2.50}$$

3 PROBLEM DEFINITION

The objective of any Weight-in-motion system is to obtain accurate axle load and gross vehicle weight information. They are introduced as an alternative to conventional static weight stations.

B-WIM systems convert an instrumented bridge in a weighing mechanism for passing vehicles. It offers advantages when compared to other WIM alternatives since its installation is non-destructive and its operation does not disrupt traffic. This allows for unbiased data acquisition with respect to traffic behavior and permits the system to be moved from one bridge to another. Moreover, the usefulness of the system comes not simply from its weighing capability. It also offers general parameters related to the bridge it is installed. Considering this data, such as the estimated bridge influence line, one can estimate the current structural condition of the bridge, its service life and the stresses developed given a hypothetical heavy vehicle (LYDON *et al.*, 2016; JUNGES *et al.*, 2017; LANSDELL *et al.*, 2017).

The two main parts needed for estimating weights are the strain measurements and axle detection. This study concerns itself with the form of B-WIM known as nothing on road (NOR). In the NOR B-WIM formulation, the axle detection, which often relied on detectors on the road surface, is conducted by Free of axle detector (FAD) sensors attached beneath the bridge slab. This contributes to avoiding traffic disruption by positioning all sensors hidden from plain sight.

3.1 POINTS OF IMPROVEMENT

After an initial evaluation of the methods employed in literature for weight estimation, which are briefly discussed in chapter 4 and resulted in the published paper in Appendix A, two main points of improvement were identified. Both could be seen as a consequence of the lack of information associated with either model description or solution. Firstly, most methods employed do not directly address the dynamic effects. This, along with other systematic forms of errors such as pavement temperature and lack of accuracy in axle detection, were found to be detrimental for the overall performance of the system (YU *et al.*, 2016). The approach of better dynamic considerations was recently addressed in literature in the work of Gonçalves *et al.* (2021a).

Secondly, the other major aspect that influences the system performance is related to the ill-posedness of the inverse problem. Figure 12 illustrates how multiple vehicles can lead to the same structural response. In the figure, two vehicles are shown, the first representing a 4-axle vehicle with equal axle weights, the second an arbitrarily generated

vehicle with random weights and five axles. As shall be later discussed, the process of finding weights from B-WIM measurements involves comparing generated bending moments on the bridge. Thus, the ill-posedness of the problem becomes clear since reversing the process with the moment data may lead to an infinite number of valid solutions.

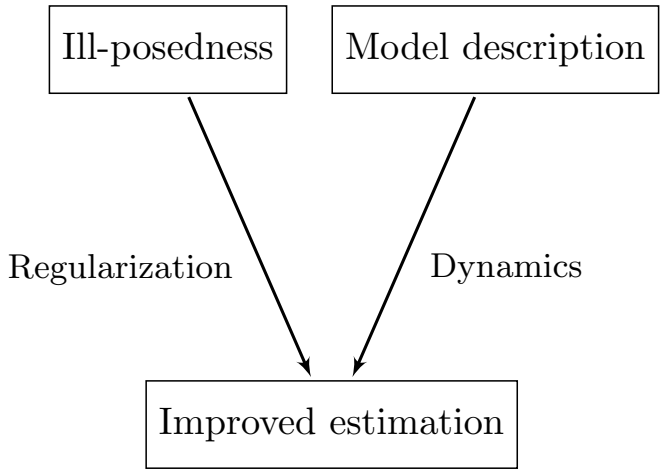


Figure 11 – Points of improvement

Despite common acknowledgment of this issue in literature, few studies have looked for forms of overcoming it. Considering different inputs (vehicle weight) can generate the same output (strains), when working backward on the inverse problem one could aid the solution procedure by introducing a bias from prior knowledge. The alternative that shall be explored is this thesis is regularizing the solution. In order to regularize the solution, one needs to introduce generic constraints for restricting the solution space in an appropriate manner. This is the main premise of this study and something that has not been thoroughly explored in literature. Indeed, regularization techniques were naively applied to this problem, as pointed out in the paper presented in Appendix A. For example, in O'Brien *et al.* (2009) only Tikhonov regularization was considered without discussing forms of obtaining the regularization parameter and detailing the L-curve procedure applied.

As should be detailed throughout this section, it is possible to view the problem under a convolution perspective. This perspective results in further possibilities for incorporating relevant knowledge to the problem. This information shall be used for either constraining the problem in a meaningful way or removing constraints that were harmful to the estimation procedure. The main purpose of this step is to actually use the available information about the problem at hand. For example, the consideration of inaccurate axle spacing, which relaxes the problem and is in accordance with what is generally seen on field. Moreover, knowing that no force is applied on the bridge between axles, induces a sparse model representation that may aid constraining the solution space. Further details on this regard will be discussed after a more in-depth problem presentation.

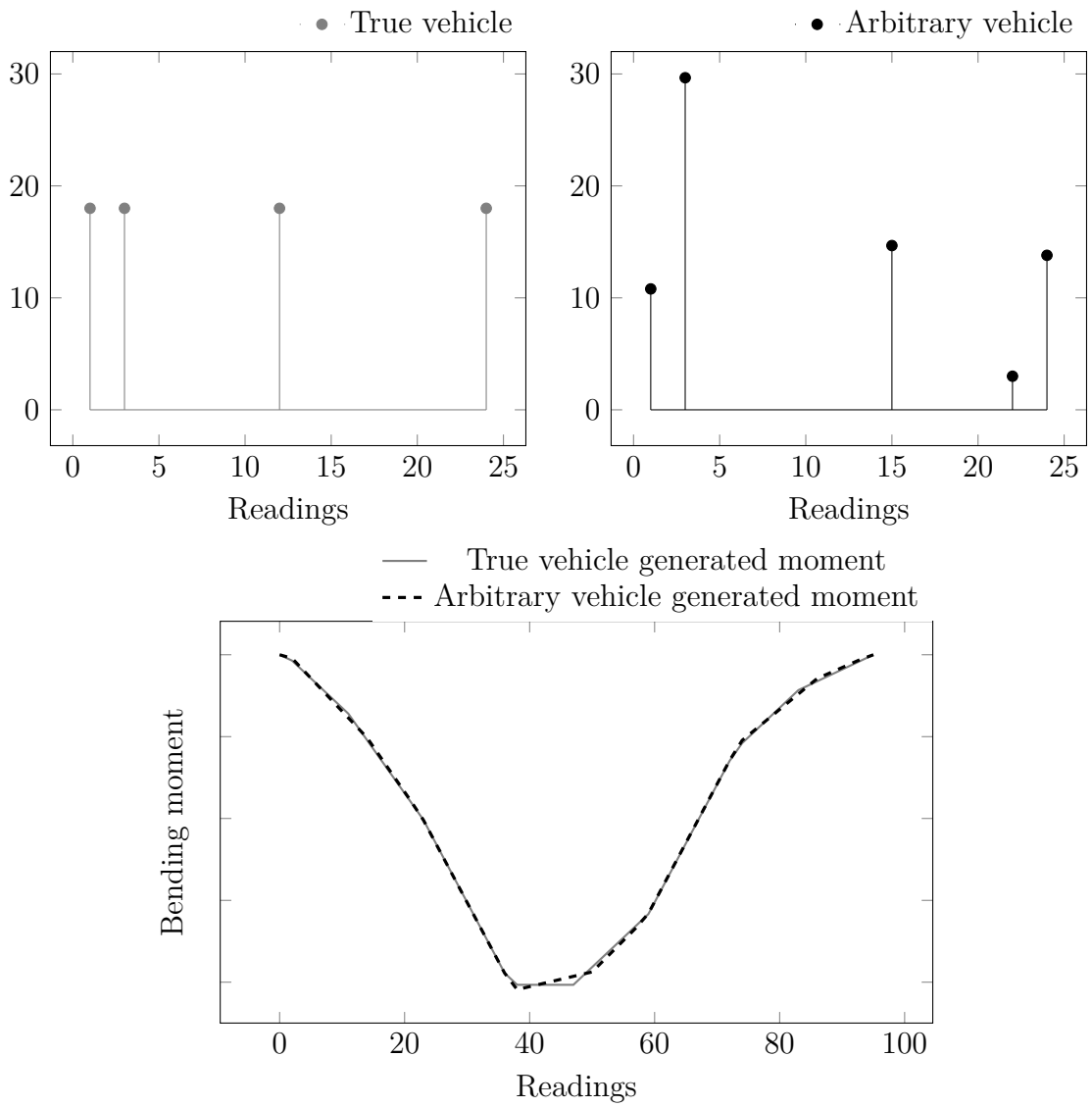


Figure 12 – Illustration of the problem ill-posedness

3.2 MOSES AND MATRIX METHOD

Moses (1979) verified the availability of using strain measurements for vehicle weight estimation for the first time. The standard solution procedure for obtaining vehicle axle weights consisted of a least squares approach, relating measured and expected bending moments calculated from the influence line. The influence line is a curve that represents values of a certain function given the passage of a unit load over the member being considered, in this case, the bridge. For this particular application, the function of interest is the bending moment at the mid-span, where the maximum strains occur. Figure 13 illustrates the relation between theoretical and measured moments.

Influence lines derived from purely theoretical considerations, such as support conditions, were often seen as inaccurate for the weighing procedure (ZHAO; UDDIN, 2010). Considering that information for modeling and extracting bridge influence lines may

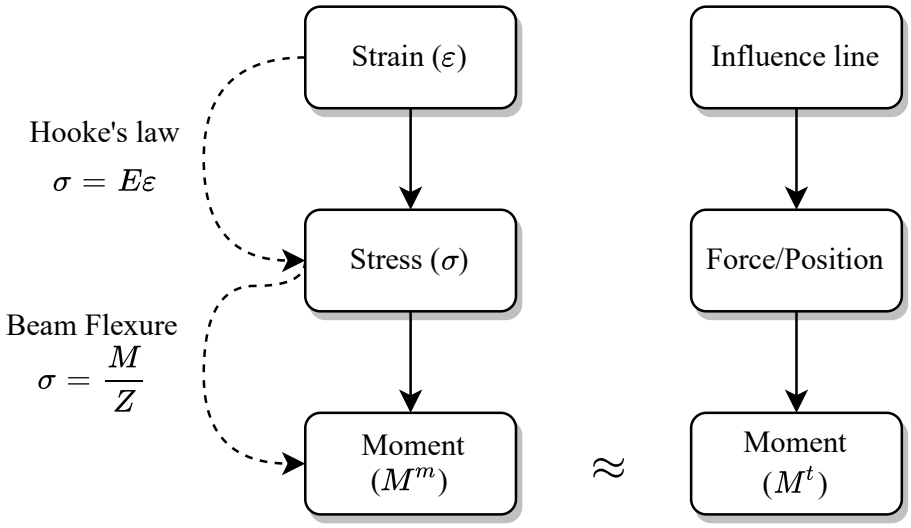


Figure 13 – Schematic representation of data relations in B-WIM

be unavailable or unreliable, O'Brien *et al.* (2006) proposed to overcome these issues with the introduction of the Matrix method. The idea behind the Matrix method is obtaining an influence line that better represents the actual bridge condition with a calibration procedure. A vehicle with known weight traverses the bridge and that measured response is employed for calibrating the influence line which minimizes the difference between measured and expected bending moments.

The difference between signals is considered in a least-squares sense with the introduction of an error function R. It is comprised of the sum of the squares of differences between the measured bending moment M^m and the theoretical bending moment M^t. Despite being usual, the formulation equating moments is not unique for solving the problem. One could carry a similar derivation calibrating the system by only relating strains. The results are the same for both forms of formulating the problem since moments and strains are only related by a multiplicative constant EZ. Following with the derivation as usually presented, the measured term for an instant k is given by:

$$M_k^m = \sum_{g=1}^{G} E_g Z_g \varepsilon_g, \tag{3.1}$$

where G represent the number of girders and E_g, Z_g, and ε_g are the elastic modulus, section modulus and measured strain of the *g-th* girder, respectively. The theoretical term introduces the influence line and reads:

$$M_k^t = \sum_{j=1}^{J} W_j I\ell_{(k-C_j)}, \tag{3.2}$$

where:

$$C_j = \frac{d_j f}{v}, \tag{3.3}$$

and J is the number of axles, W_j is the weight of the *j-th* axle, $I\ell_{(k-C_j)}$ is the influence ordinate at the position of the *j-th* axle, d_j is the distance between the first axle and *j-th* axle, C_j is the number of scans corresponding to d_j , f is the sampling frequency and v is the vehicle velocity. It is worth pointing out that if $k - C_j$ results in an index that does not match an influence line ordinate, it is attributed the value of zero.

The sum of squares of the residual reads:

$$R = \|M^m - M^t\|_2^2. \tag{3.4}$$

The derivation of the influence from Equation 3.4 is usually presented in literature in a very cumbersome manner, hereby exemplified with the deduction for a three axle vehicle. In the case of a three-axle truck passing over the bridge, the least-squares solution is obtained when the gradient of the error function with respect to the influence line is null. The error function is rewritten as:

$$R = \sum_{k=1}^{K} \left[M_k^m - (W_1 I\ell_{k-C_1} + W_2 I\ell_{k-C_2} + W_3 I\ell_{k-C_3})\right]^2, \tag{3.5}$$

while the gradient with respect to an arbitrary influence line ordinate $I\ell_a$ reads:

$$\begin{aligned} \frac{\partial R}{\partial I\ell_a} =& 2[M_a^m - (W_1 I\ell_a + W_2 I\ell_{a-C_2} + W_3 I\ell_{a-C_3})](-W_1) + \\ & 2[M_{a+C_2}^m - (W_1 I\ell_{a+C_2} + W_2 I\ell_a + W_3 I\ell_{a+C_2-C_3})](-W_2) + \\ & 2[M_{a+C_3}^m - (W_1 I\ell_{a+C_3} + W_2 I\ell_{a-C_2+C_3} + W_a I\ell_a)](-W_3) = 0 \end{aligned} \tag{3.6}$$

rearranging the terms it is possible to obtain:

$$\begin{aligned} I\ell_{a-C_3} W_1 W_3 + I\ell_{a-C_2} W_1 W_2 + I\ell_{a+C_2-C_3} W_2 W_3 + I\ell_a (W_1^2 + W_2^2 + W_3^2) + \\ I\ell_{a-C_2+C_3} W_2 W_3 + I\ell_{a+C_2} W_1 W_2 + I\ell_{a+C_3} W_1 W_3 \\ = W_1 M_a^m + W_2 M_{a+C_2}^m + W_3 M_{a+C_3}^m, \end{aligned} \tag{3.7}$$

which, when converted to matrix form, becomes:

$$A\ I\ell = L. \tag{3.8}$$

Then, solving for the influence line can be written as:

$$I\ell = A^{-1}L. \tag{3.9}$$

The drawback of this form of presentation is that a common form for the A matrix is not identified. Here, the case for a hypothetical 3-axle vehicle was considered. However, the procedure would need to be reapplied for considering any other vehicle with a distinct number of axles, resulting in a repetitive and isolated derivation for the consideration of each different vehicle type that needs to be analyzed.

After the influence line based on direct measurements is found, one can proceed as usual with the Moses method for finding the unknown weights of vehicles passing over the bridge. The usual literature presentation for the weighing procedure involves the definition of a so-called influence line matrix Λ (ZHAO; UDDIN, 2010). This matrix is of size $K \times J$, and is constructed as to indicate the influence line ordinate for all axles for every *k-th* measurement, *i.e.*:

$$\Lambda_{kj} = I\ell_{k-C_j}. \tag{3.10}$$

It can be noted that the matrix terms take into account the fixed spacing between axles represented by the quantity C_j for establishing each of its lines. The characteristic shifts in matrix Λ can have a deeper meaning assigned, related to the convolution procedure. This aspect is further discussed in section 3.3. Given this matrix, one can write:

$$M^t = \Lambda W. \tag{3.11}$$

Employing this form on the error function and repeating the derivation, where now the gradient is taken with respect to the vector W, the resulting least squares solution becomes:

$$W = (\Lambda^T \Lambda)^{-1} \Lambda^T M^m, \tag{3.12}$$

where W has a component representing each axle weight. Thus, the Gross Vehicle Weight (GVW) of the vehicle can be computed as:

$$GVW = \sum_{j=1}^{J} W_j. \tag{3.13}$$

In a B-WIM system, the GVW accuracy is regarded as one of the most important features of the system. This measurement is employed when comparing different weight estimation methods and it serves as a basis to determine possible applications of the system such as overweight truck law enforcement.

It is important to highlight here the peculiar asymmetry on Equations 3.9 and 3.12 in the form they are usually presented in literature. The reader should recognize that finding the influence line given the weights and finding the weights given the influence line should be regarded as dual problems. Therefore, why the resulting solutions for both are seemingly distinct in form? The insight that permits a symmetrical view of both problems is further explored in the following section. The possibility of viewing the B-WIM problem under this different perspective is one of the main points of this study, since it leads to insights for further improving the B-WIM estimation procedures.

3.3 CONVOLUTION INTERPRETATION

A recent study from Frøseth *et al.* (2017) enabled a novel interpretation for the use of influence lines for obtaining axle weights. Initially, one can idealize the vehicle passing over the bridge as an impulse vector, where each impulse is spaced according to the spacing between axles and the impulse magnitudes correspond to axle weights. Then, it is possible to view the resulting vector of any bridge property associated with an influence line as the convolution between the impulse vector and the influence line.

Regarding bridge safety, the evaluation of the mid-span bending moment is the measure of interest. An illustration of the convolution is shown in Figure 14. In this figure, Ai represents the weight of the i-th axle and Lj the influence line associated with the j-th measurement.

As a simple example, suppose a two-axle truck is to be modeled as an impulse vector A with the form:

$$A = \{A_1, A_2\}^T, \tag{3.14}$$

where the numerical values of A_1 and A_2 represent the weight of the first and second axle. Moreover, suppose the already calibrated influence line of the bridge is represented by the vector L:

$$L = \{L_1, L_2, L_3, L_4, L_5\}^T. \tag{3.15}$$

Given the mentioned interpretation it is possible to compute the vector representing the bending moment of the mid-span of the bridge as the following convolution:

$$M^t = A * L, \tag{3.16}$$

recalling that $*$ is the convolution operator. The above relation is graphically presented in Figure 14. The usual convolution pattern can be noticed, where convolving a vector of

size n with another of size m, results in a vector of size $n + m - 1$.

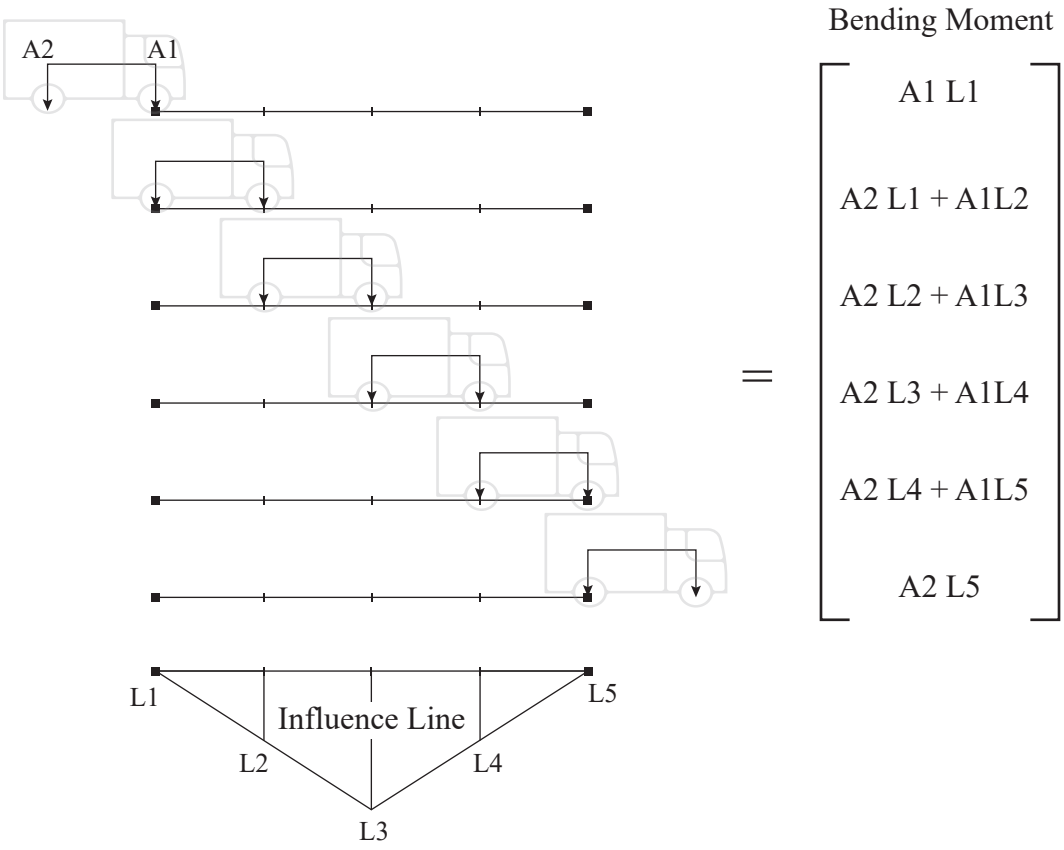


Figure 14 – Convolution visualization

As seen in subsection 2.12.1, the convolution operation can be performed by a matrix multiplication, converting one of the vectors to a matrix in Toeplitz form. For this simple two-axle vehicle the form of the matrix, here called τ, becomes:

$$\tau = \begin{bmatrix} A_1 & 0 & 0 & 0 & 0 \\ A_2 & A_1 & 0 & 0 & 0 \\ 0 & A_2 & A_1 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & 0 \\ 0 & 0 & 0 & A_2 & A_1 \\ 0 & 0 & 0 & 0 & A_2 \end{bmatrix}, \tag{3.17}$$

where the matrix is formed with the consideration of vector A padded by zeros, equating the number of columns to the number of measurements taken or the number of influence line ordinates considered in the analysis.

The gradient with respect to the influence line vector L is derived based on an error measure e, corresponding to the difference between measured a theoretical bending

moment response. The error function R_e, which shall be minimized, comprises the sum of the squares of deviations e. The derivation follows:

$$e = \tau L - M^m \tag{3.18}$$

$$R_e = \|e\|_2^2 = e : e \tag{3.19}$$

$$dR_e = 2e : de \tag{3.20}$$

$$dR_e = 2(\tau L - M^m) : \tau dL \tag{3.21}$$

$$dR_e = 2\tau^T(\tau L - M^m) : dL \tag{3.22}$$

$$\frac{\partial R_e}{\partial L} = 2\tau^T(\tau L - M^m). \tag{3.23}$$

Setting the above gradient to zero, one obtains the expression for the influence line vector in the convolution interpretation of the B-WIM problem:

$$0 = \tau^T\tau L - \tau^T M^m \tag{3.24}$$

$$L = (\tau^T\tau)^{-1}\tau^T M^m. \tag{3.25}$$

Here it is possible to notice how this new relation for computing the influence line based on known measured moments (M^m) and axle weights (τ) has a similar form, as the form presented by weight estimation procedure (Equation 3.12).

Since the convolution operation is commutative, it permits for exchanging the order of the convolution without loss of generality. This allows one to access the duality of the problem. The conversion between weighing and obtaining the influence line is readily evident when the order of the convolution terms is reversed. The Toeplitz form for the matrix relating the influence line ordinates becomes:

$$\Psi = \begin{bmatrix} I\ell_1 & 0 \\ I\ell_2 & I\ell_1 \\ I\ell_3 & I\ell_2 \\ I\ell_4 & I\ell_3 \\ I\ell_5 & I\ell_4 \\ 0 & I\ell_5 \end{bmatrix} \tag{3.26}$$

where the theoretical bending moment vector at mid-span is now written as:

$$M^t = \Psi A. \tag{3.27}$$

Considering now the unknown as being the weight and not the influence line, the moment difference can be written as:

$$e = \Psi A - M^m, \tag{3.28}$$

and applying the same procedure detailed above for deriving the minimal squared residual solution, one arrives at the following expression for axle weight estimation expression:

$$W = (\Psi^T \Psi)^{-1} \Psi^T M^m. \tag{3.29}$$

Apparently, given the presented derivation, one arrived with Equation 3.29 at the same result as the traditional B-WIM equation displayed in Equation 3.12. However, this is not the case, in general. For this particular case, the impulse vector consisted of only two terms, both of them being axle weights. In other words, the impulse vector and the vector of axle weights are the same. Therefore, for this specific case, both equations yield the same result. Nevertheless, Equation 3.29 is more general, since it allows for recovering the full impulse vector. It also takes into account the spacing between axles by appending zeros between axle weight terms, which correspond to the number of measurements taken up to the point where the next axle passes at the influence line ordinate.

As discussed earlier, this presentation, differently from what is derived in literature, enables a more symmetric view of the problem. The duality of the problem is readily identified from inspecting equations Equation 3.25 and Equation 3.29. It is also possible to identify that matrix A in Equation 3.9 is equivalent to $\tau^T \tau$ in Equation 3.25. This enables an easy computation of matrix A for any number of axles by employing the Toeplitz form on the impulse vector.

For a long truck, for example, there may be numerous strain and, consequently, bending moment measurements that are taken in the time span from between the first and second axles entering the bridge. When trying to recover the whole impulse vector, these inter-axle ordinates should all be zero. However, this is not the case when trying to undo the convolution (deconvolution) by simply applying the least-squares procedure. Figure 15 illustrates this issue by showing the original and recovered impulse vector. One can notice that when working on this problem by the standard least-squares approach there is a considerable amount of noise on the resulting impulse vector. Although the axle peaks can be identified close to their correct positions, there are also a number of other spurious peaks in the region between axles. Moreover, the inter-axle region which was supposed to be flat at zero, contains values different from zero, even including negative loads which do not make physical sense.

Considering the exposed discussion, it is possible to see that there is an unexplored domain in B-WIM literature with respect to this form of solution. Only recently did Frøseth *et al.* (2017) expose this other form of viewing the problem which has been neglected since the early work of Moses (1979).

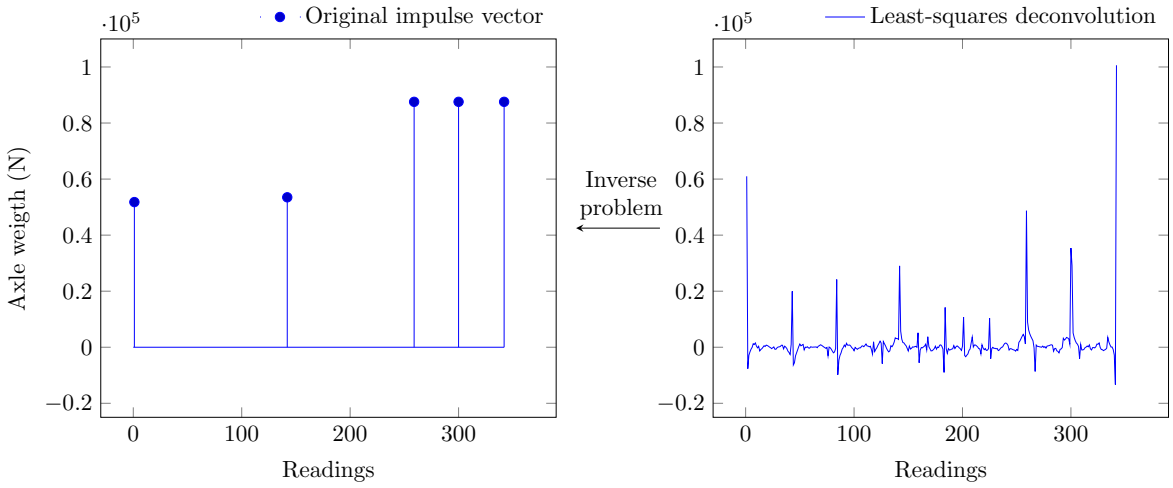


Figure 15 – Distorted impulse vector recovered from least squares

Exploring the problem under the convolution interpretation allows another view on the general ill-posedness of the weight estimation problem. Moreover, it opens an opportunity for further gains in accuracy by proposing approaches which exploit the known expected properties of the solution. Despite the contribution of the study of Frøseth *et al.* (2017), no further developments on B-WIM estimation were achieved by them, where their approach was presented only as an alternative with better computational performance. This thesis aims to use the bridge and analogy that the convolution interpretation promoted in order to propose better solutions to the problem, bringing together the already established knowledge applied in correlated literature topics such as signal processing and image processing. In this regard, this study proposes the use of regularization procedures for enforcing the constraints which are seen to be violated on the standard least-squares solution. This aspect was briefly explored in B-WIM literature by the work of O'Brien *et al.* (2006), which employed Tikhonov regularization for computing weighing estimates. Despite the study claims that the method achieved good performance, few information was given on how the regularization procedure was conducted, especially regarding the form employed for obtaining the regularization parameters. In the preliminary study that originated the published paper (Appendix A), comparisons relating the regularization procedure with the standard Matrix Method have shown very similar results obtained for both methods in simulated and real-world data. One of the points that may explain this similarity when not working for recovering the full impulse vector is that the resulting system of equations is "stiffened" in such a way that it becomes well-posed. This rigidity originates from incorporating into the formulation the spacing between axles and neglecting all inter-axle data. As a matter of fact, numerical experiments emphasized this intuition given the difficulty in finding the corner of L-curve for estimating the regularization parameter. The L-curve observed was, in fact, not L-shaped at all. Therefore it is proposed by this study to work directly with the recovery of the impulse vector which more easily conveys the

regularization necessity.

O'Brien *et al.* (2006) have not explored in depth the possibilities provided by distinct regularization approaches. Despite ℓ_2 or Tikhonov regularization being one of the most widely employed procedures, the theoretical soundness of employing this metric was not justified by the authors. Indeed, this thesis proposes to initially study a regularization form with better properties than that employed for the problem being analyzed.

Considering the prior knowledge of the impulse vector form, it is evident that the solution being sought is sparse. In general, B-WIM systems operate with signal acquisition rate of 512 or 1024 measurements per second. This is often enough to incorporate numerous measurements in-between axles, which corresponds to more null terms in the impulse vector. Thus, it becomes advantageous to explore regularization procedures which could better handle the characteristics of the solution.

Given the available regularization methods in literature, ℓ_1 or Lasso regularization seems to be a rational choice when considering the characteristics of the B-WIM problem. With the small number of non-zero terms on the original vehicle impulse vector, ℓ_1 regularization should lead to improved results, given its sparseness promoting formulation.

The implementation of regularization techniques other than the common ℓ_2 regularization often results in expressions where no closed-form solution exists. Thus, finding the solution may involve some form of optimization procedure. This may result in higher computational costs for obtaining a solution. Since it may be desirable for the B-WIM system to have real-time monitoring capabilities, one must be careful in designing robust and efficient solution procedures. Given the possibility of the proposed solution being too computationally expensive, this study also aims at evaluating forms of performing the regularization procedure in frequency domain. As already discussed, performing a convolution in frequency domain is equivalent to a point-wise multiplication between two vectors. This, combined with the existence of fast algorithms for transferring the problem to the frequency domain (FFT) may allow possible computational drawback to be overcome.

Other exploratory alternatives may also arise as a consequence of the previously mentioned parallel with signal and image processing literature. Viewing the problem as a signal decovolution establishes a direct relation with a number of common problems in domains such signal filtering and image deblurring/reconstruction (RUDIN *et al.*, 1992). Given the recency of the establishment of this view, authors in B-WIM literature have not yet explored the adequacy of the application of methods from those domains estimation process. Therefore, as a long-term goal, this study intends to explore additional solutions procedures exploiting the analogy.

Thus the main topics that are novel in literature and are adressed in this study

can be summarized in:

- Look for the full impulse vector instead of each axle separately;
- Employ regularization for addressing the problem ill-posedness;
- Enforce the expected sparseness of the impulse vector;
- Exploit the structure of the problem model;
- Enable variations on axle spacing estimates;
- Explore the parallel with signal and image processing literature.

4 ANALYSIS AND DISCUSSION OF STATIC BWIM SOLUTION METHODS

Numerous studies have proposed methods for the solution of the BWIM problem. In this chapter, a subset of static algorithms present in literature will be discussed and their theoretical aspects will be analyzed.

4.1 MAXIMUM LIKELIHOOD ESTIMATE (MLE)

In statistics, maximum likelihood estimate (MLE) consists of a method for estimating the parameters of a probability distribution based on a set of observed data. The goal of MLE is to find the set of parameters that maximize the likelihood of the observed data given the probability distribution. Therefore, MLE finds the set of parameters which makes the observed data most probable (TARANTOLA, 2005).

In the B-WIM context, the MLE method proposed by Ieng (2015) follows the idea of Moses method, employing the influence line for weighing an unknown vehicle passing over the bridge. However, the main difference of this method is the way it employs data from multiple calibration vehicles for obtaining an estimated influence line.

It is assumed in MLE that the measurements are corrupted by ϵ, a zero mean multivariate normal random variable:

$$M^m = M^t + \epsilon, \tag{4.1}$$

where M^m and M^t are the vectors of measured and modeled moments, respectively, such that:

$$M^t = AI\ell, \tag{4.2}$$

where IL is a vector with the ordinates of the influence line and A is a Toeplitz matrix of the loads. The matrix A is based on the impulse load vector $\mathcal{W}$:

$$\mathcal{W}_i = \begin{cases} W_j, & \text{if } i = C_j + 1 \\ 0, & \text{otherwise} \end{cases}, \tag{4.3}$$

where W_j is the weight of the *j-th* axle and C_j is analogous to that one defined in Equation 3.3. This impulse vector represents the whole vehicle, with each axle load at their respective axle position. The A matrix is formed by shifting $\mathcal{W}$ in each line of A, which corresponds to a discrete convolution, where each line relates to a time step of the vehicle passing over the bridge:

$$A = \begin{bmatrix} W_1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ W_J & \dots & W_1 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & W_J & \dots & W_1 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & W_J \end{bmatrix}_{(K \times K - C_J)} , \tag{4.4}$$

where K and J are the total number of scans and axles, respectively.

The modeling premise of this methods is that the series of N events are independent. Then, the likelihood P of observing the data is the product of the probabilities of observing each data point individually, *i.e.*, the product of the marginal probabilities:

$$P = \prod_{i}^{N} \text{pdf}(\epsilon_i | I\ell) \tag{4.5}$$

where pdf is the probability density function of the error ϵ. In order to obtain the maximum likelihood estimate, it often makes calculations easier by rewriting the problem as the equivalent minimization of the negative log-likelihood:

$$\arg\max_{I\ell} P = \arg\min_{I\ell} -\sum_{i}^{N} \log(\text{pdf}(\epsilon_i | I\ell)) = \arg\min_{I\ell} g(I\ell) \tag{4.6}$$

4.1.1 Equivalence relation to least squares

Assuming that data comes from n independent and identically distributed samples, the likelihood can be written as:

$$p(D|w) = \prod_{i=1}^{n} p(D_i|w) \tag{4.7}$$

It is often useful express the formula in terms of minimization, changing its sign and introducing the logarithm to convert the products into a summation.

$$w_{\text{MLE}} = \arg\max_{w} \left[\prod_{i=1}^{n} p(D_i|w) \right] \equiv \arg\min_{w} \left[-\sum_{i=1}^{n} \log p(D_i|w) \right] \tag{4.8}$$

$$p(\epsilon_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\epsilon_i^2}{2} \right) \tag{4.9}$$

The Gaussian likelihood term can be written as

$$p(y_i|x_i, w) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(w^T x_i - y_i)^2}{2}\right) \tag{4.10}$$

Thus, instead of maximizing the likelihood, one can proceed with the equivalent minimization of the negative log-likelihood:

$$f(w) = -\sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(w^T x_i - y_i)^2}{2}\right)\right) \tag{4.11}$$

$$= -\sum_{i=1}^{n} \left[\log\left(\frac{1}{\sqrt{2\pi}}\right) + \log\left(\exp\left(-\frac{(w^T x_i - y_i)^2}{2}\right)\right)\right] \tag{4.12}$$

$$= (\text{constant}) + \frac{1}{2}\sum_{i=1}^{n}(w^T x_i - y_i)^2 \tag{4.13}$$

$$= (\text{constant}) + \frac{1}{2}\|Xw - y\|^2. \tag{4.14}$$

Ignoring the scaling fraction and the constant terms, the expression for the maximum likelihood estimate becomes:

$$w_{\text{MLE}} = \arg\min_w f(w) \equiv \arg\min_w \|Xw - y\|^2 \tag{4.15}$$

Thus, its possible to observe the equivalence of the Least Squares solution with the MLE when assigning a likelihood which follows a Gaussian distribution.

4.2 PROBABILISTIC B-WIM (PBWIM)

The pBWIM method proposed by O'Brien *et al.* (2018) aims to incorporate a deeper probabilistic reasoning into the problem formulation. It applies data from multiple vehicle passages in order to construct an influence line. The main assumption of the method is that ordinates of the influence line are random variables following a Gaussian distribution each with a certain mean and variance to be determined.

The matrix method is employed to derive the parameters of these distributions, based on a single influence line for each event. Finally, the estimated axle weights are those with the highest probability of occurrence among all possible combinations.

As usual, the method starts with the Moses approach aiming to relate theoretical and measured bending moments. The theoretical response M_k^t for each scan k is taken as the sum of products of axle weights and influence line ordinates:

$$M_k^t = \sum_{j=1}^{J} W_j I\ell_{k-C_j} \tag{4.16}$$

where j is the number of axles, W is the vector of axle weights, $I\ell_{k-C_j}$ is the influence line value at ordinate $k - C_j$ and C_j is the offset distance between the ordinates of the influence line related to axle j.

It can observed from Equation 4.16 that the expected response is a linear combination of the random variables associated with the influence lines and the constant weight of the axles. That is:

$$R_k = e + \sum_k W_j I\ell_{k-C_j} \tag{4.17}$$

where the e term corresponds to the zero mean random measurement noise.

As seen from probability theory, the sum of independent Gaussian random variables also follows a Gaussian distribution (DEGROOT; SCHERVISH, 2012). Characterizing the random variables associated to each measured influence line ordinate as having mean μ_k and standard deviation σ_k, the random variable associated with a particular weight vector combination can be characterized by the following equations:

$$\mu^c_{r,k} = \sum_{k=1}^{K} \sum_{j=1}^{J} W^c_j \mu^c_{k-C_j} \tag{4.18}$$

$$\sigma = \sqrt{\sum_{i=1}} \tag{4.19}$$

where the superscript c identifies a possible combination of weights for the W vector.

In order to compute the weights the original pBWIM method employs a grid search, which increases the computational burden of the method, especially for higher number of axles. Recent research from Gonçalves *et al.* (2021b) has shown that these weights may be more efficiently estimated by employing a gradient based optimization procedure.

4.3 CONSIDERATION OF TRANSVERSE WEIGHT DISTRIBUTION

The consideration of the transverse weight distribution of axle loads on the bridge girders is a innovative augmentation of the Moses method proposed by Zhao *et al.* (2014). In usual practice, the measured responses from the sensors installed beneath the bridge are averaged over each girder. This study proposes that the responses remain separated, deriving an individual influence line for each girder of the bridge.

One of the possible issues that arrive from trying to consider the bridge as two dimensional structure is the need for establishing the fraction of the total vehicle load that gets transmitted to each girder. The study from Zhao *et al.* (2014) proposes the introduction of a Q term that tries to account for vehicle weight distribution across girders.

The transverse load distribution factor of the wheel load for the j-th girder and scan k is given by:

$$Q_{j,k} = \frac{\epsilon_{j,k}}{\sum_{j=1}^{J} \epsilon_{j,k}} \tag{4.20}$$

Note the essentially the Q coefficient is the ratio between the currently analyzed girder strain and the total observed strain from all sensors at any given instant. Inherently this is a quantity subject to variations across the longitudinal direction of the bridge. Ideally, the tranverse distribution factor should be constant, in order to decompose independent influence lines for each girders. In order to enforce this condition, the author employs an arbitrary criterion, assuming it as a constant calculated as the average Q value from the set of the 50 largest strain measurements.

The strain from each girder is then calculated by summing the contributions of the vehicle axles for each girder separately:

$$\varepsilon_{g,k}^{t} = \frac{1}{E_g Z_g} \sum_{j=1}^{J} W_j Q_g I_{g,(k-C_j)} \tag{4.21}$$

where $W_j Q_g$ is the weight contribution of the *j-th* axle on the *g-th* girder.

the influence line for each girder could be calculated as:

$$IL_g = \left((AQ_g)^T (AQ_g)\right)^{-1} (AQ_g)^T E_g Z_g \varepsilon_g^m, \tag{4.22}$$

where IL_g is the vector of influence line ordinates for the girder g, A is the Toeplitz matrix of loads described in Equation 4.4.

The weighing procedure is based on a least square minimization between predicted and measured moments, considering the axle weights as variables of interest. In order to maintain consistency with previous formulation, the weighing procedure may be written as:

$$W = \left((\Lambda_Q)^T \Lambda_Q\right)^{-1} \Lambda_Q \sum_{g=1}^{G} (E_g Z_g \varepsilon_g^m) \tag{4.23}$$

$$= \left((\Lambda_Q)^T \Lambda_Q\right)^{-1} \Lambda_Q M^m, \tag{4.24}$$

where Λ_Q is a matrix of influence line ordinates defined by:

$$\Lambda_Q = IL_G \times_n Q^T, \tag{4.25}$$

where Q is the vector grouping the transverse distribution parameters of all girders, $\times_n$ is the n-mode tensor product (KOLDA; BADER, 2009) of the third-order tensor IL_G with the vector Q, with IL_G defined as:

$$IL_G^{kjg} = \Lambda_g^{kj} \tag{4.26}$$

with Λ_g defined by Equation 3.10 for every girder g. The tensorial product mentioned performs a parallel operation, scaling each of the third-order tensor inner matrices by the contribution factor and aggregating them, which results in a single matrix Λ_Q.

In order to verify its application conditions it is important to identify how this version is related to usual Matrix Method. This relation becomes more clear by rewriting Equation 4.22 and isolating the terms related to the contribution factor:

$$\begin{aligned} IL_g &= \left((AQ_g)^T(AQ_g)\right)^{-1}(AQ_g)^T E_g Z_g \varepsilon_g^m \\ &= Q_g^{-2}(A^TA)^{-1}Q_g A^T E_g Z_g \varepsilon_g^m \\ &= Q_g^{-1}(A^TA)^{-1}A^T E_g Z_g \varepsilon_g^m. \\ &= Q_g^{-1}(A^TA)^{-1}A^T M_g^m. \end{aligned} \tag{4.27}$$

As it can be seen, the resulting influence line is the same as the one obtained by the usual Matrix method but divided by the contribution factor.

In fact, under some conditions, the modified 2D Moses method and Matrix method are identical. For the case of constant transverse distribution along the bridge, Q_g can be written as:

$$Q_g = \frac{\varepsilon_g^m}{\sum_{g=1}^{G}\varepsilon_g^m}. \tag{4.28}$$

Substituting Equation 4.28 into Equation 4.27 one can write:

$$\begin{aligned} IL_g &= \left(\frac{\varepsilon_g^m}{\sum_{g=1}^{G}\varepsilon_g^m}\right)^{-1}(A^TA)^{-1}A^T E_g Z_g \varepsilon_g^m \\ &= (A^TA)^{-1}A^T E_g Z_g \sum_{g=1}^{G}\varepsilon_g^m. \end{aligned} \tag{4.29}$$

Therefore, Equation 4.29 shows that when considering girders with equal properties, that is, the same values for E_g and Z_g, the Modified 2D method results in equal influence lines for each girder. Moreover, this influence line is precisely the one that would be obtained

if one employs the Matrix method. Thus, under this assumption, the methods can be considered equivalent for extracting an influence line.

Under the same assumptions, this analysis can be extended for the weighing procedure. Recalling the equation relating theoretical and measured moments:

$$\sum_{g=1}^{G} E_g Z_g \varepsilon_g^t = \sum_{g=1}^{G} W Q_g \Lambda_g. \tag{4.30}$$

Again, assuming that E_g and Z_g are the same for all girders, namely E and Z, it follows that the influence lines IL_g are all equal, as previously discussed. Thus, Λ_g will also become constant, since it is a matrix constructed from of the influence lines of each girder. Dropping to subscript from Λ to indicate the constant value, one can write:

$$EZ \sum_{g=1}^{G} \varepsilon_g^t = W\Lambda \sum_{g=1}^{G} \left(\frac{\varepsilon_g^m}{\varepsilon^m} \right) \tag{4.31}$$

$$\sum_{g=1}^{G} E_g Z_g \varepsilon_g^t = W\Lambda. \tag{4.32}$$

Thus, the predicted moments for Matrix method and modified 2D Moses are the same, in this case. Therefore, one can conclude that, under the hypotheses of constant weight distribution and same mechanical properties for each girder, the weighing procedure also recovers the same solution found when applying the Matrix method, becoming independent of the distribution parameters.

4.4 TIKHONOV REGULARIZATION

The use of regularization procedures within B-WIM has been applied by O'Brien *et al.* (2006). The author proposes the use of the form known as Tikhonov regularization. In this form, the L2 norm is incorporated into the linear equation which defines the problem. The reason for applying regularization is the ill-possedness of the B-WIM problem.

Since one is dealing with an inverse problem, inputs are sought by its model output only. However, the output data may be corrupted by noise. In the presence of noise, the recovery processes may become unstable, and noise can be largely amplified in the solution. Thus, regularization aims to provide a more stable solution by rewriting the problem and constraining solutions towards less flexible models, avoiding over-fitting. In the case of Tikhonov regularization, solution terms are shrinked employing the L2 norm as in:

$$f = \|r\|_2^2 + \lambda \|W\|_2^2 = r : r + \lambda W : W. \tag{4.33}$$

The parameter λ in the above equation is called the regularization parameter. It is the variable responsible for controlling the trade-off between the original problem and the regularization effect. That means the

Given the proposed formulation, it is straightforward to compute the differential and gradient of the function being minimized as follows:

$$\begin{aligned} df &= 2r : dr + 2\lambda W : dW \\ &= 2(M^m - \Lambda W) : -\Lambda dW + 2\lambda W : dW \\ &= 2\Lambda^T(\Lambda W - M^m) : dW + 2\lambda W : dW \\ &= (2\Lambda^T(\Lambda W - M^m) + 2\lambda W) : dW \\ \frac{\partial f}{\partial W} &= 2\Lambda^T(\Lambda W - M^m) + 2\lambda W. \end{aligned} \tag{4.34}$$

Setting this gradient to zero, one can find an expression for the optimal weight matrix, which minimizes the augmented function f:

$$\begin{aligned} 2\Lambda^T(\Lambda W - M^m) + 2\lambda W &= 0 \\ \Lambda^T\Lambda W - \Lambda^T M^m + \lambda W &= 0 \\ (\Lambda^T\Lambda + \lambda I)W &= \Lambda^T M^m \\ W &= (\Lambda^T\Lambda + \lambda I)^{-1}\Lambda^T M^m, \end{aligned} \tag{4.35}$$

where the matrix I represents the identity matrix.

The expression for the weight can be seen as a function of the regularization parameter λ. When λ is zero, the regularized solution becomes the least squares solution, which means no importance is being given for how the L2 norm of the solution behaves. On the other hand, when one increases this regularization parameter, the augmented function seeks more aggressively to minimize the norm of the solution, up to a extreme where the solution is shrinked into a vector of zeros.

Since the choice on the regularization parameter λ leads to solutions W_λ with distinct characteristics, it is important to define a strategy for selecting this parameter. That is non-trivial task, which has been addressed in multiple ways by literature. One can employ, for example, cross-validation, ridge trace or the L-curve method. In the study of O'Brien *et al.* (2009), the L-curve method is applied. This method tries to establish a middle ground between two defined norms. The first one is the residual norm of the solution error, called E_{norm} and calculated as:

$$E_{norm} = \sqrt{(M - TW_\lambda)^T(M - TW_\lambda)}. \tag{4.36}$$

The second norm is called F_{norm} and takes into account the norm of the solution associated with the regularization parameter:

$$F_{norm} = \sqrt{W_\lambda^T W_\lambda} \tag{4.37}$$

The L-curve is then the curve obtained from plotting on a log-log scale the points (E_{norm}, F_{norm}) for a range of λ values. The corner of this curve, is assumed to represent a good trade-off between bias and variance on the system approximation. In Figure 16 is shown the expected L-curve formed by the method.

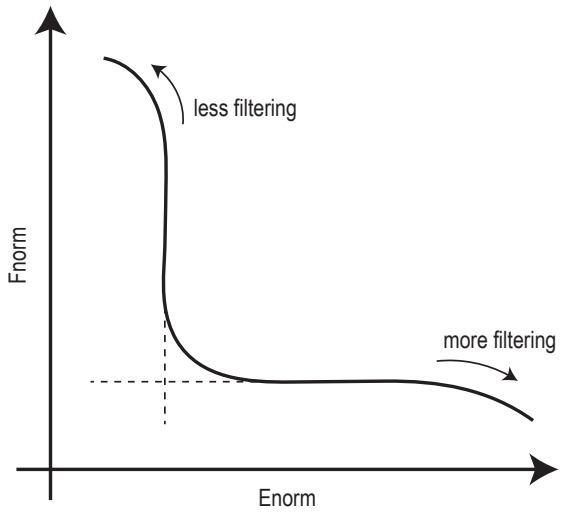


Figure 16 – The generic form of the L-curve plotted in double-logarithmic scale, adapted from: (HANSEN, 1998)

Solving for the λ corresponding to the corner is also not a simple task, specially for the cases where the transition between the effects of the two norms is not smooth. The corner, then, may not be well defined, and the "optimal" lambda may depend on which method was employed for obtaining it. Some methods in literature include the Spline-based Curvature Method (HANSEN; O'LEARY, 1993), Triangle Method (CASTELLANOS *et al.*, 2002) or Adaptive Pruning (HANSEN *et al.*, 2007).

4.4.1 Bayesian interpretation

An interesting parallel exists between regularization procedures and Bayesian statistics. The general goal of a Bayesian inference procedure is to encode the problem uncertainty in terms of probability distributions (GREGORY, 2005). These distributions are then manipulated and updated taking into account newer collected data, and then used for making predictions about the modelled problem. That is, by the Bayesian perspective, initial probability statements are updated, providing a posterior distribution that combines both prior knowledge and the data at hand (CONGDON, 2007).

It is possible to derive a single estimate from such posterior as, for instance, its maximum *a posteriori* (MAP). The prior distribution is independent from the results of measurements, and could be selected in order to better encode the properties of the problem (TARANTOLA, 2005). In a BWIM application, for example, the prior may enforce the knowledge about the non-negativity of predicted axle weights.

MLE and MAP are connected by the Bayes rule:

$$\underbrace{p(w|D)}_{\text{posterior}} = \frac{p(D|w)p(w)}{p(D)} \propto \underbrace{p(D|w)}_{\text{likelihood}} \underbrace{p(w)}_{\text{prior}} \tag{4.38}$$

The term in the denominator, called evidence or marginal likelihood does not depend on w. It can be considered a scaling constant, which ensures that the integral of the posterior probability distribution has unit value. Since it is a constant in w, this term is often omitted from the maximization procedure.

Therefore, MAP seeks to find the parameters which maximize the product of the likelihood and the prior. In this context, the prior is related the belief that the parameters are correct before seeing the data.

$$\begin{aligned} w_{\text{MAP}} &= \arg\max_w P(Y|X,w)P(w) \\ &= \arg\max_w \Big[\underbrace{\log P(Y|X,w)}_{log-likelihood} + \underbrace{\log P(w)}_{\text{log-prior}}\Big] \end{aligned} \tag{4.39}$$

The logarithm is a monotonic increasing function. Applying it to the objective function being maximized does not shift the location of the maximum. Therefore, working with logs is often preferred as it converts the product of probabilities into a summation. From Equation 4.39 it can be seen that the MAP estimate can be decomposed in a log-likelihood term and a log-prior term, in which the log-prior acts as a penalty for optimization procedure. Thus, several regularization methods that consists of including a penalty term have a Bayesian interpretation related to the assigned log-priors distribution.

For example, assuming the Gaussian prior with zero mean and variance σ^2:

$$p(w_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w_i^2}{2\sigma^2}\right). \tag{4.40}$$

Applying the logarithm to the prior and ignoring constant terms:

$$\log p(w) \propto \log\left(\exp\left(-\frac{1}{2\sigma^2}\sum_i^n w_i^2\right)\right). \tag{4.41}$$

Rewriting the expression in terms of a parameter $\lambda = -\frac{2}{\sigma^2}$:

$$\log p(w) \propto \log\left(\exp\left(\lambda \sum_{i}^{n} w_i^2\right)\right) = \lambda \|w\|^2. \tag{4.42}$$

Therefore, the MAP estimate found when considering a Gaussian prior and Gaussian likelihood, matches the formulation employed in ℓ_2 regularization, where the regularization parameter λ is proportional to the constant variance attributed to the prior.

4.5 OTHER APPROACHES FOUND IN LITERATURE

Regarding the methods that are focused in this paper, it is worth to point out that there are some proposed methodologies that address the same problem, however with characteristics that prevent a suitable comparison. In what follows, some of them are summarized and such aspects are discussed.

An approach to construct a more realistic influence line was proposed by Zhao *et al.* (2017). Their research intended to overcome the limitations of the theoretical influence line used by Moses in his first work. The bridge was modeled with semi-rigid connections and horizontal springs as boundary conditions. Moreover, it was included in the formulation the transverse load distribution, as proposed by Zhao *et al.* (2014). To completely define the model, it was necessary to estimate the values of some stiffness coefficients. In order to find this values, the authors performed a trial and error procedure comparing measured and modeled values. Two algorithms were proposed: semi-rigid approach, adjusting end moments and semi-rigid approach, using moments of the whole bridge. Such methods were compared with Moses, employing the theoretical influence line. The experimental procedure used two trucks to calibrate and validate the method. The results showed that both proposed approaches achieved better results, specially the second method, which reached the lowest error among all. However, the procedure for finding the stiffness parameters that defines the model does not have a clear definition. Thus, as such analysis could be user dependent, performed comparisons may be inconclusive.

Kim *et al.* (2009) proposed an approach to weigh vehicles using deformation measures as inputs to neural networks, obtaining the weights of each axle as the output. The error for both gross vehicle weight (GVW) and weight by axle were considerably low, indicating that the proposed approach could be applied in real situations. Nevertheless, the training process needed numerous training examples, preferably containing vehicles with distinct number of axles, which is not available in most practical cases. The advantage of this method was that it could be applied in cases where the traditional approach have some difficulties, which is not the focus of the present work.

In the work of Helmi *et al.* (2014) three weigh methods were compared utilizing data of a real bridge in Canada. The first two methods were developed by the authors and consisted in the creation of an equivalent uniform distributed load to represent the axle

loads, considering the influence line of a simply supported beam. The authors tried to find the fraction of the bridge span, corresponding to the length of the equivalent distributed load, which causes the maximum moment in the bridge. Thus, GVW could be calculated as the ratio between the maximum moment and this length. Nevertheless, both of the authors' proposed methods performed worse than the third alternative tested, namely the Beta method from Ojio e Yamada (2002). This method used the area under the moment or strain curve to calculate GVW, where errors of less than 5% were observed. However, none of such methods is able to distinguish the weight contribution of each axle, which is a parameter of comparison in the present paper.

The work of Frøseth *et al.* (2017) intended to overcome issues related to implementation complexity and computational cost through the realization that the response of the structure is the convolution of the influence line and the loading. Thus, instead using the well-established matrix method, the author suggested that the convolution could be performed in frequency domain, since the convolution integral transforms into an element-wise multiplication operation, which is very efficiently handled. The reported gains in computational time were, in general, of one order of magnitude, at least. Another advantage of viewing the problem under the proposed aspect, was that the matrices utilized in the least-squares approach in O'Brien *et al.* (2009) were straightforward to generalize for arbitrary number of axles. It is worth to mention that it may be necessary to apply a Tikhonov regularization in order to perform the deconvolution. The reason is that the system solution could result in an ill-posed problem in the frequency domain, for example, when the passing vehicle has two axles with identical loads. The authors concluded that the obtained influence line provides virtually identical results in comparison with the matrix method. Therefore, the main practical utility of this approach was not precision, but computational gains. As computational complexity is not addressed in the present paper, no further analyses are performed for this method.

5 NUMERICAL EXPERIMENTS

Along with the unifying theoretical analysis regarding the premises and modelling considerations conducted for the surveyed methods, numerical experiments were also performed. The main objective is to provide a numerical framework for testing and validating the theoretical findings and analyses presented in the previous chapter. Moreover, the framework establishes a reproducible and comprehensive data-set for evaluating any other applicable solution methods.

Recalling, the methods evaluated are the Matrix method (O'BRIEN *et al.*, 2006), maximum likelihood (MLE) (IENG, 2015), pBWIM (O'BRIEN *et al.*, 2018), Regularization (O'BRIEN *et al.*, 2009) and Modified 2D Moses (ZHAO *et al.*, 2014). The procedure employed for finding the corner of the L curve for the regularization approach consisted on minimizing the Euclidean distance between adjacent points, as suggested in Frasso e Eilers (2012). This is done utilizing all the pairs (E_{norm}, F_{norm}) employed for plotting the L-curve.

Table 1 – Literature method comparisons

Method	Comparison	Data	IL	Weight
Matrix	Measured/Predicted	Real	✓	-
MLE	Matrix method	Real	✓	-
PBWIM	Matrix method	Real	-	✓
Regularization	Matrix method	Synthetic	-	✓
Modified 2D Moses	Measured/Predicted	Real	✓	✓

5.1 SIMULATED VEHICLE DATA

In order to evaluate the analyzed methods performance on distinct conditions, a dataset comprised of algorithmically generated sensor data was developed. The data was generated simulating several vehicles crossing a multi-girder bridge at different levels of measurement noise and surface roughness. The model simulated each vehicle as series of moving sprung-masses, each of them representing a vehicle axle and the bridge girders as Euler-Bernoulli beams with distinct properties.

The axle arrangement employed aimed at simulating real traffic, where distinct vehicle classes were selected based on real data reports from the brazilian infrastructure department (DEPARTAMENTO NACIONAL DE INFRAESTRUTURA E TRANSPORTES, 2006). The classes are defined by the weight of the truck, the number of axles and their spacing. Since this standard classification only establishes a lower bound for the axle spacing on some classes, the value of 5 meters was adopted for constraining the upper

bound and allowing generating vehicles assuming an uniform distribution. Table 2 presents all classes of vehicle employed as well as their number of axles. In order to allow a more concise presentation, axle spacing and weight of each employed truck are omitted. For assessing such values, the reader is referred to Departamento Nacional de Infraestrutura e Transportes (2006).

Table 2 – Vehicle types employed in the study

Vehicle Class	Axles
2C	2
3C	3
4CD, 2S2	4
3I2, 2S3, 3S2	5
2R4, 3S3, 3D3, 3N3	6
3D4 , 3N4	7
3D5	8
3M6 , 3Q6	9

Despite the standard classification employed, the dynamical attributes of axles which is applied to the vehicle-bridge interaction model, comprised of moving sprung masses, still have to be defined. In this regard, axles were divided in three categories: rear, front and semi-trailer. The stiffness and damping coefficients are distinctly modeled as uniformly distributed random variables. The mean values of the random variables are shown in Table 3. For the lower and upper bound of these variables, factors of 0.5 and 1.5 times the mean value were employed, following the studies of Fancher (1986) and Nosseir *et al.* (1982).

Table 3 – Mean values of stiffness and damping coefficients for each kind of axle

Axle type	Stiffness (N/m)	Damping (Ns/m)
Rear	1000000	10000
Front	400000	10000
Semi trailer	750000	10000

After assigning the properties to the model, the vehicle-bridge interaction model is solved, whereas the solution procedure follows the studies of of Biggs (1964), Yang *et al.* (2004) and Yang e Lin (2005). The solution consists on the application of model decomposition on the equation of motion of the bridge, which results in the following equation:

$$\ddot{q}_i + 2\xi_i\omega_i\dot{q}_i + \omega_i^2 q_i = \sum_{j=1}^{N} \frac{2P_j}{\bar{m}L} \sin\frac{i\pi Vt}{L}, \tag{5.1}$$

where N is the number of axles, q_i, ω_i and ξ_i are the modal coordinates, natural frequency and damping of the i mode, respectively. Furthermore, dotted variables represent the derivative with respect to time, P_j is the load of the j axle, V the speed of the vehicle, L is the length of the bridge and $\bar{m}$ is the mass per unit length of the bridge.

The equations of motion for each sprung-mass system is given by Equation 5.2:

$$M_{vj}\ddot{z}_j + C_{vj}\dot{z}_j + K_{vj}z_j = K_{vj}\left(u + y\right)|_{x=Vt} + C_{vj}\left(\dot{u} + Vy'\right)|_{x=Vt}, \tag{5.2}$$

where M_{vj}, C_{vj} and K_{vj} are the mass, damping and stiffness of axle j. Furthermore, u represents the bridge vertical displacement, z_j is the vertical position of the j axle, y is the pavement elevation ordinate and a prime denotes the derivative with respect to x.

Both bridge displacement and its first derivative could be computed directly with the modal coordinates and mode shapes of the simply supported beams:

$$u = \sum_{j=1}^{m} \sin\left(\frac{j\pi x}{L}\right) q_i, \tag{5.3}$$

$$\dot{u} = \sum_{j=1}^{m} \sin\left(\frac{j\pi x}{L}\right) \dot{q}_i, \tag{5.4}$$

where m is the number of modes applied, which is equal to 5 in all simulations.

The simulation of pavement roughness applied here is a common approach in studies in the same subject (MÚČKA, 2017; MIGUEL *et al.*, 2016; WANG *et al.*, 2017). The ordinates of the pavement irregularities are modeled as a random process, with a specific power spectral density function (PSD):

$$\Phi(\Omega) = \Phi(\Omega_0)\left(\frac{\Omega}{\Omega_0}\right)^{-2}, \tag{5.5}$$

where $\Phi(\Omega_0)$ is the amplitude coefficient, analyzed in previous sections, measured in $(10^{-6}m^3/\text{cycle})$ and Ω_0 is a reference spatial frequency. Thus, the road profile is generated by sampling from this PSD, using the method of superposition of harmonics (DHARANKAR *et al.*, 2017):

$$y(x) = \sum_{i=1}^{n\Omega} \sqrt{2\Delta\Omega\Phi(\omega_i)}\cos\left(2\pi\Omega_i x - \phi_i\right) \tag{5.6}$$

where $y(x)$ is the generated road vertical profile, ϕ_i is a random uniform phase angle between 0 and 2π, $\Delta\Omega$ is a constant increment, $n\Omega$ is the total number of frequency increments in the interval and Ω_i is a frequency uniformly distributed in the range of

Ω_{min} and Ω_{max}. The parameters adopted are $\Omega_0 = 0.01$ cycle/m, $\Omega_{min} = 0.001$ cycle/m, $\Omega_{max} = 4$ cycle/m and $\Delta\Omega = 0.001$ cycle/m.

It is worth to mention that for each run a distinct road profile is generated, since in the practical scenario lateral deviations occur. Moreover, a moving average with total size of 30 cm is employed to approximate the real contact between tire and pavement (MIGUEL *et al.*, 2016).

The interaction between bridge and vehicle clearly appears in Equation 5.2, by means of the displacement term related to the beam at the current axle position. Moreover, such interaction also occurs in Equation 5.1, due to the P_j term. This relation is remarked in Equation 5.7:

$$P_j = p_j \delta(x - Vt) \left(H\left(t - t_j\right) - H\left(t - t_j - \frac{L}{V}\right)\right), \tag{5.7}$$

$$p_j = -M_{vj} g + K_{vj}(z - (u + y)|_{x=Vt}) + C_{vj}(\dot{z} - (\dot{u} + V y')|_{x=Vt}), \tag{5.8}$$

with $\delta()$ and $H()$ representing the Dirac delta and Heaviside functions, respectively. Moreover, g is the acceleration of gravity and t_j is the time that the j axle arrives the bridge.

The equations of motion are solved independently, by a decoupled approach. Both bridge and vehicle equations are solved numerically by applying the Newmark-β method, with 1400 time steps. The time window begins when the first vehicle axle enters the bridge and ends when the last axle leaves it. The problem is solved iteratively, since the interaction force in the bridge-vehicle system changes with the displacement of both beam and sprung mass. An initial guess of interaction force is given to the beam equations, where the beam displacement is calculated. Such a displacement is then enforced to the vehicle model and a new interaction force is calculated. This procedure continues until the change in the interaction force reaches a small tolerance (10^{-5}). Usually such procedure converges rapidly, within 5 iterations. The midspan strains (s), which are the main output of simulation, are also updated with this interaction. Adopting, without loss of generality, a unit vertical distance from the neutral axis, and utilizing the fact that strains are related with the second derivative of displacement with respect to x, the strains can be written as:

$$s = -\sum_{i=1}^{m} \left(\frac{i\pi}{L}\right)^2 q_i \sin\left(\frac{i\pi}{2}\right). \tag{5.9}$$

For the case of multiple beams, a transverse distribution factor (Q) is provided and applied to divide the axle loads for each beam. Thus, each beam is simulated independently.

In order to simulate the inherent imperfections of the measured signal due to all possible aspects in the measurement field, noise is applied to the simulated response. The

noise applied consists of a white Gaussian random noise, with a constant signal to noise ratio of 20.

Numerical simulation aims to evaluate not only multiple methods on equal measured data but also how bridge characteristics influence the overall results for estimating gross vehicle weights. The two bridge characteristics accounted for are the bridge span length and pavement roughness. Three distinct road profiles are evaluated, with Power Spectral Density (PSD) amplitudes of zero (no roughness), 4 and 16, where roughness increases with the amplitude. For the bridge span parameter, three cases are considered, with spans of 10, 20 and 30 meters. Thus, the combination of every case of road profile and bridge span results in 9 distinct cases. Furthermore, a Gaussian noise with signal to noise ratio of 20 is added to every signal, aiming to incorporate measurement errors due to other sources than pavement roughness. The choice of values adopted for these parameters follows the recommendations from Jacob (1999) for B-WIM sites, hence, reproducing practical cases of interest. The combination of road profile and bridge length results in the following 9 simulated situations:

1. bridge span length of 10m and no road roughness
2. bridge span length of 10m and road roughness class A
3. bridge span length of 10m and road roughness class B
4. bridge span length of 20m and no road roughness
5. bridge span length of 20m and road roughness class A
6. bridge span length of 20m and road roughness class B
7. bridge span length of 30m and no road roughness
8. bridge span length of 30m and road roughness class A
9. bridge span length of 30m and road roughness class B

5.1.1 Numerical results

The algorithms are evaluated in a set of different scenarios for assessing performance aspects and sensitivity of the methods. The comparison criterion consists of the mean absolute percentage error related to the known GVW. The results are shown in Figure 17, as a function of roughness amplitude and bridge length. In order to facilitate the distinction among methods performance, these values are also presented in Table 4, where some differences appears more evidently.

The simultaneous presentation of performance evolution with respect to both parameters, in Figure 17 and Table 4, allow for quite interesting remarks. Firstly, all

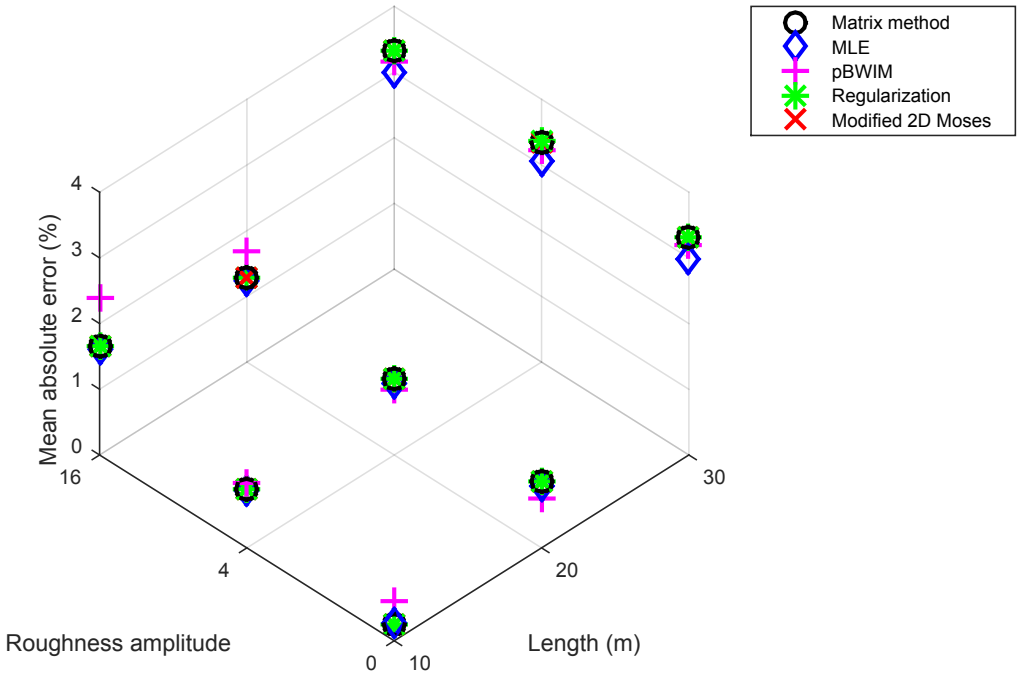


Figure 17 – Mean absolute error of each method as a function of bridge length and roughness amplitude

Table 4 – Mean absolute error for each method and case

Amplitude	Length (m)	Matrix	MLE	pBWIM	Regularization	Modified 2D Moses
0	10	0.26	0.26	0.61	0.26	0.26
4	10	0.88	0.88	0.99	0.88	0.88
16	10	1.65	1.62	2.39	1.65	1.65
0	20	1.02	0.93	0.75	1.02	1.02
4	20	1.17	1.09	0.98	1.17	1.17
16	20	1.28	1.24	1.69	1.28	1.28
0	30	3.31	2.99	3.19	3.31	3.31
4	30	3.35	3.06	3.21	3.36	3.35
16	30	3.32	3.00	3.15	3.32	3.32

methods showed similar performance, mainly Matrix method, Regularization and Modified 2D Moses. The MLE method also shared a same trend, however its performance surpassed the other methods when bridge span increases. This fact is observed independently of the roughness amplitude applied. Thus, although the difference is not so remarkable, MLE method can be argued as the most accurate method for this data set.

On the other hand, pBWIM method showed similar performance for lower roughness amplitude, becoming worse than the others with the increase in this parameter. It is worth mentioning that the performance loss occurs in conjunction with a more computationally

expensive prediction procedure.

Since the whole bridge is simulated as a unique beam, it is already expected for Modified 2D Moses and Matrix method results to be numerically identical. Indeed, in all evaluated cases and runs, the results are exactly the same. However, Regularization also shows similar behavior, presenting only slight deviations from the trends in the Matrix method. Such aspects can be seen in Table 5, where the maximum deviation of each method with respect to the Matrix method is presented. One can conclude that, in the conditions of this analysis, the difference in performance among Matrix method and Regularization is not significant. The reason is that the parameters found were usually close to zero, which promotes solutions very close to the least squares in the Matrix method. Nevertheless, it does not mean that the regularization approach is not useful, however, in the examples, the application of the L-curve corner lead to similar results. Therefore, it did not have a large influence in the cases analyzed here.

In order to correctly analyze the method presented by Zhao *et al.* (2014), it is necessary to take into account multiple beams in the bridge, which is addressed in the following sections.

Table 5 – Max of absolute difference between each method and the Matrix method

Amplitude	Length (m)	MLE	pBWIM	Regularization	Modified 2D Moses
0	10	0.12	1.32	0.02	0.00
4	10	0.30	5.44	0.02	0.00
16	10	0.53	18.26	0.03	0.00
0	20	0.52	2.43	0.01	0.00
4	20	0.93	6.30	0.01	0.00
16	20	1.54	15.21	0.01	0.00
0	30	1.22	5.24	0.01	0.00
4	30	1.30	6.15	0.01	0.00
16	30	1.38	5.35	0.01	0.00

When comparing the evolution of mean absolute error regarding bridge length and roughness amplitude separately, one can conclude that the first is more problematic. The difficulty of static methods in dealing with long span bridges is a fact already well known on BWIM literature (JACOB, 1999; KIM *et al.*, 2009). However, the comparison of variation of bridge length and roughness amplitude shows an interesting aspect. Analyzing the results, it is clear that bridge length has a higher impact on the overall result than the roughness amplitude, since the cases of smooth profile and 30 m bridge span easily surpasses the error of the 10 m bridge span and roughness amplitude of 16.

It is interesting to notice that the increase in bridge length spreads the effect of roughness in the prediction accuracy. Thus, the result is more sensible to the road pavement profile for short span bridges. The opposite is not true, in other words, independently of

Table 6 – Load distribution factors for each beam (%)

Case	Beam 1	Beam 2	Beam 3
Q1	33.3	33.3	33.3
Q2	20.0	40.0	40.0
Q3	25.0	50.0	25.0

the roughness amplitude, the increase in bridge length decreases the accuracy of weight prediction.

5.1.2 Multiple girder analysis

Given that Modified 2D Moses and Matrix method resulted in rigorously the same predictions in previous analyses, it becomes necessary simulating cases where the bridge response is modeled considering multiple beams. In this section, the bridge structure is comprised of 3 distinct beams. Moreover, the same vehicles and properties previously defined are adopted in this section. Three distinct cases of transverse distribution (Q) are considered, defined by Table 6.

Since pBWIM, MLE and Regularization methods do not make any assumption regarding transverse distribution of loads, their evaluation for this new case should not bring any new insight. Thus, only as a comparison criterion, the Matrix method is jointly evaluated with Modified 2D Moses. As a result of this analysis, Figures 18, 19 and 20 present the influence of roughness amplitude and transverse distribution case for each bridge span.

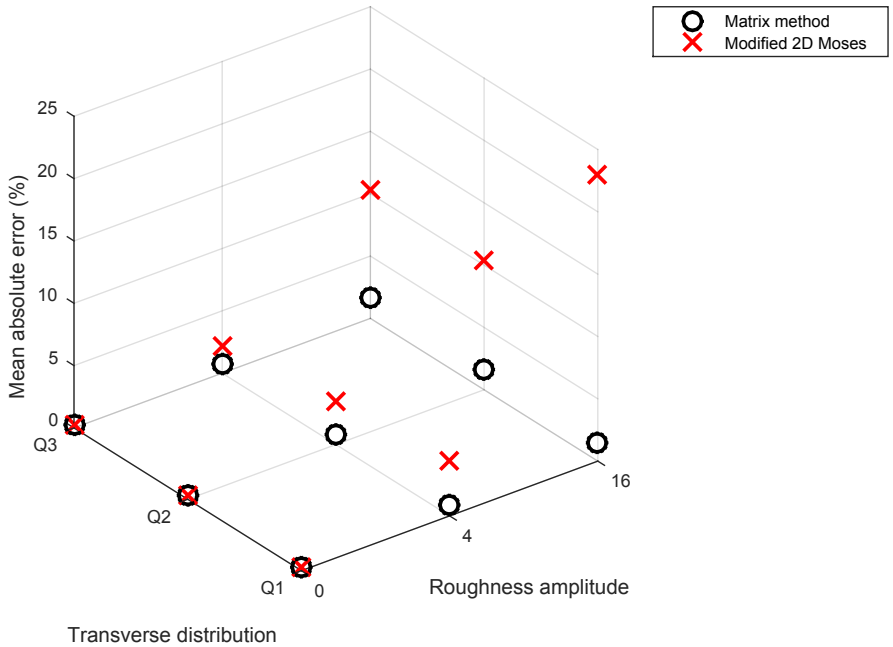


Figure 18 – Comparison between Matrix method and Zhao for bridge span of 10 meters

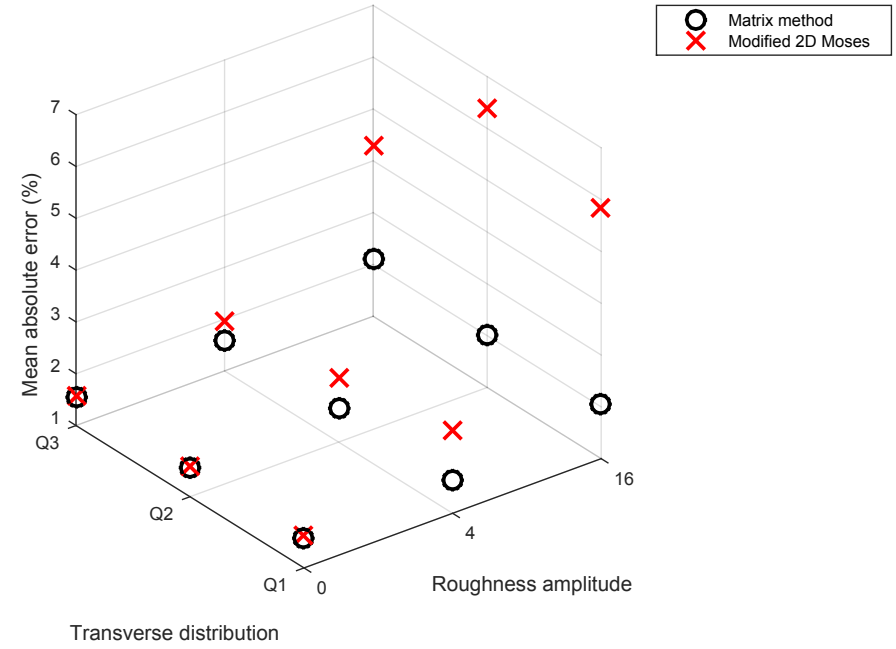


Figure 19 – Comparison between Matrix method and Modified 2D Moses for bridge span of 20 meters

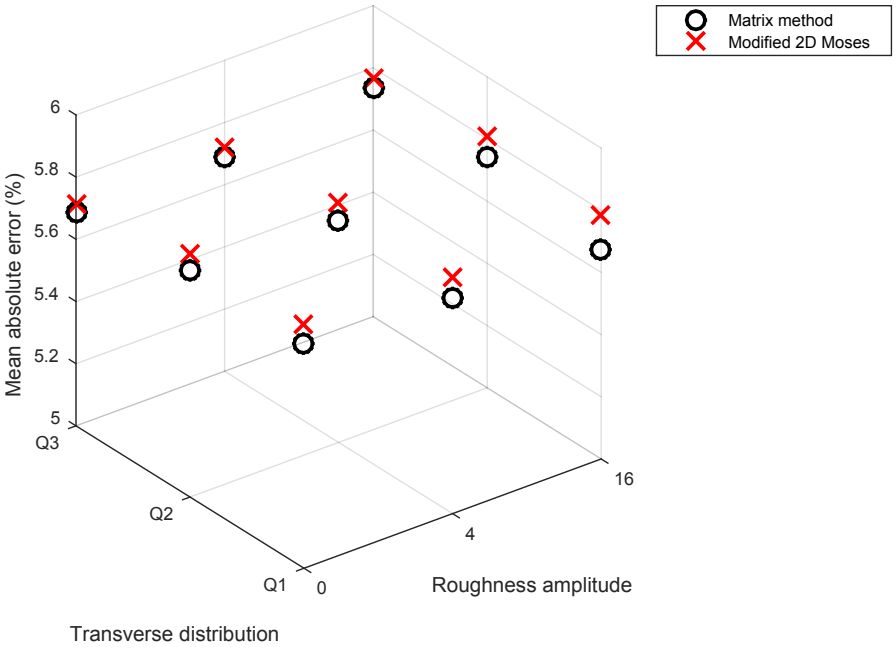


Figure 20 – Comparison between Matrix method and Modified 2D Moses for bridge span of 30 meters

In all cases, Modified 2D Moses performed equally or worse than Matrix method. In the case of low roughness amplitude, both methods could be argued as similar. When roughness amplitude increases, otherwise, there is a trend for the Modified 2D Moses method to present higher errors, for all analyzed bridge spans. Such performance gap is

specially remarkable in the 10 m bridge. This last statement is in agreement with the previous section, where the effect of road profile showed higher impact for short span bridge cases.

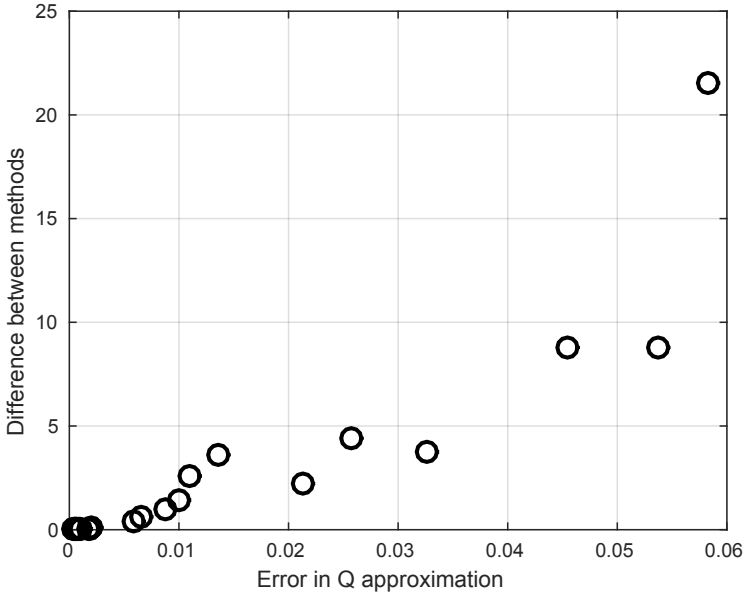


Figure 21 – Influence of error in transverse distribution factor on the difference between methods

One possible explanation for the performance gap can be seen in Figure 21. It shows the influence of the error in approximating Q (measured as the mean of Euclidean distance among predicted and real Q values) and the percentage difference between the results of the two methods. This figure makes clear that the accuracy in Q estimate is directly related to the difference in performance when compared with the Matrix method. Therefore, if transverse distribution factors can be accurately predicted, Modified 2D Moses approximates Matrix method performance. This fact is in accordance with section 4.3, since Q values are kept constant in the simulation. On the other hand, the increase in difference of such factors also increases the likelihood that weight prediction is corrupted.

5.2 FIELD-TESTING

The application of B-WIM technology in Brazil is still incipient. Trial experiments were conducted in bridges from the state of Goiás in a research project partnership between DNIT and UFSC. Data collected from the Itinguijada bridge instrumented with strain and free-of-axle (FAD) sensors and located at the city of Uruaçu is utilized in this study. The objective is to apply the previously discussed methods on non-simulated data and evaluate if the findings and general trends observed in both the theoretical investigation and simulated analysis carry over for the practical engineering problem.

The Itinguijada bridge, shown in Figure 22, is comprised by two girders and five cross beams, with a total length of 29.0 m. Figures 23 and 24 show the main dimensions of the cross section and the lateral view, respectively. One FAD sensor is installed in the mid-span of the bridge, while the other is spaced 4 meters longitudinally from the first. Two strain sensors are attached to each girder, at the mid-span.

Figure 22 – Itinguijada bridge

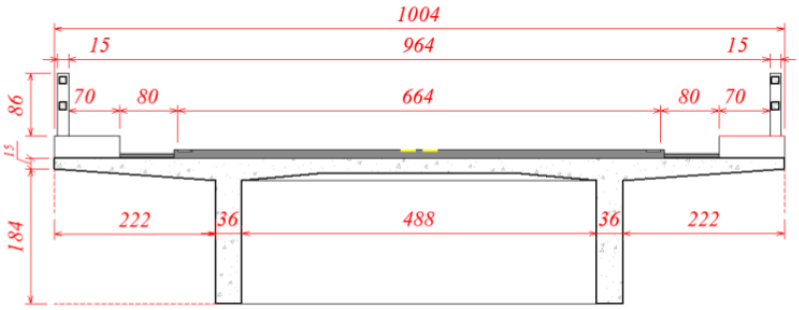


Figure 23 – Mid-span cross section dimensions

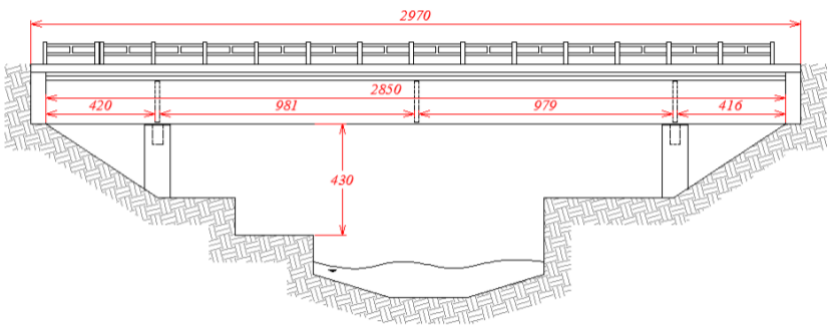


Figure 24 – Lateral view dimensions

5.2.1 Numerical results

In this section, following what is done in practice, as reported in Lydon *et al.* (2016) and Yu *et al.* (2018), one influence line is constructed for each lane. The analyses are based on prediction for three distinct quantities, defined by Jacob (1999): GVW, single axle and group of axles. The mean absolute error of such quantities for each method on the whole data set is presented in Table 7.

Table 7 – Mean absolute error

	Matrix method	MLE	pBWIM	Regularization	Modified 2D Moses
GVW (%)	4.47	4.08	4.91	4.40	4.41
Single axle (%)	17.79	15.26	36.46	15.35	16.91
Group of axles (%)	6.62	5.38	5.78	6.22	6.52

All methods showed higher errors for single axle prediction when compared to the group counterparts, which is in accordance with most studies in this subject, as in Zhao *et al.* (2014) and O'Brien *et al.* (2009). The prediction for GVW and group of axles weight showed reasonable results, with mean absolute errors always smaller than 7%. However, single axle prediction did not present the same level of performance, achieving values as high as 36% for pBWIM method.

MLE reached the best results, independently of the quantity being measured. However, for almost all methods and quantities measured, the mean absolute errors reported remained at a quite similar level. Since all methods disregarded dynamic effects, it is expected that the higher such effects are, the lower the suitability of all approaches are. As in this bridge the dynamic behavior is not remarkable, the results for MLE suggests that this method should perform better in cases where the model seems to be more suitable.

The similarity of measured errors previously discussed justify a more detailed analysis on it. Table 8 focus on this statement, based on the absolute difference in GVW, for each event, between all methods and the Matrix method, taken as reference here. The parameters presented are the mean, standard deviation and maximum value of the absolute difference.

Table 8 – Difference between each approach and the Matrix method

	MLE	pBWIM	Regularization	Modified 2D Moses
Mean (%)	1.2101	4.2554	0.3860	0.1526
Std (%)	0.6351	3.6975	0.2901	0.1153
Maximum (%)	2.5673	16.6928	1.2900	0.6864

From the five analyzed methods, two can be seen as almost identical to the Matrix

method. The first of them is the Regularization method, whose maximum difference in GVW for all events does not surpass 1.29 %. This lack of difference, as previously discussed, is caused by small regularization parameters obtained from the application of L-curve method. The second method quite similar to the Matrix method is Modified 2D Moses, in which maximum difference did not reached 1%. As already discussed, all girders have the same mechanical properties, which remain constant along the span. Thus, it is possible that the contribution factor Q_g resulting for each girder and event be approximately a constant value. As presented in section 4.3, in such a case it is already expected for results of Matrix method and modified 2D Moses to be similar.

The remaining two methods, MLE and pBWIM, showed more distinct values, rendering them as alternative approaches for the Matrix method. Observing these three methods by a probabilistic point of view, these differences are remarked. The Matrix method is based on least squares, which assumes that errors are uncorrelated, normally distributed and with the same variance, using all the measurements of one event as realization of only one normal random variable. The MLE method also applies the least squares approach, however now considering a multivariate normal distribution whose variables are moments measured in each ordinate, independently of the other ordinates. The pBWIM relaxes the assumption of equal variance, allowing for each ordinate to have a standard deviation estimated by the value of each influence line previously calculated from calibration events. Therefore, in some cases, namely when the data from both distinct events and the whole bridge have the same normal distribution, all methods should present virtually the same results. However, considering the real-world examples presented, this is not the case in practical applications.

One could say that pBWIM should achieve the best result, since this method has the least restrictive formulation. However, one must realize that this can only occur if the assumptions are met, and this is not the case for the pBWIM formulation, as demonstrated in section 4.2. It could be observed that this method assumes uncorrelated error moments. Yet, the formulation based on uncertainties in influence line will necessarily results in banded correlation of the resulting moments random variable, violating this assumption.

5.3 FURTHER REMARKS

Among five tested methods, three of them showed almost identical results for all events evaluated, namely, the Matrix method, Regularization and Modified 2D Moses. Although the pBWIM method could be argued as an alternative approach for weight prediction, both theoretical and practical results are not promising. Comparing MLE and Matrix method, the results for the simulated runs were very similar, however with MLE being slightly better. Furthermore, when field-testing was conducted, MLE approach showed even better performance. These points, together with its more suitable statistical

background, indicates that MLE is the most promising method evaluated in the present study. Some suggestions for improving these methods are included in the following.

The formulation employed for Regularization leads to a solution that recovers only weights associated to each axle. However, it is possible that the regularization approach can have improved results when trying to recover the whole impulse vector, as more information about its characteristics may be incorporated and enforced by the regularization procedure. Furthermore, it may be worth to employ different regularization procedures instead of Tikhonov, such as Lasso, which uses the L1 norm instead of the usual L2 (TIBSHIRANI, 1996). As a further step, the non-negativity of the vector can be enforced as in the work of Hummelsheim (2014).

Discussing specifically the pBWIM method, it could be concluded that an assumption made in the approach formulation is conceptually violated. Thus, it is useful to change such a methodology to account for a covariance matrix that is not diagonal. Furthermore, the procedure suggested to find the most likely weights in the original paper is a quite time consuming task. Thus, applying an optimization procedure is necessary for achieving a reasonable computational cost.

The procedures applied are clearly divided in two steps: building influence line and weighing the vehicles. From this perspective, one could easily apply distinct methods for each step. For instance, MLE could address the former and regularization the latter. Since it was not the focus of this work, coupling methods was not evaluated. However, it has potential to improve the accuracy of predicted axle weights.

Overall, it becomes clear that the main point that should be addressed is the incorporation of additional knowledge to the model. One aspect is related to the dynamic effects. The results could be improved by a more suitable consideration on the dynamic behavior of the vehicle-bridge system model. A second aspect is related to the form of the solution. The knowledge that a vehicle is modeled by point loads on each axle imposes constraints on the form of the impulse vector and the sign of the resulting weights. Nevertheless, relaxing assumptions that over-constrain the model, such as the correctness of axle spacing measurement, may improve flexibility and robustness to the methods.

For all evaluated methods, the weigh procedure disregarded some prior information that could lead to more reliable estimates. As examples, one could cite the non-negativity of weights and the trend for axle weights having values at the same order of magnitude. As discussed previously, a Bayesian approach could be employed in such a case, using these prior information to create improved estimators for axle weights.

When observing methods by the statistical point of view, the assumptions implicitly made in the development of such formulations arose in a more clear fashion. Thus, the possibilities of relaxing some of those assumptions could be seen as a good initial point for

the development of new methods. Furthermore, by knowing in advance which statistical assumption was made, it becomes possible to easily predict for which real cases the new methods could perform better. As an example, allowing errors to be correlated should perform better in problems where model fits poorly, since missing information on formulation could be seen as correlated errors. Thus, enabling the model to allow correlated error is an alternative to address the dynamic effects.

The comparison criterion of the evaluated methods was done without a clear performance threshold. Thus, although it is possible to verify which method is best in the tested cases, nothing can be said about robustness or suitability of overall results. Therefore, it motivates the development of a methodology for extracting more useful information from such results, mainly regarding robustness in practical applications.

6 EXPLOITING PROBLEM KNOWLEDGE

As discussed in chapter 1, one of the aims of this study is to apply the knowledge of specific characteristics of the B-WIM problem in order to devise a general and robust weight recovery procedure. As shown in chapter 3 the B-WIM problem in its original form proposed by Moses (1979) is viewed from a pure least squares perspective, relating theoretical and measured bending moments. This linear system assumes that the observations in vector b are corrupted by an additive noise vector e in the form of:

$$b + e = Ax. \tag{6.1}$$

Assuming that information on axle location is not enforced, the solution of this system is not sparse, as shown in Figure 15 from chapter 3. The enforcement of axle information consists of solving the least squares problem in Equation 6.1 by considering only a subset of the columns of the matrix A, where each column considered corresponds to the inferred axle spacing. This solution procedure can be regarded as performing a regression subset selection (GATU; KONTOGHIORGHES, 2013):

$$x_s = \underset{x_s \subseteq x}{\arg\min} \|A_s x_s - b_s\|_2^2, \tag{6.2}$$

where the subscript s consists of the relevant indices. In the B-WIM case, it means those related to the nonzero entries of the recovered impulse weight vector. Since the sparsity information is obtained from axle detection sensors either on the pavement or beneath the bridge slabs, the resulting information is subject to measurement noise. Thus, the assumption of perfect knowledge on the sparse entries of the impulse vector has consequences for weight estimation. In a sense, it effectively constraint the possible recovered solutions, ignoring most of the measured data from sensors. For example, considering a vehicle with 6 axles passing over a instrumented bridge and 1000 measured moments are collected during the crossing. In the formulation employing the restricted least squares, b_s would only have 6 elements, that is, only 6 of 1000 measurements would be employed to estimate the vehicle weight. Moreover the location of those points would be fixed and unrelated to the measured strains and moments, depending only on the accuracy of the axle detection. Therefore, it can be beneficial the employment of a method that results to sparse solutions without assuming perfect sparsity knowledge and also utilizing all available data in order to try to improve the estimation procedure.

6.1 SPARSENESS

The simplest and most direct form for incorporating the sparseness constraint into the least squares formulation is by considering the ℓ_0 norm:

$$x = \arg\min_x \|b - Ax\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq s \tag{6.3}$$

where $\|x\|_0$ represents the ℓ_0 norm of x and s is an unknown non-negative integer. Recalling from chapter 2, the ℓ_0 effectively counts the number of non-zero entries. Thus, the problem constraint establishes a minimum level of desired sparsity.

Solving the problem involving the ℓ_0 norm directly is often a computationally difficult task (DONOHO *et al.*, 2006). As a form of problem relaxation, the introduction of the ℓ_1 norm can accomplish the same task of delivering sparse solutions while having better properties for the development of fast solution methods such as convexity. It has been shown that, under certain high probability conditions, the ℓ_1 norm approximation is a suitable alternative, reaching near-oracle performance (CANDES; ROMBERG, 2007; BICKEL *et al.*, 2009).

In view of its direct application to promote sparseness and simple formulation, Lasso or ℓ_1 regularization procedures have become very popular over the years, having received ample attention from researchers (TIBSHIRANI, 2011). Due to the non-convex nature of the problem, there is not a single solution procedure available. In fact, a considerable number of variants and alternative solutions procedures have been proposed in literature. This may become a problem when trying to choose a solution method since there are many options and some methods are not robust enough to deal with all types of regularization demands.

6.1.1 ℓ_1 Solution Toolbox

Given the importance of finding sparse solutions for this study and the unclear picture in literature about which are most effective solutions methods, a subset of methods for the solution of the ℓ_1 were evaluated. The methods were implemented by the author in a unifying manner. This manner of implementation using equal variables for the same concepts, which is not often the case when dealing with multiple authors, and employing shared functional procedures enabled the transformation of the implementation of multiple isolated methods into an integrated lasso optimization toolbox with different algorithms available and source code possessing interesting educational value.

6.1.1.1 ISTA/FISTA/RFISTA

ISTA and the FISTA/RFISTA variants consist on the application of proximal methods and can be used for the solution of ℓ_1 regularization problem. Proximal methods are based on iterative optimization methods where each iterate is assembled from a linear approximation and a quadratic or proximal term on the neighborhood of the iterate. For a typical nonsmoth convex optimization one wants to solve:

$$\arg\min_x J(x) = f(x) + h(x), \tag{6.4}$$

where $f(x)$ is convex and differentiable and $h(x)$ is convex but nonsmooth, such as ℓ_1 norm.

The proximal operator of a convex function $f(x)$ is defined as:

$$\text{prox}_{f,t}(u) = \arg\min_x \left(f(x) + \frac{1}{2t}\|x-u\|_2^2 \right), \tag{6.5}$$

where $t > 0$ is a scaling parameter. This mapping always exists and is unique for all x. Another important property of this mapping is the following correspondence:

$$x = \text{prox}_f(u) \iff x - u \in \partial f(x). \tag{6.6}$$

In other words, it is possible to characterize elements of the sub-gradient of the function knowing its proximal mapping.

Consider the following quadratic approximation of a function at a point x_k:

$$f(x) = f(x_k) = (x-x_k)^T \nabla f(x_k) + \frac{1}{2}(x-x_k)^T \nabla^2 f(x_k)(x-x_k) \tag{6.7}$$

$$\approx f(x_k) + (x-x_k)^T \nabla f(x_k) + \frac{L}{2}\|x-x_k\|_2^2. \tag{6.8}$$

Here, second order information contained in $\nabla^2 f(x_k)$ is approximated by a diagonal matrix LI, with L being known in literature as the Lipschitz constant.

Replacing this approximation back into Equation 6.4 and using the concept of proximal mapping it is possible to obtain the following iterative scheme:

$$x_{k+1} = \arg\min_x (f(x) + h(x)) \tag{6.9}$$

$$\approx \arg\min_x \left(f(x_k) + (x-x_k)^T \nabla f(x_k) + \frac{L}{2}\|x-x_k\|_2^2 + h(x) \right) \tag{6.10}$$

$$= \arg\min_x \left(h(x) + \frac{L}{2} \left\| x - \left(x_h - \frac{1}{L}\nabla f(x_k) \right) \right\|_2^2 \right) \tag{6.11}$$

$$= \text{prox}_{h,L^{-1}} \left(x_k - \frac{1}{L}\nabla f(x_k) \right). \tag{6.12}$$

The Lipschitz constant associated to the problem is often unknown. In practice, one may interpret it as step s which may accelerate or slow down the convergence of the method. Therefore, the basic iteration scheme becomes:

$$x_{k+1} = \text{prox}_{h,s}(x_k - s\nabla f(x_k)). \tag{6.13}$$

This is the core of the Iterative Soft-Thresholding Algorithm (ISTA) algorithm for the solution of the ℓ_1 regularization problem. In this case $f = \|b - Ax\|_2^2$ and $h = \lambda\|x\|_1$. The proximal mapping for $h(x) = \lambda\|x\|_1$ is called the soft-thresholding operator. Thus:

$$x_{k+1} = \text{soft}_\lambda(x_k - s\nabla f(x_k)), \tag{6.14}$$

where the soft-thresholding operator defined as:

$$\text{soft}_\lambda(x) = \begin{cases} x + \lambda & x \leq -\lambda \\ 0 & -\lambda \leq x \leq \lambda \\ x - \lambda & x > \lambda \end{cases} \tag{6.15}$$

or in the more compact form:

$$\text{soft}_\lambda(x) = \text{sign}(x)\max(0, |x| - \lambda). \tag{6.16}$$

Despite being a classic algorithm and and of simple implementation, this method is often considered to be of low practical applicability due to the generally slow convergence. One possible improvement, which was included into the toolbox implementation of this method is adaptive stepping. As discussed, it is possible to tweak convergence of the algorithm by choosing an appropriate step s. The strategy consist of a backtracking line search, where at each iteration one starts from a small step and keeps increasing it while $J(x_k) > Q(x_k + 1)$, where Q is the quadratic approximation of the function (ARORA, 2004). Since at each iteration the line search allows for big steps while still decreasing the objective function, the convergence tends to accelerate.

As for another take on improving the convergence performance of proximal methods applied to ℓ_1 regularization, Beck e Teboulle (2009) derived a fast variant of the ISTA called Fast Iterative Soft-thresholding algorithm (FISTA). FISTA employs the accelerated gradient theory first proposed by Nesterov (1983). These class of methods are known as momentum-based optimization methods. The reason for this is that they make use of information from previous iterations, as a form of memory, in order to accelerate the solution procedure. In the case of FISTA, an auxiliary y variable is defined which combines the x value of the last two iterations:

$$y_k = x_k + \frac{t_{k-1} - 1}{t_k}(x_k - x_{k-1}) \tag{6.17}$$

$$x_k = \text{prox}_{s,h}(y_k - s\nabla f(y_k)) \tag{6.18}$$

$$t_k = \frac{\sqrt{4t_k^2 + 1} + 1}{2}. \tag{6.19}$$

Differently from the classic ISTA, FISTA is not a monotone algorithm, *i.e.*, $J(x_{k+1})$ may be higher than $J(x_k)$. As with ISTA, FISTA also allows dynamic stepping by employing a backtracking line search.

Based on the work of O'donoghue e Candes (2015), further improvements can be made in order to make FISTA converge faster. The RFISTA or Restarted FISTA is an interesting optimization to the FISTA method. It consists of performing restarts, which essentially erase the momentum memory of the optimization procedure when a certain condition is achieved. More then one condition exists, where they can be fixed at a certain number of iterations or adaptive based on the ongoing optimization. For the implementation in the toolbox, the gradient condition was chosen due to the better reported performance. In this case, the restarts occur whenever:

$$\nabla f(y_{k-1})^T(x_k - x_{k-1}) > 0. \tag{6.20}$$

This restart scheme relies is a computationally cheap observation that allows for a significantly faster descent, as seen in a example problem in Figure 25.

6.1.1.2 ADMM

The alternating direction method of multiplier (ADMM) is prominent optimization framework for the solution of problems involving the sum of convex functions and linear constraints proposed by Boyd *et al.* (2011). It can solve problems of the form:

$$\arg\min_x f(x) + h(x) \iff \arg\min_{x,z} f(x) + h(z) \quad \text{s.t.} \quad x = z. \tag{6.21}$$

The solution method derives from the theory of Lagrange multipliers (ARORA, 2004). It constructs an augmented Lagrangian incorporating the constraints and performs Gauss-Seidel updates on primal and dual variables for iteratively computing approximations of the solution. The augmented Lagrangian, with an assumed parameter ρ and Lagrangian multipliers w is written as:

$$L_\rho(x, z, w) = f(x) + h(z) + \frac{\rho}{2}\|x - z + w\|_2^2 + \frac{\rho}{2}\|w\|_2^2. \tag{6.22}$$

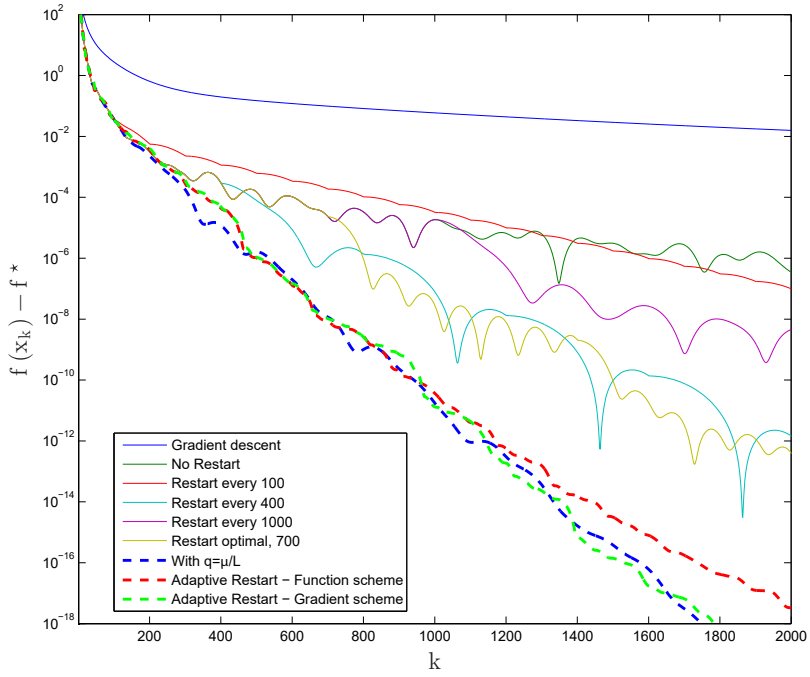


Figure 25 – Comparison of RFISTA restart schemes

Source: Adapted from (O'DONOGHUE; CANDES, 2015)

This expression is employed for the rotation of three basic steps:

1. minimize the Lagragian over x with z and w fixed
2. minimize the Lagrangia over z with x and w fixed
3. update the lagrangian multipliers w.

For the ℓ_1 regularization problem one can write:

$$\arg\min_x \left\{ \|Ax - b\|_2^2 + \lambda \|z\|_1 \quad \text{s.t.} \quad x - z = 0 \right\}, \tag{6.23}$$

where the Lagrangian minimization results in the following updates:

$$x_{k+1} = (A^T A + \rho I)^{-1} \left(A^T b + \rho \left(z - \frac{w_k}{\rho} \right) \right) \tag{6.24}$$

$$z_{k+1} = \text{soft}_{\lambda/\rho} \left(x_{k+1} + \frac{w_k}{\rho} \right) \tag{6.25}$$

$$w_{k+1} = \frac{w_k}{\rho} + x_{k+1} - z_{k+1}. \tag{6.26}$$

As seen, all update terms depend on the parameter ρ. This parameter can significantly influence the convergence of ADMM. A too large value of ρ means not enough emphasis

on minimizing $f + h$, while a too low value does not emphasize enough feasibility. In order to improve the method performance, and address the need for setting the parameter ρ a variant of the method was implemented. The variant allows for and attentively set ρ and and relaxation parameter aimed at improving the solution speed. The method is called Adaptive Relaxed ADMM and was proposed in the work of (XU *et al.*, 2017).

6.1.1.3 Orthantwise Enriched Second Order Method

This method, proposed by Reyes *et al.* (2017), consists of a Newton-like second order optimization. The term orthant is employed to refer to a specific subset of the n-dimensional space where all variables are non-negative. The orthant is connstructed in a way where it defines a feasible region for progressing with the optimization. The method aims to minimize a function ϕ corresponding to the sum of least squares and regularization terms:

$$\arg\min_{x} \phi(x) = \|Ax - b\|_2^2 + \lambda\|x\|_1. \tag{6.27}$$

As stated earlier, the ℓ_1 norm is discontinuous, therefore one cannot compute a gradient directly. However, following the work of Byrd *et al.* (2012), it is possible to compute a discontinuous pseudo-gradient that can be used in a steepest descent optimization procedure. It is written as:

$$\nabla_i \tilde{\phi}(x) = \begin{cases} (A^TA)^{-1}A^Tb_i + \lambda\,\text{sign}(x_i) & \text{if } x_i \neq 0 \\ (A^TA)^{-1}A^Tb_i + \lambda \text{ if } x_i = 0 \text{ and } (A^TA)^{-1}A^Tb_i < -\lambda \\ (A^TA)^{-1}A^Tb_i - \lambda \text{ if } x_i = 0 \text{ and } (A^TA)^{-1}A^Tb_i > \lambda \\ 0, \quad \text{otherwise.} \end{cases} \tag{6.28}$$

The solution, then, employs this pseudo-gradient combined with a quasi-Newton Hessian approximation in order to compute the descent direction. The form of the Hessian comprises the BFGS matrix summed to a generalized Hessian diagonal matrix computed from the generalized derivative of the ℓ_1 norm. The expression for computing the directions reads:

$$(B^k + \Gamma^k)d^k = -\tilde{\nabla}\phi(x^k) \tag{6.29}$$

where k is the iteration number, B is the BFGS Hessian approximation at iteration k, Γ is a ℓ_1 second order approximation term and d are the search directions.

The linear system in Equation 6.29 is solved for the search direction. Once this direction is found a new iterate is calculated using a projected line-search whereas the

projection ensures that the new iterate lies within the solution orthant. The orthant information allows the propagation of the information of the active sets, that is, the variables that have been zeroed and those that can still be updated according to the optimality conditions of the problem.

The orthant directions are defined as:

$$z_i(x) = \begin{cases} \text{sign}(x) & \text{if } x_i \neq 0 \\ 1 & \text{if } x_i = 0 \text{ and } \nabla_i f(x) < -\lambda, \\ -1 & \text{if } x_i = 0 \text{ and } \nabla_i f(x) > \lambda, \\ 0, & \text{otherwise.} \end{cases} \tag{6.30}$$

Following only the orthant directions does not guarantee that all restrictions imposed by the ℓ_1 norm are satisfied. In order to ensure the constraints, an additional projection step is applied to the b vector. Thus, this method could be classified as a projected gradient descent method (ARORA, 2004). The orthogonal projection applied at each component of b can be written as:

$$\mathcal{P}(y)_i = \begin{cases} b_i & \text{if } \text{sign}(b_i) = \text{sign}(z_i), \\ 0, & \text{otherwise.} \end{cases} \tag{6.31}$$

Thus, the iteration scheme can be written as:

$$x_{k+1} = \mathcal{P}(x_k + s_k d_k), \tag{6.32}$$

where d_k is the direction computed from Equation 6.29 and s_k is the step length. In order to improve the efficiency of the method, the step length is calculated employing a projected backtracking line search.

6.1.1.4 Coordinate Descent

The coordinate descent (CD) method exploits the fact that the lasso problem can decomposed into separable uni-dimensional sub-problems. The optimality conditions for the problem can be stated in terms of individual coordinates as:

$$\sum_{i=1}^{n}(b_i - \sum_{k=1}^{p} a_{ik}x_k)a_{ij} + \lambda\,\text{sign}(x_j) = 0, \tag{6.33}$$

where n and p are the number of rows and columns from the matrix A.

Given this formulation, coordinate descent proceeds cyclically solving these equations for all j. In order to simplify the solution statement, the partial residual r is defined as:

$$r_i = y_i - \sum_{k \neq j} a_{ik} x_k. \tag{6.34}$$

This auxiliary variable removes from the outcome the current fit from all but the j-th predictor. Then, the solution to the lasso sub-problem satisfies:

$$x_j = \frac{\text{soft}_\lambda(\sum_{i=1}^{n} r_i x_{ij})}{\sum_{i=1}^{n} a_{ij}^2}. \tag{6.35}$$

Therefore, a solution is obtained after cyclically applying this update to each vector index until convergence.

Seeking further efficiency improvements for the toolbox implementation, the ideas of screening present in the work of (FERCOQ *et al.*, 2015) were also utilized. The basic idea of screening is weeding out from the optimization procedure coordinates that will no longer be active. The author is able to derive safe rules, considering the dual problem and duality gap of the current solution in order to safely exclude certain vector indices from the search space. This variant of the coordinate descent algorithm is known as CD-GAP.

6.1.1.5 Block pivoting

The block-pivoting method of solution is based on the work of (KIM; PARK, 2010). The main idea of the method is to employ an active set strategy considering the optimality conditions of the dual problem. If the main problem is:

$$\arg\min_x J(x, \lambda) = \frac{1}{2}\|b - Ax\|_2^2 + \lambda\|x\|_1 \tag{6.36}$$

then, following a derivation presented in Kim *et al.* (2007), a dual problem can be written as:

$$\arg\min_r J(r, \lambda) = \frac{1}{2} r^T r - b^T r \quad \text{s.t. } \|A^T r\|_\infty \leq \lambda. \tag{6.37}$$

Note that working with the dual problem has the advantage of the objective function being smooth and the constraint being a simple inequality. The method then applies an active set strategy with multiple candidate variables aiming to satisfy the KKT optimality conditions. The KKT conditions for the dual problem and its implications to the primal problem can be written as:

$$w^* = A^T r^* = A^T b - A^T A x^* \tag{6.38}$$

$$-\lambda \leq w^* \leq \lambda \tag{6.39}$$

$$-\lambda < w_i^* < \lambda \implies x_i^* = 0 \tag{6.40}$$

$$w_i^* = \lambda \implies x_i^* \geq 0 \tag{6.41}$$

$$w_i^* = -\lambda \implies x_i^* \leq 0, \tag{6.42}$$

where the star superscript indicates the optimal value. The method then defines three disjoint subsets $(\mathcal{H}, \mathcal{F}_-, \mathcal{F}_+)$ of all p variable indices. Subset $\mathcal{H}$ holds the feasible indices, while the subsets $\mathcal{F}_-$ and $\mathcal{F}_+$ hold the unfeasible indices according to each of the inequality constraints imposed by the infinity norm in Equation 6.37. For convenience it is possible to call the unfeasible indices set as $\mathcal{F} = \mathcal{F}_- \cup \mathcal{F}_+$. The method starts with $\mathcal{H} = \{1, .., p\}$ and $\mathcal{F} = \varnothing$ and employs the following initial assumptions:

$$x_{\mathcal{H}} = 0, \tag{6.43}$$

$$w_{\mathcal{F}_+} = \lambda \tag{6.44}$$

$$w_{\mathcal{F}_-} = -\lambda. \tag{6.45}$$

Then, it is possible to calculate the remaining elements using the KKT conditions as:

$$x_{\mathcal{F}} = (A_{\mathcal{F}}^T A_{\mathcal{F}})^{-1}((A^T b)_{\mathcal{F}} - w_{\mathcal{F}}) \tag{6.46}$$

$$w_{\mathcal{H}} = (A^T b)_{\mathcal{H}} - A_{\mathcal{H}}^T A_{\mathcal{F}} x_{\mathcal{F}}. \tag{6.47}$$

From this point, some restrictions may be violated and in this case some exchange among the working subsets indices is required. If all values obtained satisfy the relations:

$$-\lambda \leq w_{\mathcal{H}} \leq \lambda \tag{6.48}$$

$$x_{\mathcal{F}_+} \geq 0 \tag{6.49}$$

$$x_{\mathcal{F}_-} \geq 0, \tag{6.50}$$

then the value of r corresponding to w is optimal for the dual problem and due to strong duality x is the solution to the primal problem. On the contrary, if there are still unfeasible variables, an update step on all subsets is needed. Defining the set of unfeasible variables as:

$$\mathcal{J} = \mathcal{J}_1 \cup \mathcal{J}_2 \cup \mathcal{J}_3 \cup \mathcal{J}_4 \tag{6.51}$$
$$\mathcal{J}_1 = i \in \mathcal{H} : w_i > \lambda \tag{6.52}$$
$$\mathcal{J}_2 = i \in \mathcal{H} : w_i < -\lambda \tag{6.53}$$
$$\mathcal{J}_3 = i \in \mathcal{F}_+ : x_i < 0 \tag{6.54}$$
$$\mathcal{J}_4 = i \in \mathcal{F}_- : x_i > 0, \tag{6.55}$$

the update rule is written as:

$$\mathcal{H} = (\mathcal{H} - (\mathcal{J}_1 \cup \mathcal{J}_2)) \cup (\mathcal{J}_3 \cup \mathcal{J}_4) \tag{6.56}$$
$$\mathcal{F}_+ = (\mathcal{F}_+ - \mathcal{J}_3) \cup \mathcal{J}_1 \tag{6.57}$$
$$\mathcal{F}_- = (\mathcal{F}_- - \mathcal{J}_4) \cup \mathcal{J}_2, \tag{6.58}$$

where the $-$ sign is interpreted as set difference. The method exchange all variables in $\mathcal{J}$ at each iteration, which gives the name "block" to the method as opposed to single pivoting schemes where a single variable is exchanged at each iteration. Swapping multiple variables at once dilutes the computational costs matrix multiplication and inversion steps and improves convergence. One drawback of this form is that finite termination is not guaranteed since cycles may occur. In order to remedy this rare yet possible issue, a backup rule is enforced which momentarily reverts the algorithm to single swapping if the number of unfeasible variables does not decrease for more than 5 iterations.

An interesting aspect of this form of active set solution procedure is that the there is a clear stopping criterion. Once the number of indices in $\mathcal{J}$ is zero, the algorithm can safely stop knowing that the solution found is optimal. This contrasts with other iterative algorithms which operate on the x vector as whole, such as FISTA, where depending on the desired accuracy the algorithm may stall performing many iterations each of them with very small improvements to the accuracy of the solution.

6.1.1.6 Homotopy

The Homotopy method, proposed by Donoho *et al.* (2006), exploits the fact the ℓ_1 regularization problem has piece-wise linear solutions among different values of the regularization parameter (OSBORNE *et al.*, 2000). Similarly to the Block Pivoting, it belongs to the class of active set methods. The method starts with the solution for the biggest regularization parameter, that is, the all zeros solution and tracks down the breakpoints where variables enter or leave the active set. The method derivation employs the concept of subgradients. The subdifferential of the lasso problem can be written as:

$$\partial f(x) = -A^T(b - Ax) + \lambda \partial \|x\|_1, \tag{6.59}$$

where the subgradient of the ℓ_1 norm term can be written as:

$$\partial \|x\|_1 = \begin{cases} u_i = \text{sign}(x), & x_i \neq 0, \\ u_i \in [-1, 1], & x_i = 0 \end{cases} \tag{6.60}$$

Denoting the support by the index set $\mathcal{I} = \{i : x_i \neq 0\}$ and calling $w = A^T(b - Ax)$ the vector of residual correlations, one can rewrite the expression for the subdifferential as the following two conditions:

$$w_{\mathcal{I}} = \lambda \,\text{sign}(x_{\mathcal{I}}), \tag{6.61}$$

$$|w_{\mathcal{I}^c}| \leq \lambda, \tag{6.62}$$

where $\mathcal{I}^c$ consists of the complementary index set from I. These relations state that residual correlations on the support must all have magnitude equal to λ and signs must match the corresponding elements of x, whereas residual correlations not present in the support must have magnitude less than or equal to λ. Note that the two conditions obtained from the subdifferential are essentially the same derived from the KKT conditions for the dual problem in Block Pivoting strategy described in subsubsection 6.1.1.5, only written in a more compact form. The solution strategy here differs, however, making use of a more geometrical approach. The Homotopy algorithm traces the optimal x path for a certain λ while maintaining both conditions. The algorithm starts with inital all zeros solution maintain the $\mathcal{I}$ index set satisfying:

$$\mathcal{I} = \{j : |w_j| = \|w\|_\infty = \lambda\}, \tag{6.63}$$

which is directly implied by the subdifferential conditions. Then, Homotopy computes an update direction d by solving:

$$A_{\mathcal{I}}^T A_{\mathcal{I}} d_{\mathcal{I}} = \text{sign}(w_{\mathcal{I}}), \tag{6.64}$$

for $d_{\mathcal{I}}$ and setting $d_{\mathcal{I}^c}$ to zero. This direction is constructed in order to ensure equal decrease to the residual correlations of all variables in the active set. As usual with descent algorithms, once there is a direction there must be a step length. The step here is cleverly calculated as the breakpoint where either a new variable enters the active set or leaves it. The first case the may occur is when an inactive element of w increase the magnitude beyond λ, which violating the conditions. This first occurs when:

$$s^+ = \arg\min_{i \in \mathcal{I}_c} \left(\frac{\lambda - w_i}{1 - a_i^T A_{\mathcal{I}} d_{\mathcal{I}}}, \frac{\lambda + w_i}{1 - a_i^T A_{\mathcal{I}} d_{\mathcal{I}}} \right), \tag{6.65}$$

where a_i is the i-th column of the matrix A. The other possibility occurs when an active coordinate crosses zero and violates the sign agreement. This first occurs when:

$$s^- = \arg\min_{i \in \mathcal{I}} -\frac{x_i}{d_i}. \tag{6.66}$$

Once both steps have been computed, the chosen one corresponds to the minimum step, *i.e.*:

$$s = \min(s^+, s^-). \tag{6.67}$$

From here, the index associated with the chosen step is employed in updating the active set, either by appending it to $\mathcal{I}$ if the smaller step was s^+ or removing it if the smaller step was s^-. The next solution for each iteration k is computed as:

$$x_{k+1} = x_k + s_k d_k, \tag{6.68}$$

which is run until $\|w\|_\infty = 0$, which marks that the optimal value of x has been found.

6.1.1.7 Dual Projected Newton Method

The Dual Projected Newton Method (DPNM) is a ℓ_1 solution procedure based on convex duality proposed by Gong e Zhang (2011). The corresponding dual problem of a convex optimization problem is obtained by replacing the inequality constraints with equality constraints, and replacing the variables with their corresponding dual variables. The objective function is also replaced by the dual objective function, which is the pointwise infimum of the original objective function with respect to the primal variables.

In order to define the dual problem, the method first starts introducing an auxiliary variable to the primal problem:

$$\arg\min_{x,y} \frac{1}{2}\|b - Ax\|_2^2 + \lambda\|y\|_1 \quad \text{s.t. } x = y \tag{6.69}$$

From here, one can incorporate the constraint into the objective function employing a vector of Lagrange multipliers μ and write the following Lagrangian function:

$$L(x, y, \mu) = \frac{1}{2}\|b - Ax\|_2^2 + \lambda\|y\|_1 + \mu^T(x - y) \tag{6.70}$$

By minimizing the Lagrangian in Equation 6.70 with respect to variables x and y, one can obtain the dual objective function:

$$D(\mu) = \arg\min_x \left(\frac{1}{2}\|b - Ax\|_2^2 + \mu^T x\right) - \arg\max_y(\mu^T y - \lambda\|y\|_1). \tag{6.71}$$

Utilizing the concept of dual norms introduced in Boyd e Vandenberghe (2004), which proves that the dual norm of the ℓ_1 norm is the infinity norm, one can rewrite the second term of the expression as:

$$\arg\max_y(\mu^T y - \lambda\|y\|_1) = \begin{cases} 0, & \|\mu\|_\infty \leq \lambda \\ +\infty, & \|\mu\|_\infty > \lambda \end{cases}. \tag{6.72}$$

Plugging the above expression into Equation 6.71 and maximizing it with respect to the dual variable μ, the following dual problem is obtained:

$$\arg\min_\mu D(\mu) = \left(\frac{1}{2}\mu^T H\mu + (HA^T b)^T\mu + const\right) \quad \text{s.t. } \|\mu\|_\infty \leq \lambda. \tag{6.73}$$

where $H = (A^T A)^{-1}$. The solution method is based on a Newton method, aiming to achieve quadratic convergence rate:

$$\mu_{k+1} = (\mu_k - sH^{-1}\nabla D(\mu_k))^+, \tag{6.74}$$

where s is a step size and $(.)^+$ is the Euclidean projection of x onto the box constraint, written as:

$$(x)^+ = \text{sign}(x)\min(\lambda, |x|). \tag{6.75}$$

Once convergence is reached and an optimal μ is found, the primal value of x^* can be found as:

$$x^* = H(A^T b - \mu^*). \tag{6.76}$$

6.1.1.8 Performance of implemented methods

The performance of the implemented sparseness-inducing regularization procedures was evaluated on data samples similar to real-world B-WIM signals. The methods were compared in terms of number of iterations and time taken to achieve machine precision accuracy of the solution on 150 logarithmically spaced regularization parameters. The objective was to see which methods performed better and which were sensitive to increase or

decrease in the regularization parameter. The overall relative performance of the methods for solving the problem can be assessed in Figures 26-29.

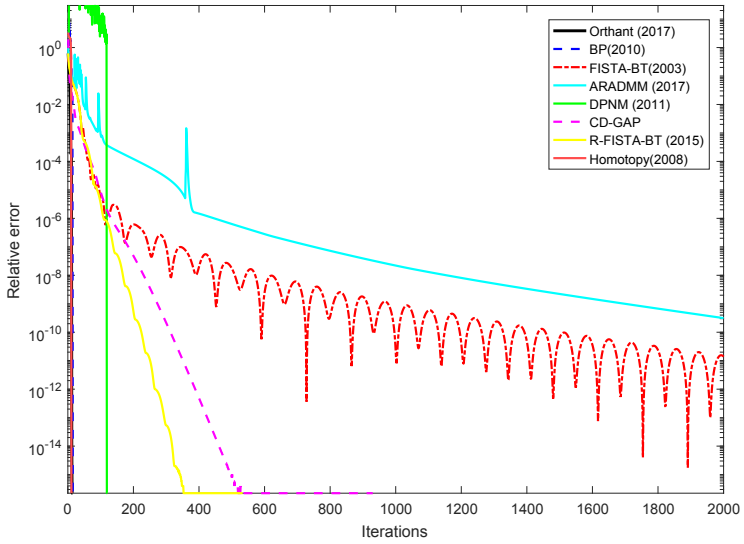


Figure 26 – Change in relative error with iterations

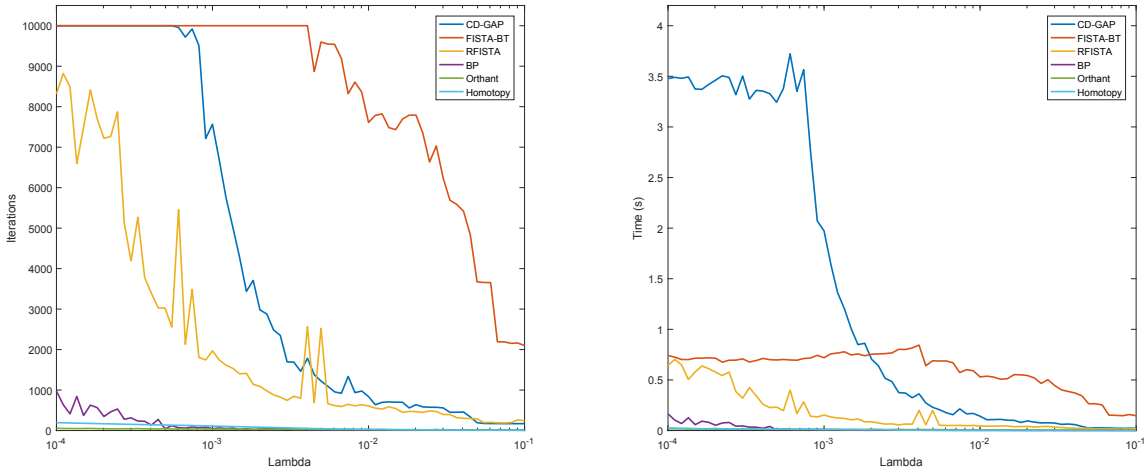


Figure 27 – Variation of number iterations and time with λ

Among proximal methods, ISTA presented the worst performance, being surpassed by the fast variants. The use of a backtracking line search for selecting a step has shown improved performance on all proximal methods and was always activated. Due to low performance and high number of iterations required to achieve the desired precision, ISTA was not included in the presented statistics. The restart scheme proposed RFISTA has shown to significantly improve the search speed for the optimal solution.

The Coordinate Descent method presented poor performance and was the slowest methods even with the implementation of gap-safe rules for ruling out some variables. This is probably due to need to cyclically iterate on all variables and each of the updates

relies on a matrix by vector multiplication which may be expensive depending on the size of the matrices.

The DPNM presented drawbacks related to its dual problem nature. Since it exploits duality to solve an easier problem, each iteration does not correspond to an estimate of the x solution, rather an estimated of the solution to the dual problem. This can be seen from the graph in Figure 26 where during many iterations the algorithm has high error measurement and has a sudden jump to the correct solution. This is due to the fact that only at the optimal point of the dual function, given the strong duality property, one can recover the optimal x. Therefore, strategies where early termination is employed once a certain tolerance is reached may not be possible.

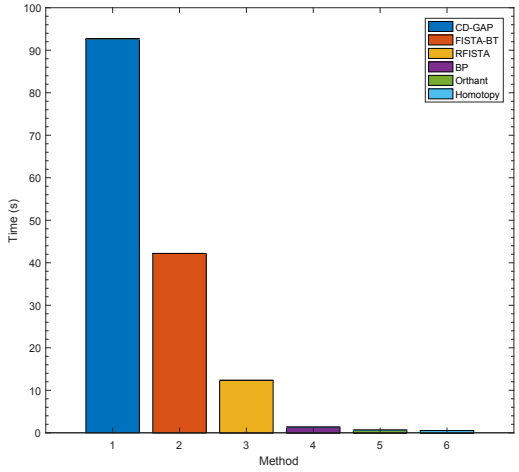


Figure 28 – Total time of across a range of λ

The ADMM implemented in its ARADMM variant has show a slow convergence curve, even slower than FISTA. This, combined with the fact that the method relies on good internal parameters such as ρ and the relaxation constant, makes the method be considered one of the worst methods ℓ_1 solution methods from those being evaluated.

The Orthant method enjoys fast convergence due to it belonging to the class of second order methods. Nevertheless, numerical experimentation has shown that it is a method that may get stuck and not converge. This is due to the fact second order information is obtained from an approximation of the Hessian and it needs to be inverted. This inaccurate Hessian inversion, especially considering the ill-posed matrices associated with regularization problems, may result in numerical instability and lack of convergence. The same drawback also applies to DPNM.

The BP method is very distinct from the other methods as it belongs to the class of active set methods. This property allows for an efficient implementation based on boolean arrays for performing the necessary swaps between each of the index sets. In terms of solution performance it has shown be very competitive across most λ only starting to slow

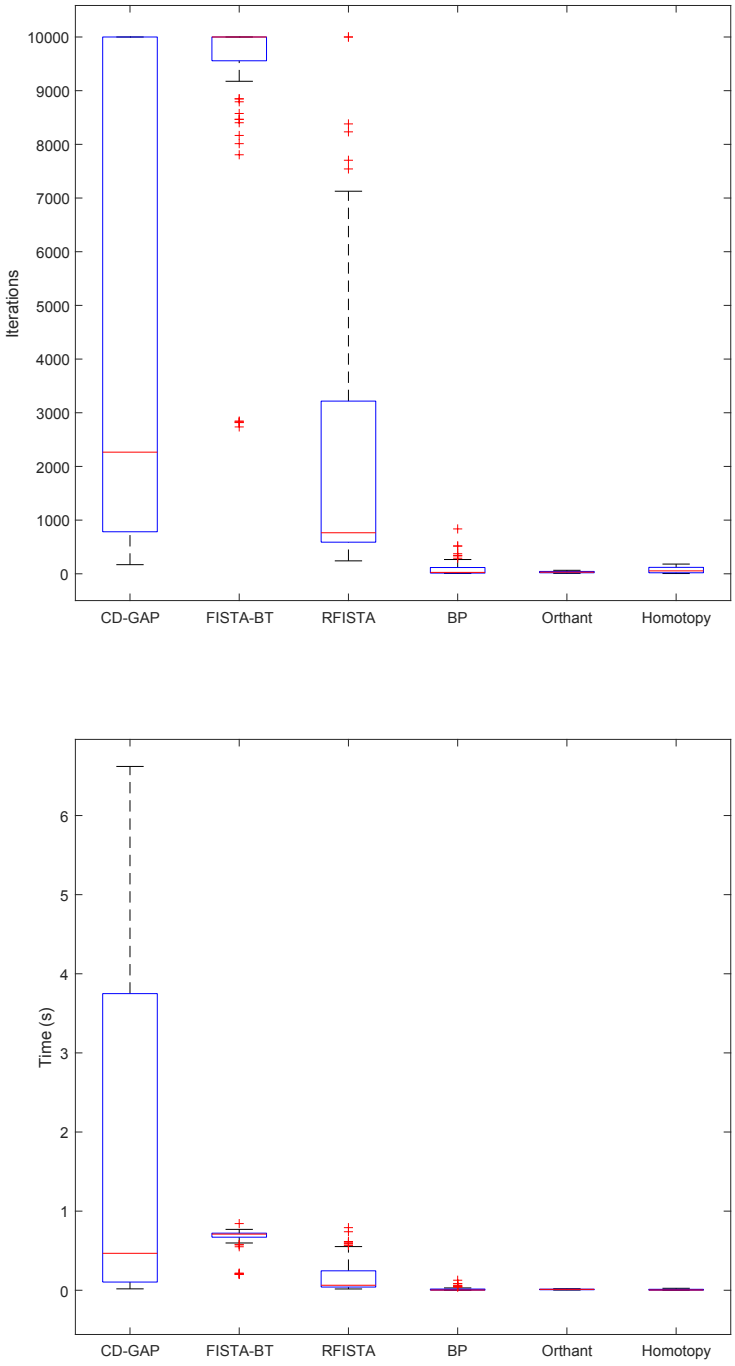


Figure 29 – Box plot of the number of iterations and time

down with small regularization parameters which ends up in a higher amount of swaps required.

The Homotopy method achieved the best performance across the evaluated methods, achieving the optimal solution quickly across the complete range of regularization parameters.

From the conducted experiments, both Block Pivoting and Homotopy algorithms

were considered the most advantageous for real-world applications, mainly due to robustness, lack of tuning parameters, ability to give highly accurate solutions up to machine precision and evenness in terms of computational costs for either high or low regularization parameters. In the remainder of numerical experiments, the Homotopy method was employed for solving regularization problems. The Orthant methods also showed impressive performance, however, it was deemed not as reliable and prone to failure depending on the matrix conditioning.

These results demonstrate the effectiveness of the toolbox and its potential for solving L1 regularized problems in a wide range of engineering applications.

6.1.2 Impulse Vector Identification in BWIM

After having at disposal a good set of tools for finding sparse solutions employing regularization procedures, it becomes important choosing an appropriate sparse solution. In this regard, having a fast and reliable method for finding the parameter λ which allows for recovering the B-WIM impulse vector becomes important.

As discussed in section 2.8 there are multiple ways for finding an optimal λ value to use in regularization. In the context of B-WIM application those methods were implemented and compared against B-WIM-like signals. The methods chosen were the 10-fold cross-validation and, due to cheaper computational performance, the information based models. The L-curve method was not considered to the difficulties associated with finding the corner of the curve reliably for different situations. After initial experiments the AIC method was seen to perform poorly. It is indicated in literature that AIC may perform poorly when the number of non-zero elements is low. For this reason it was substituted by a variant called AICc, which is a bias-correct derivation accounting for small samples developed in (HURVICH; TSAI, 1989). From further experiments, the results from BIC and AICc were identical.

The testing started with the a known impulse vector, called Oracle, and from this a regularization procedure was performed on the respective convolution signal for each λ on a grid of spaced logarithmically regularization parameters. The result in Figure 30 comprises the information-based methods and 10-fold cross validation. It is possible to see that in B-WIM-like data the correct support is identified in all methods however with the presence of spurious smaller magnitude variables. The cross-validation was the method that introduced the bigger amount of extra variables. Considering the more theoretically advanced properties of the EBIC criterion over its counterparts, which generally implies more sparse solutions, the criterion was chosen as default for B-WIM application.

Nevertheless, a procedure was conducted in order the remove the spurious smaller variables. Considering the real problem of finding weights of axles in a vehicle it is not often that one axle is much heavier than the others. Thus, it is possible to apply a thresholding

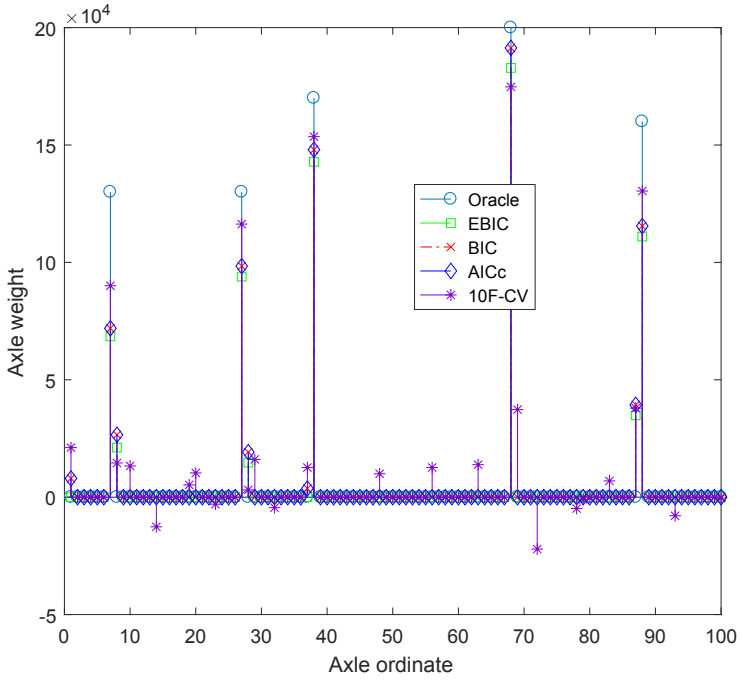


Figure 30 – Lasso solutions for multiple λ selections strategies

procedure based on the maximum vehicle axle in order to remove those spurious smaller magnitude variables and obtain the correct support. This idea has ground in literature, specifically the work of Slawski e Hein (2012) which shows that a useful way to obtain a sparse solution without regularization is to first employ a non-negative least squares procedure followed by a thresholding step to remove the small values.

6.2 ERROR IN VARIABLES

Most solution methods reviewed in this study assume noise only at the observation vector. However, as discussed in section 3.3, the convolution interpretation allows for a dual view of the B-WIM problem, that is, both weighing unknown vehicles assuming perfect knowledge of the influence line as well as obtaining an influence line assuming perfect knowledge of the vehicle characteristics are complementary sides of the same problem. In reality, however, this perfect knowledge is not guaranteed. Influence lines depend on the current bridge conditions and are estimated from a limited set of calibration vehicles with all the inherent assumptions of the model such as constant speed of the vehicle, Euler-Bernoulli bridge behaviour, force application varying only along the direction of the crossing vehicle, etc. Thus, in B-WIM theory there is a situation where data matrix is assumed to be error-free while it is known that such matrix follows a certain structure constructed from data vectors that are merely estimated to begin with. From a theoretical point of view, it can beneficial to newer solutions methods to tweak these assumptions in order to allow for a model that more closely corresponds to reality.

One of the ways statistical literature addresses the issue of errors on both output and data matrix is employing the method known as Total Least Squares (TLS) (GOLUB; LOAN, 2013). The "total" in the name refers to the fact that all variables are assumed to contain errors, which is a more realistic premise for real-world engineering applications such as the B-WIM problem. This method consists of a generalization of the least squares procedure where another error variable is introduced in the form of a correction matrix. It has been applied in different fields of science such as system identification, astronomy, computer vision, and econometrics (MARKOVSKY; HUFFEL, 2007). For the usual $Ax = b$ system this becomes:

$$(A + E)x = b + e \tag{6.77}$$

where E and e express the error data matrix and and error data vector, respectively. The formulation for the TLS then can be written as:

$$x_{TLS} = \underset{x,e,E}{\arg\min} \|[E, e]\|_F^2 \quad \text{s.t.} \quad (A + E)x = b + e \tag{6.78}$$

where $\|x\|_F$ is the Frobenius norm.

The total least squares problem can be reformulated in terms of finding a rank-deficient matrix C which approximates the matrix A under the Frobenius norm:

$$\underset{C,y}{\arg\min} \|A - C\|_F, \quad \text{s.t. } Cy = 0, \quad y^T y = 1. \tag{6.79}$$

Recasting the problem on this form allows to show that the solution for Equation 6.79 can be obtained from SVD. The solution is obtained from the singular triplet (x, y, τ), where τ is the smallest singular value obtained from the SVD operation (GOLUB; LOAN, 1980). Therefore, SVD can be directly applied to solve the TLS optimization . One drawback of the TLS method is that for applications where the problem has an inherent structure, the solution may be inappropriate since it does not preserve any existing structure in E (JIANG; BERRY, 2000).

6.3 STRUCTURE

As seen throughout this study, matrices are employed almost everywhere, to model the B-WIM problem and analyze data. An important aspect of matrices is their structure, which can greatly impact the efficiency and accuracy of algorithms used to solve problems. The specific structure of a matrix depends on the problem being solved and the underlying mathematical model. For the case of B-WIM, literature has not yet explored the potential for exploiting the matrix model structure in order to propose novel solution procedures.

As discussed in section 3.3, under the convolution interpretation, the data matrices involving the B-WIM problem follow a definite structure. Consider for example a vehicle traversing over the bridge. It applies a constant force on the bridge while it moves along the bridge end direction. Assuming constant speed, at each time step this force gets displaced by a constant amount. In terms of the impulse matrix, it results in a matrix with constant terms along diagonals, *i.e.* a Toeplitz matrix. Knowledge on the structure of the matrix employed for performing the inverse operation, the deconvolution, is something that has not been addressed in current solution methods.

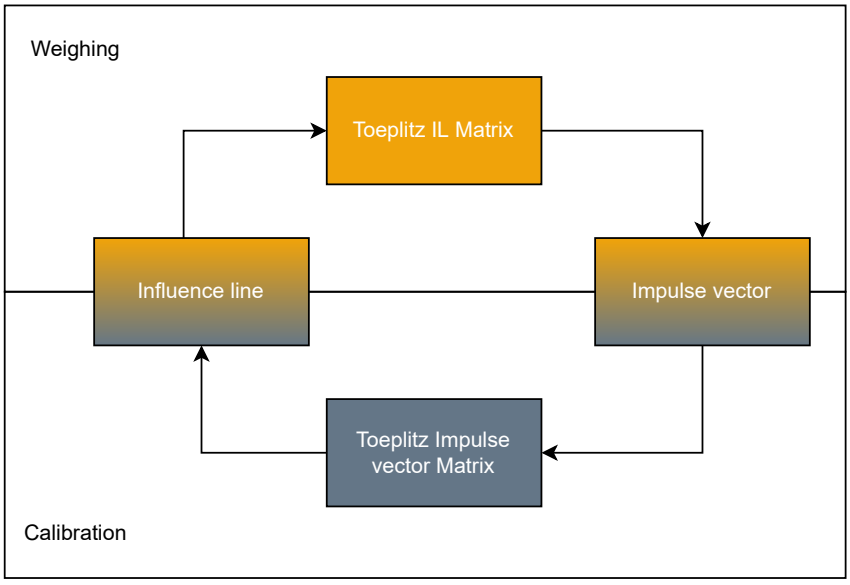


Figure 31 – Toeplitz structure present in B-WIM complementary problems

The Toeplitz structure is part of both the weighing and calibration procedures, as seen in Figure 31. Thus, it can be useful to restate the B-WIM problem in terms of the Toeplitz restriction on the convolution matrix, as well as accounting for possible errors in this matrix and the on the measured data. That corresponds to a formulation known as Structured Total Least Squares (STLS) (MOOR, 1993; LEMMERLING *et al.*, 2003). The problem formulation is very similar to the TLS with the additional constraint on the structure of either the data matrix or observations. That is:

$$x_{STLS} = \underset{x,z,E}{\arg\min} \|[E, z]\|_F \quad \text{s.t.} \quad (A + E)x = b + z \tag{6.80}$$

and [E z] with the same linear structure as [A y] (6.81)

The solution results by taking $[Ez]$ as the matrix with smallest Frobenius norm that added to $[Ab]$ makes the result singular. As such, it is related to the smallest singular value of $[Ab]$ (DUMITRESCU, 2013).

6.4 COMBINING AVAILABLE KNOWLEDGE

As seen in previous sections, considerations on noise form, sparseness, and structure of the problem may be exploited for providing a more comprehensive and flexible model for the solution of the B-WIM problem.

Up to this point, no study has proposed a combined effort of incorporating error in variables, sparseness-promoting metrics and structure preserving solution procedures for addressing neither weighing nor calibration on B-WIM data. This study aims to employ such methods and evaluate their performance for the same numerical data provided in chapter 5.

The work of Kazemi (2018) proposes a procedure for the solution of a inverse problem with similar characteristics to the one found in B-WIM applications. It consists in finding a a reflexivity series given noise-corrupted seismic recordings. It also aims at incorporating the prior knowledge of the sparseness of the reflexivity series. Such a description has an analogous counterpart to finding the sparse impulse vector given the moment measure readings in B-WIM.

The problem statement consists on a variant of the STLS problem, where an additional sparseness promoting ℓ_1 term is included into the problem. It reads:

$$x_{TSTLS} = \underset{x,E}{\arg\min} \|(A+E)x - y\|_2^2 + \lambda\|x\|_1 \quad \text{s.t.} \quad Ex = Xe \tag{6.82}$$

where X is the Toeplitz-strucutured matrix associated to the x vector and e is the vector associated to the Toeplitz-strucutred E matrix. In order to solve the problem in Equation 7.1 one can expand the cost function around x and E, ignore higher-order terms and incorporate the constraint into the cost function. With this modification the problem statement reads:

$$\{\hat{\Delta}x, \hat{\Delta}E\} = \underset{\Delta x,\Delta E}{\arg\min} \|r - \Delta Ex - P\Delta x\|_2^2 + \lambda\|x + \Delta x\|_1 \tag{6.83}$$

where Δx is the perturbation in x, ΔE is a perturbation in E, $P = A + E$ and $r = Px - y$. Using the structure constraint, one can replace ΔEx by $X\Delta e$, which results in the following unconstrained minimization problem:

$$\{\hat{\Delta}x, \hat{\Delta}e\} = \underset{\Delta x,\Delta e}{\arg\min} \|res - X\Delta e - P\Delta x\|_2^2 + \lambda\|x + \Delta x\|_1. \tag{6.84}$$

Since there two variables being optimized, one of the most straightforward ways to solve it is to perform an alternating minimization procedure. Considering Δx a constant, one can take the gradient of the expression and obtain a expression for the optimal Δe

$$\Delta e = (X^T X)^{-1} X^T (res - P\Delta x), \tag{6.85}$$

with this estimate one can consider now this quantity fixed and solving Equation 6.84 for Δx, which requires a ℓ_1 regularization solver such as those presented in subsection 6.1.1. After obtaining an estimate for Δx is it possible to update the initial values of x and e and proceed with alternating procedure obtaining newer estimates for each variable until convergence.

The method proposed by Kazemi (2018) was implemented and tested for solving the B-WIM problem on the analytical database. However, issues related to convergence were observed when applying to the B-WIM data. One of the observed drawbacks of the method is the joint optimization of both sparsity constraints and structured total least squares constraints. Considering that the sparsity, *i.e.*, the number of nonzero elements on the impulse vector keeps changing throughout the iterative procedure, the solutions being employed on each of the alternating steps of the optimization may change drastically, which can lead into further digression from the starting points, causing degenerate solutions and infinite loops. Nevertheless, the ideas introduced by Kazemi (2018) in the field of seismic response are theoretically promising. This fact prompted the author to further investigate ways to adapt this idea into the B-WIM context.

7 PROPOSED METHOD: SPARSE AND STRUCTURED TOTAL LEAST SQUARES

Since empirical testing demonstrated that the joint optimization the impulse vector sparseness, structure and error in variables modeling was a hard task in the context of the usual B-WIM data, one practical idea explored was the development of a two-phase algorithm. The idea consists of first addressing the sparseness consideration, by finding the correct support. Once the support, the location of non-zero elements is found, then the main objective is proceed with the structured optimization, taking into account the inherent Toeplitz structure established from the convolution interpretation and accounting for the error in variables modeling.

As seen in section 6.1, ℓ_1 regularization aided by a parameter selection procedure and posterior threshold post-processing can be employed for recovering an impulse vector support. Therefore, in the proposed two-phase algorithm, the first phase consists of finding the support by ℓ_1 regularization. The decoupling of sparseness search from the full impulse vector recovery procedure offers an advantage. The regularization procedure does not need to find the correct impulse values, only the correct location of the nonzero elements. This eases an issue often found in ℓ_1 regularization literature which is the shrinking of iterates (MEINSHAUSEN; YU, 2009). Since initially one is interested only on the location of the sparse items, it does not matter if the values obtained are smaller than they should be.

Once the support has been selected by ℓ_1 regularization, the proposed algorithm starts its second phase. In this phase, the two points of improvement identified are addressed: error in variables and structure. For this, once more the STLS method may be applied. However, at this time the sparseness has already been considered which allows for a slightly different formulation. For each estimate of the new x in the STLS algorithm, only the subset of non-zero elements is considered. From this an estimate of the vector of structured errors e may be found for starting an alternating optimization procedure. In order to find e however, a new derivation must be made taking into account the underlying model constraints. The proposed formulation aims to minimize both the residual and the error in variables, enforcing structured variable errors:

$$x_{STLS} = \arg\min_{x,E} f = \frac{1}{2}\|(A+E)x - b\|_2^2 + \|e\|_2^2 \quad \text{s.t.} \quad Ex = Xe. \tag{7.1}$$

writing $r = b - (A+E)x$ and the differential of f:

$$f = \frac{1}{2}(r : r + e : e) \tag{7.2}$$
$$df = r : dr + e : de \tag{7.3}$$
$$= r : -Xde + e : de \tag{7.4}$$
$$= (-X^T r + e) : de. \tag{7.5}$$

Therefore:

$$\frac{\partial f}{\partial e} = -X^T r + e. \tag{7.6}$$

Equating the gradient to zero and solving for e it is possible to obtain:

$$e = X^T r \tag{7.7}$$
$$e = X^T(b - Ax - Ex) \tag{7.8}$$
$$e = X^T(b - Ax - Xe) \tag{7.9}$$
$$e = X^T(b - Ax) - X^T Xe \tag{7.10}$$
$$(X^T X + I)e = X^T(b - Ax) \tag{7.11}$$
$$e = (X^T X + I)^{-1} X^T(b - Ax). \tag{7.12}$$

From this expression for e which enforces the Toeplitz structure into the TLS formulation, one can derive an estimate for x considering the sparse subset of indices:

$$df = r : -(A + E)_s dx \tag{7.13}$$
$$= -P_s^T r : dx. \tag{7.14}$$

where the s subscript indicates the matrices columns associated with the non-zero values and $P = A + E$. Thus:

$$\frac{\partial f}{\partial x} = -P^T r. \tag{7.15}$$

Equating the gradient to zero and solving for x_s:

$$-P_s^T(b - P_s x) = 0 \tag{7.16}$$
$$-P_s^T b + P_s^T P_s x = 0 \tag{7.17}$$
$$x = (P_s^T P_s)^{-1} P_s^T b. \tag{7.18}$$

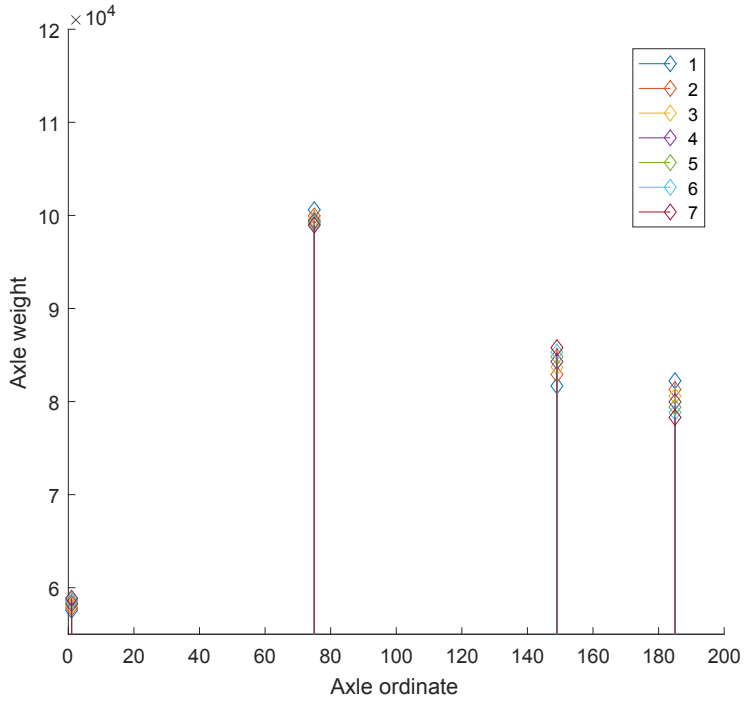


Figure 32 – Range of change across alternating iterations

Thus, in the second phase of the algorithm alternate estimates for e and x are performed. A new estimate for e augments the matrix P which contains the matrix E. The matrix E is formed by the concatenation of errors in the structured form, *i.e.*, for the B-WIM case, $E = \text{Toeplitz}(e)$. Each new estimate of x is sparse vector estimated from the augmented matrix P. One issue that may arise in this setup is the fact that despite converging, each new estimate of x may trade individual-point accuracy for global accuracy. In terms of B-WIM this means that the resulting impulse vector may be further increasing errors by axle at the expense of slightly tweaked GVW. This principle can be observed in Figure 32. The figure illustrates how the impulse vector changes across 7 alternating iterations. Note that the last 2 pulses present similar range of change, however at opposing directions. The optimal x value may be in between these ranges, such that a form of early termination to the procedure may be beneficial for more accurately estimating individual axles.

Considering that individual axle estimates are important in context of correct vehicle representation and classification, some form of selection criterion may be employed for choosing a good a trade-off between generality of the solution and accuracy. In this regard it is possible to adapt the GCV idea employed in the work of Dupé *et al.* (2009). There the authors present of form of generalized cross validation applicable for sparse signal recovery. In that study it is employed in the context of finding the optimal regularization parameter λ. However, since every alternating step incurs in a deviation from the starting least squares solution obtained when $e = 0$, then early termination of the algorithm can

be viewed as a form of regularization. In fact, early termination has been employed as a strategy by several authors in sparse recovery literature (GAZZOLA; NAGY, 2014). Therefore, the optimal x is considered the one which minimizes the following measurement:

$$GCV(k) = \frac{\|b - A\ x_0 - X\ e_k - (A + E)\ x_k\|_2^2}{n - \|x\|_0}, \tag{7.19}$$

where k is the current iteration, n is the number of rows in A and $\|.\|_0$ is the ℓ_0, or the number of non-zero elements on the vector.

7.1 NUMERICAL EXPERIMENT

Since the proposed method can be seen as a variant of the least-squares method with improved robustness and theoretical considerations it is worth comparing the performance of the proposed method by numerical experiments.

The method was evaluated on the publicly available vehicle simulation dataset described in subsection 5.1.1 (GONÇALVES *et al.*, 2019). The evaluation is applied to all the 360 calibration vehicles in the dataset and comprises both GVW and axle calculations for three conditions of bridge length and roughness amplitudes. The form of error considered is the absolute mean across all simulations available. The absolute value is employed in order to better verify the changes in the mean between methods, avoiding the effects of negative and positive errors canceling each other. The results are presented in the form of boxplots, in order to better highlight both the average variation, amplitude of errors and presence of outliers. In Figures 33 and 34 the results consider the GVW recovery across different road class and bridge length scenarios. Overall it is possible to say that the performance of the proposed STLS optimization is very similar to that of Matrix Method for GVW estimation. It is important to notice that a boxplot with a length fixed conveys all simulations for all distinct surface roughness. This explains how in Figure 33 the average results for a length 20m bridge are smaller than those for the length 10 m bridge. This is caused by increased sensitivity of the higher roughness on the small bridge as discussed in Carraro *et al.* (2019). In terms of the STLS improvements, slight decreases in the amplitude of errors and presence of outliers have been observed.

More noticeable improvements can be seen when considering the error by axle. From Figure 35 it is possible to notice slightly worse performance on the 10m bridge, probably caused by the amplification of dynamic effects caused by surface roughness consideration which are not diluted across a bigger span and are not directly taken into account by the model. Nevertheless, continuing the analysis for the 20 and 30m bridge it is possible to observe significant decrease on the absolute errors both at average value and dispersion. A similar remark can be made when considering the distinct road profiles in Figure 36. For the RoadClassI which is the situation of amplitude zero, not taking

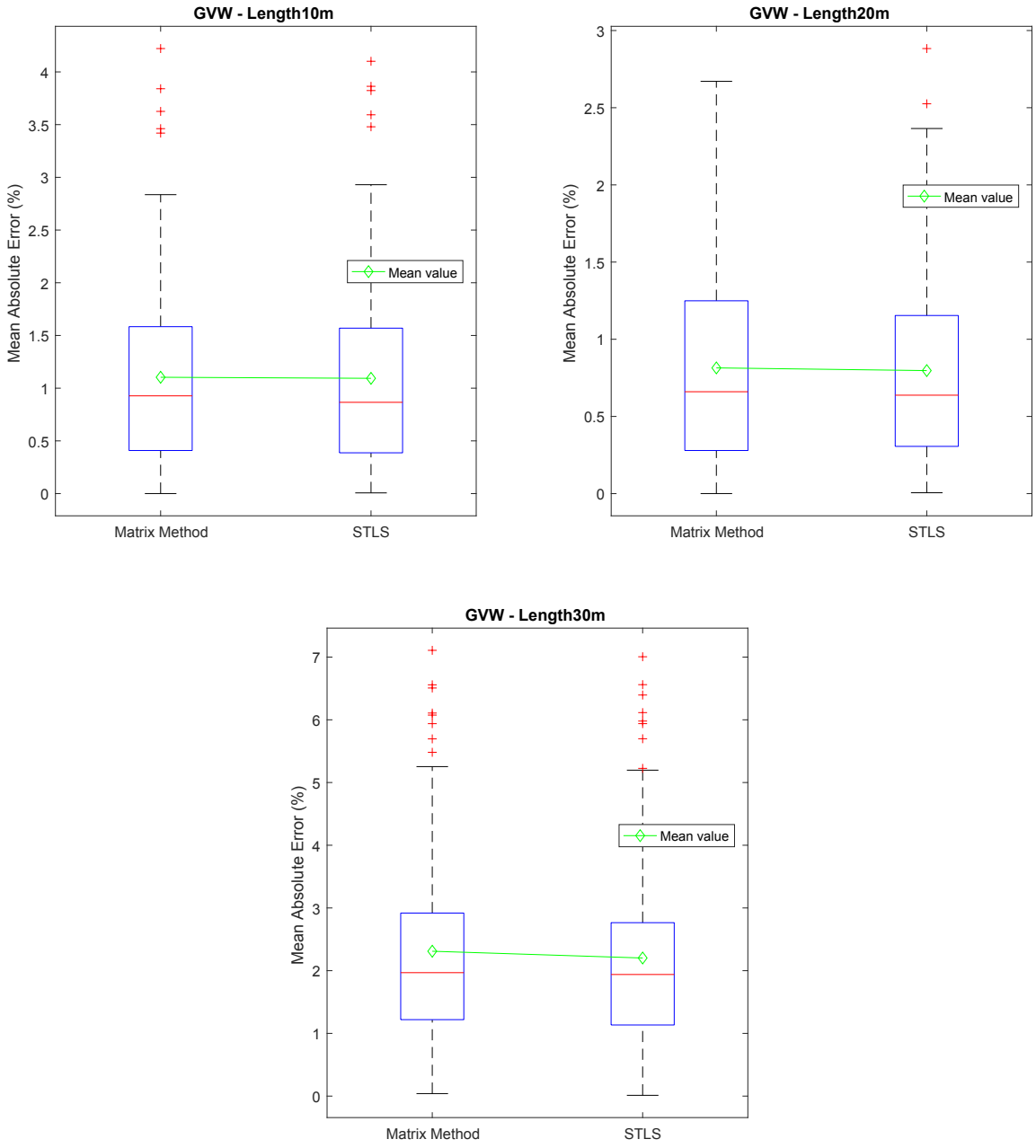


Figure 33 – Average absolute error for GVW - Lengths

into account the roughness, it is possible to observe that the outliers seem to cluster at a lower value, which is beneficial when considering the minimization of the maximum error. Moreover, for RoadClassA and RoadClassB a drop in the mean average absolute error (green line) is clear.

Lastly, the overall trend of the STLS method can be observed from Figures 37 and 38. It can be seen from Figure 37 that the GVW results are not much affected by the introduction of the proposed STLS method. On the other hand, in Figure 38 it is possible to notice improved single-axle performance when performing the STLS scheme and taking into account the structure of the problem. Thus, it is possible to conclude that

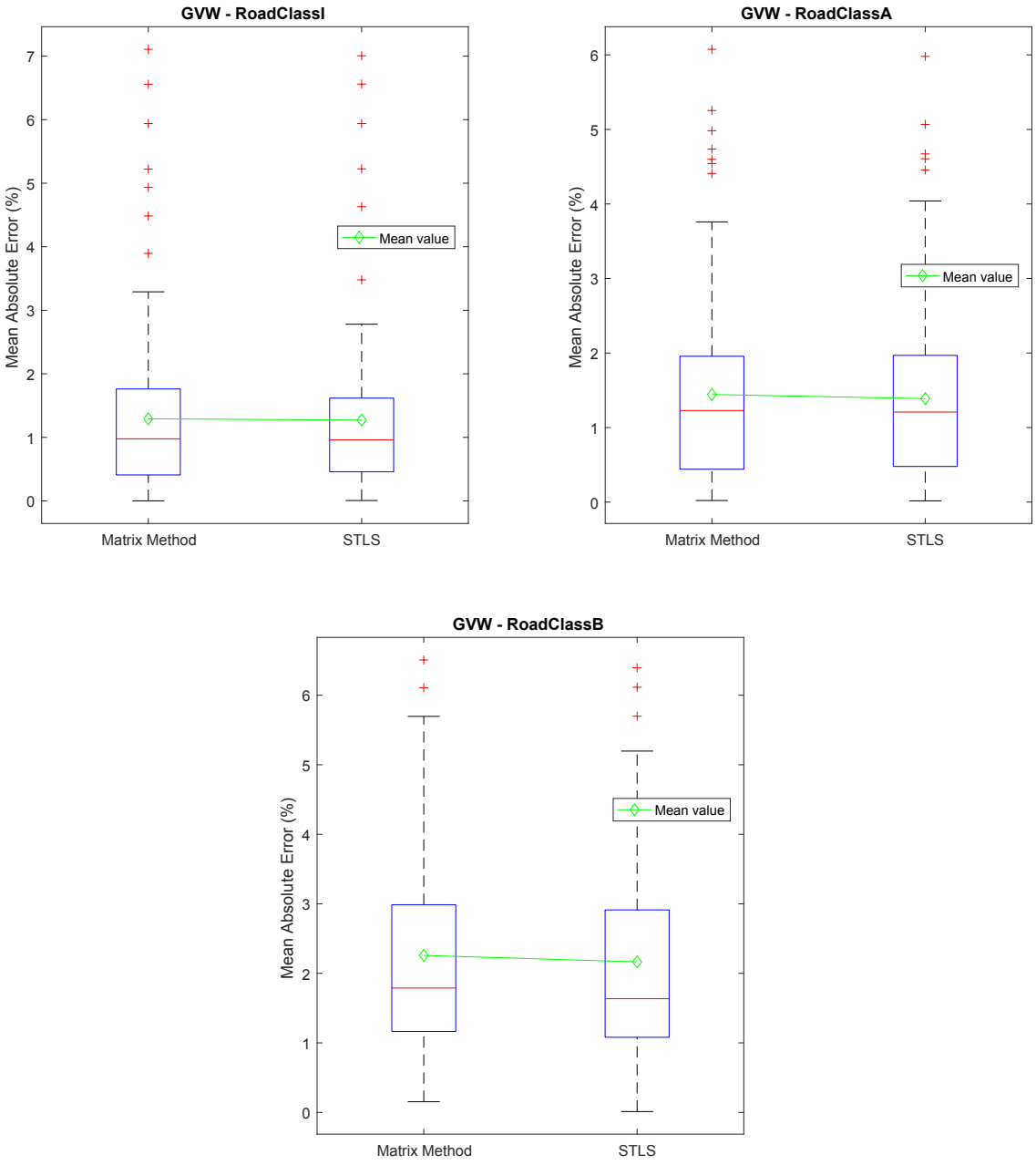


Figure 34 – Average absolute error for GVW - RoadClass

the proposed method, comprised of computationally efficient sub-algorithms is able to improve one of the main drawbacks of traditional B-WIM solutions, which is single-axle accuracy and doing this without sacrificing the overall performance for GVW estimation.

7.2 FIELD-TESTING DATA

Additionally to the numerically simulated experiments, data from a real instrumented bridge in Goiás was evaluated. The Itinguijada bridge, already employed in the comparison of static weighting algorithms, was utilized. In order to validate the algorithm

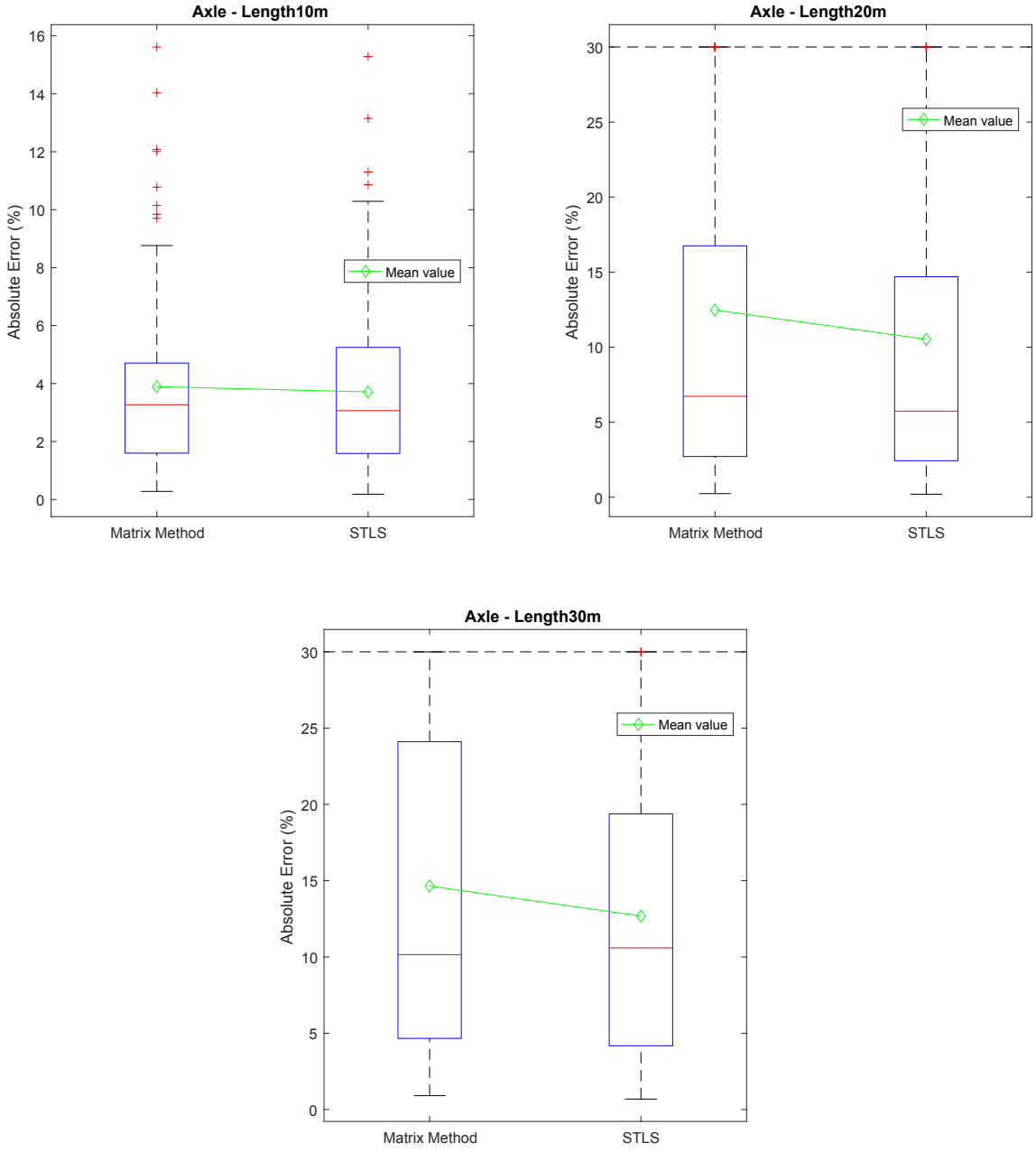


Figure 35 – Average absolute error for Axle - Lengths

one needs conduct the analysis knowing in advance the characteristics and the weight of the vehicle traversing the bridge. For this reason, data from B-WIM calibration and influence line extraction was employed. In this setting, multiple runs of two known vehicles were utilized. Two vehicles were employed in calibration with 3 and 5 axles. A total of 31 calibration runs were recorded. Due to the small number of events, further research should be conducted in order to more assertively assess the performance of the method. In Figure 39 the absolute average error is displayed. It can be noticed that the overall trend of the simulated events is still present. The novel proposed STLS method is able to provide similar performance in the case of GVW estimation and offers a decrease in the variability

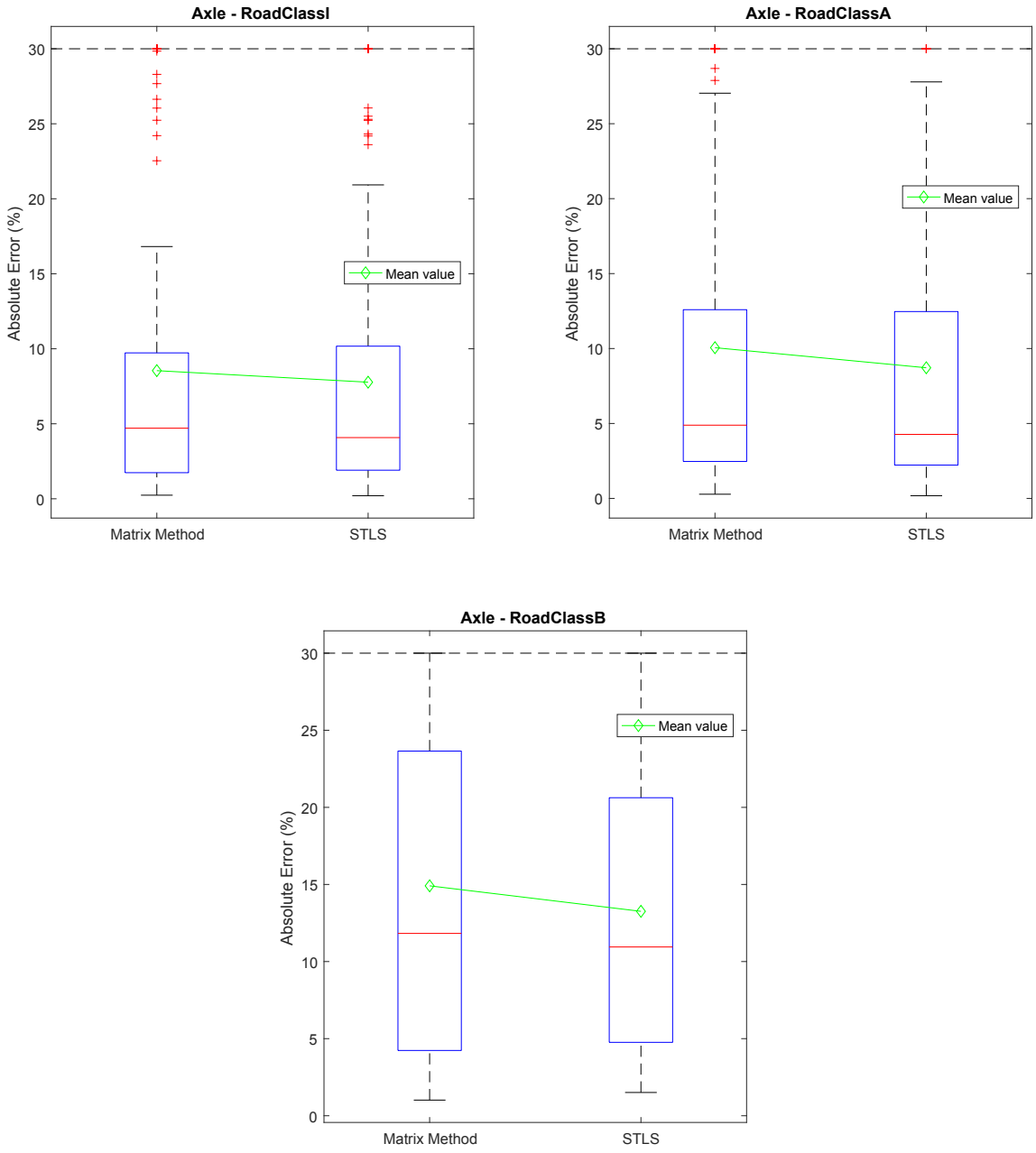


Figure 36 – Average absolute error for Axle - RoadClass

of results for single axle estimation. In can be seen that for GVW there is a slight increase in the mean absolute error, however, the amplitude of the outliers is decreased. Without considering the outliers, the red line representing the 50% quantile is on the same level for both methods. Distinctly, when considering the response by axle, a decrease in amplitude of errors for the STLS method is visible. Moreover, the red line representing the median without accounting for outliers is also on a smaller level, indicating a better individual axle recovery performance.

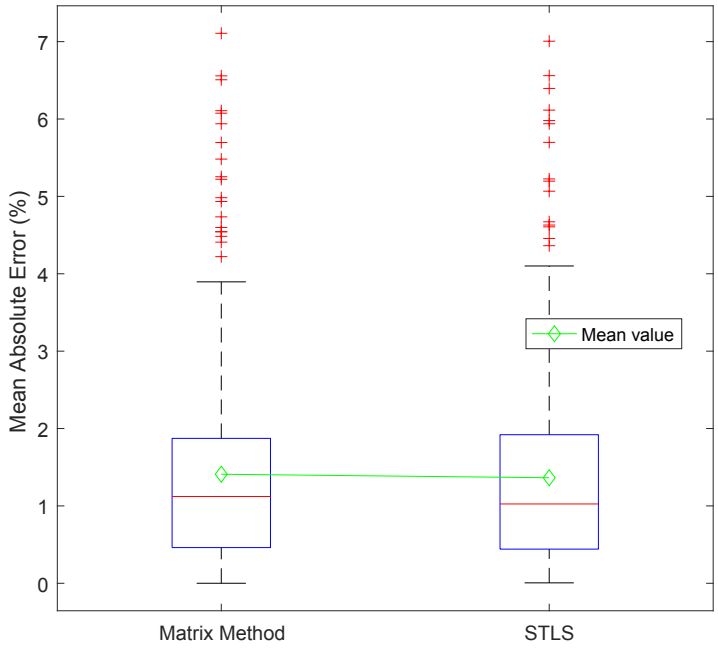


Figure 37 – Overall average absolute error for GVW

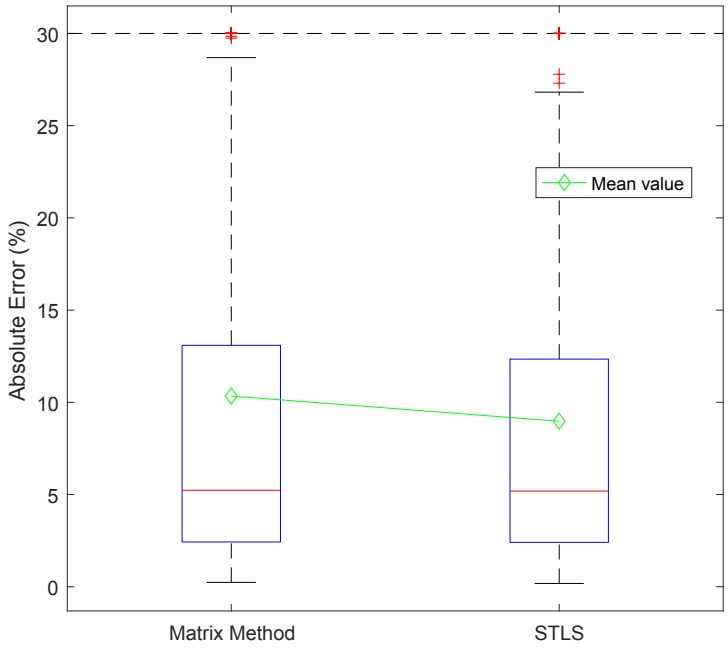


Figure 38 – Overall average error for individual axles

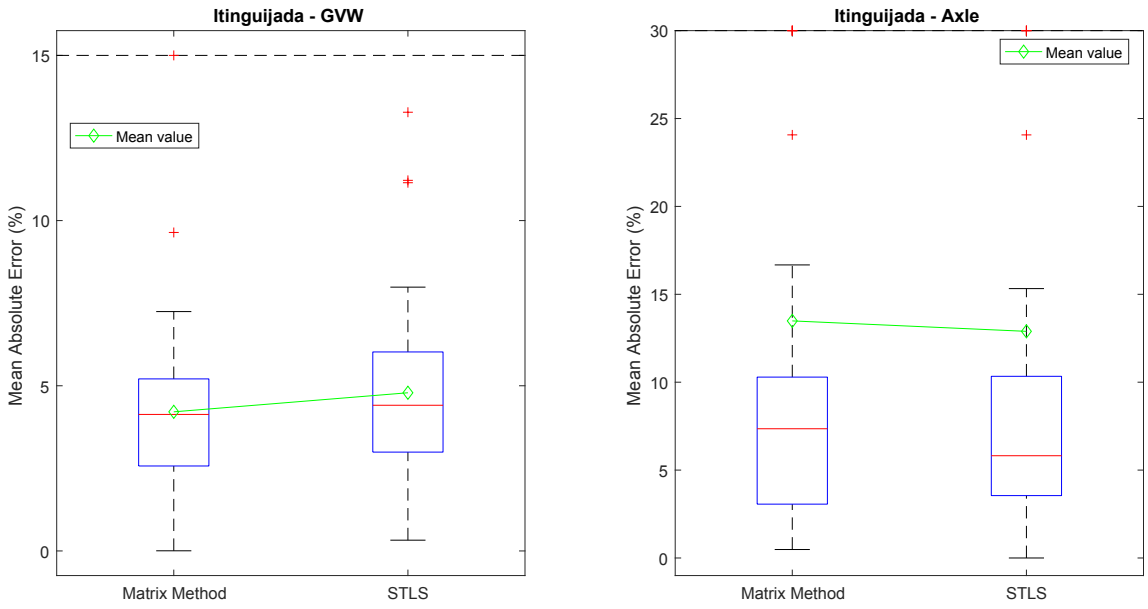


Figure 39 – Average absolute error for GVW and axles in Itinguijada bridge

8 CONCLUSIONS

This work presented a comparative study regarding some methods applied in B-WIM systems to derive bridge influence line and predict vehicle weights. In order to fairly compare the analyzed methods, only those which do not explicitly address the dynamic behavior of the vehicle-bridge system were considered. Furthermore, the methods were interpreted from a statistical point of view, where their assumptions were highlighted, remarking their theoretical differences.

In addition to the theoretical comparisons, the methods were implemented and numerically compared, addressing the lack of comparisons found in literature. In order to assess the similarity or the advantage of a method compared to the others, synthetic generated and real data were considered. Furthermore, the influence of some main factors that affects B-WIM systems were evaluated in simulated analyses, such as road profile and bridge length.

The analysis using simulated data showed that the increase in bridge length had more impact in corrupting the overall results than the roughness amplitude. The influence of road profile increased when decreasing the bridge span. When analyzing performance of the methods, results indicated that pBWIM and Modified 2D Moses were surpassed by the other approaches. Indeed, pBWIM usually provided worse weight estimation performance, together with a considerably higher computational cost. Modified 2D Moses was, for all analyzed cases, equal or worse than Matrix method. The remaining methods, MLE, Matrix method and Regularization, presented very similar results, with MLE being slightly better. The results regarding real data followed the trend already observed in numerical simulation, however with MLE showing superior performance. Thus, from both theoretical and practical perspectives, the MLE could be argued as the most promising method evaluated in this study.

Overall, the reformulation and reinterpretation of methods under a common statistical point of view allowed a better understanding of the underlying assumptions and theoretical hypothesis. This fact enabled a better comparison among methods and generated an opportunity for further improvements. For example, developing novel methods by relaxing the model statistical assumptions or applying a Bayesian approach to better account for prior information.

The exploration of the convolution interpretation of the B-WIM problem has given insights for the development of new solutions methods. The parallels established with the fields of signal and image processing allowed transporting ideas from those distinct science fields for the application on a practical engineering problem. Given this new

view, multiple improvement directions were explored. Recognizing B-WIM as an inverse problem with sparse characteristics, allowed exploring sparse recovery techniques, such as ℓ_1 regularization. Also, recognizing B-WIM as a "dual problem", in the sense that it possess the complementary processes of weighing and calibration, allowed the identification that a formulation accounting for error in variables such as the one in TLS was more in line with the real-world application. Finally, recognizing B-WIM as deconvolution problem allowed the Toeplitz structure of the problem to be exploited in order to introduce a consistent error vector.

Given the idea of exploiting known but underutilized knowledge about the B-WIM problem, a novel solution strategy was formulated and implemented based on different ideas from multiple authors and tweaked for practical considerations observed in the practical problem. The proposed method achieved comparable performance for GVW identification and improved performance for single axle estimation when compared to the classic Matrix method. The approach offers a better theoretical foundation and allows for inaccuracies in influence line data obtained from calibration, as well as indirect assessment of non-zero entries in the impulse vector.

9 FURTHER STUDIES

The current study has opened the venue for further research in the area of exploiting existing knowledge in order to propose better solution methods for the B-WIM problem.

As seen in this study, the proposed solution method involved a two-step optimization procedure that aimed at finding the support for the sparse solution first ant then conducting the weighing procedure exploiting the existing knowledge. One area of further research could be to investigate the potential for applying a joint optimization for noisy, sparse and structured recovery.

As another alternative, researchers could investigate a more practical aspect of B-WIM related to real-time applications. As seen in the Lasso optimization section, not all methods are suitable for quick weight estimation. There is an important trade-off between having good and complete bridge models and having the capacity of reliably offer real time estimates based on those models. This same remark is applied for models involving dynamic bridge simulations, which was not evaluated in this study.

On the subject of sparse recovery, another area of research could be to investigate the use of distinct sparseness inducing norms such the combination of Lasso and Euclidean norms employed in the Elastic Net method. Moreover, the whole class of greedy methods, which often have significant performance advantages when compared to classical norm methods, could be another venue for improving the overall computational efficiency of the solution procedure and amplifying the application range of the methods.

Concluding, another interesting aspect that may be addressed by further research is the computational improvements allowed by viewing the B-WIM problem as deconvoluition. As seen in previous chapters, the solution procedures often involve a linear system of the type $b = Ax$. However, when seeing the problem as a convolution the matrix A can be rewritten as a simple vector a which is convolved with the vector x. Therefore, following the approach which was already employed in similar problems that arose in literature, one could exploit the Fast Fourier Transform, avoiding the creation of matrix A and improving the efficiency of the solution procedure.

REFERENCES

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. *Selected papers of hirotugu akaike*, Springer, p. 199–213, 1998. Cited in page 28.

ARCHER, G.; TITTERINGTON, D. On some bayesian/regularization methods for image restoration. *IEEE Transactions on Image Processing*, IEEE, v. 4, n. 7, p. 989–995, 1995. Cited in page 20.

ARORA, J. S. *Introduction to optimum design*. New York, USA: Elsevier, 2004. Cited 5 times in pages 21, 25, 78, 79, and 82.

ASTER, R. C.; BORCHERS, B.; THURBER, C. H. *Parameter estimation and inverse problems*. New York, USA: Elsevier, 2018. Cited 2 times in pages 23 and 25.

BECK, A.; TEBOULLE, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, SIAM, v. 2, n. 1, p. 183–202, 2009. Cited 2 times in pages 30 and 78.

BHAT, H. S.; KUMAR, N. On the derivation of the bayesian information criterion. *School of Natural Sciences, University of California*, Citeseer, v. 99, 2010. Cited in page 29.

BICKEL, P. J.; RITOV, Y.; TSYBAKOV, A. B. Simultaneous analysis of lasso and dantzig selector. *The Annals of statistics*, Institute of Mathematical Statistics, v. 37, n. 4, p. 1705–1732, 2009. Cited in page 76.

BIGGS, J. M. *Introduction to structural dynamics*. [S.l.]: McGraw-Hill College, 1964. Cited in page 61.

BOYD, S.; PARIKH, N.; CHU, E.; PELEATO, B.; ECKSTEIN, J. *et al.* Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, Now Publishers, Inc., v. 3, n. 1, p. 1–122, 2011. Cited in page 79.

BOYD, S.; VANDENBERGHE, L. *Convex optimization*. [S.l.]: Cambridge university press, 2004. Cited in page 88.

BYRD, R. H.; CHIN, G. M.; NOCEDAL, J.; WU, Y. Sample size selection in optimization methods for machine learning. *Mathematical programming*, Springer, v. 134, n. 1, p. 127–155, 2012. Cited in page 81.

CANDES, E.; ROMBERG, J. Sparsity and incoherence in compressive sampling. *Inverse problems*, IOP Publishing, v. 23, n. 3, p. 969, 2007. Cited in page 76.

CARRARO, F.; GONÇALVES, M. S.; LOPEZ, R. H.; MIGUEL, L. F. F.; VALENTE, A. M. Weight estimation on static b-wim algorithms: A comparative study. *Engineering Structures*, Elsevier, v. 198, p. 109463, 2019. Cited in page 101.

CASELLA, G.; BERGER, R. L. *Statistical inference*. California, USA: Duxbury Pacific Grove, CA, 2002. v. 2. Cited in page 23.

CASTELLANOS, J. L.; GÓMEZ, S.; GUERRA, V. The triangle method for finding the corner of the l-curve. *Applied Numerical Mathematics*, Elsevier, v. 43, n. 4, p. 359–373, 2002. Cited in page 56.

CHATTERJEE, S.; HADI, A. S. *Regression analysis by example*. Hoboken, New Jersey, USA: John Wiley & Sons, 2015. Cited in page 23.

CHEN, J.; CHEN, Z. Extended bayesian information criteria for model selection with large model spaces. *Biometrika*, Oxford University Press, v. 95, n. 3, p. 759–771, 2008. Cited in page 29.

COMBETTES, P. L.; PESQUET, J.-C. Proximal splitting methods in signal processing. In: *Fixed-point algorithms for inverse problems in science and engineering*. New York, USA: Springer, 2011. p. 185–212. Cited in page 30.

CONGDON, P. *Bayesian statistical modelling*. [S.l.]: John Wiley & Sons, 2007. v. 704. Cited in page 56.

DEGROOT, M. H.; SCHERVISH, M. J. *Probability and statistics*. [S.l.]: Pearson Education, 2012. Cited in page 51.

DEPARTAMENTO NACIONAL DE INFRAESTRUTURA E TRANSPORTES. *Manual de estudos de tráfego*. Rio de Janeiro, Brazil, 2006. Cited 2 times in pages 60 and 61.

DHARANKAR, C. S.; HADA, M. K.; CHANDEL, S. Numerical generation of road profile through spectral description for simulation of vehicle suspension. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, Springer, v. 39, n. 6, p. 1957–1967, 2017. Cited in page 62.

DOICU, A.; TRAUTMANN, T.; SCHREIER, F. *Numerical regularization for atmospheric inverse problems*. New York, USA: Springer Science & Business Media, 2010. Cited in page 27.

DONOHO, D.; ELAD, M.; TEMLYAKOV, V. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, v. 52, n. 1, p. 6–18, 2006. Cited 2 times in pages 76 and 85.

DUMITRESCU, B. Sparse total least squares: analysis and greedy algorithms. *Linear Algebra and its Applications*, Elsevier, v. 438, n. 6, p. 2661–2674, 2013. Cited in page 95.

DUPÉ, F.-X.; FADILI, J. M.; STARCK, J.-L. A proximal iteration for deconvolving poisson noisy images using sparse representations. *IEEE Transactions on Image Processing*, IEEE, v. 18, n. 2, p. 310–321, 2009. Cited in page 100.

EFRON, B.; HASTIE, T.; JOHNSTONE, I.; TIBSHIRANI, R. *et al.* Least angle regression. *The Annals of statistics*, Institute of Mathematical Statistics, v. 32, n. 2, p. 407–499, 2004. Cited in page 30.

FANCHER, P. S. A factbook of the mechanical properties of the components for single-unit and articulated heavy trucks. phase i. final report. 1986. Cited in page 61.

FERCOQ, O.; GRAMFORT, A.; SALMON, J. Mind the duality gap: safer rules for the lasso. In: PMLR. *International Conference on Machine Learning*. [S.l.], 2015. p. 333–342. Cited in page 83.

FLORENS, J.-P.; SIMONI, A. Regularized posteriors in linear ill-posed inverse problems. *Scandinavian Journal of Statistics*, Wiley Online Library, v. 39, n. 2, p. 214–235, 2012. Cited in page 20.

FOX, C.; NICHOLLS, G. K.; TAN, S. M. *An Introduction To Inverse Problems*. Dunedin, New Zealand, 2010. Cited in page 16.

FRASSO, G.; EILERS, P. Smoothing parameter selection using the l-curve. In: *27th International Workshop on Statistical Modelling, Prague, Czech Republic, 2012, Proceedings*. [S.l.: s.n.], 2012. Cited in page 60.

FRøSETH, G. T.; RøNNQUIST, A.; CANTERO, D.; ØISETH, O. Influence line extraction by deconvolution in the frequency domain. *Computers & Structures*, v. 189, p. 21 – 30, 2017. ISSN 0045-7949. Cited 5 times in pages 8, 41, 44, 45, and 59.

GATU, C.; KONTOGHIORGHES, E. J. A fast algorithm for non-negativity model selection. *Statistics and Computing*, Springer, v. 23, n. 3, p. 403–411, 2013. Cited in page 75.

GAZZOLA, S.; NAGY, J. G. Generalized arnoldi–tikhonov method for sparse reconstruction. *SIAM Journal on Scientific Computing*, SIAM, v. 36, n. 2, p. B225–B247, 2014. Cited in page 101.

GOLUB, G.; KAHAN, W. Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis*, SIAM, v. 2, n. 2, p. 205–224, 1965. Cited in page 20.

GOLUB, G. H.; HEATH, M.; WAHBA, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, Taylor & Francis Group, v. 21, n. 2, p. 215–223, 1979. Cited in page 28.

GOLUB, G. H.; LOAN, C. F. V. An analysis of the total least squares problem. *SIAM journal on numerical analysis*, SIAM, v. 17, n. 6, p. 883–893, 1980. Cited in page 94.

GOLUB, G. H.; LOAN, C. F. V. *Matrix computations*. [S.l.]: JHU press, 2013. Cited in page 94.

GONÇALVES, M. S.; CARRARO, F.; LOPEZ, R. H. *Vehicle-bridge Dynamics simulation*. 2019. Mendeley Data. V1. Disponível em: <http://dx.doi.org/10.17632/kt48wf5vjz.1>. Cited in page 101.

GONÇALVES, M. S.; CARRARO, F.; LOPEZ, R. H. A b-wim algorithm considering the modeling of the bridge dynamic response. *Engineering Structures*, Elsevier, v. 228, p. 111533, 2021. Cited in page 35.

GONÇALVES, M. S.; CARRARO, F.; LOPEZ, R. H. A gradient based optimization procedure for finding axle weights in probabilistic bridge weigh-in-motion method. *Canadian Journal of Civil Engineering*, NRC Research Press 1840 Woodward Drive, Suite 1, Ottawa, ON K2C 0P7, v. 48, n. 5, p. 570–574, 2021. Cited in page 51.

GONÇALVES, M. S.; LOPEZ, R. H.; OROSKI, E.; VALENTE, A. M. A bayesian algorithm with second order autoregressive errors for b-wim weight estimation. *Engineering Structures*, v. 250, p. 113353, 2022. ISSN 0141-0296. Cited in page 20.

GONG, P.; ZHANG, C. A fast dual projected newton method for l1-regularized least squares. In: *Twenty-Second International Joint Conference on Artificial Intelligence*. [S.l.: s.n.], 2011. Cited in page 87.

GREGORY, P. *Bayesian logical data analysis for the physical sciences: a comparative approach with mathematica® support*. [S.l.]: Cambridge University Press, 2005. Cited in page 56.

GRENANDER, U.; SZEGÖ, G. *Toeplitz forms and their applications*. Chelsea, England: Univ of California Press, 1958. Cited in page 33.

GUERRA, V.; HERNANDEZ, V. Numerical aspects in locating the corner of the l-curve. In: *Approximation, Optimization and Mathematical Economics*. Heidelberg, Germany: Springer, 2001. p. 121–131. Cited in page 31.

HADAMARD, J. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, p. 49–52, 1902. Cited in page 19.

HANSEN, P. *Rank-Deficient and Discrete Ill-Posed Problems*. Philadelphia, USA: Society for Industrial and Applied Mathematics, 1998. Cited 2 times in pages 27 and 56.

HANSEN, P. C.; JENSEN, T. K.; RODRIGUEZ, G. An adaptive pruning algorithm for the discrete l-curve criterion. *Journal of computational and applied mathematics*, Elsevier, v. 198, n. 2, p. 483–492, 2007. Cited 2 times in pages 27 and 56.

HANSEN, P. C.; NAGY, J. G.; O'LEARY, D. P. *Deblurring images: matrices, spectra, and filtering*. Philadelphia, USA: Siam, 2006. v. 3. Cited in page 31.

HANSEN, P. C.; O'LEARY, D. P. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, SIAM, v. 14, n. 6, p. 1487–1503, 1993. Cited 2 times in pages 27 and 56.

HANSEN, P. C.; PEREYRA, V.; SCHERER, G. *Least squares data fitting with applications*. Baltimore,USA: JHU Press, 2013. Cited in page 22.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, USA: Springer, 2009. (Springer series in statistics). ISBN 9780387848846. Cited 3 times in pages 24, 28, and 30.

HASTIE, T.; TIBSHIRANI, R.; WAINWRIGHT, M. *Statistical learning with sparsity: the lasso and generalizations*. [S.l.]: Chapman and Hall/CRC, 2015. Cited in page 20.

HELMI, K.; BAKHT, B.; MUFTI, A. Accurate measurements of gross vehicle weight through bridge weigh-in-motion: a case study. *Journal of Civil Structural Health Monitoring*, Springer, v. 4, n. 3, p. 195–208, 2014. Cited in page 58.

HESTENES, M. R.; STIEFEL, E. *Methods of conjugate gradients for solving linear systems*. [S.l.]: NBS Washington, DC, 1952. v. 49. Cited in page 20.

Hummelsheim, S. Lasso and equivalent quadratic penalized models. *ArXiv e-prints*, 2014. Cited in page 73.

HURVICH, C. M.; TSAI, C.-L. Regression and time series model selection in small samples. *Biometrika*, Oxford University Press, v. 76, n. 2, p. 297–307, 1989. Cited in page 92.

IENG, S.-S. Bridge influence line estimation for bridge weigh-in-motion system. *Journal of Computing in Civil Engineering*, v. 29, n. 1, p. 06014006, 2015. Cited 2 times in pages 48 and 60.

ISWIM - International Society for Weigh in Motion. 2019. <http://www.is-wim.org>. Cited in page 14.

JACOB, B. *COST 323 Weigh in motion of road vehicles. Final Report, Appendix 1 European WIM Specification*. [S.l.]: LCPC Publications, Paris, 1999. Cited 3 times in pages 64, 66, and 71.

JENSEN, T. K. *Stabilization algorithms for large-scale problems*. Denmark: IMM, Informatik og Matematisk Modellering, Danmarks Tekniske Universitet, 2006. Cited in page 20.

JIANG, E. P.; BERRY, M. W. Solving total least-squares problems in information retrieval. *Linear Algebra and its Applications*, v. 316, n. 1, p. 137–156, 2000. ISSN 0024-3795. Special Issue: Conference celebrating the 60th birthday of Robert J. Plemmons. Cited in page 94.

JUNGES, P.; PINTO, R.; MIGUEL, L. F. B-wim systems application on reinforced concrete bridge structural assessment and highway traffic characterization. *Revista IBRACON de Estruturas e Materiais*, SciELO Brasil, v. 10, n. 6, p. 1338–1365, 2017. Cited in page 35.

KAZEMI, N. Automatic blind deconvolution with toeplitz-structured sparse total least squarestoeplitz-structured sparse tls. *Geophysics*, GeoScienceWorld, v. 83, n. 6, p. V345–V357, 2018. Cited 2 times in pages 96 and 97.

KILMER, M. E.; O'LEARY, D. P. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM Journal on matrix analysis and applications*, SIAM, v. 22, n. 4, p. 1204–1221, 2001. Cited in page 20.

KIM, J.; PARK, H. Fast active-set-type algorithms for l1-regularized linear regression. In: JMLR WORKSHOP AND CONFERENCE PROCEEDINGS. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. [S.l.], 2010. p. 397–404. Cited in page 83.

KIM, S.; LEE, J.; PARK, M.-S.; JO, B.-W. Vehicle signal analysis using artificial neural networks for a bridge weigh-in-motion system. *Sensors (Basel)*, Molecular Diversity Preservation International (MDPI), v. 9, n. 10, Oct 2009. Cited 2 times in pages 58 and 66.

KIM, S.-J.; KOH, K.; LUSTIG, M.; BOYD, S.; GORINEVSKY, D. An interior-point method for large-scale *ell*_1-regularized least squares. *IEEE journal of selected topics in signal processing*, IEEE, v. 1, n. 4, p. 606–617, 2007. Cited in page 83.

KOLDA, T. G.; BADER, B. W. Tensor decompositions and applications. *SIAM review*, SIAM, v. 51, n. 3, p. 455–500, 2009. Cited in page 53.

LANSDELL, A.; SONG, W.; DIXON, B. Development and testing of a bridge weigh-in-motion method considering nonconstant vehicle speed. *Engineering Structures*, Elsevier, v. 152, p. 709–726, 2017. Cited in page 35.

LEMMERLING, P.; MASTRONARDI, N.; HUFFEL, S. V. Efficient implementation of a structured total least squares based speech compression method. *Linear algebra and its applications*, Elsevier, v. 366, p. 295–315, 2003. Cited in page 95.

LYDON, M.; TAYLOR, S. E.; ROBINSON, D.; MUFTI, A.; BRIEN, E. Recent developments in bridge weigh in motion (b-wim). *Journal of Civil Structural Health Monitoring*, Springer, v. 6, n. 1, p. 69–81, 2016. Cited 2 times in pages 35 and 71.

MARKOVSKY, I.; HUFFEL, S. V. Overview of total least-squares methods. *Signal processing*, Elsevier, v. 87, n. 10, p. 2283–2302, 2007. Cited in page 94.

MEINSHAUSEN, N.; YU, B. Lasso-type recovery of sparse representations for high-dimensional data. *The annals of statistics*, Institute of Mathematical Statistics, v. 37, n. 1, p. 246–270, 2009. Cited in page 98.

MIGUEL, L. F.; LOPEZ, R. H.; TORII, A. J.; MIGUEL, L. F.; BECK, A. T. Robust design optimization of tmds in vehicle–bridge coupled vibration problems. *Engineering Structures*, Elsevier, v. 126, p. 703–711, 2016. Cited 2 times in pages 62 and 63.

MOOR, B. D. Structured total least squares and l2 approximation problems. *Linear algebra and its applications*, Elsevier, v. 188, p. 163–205, 1993. Cited in page 95.

MOSES, F. Weigh-in-motion system using instrumented bridges. *Journal of Transportation Engineering*, v. 105, n. 3, 1979. Cited 3 times in pages 37, 44, and 75.

MÚČKA, P. Simulated road profiles according to iso 8608 in vibration analysis. *Journal of Testing and Evaluation*, ASTM International, v. 46, n. 1, p. 1–14, 2017. Cited in page 62.

NESTEROV, Y. E. A method for solving the convex programming problem with convergence rate $o\left(\frac{1}{k^2}\right)$. In: *Dokl. Akad. Nauk SSSR,*. Moscow: Nauka, 1983. v. 269, p. 543–547. Cited in page 78.

NOSSEIR, T.; EL-GINDY, M.; EL-SAIED, F. Tire radial properties. *Journal of Periodica Polytechnica Transportation Engineering*, p. 21–28, 1982. Cited in page 61.

O'BRIEN, E. J.; QUILLIGAN, M.; KAROUMI, R. Calculating an influence line from direct measurements. *Bridge Engineering, Proceedings of the Institution of Civil Engineers*, Institution of Civil Engineers, v. 159, n. BE1, p. 31–34, 2006. Cited 5 times in pages 38, 45, 46, 54, and 60.

O'BRIEN, E. J.; ROWLEY, C. W.; GONZALEZ, A.; GREEN, M. F. A regularised solution to the bridge weigh-in-motion equations. *International Journal of Heavy Vehicle Systems*, v. 16, n. 3, p. 310–327, 2009. Cited 5 times in pages 36, 55, 59, 60, and 71.

O'BRIEN, E. J.; ZHANG, L.; ZHAO, H.; HAJIALIZADEH, D. Probabilistic bridge weigh-in-motion. *Canadian Journal of Civil Engineering*, v. 45, n. 8, p. 667–675, 2018. Cited 2 times in pages 50 and 60.

OJIO, T.; YAMADA, K. Bridge weigh-in-motion systems using stringers of plate girder bridges. In: *Third International Conference on Weigh-in-Motion (ICWIM3) Iowa State University, Ames*. Iowa, USA: ICWIM 3, 2002. Cited in page 59.

OSBORNE, M.; PRESNELL, B.; TURLACH, B. A new approach to variable selection in least squares problems. *IMA Journal of Numerical Analysis*, v. 20, n. 3, p. 389–403, 2000. Cited in page 85.

O'DONOGHUE, B.; CANDES, E. Adaptive restart for accelerated gradient schemes. *Foundations of computational mathematics*, Springer, v. 15, n. 3, p. 715–732, 2015. Cited 2 times in pages 79 and 80.

PAIGE, C. C.; SAUNDERS, M. A. Lsqr: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS)*, ACM, v. 8, n. 1, p. 43–71, 1982. Cited in page 20.

REYES, J. C. D. L.; LOAYZA, E.; MERINO, P. Second-order orthant-based methods with enriched hessian information for sparse ℓ_1-optimization. *Computational Optimization and Applications*, Springer, v. 67, n. 2, p. 225–258, 2017. Cited in page 81.

Road Traffic Technology. 2019. <https://www.roadtraffic-technology.com/contractors/test/psp/>. Cited in page 15.

RUDIN, L. I.; OSHER, S.; FATEMI, E. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, Elsevier, v. 60, n. 1-4, p. 259–268, 1992. Cited in page 46.

SCHWARZ, G. Estimating the dimension of a model. *The annals of statistics*, JSTOR, p. 461–464, 1978. Cited in page 28.

SLAWSKI, M.; HEIN, M. Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization. *Electronic Journal of Statistics*, v. 7, 05 2012. Cited in page 93.

TARANTOLA, A. *Inverse problem theory and methods for model parameter estimation*. [S.l.]: siam, 2005. v. 89. Cited 2 times in pages 48 and 57.

THEODORIDIS, S. *Machine learning: a Bayesian and optimization perspective*. [S.l.]: Academic Press, 2015. Cited in page 20.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 267–288, 1996. Cited in page 73.

TIBSHIRANI, R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 73, n. 3, p. 273–282, 2011. Cited in page 76.

TITTERINGTON, D. General structure of regularization procedures in image reconstruction. *Astronomy and Astrophysics*, v. 144, p. 381, 1985. Cited in page 20.

VOGEL, C. Optimal choice of a truncation level for the truncated svd solution of linear first kind integral equations when data are noisy. *SIAM journal on numerical analysis*, SIAM, v. 23, n. 1, p. 109–117, 1986. Cited in page 31.

WANG, H.; NAGAYAMA, T.; ZHAO, B.; SU, D. Identification of moving vehicle parameters using bridge responses and estimated bridge pavement roughness. *Engineering Structures*, Elsevier, v. 153, p. 57–70, 2017. Cited in page 62.

WANG, L.; GORDON, M. D.; ZHU, J. Regularized least absolute deviations regression and an efficient algorithm for parameter tuning. In: IEEE. *Sixth International Conference on Data Mining (ICDM'06)*. [S.l.], 2006. p. 690–700. Cited in page 20.

WASHINGTON, S. P.; KARLAFTIS, M. G.; MANNERING, F. *Statistical and econometric methods for transportation data analysis*. Boca Raton, Florida, USA: Chapman and Hall/CRC, 2010. Cited in page 23.

WIT, E.; HEUVEL, E. v. d.; ROMEIJN, J.-W. 'all models are wrong...': an introduction to model uncertainty. *Statistica Neerlandica*, Wiley Online Library, v. 66, n. 3, p. 217–236, 2012. Cited in page 28.

WRIGHT, S.; NOCEDAL, J. *et al.* Numerical optimization. *Springer Science*, v. 35, n. 67-68, p. 7, 1999. Cited in page 21.

WU, T. T.; LANGE, K. *et al.* Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics*, Institute of Mathematical Statistics, v. 2, n. 1, p. 224–244, 2008. Cited in page 30.

XU, Z.; FIGUEIREDO, M. A.; YUAN, X.; STUDER, C.; GOLDSTEIN, T. Adaptive relaxed admm: Convergence theory and practical implementation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 7389–7398. Cited in page 81.

YANG, Y.; LIN, C. Vehicle–bridge interaction dynamics and potential applications. *Journal of sound and vibration*, Elsevier, v. 284, n. 1-2, p. 205–226, 2005. Cited in page 61.

YANG, Y.-B.; YAU, J.; YAO, Z.; WU, Y. *Vehicle-bridge interaction dynamics: with applications to high-speed railways*. [S.l.]: World Scientific, 2004. Cited in page 61.

YU, Y.; CAI, C.; DENG, L. State-of-the-art review on bridge weigh-in-motion technology. *Advances in Structural Engineering*, SAGE Publications Sage UK: London, England, v. 19, n. 9, p. 1514–1530, 2016. Cited in page 35.

YU, Y.; CAI, C.; DENG, L. Nothing-on-road bridge weigh-in-motion considering the transverse position of the vehicle. *Structure and Infrastructure Engineering*, Taylor & Francis, v. 14, n. 8, p. 1108–1122, 2018. Cited in page 71.

ZHAO, H.; UDDIN, N. Algorithm to identify axle weights for an innovative bwim system. part ii. In: *Proc., IABSE-JSCE Joint Conf. on Advances in Bridge Engineering-II*. Dhaka, Bangladesh: Bangladesh Association of Consulting Engineers and Bangladesh Association of Construction Industry, 2010. v. 537, p. 546. Cited 2 times in pages 37 and 40.

ZHAO, H.; UDDIN, N.; O'BRIEN, E. J.; SHAO, X.; ZHU, P. Identification of vehicular axle weights with a bridge weigh-in-motion system considering transverse distribution of wheel loads. *Journal of Bridge Engineering*, v. 19, n. 3, p. 04013008, 2014. Cited 5 times in pages 51, 58, 60, 66, and 71.

ZHAO, Z.; UDDIN, N.; O'BRIEN, E. J. Bridge weigh-in-motion algorithms based on the field calibrated simulation model. *Journal of Infrastructure Systems*, v. 23, n. 1, p. 04016021, 2017. Cited in page 58.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, Wiley Online Library, v. 67, n. 2, p. 301–320, 2005. Cited in page 32.

APPENDIX A – PUBLISHED PAPER

The published paper mentioned in this thesis project is as follows.

Engineering Structures 198 (2019) 109463

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Weight estimation on static B-WIM algorithms: A comparative study

Felipe Carraro*, Matheus Silva Gonçalves, Rafael Holdorf Lopez, Leandro Fleck Fadel Miguel, Amir Mattar Valente

Laboratório de Transportes e Logística (LabTrans) and Center for Optimization and Reliability in Engineering (CORE), Civil Engineering Department, Federal University of Santa Catarina, Rua João Pio Duarte Silva, s/n, 88040-900 Florianópolis, SC, Brazil

ARTICLE INFO

Keywords:
Bridge weigh-in-motion
Vehicle weight prediction
Methods comparison
Inverse problems
Statistical assumptions

ABSTRACT

Bridge weigh in motion (B-WIM) comprises the use of sensors on existing bridges in order to assess the loads of passing vehicles. Although numerous methods for weight estimation on static B-WIM algorithms may be found in the literature, there is not available a comparison study among them, especially regarding accuracy and statistical assumptions. Hence, this paper provides a critical comparison on a subset of conceptually similar B-WIM methods, further extending the discussion on their theoretical assumptions, beyond what is currently available in literature. The methods are not only referenced but reinterpreted and reformulated in a unifying manner, allowing an in-depth comparison. Moreover, a parametric study on the performance and sensitivity of methods is conducted. Not only simulated but also real data are employed in the comparison, supporting conclusions.

1. Introduction

A bridge weigh in motion (B-WIM) system effectively turns a bridge in a weighing mechanism by means of recovering the live load of passing vehicles from strain information obtained through sensors. These systems are installed underneath the bridge, which does not disrupt road traffic and also improves the durability and portability of the system [1]. Therefore, it can be an efficient tool for overweight enforcement since it is able to measure vehicles weight traveling at operating speed. The information retrieved from the system can also be employed in other contexts such as maintenance planning, structural health monitoring, service life estimation and traffic network planning [2]. The cost of installing and maintaining B-WIM systems is therefore often lower due to better accessibility and synergy effects with other projects when compared to other traffic monitoring systems [3].

The main idea behind B-WIM systems, firstly introduced by Moses [4], relies on equating the bending moment on the bridge with the product of the magnitude of the applied moving load and the influence line ordinate of the bridge. By this formulation, it is possible to estimate the axle weights of passing vehicles as those which generate the best agreement between theoretical and measured bending moment response. For more aspects regarding general information about B-WIM systems, as well as some implementation considerations, the reader is referred to Lydon et al. [5], Yu et al. [6], Žnidarič et al. [7].

Differently from the theoretical influence line employed by Moses, recently developed approaches apply crossing vehicles with known weight and the corresponding measured deformations to derive the influence line of the bridge, a procedure known as calibration [8–11]. This formulation is valid as long as the static analysis is considered. In practice, however, the dynamical motions induced by vehicles may increase the difficulty in the correct influence line evaluation and consequently weight prediction [12].

Due to the dynamic behavior, B-WIM methods may be divided into two classes, related to explicitly considering or not the dynamic formulation and bridge-vehicle interaction. Several papers contributed with different approaches in the first class, with a higher emphasis on the concept of Moving Force Identification (MFI) [13–15,1,12,16]. This class of methods, however, may need to consider a full 3D model of the bridge for a suitable accuracy, thus requiring extensive computational effort [6]. Therefore, although some promising results, MFI methods are not yet able to deal with some important applications, such as real-time monitoring [17]. For the static class, which is the focus of this paper, several researchers presented novel contributions in the past decade. Ieng [8] proposed a maximum likelihood approach to estimate the bridge influence line, generalizing the method by simultaneously taking into account measurements available from as many calibration trucks as needed. Zhao et al. [9] considered the transverse distribution of wheel loads on the axle weight estimation. Kim et al. [18] trained artificial neural networks for the estimation of Gross Vehicle Weights (GVW). O'Brien et al. [10] introduced the concept of probabilistic

* Corresponding author.
E-mail addresses: felipecarraro@gmail.com (F. Carraro), matheusgoncalves.contato@gmail.com (M.S. Gonçalves), rafael.holdorf@ufsc.br (R.H. Lopez), leandro.miguel@ufsc.br (L.F.F. Miguel), amir.valente@ufsc.br (A.M. Valente).

https://doi.org/10.1016/j.engstruct.2019.109463
Received 20 December 2018; Received in revised form 6 July 2019; Accepted 30 July 2019

influence line seeking to find the most probable axle weights. O'Brien et al. [11] employed Tikhonov regularization for overcoming the ill-posedness of the B-WIM problem.

Although the performance of proposed methods is usually argued as useful for practical application, their results often disregard other recent studies in the same subject. As a consequence, it becomes hard to evaluate the state-of-the-art improvements, since each study relies on specific bridges, whose performance are not directly comparable with each other. Furthermore, although some of those methods make direct statistical assumptions, similarities with other approaches and practical consequences of these assumptions are usually not evaluated. Indeed, there has not been identified any publications comparing theoretical aspects, practical applicability as well as the performance of recent B-WIM algorithms.

In the present work, a comparative investigation, comprised of 5 state-of-the-art methods on weight estimation using static B-WIM algorithms, is conducted. The methods employed in this investigation are presented in Ieng [8], Zhao et al. [9], O'Brien et al. [10], O'Brien et al. [19] and O'Brien et al. [11], which are related to influence line acquisition, weight prediction or both simultaneously. In the proposed analysis, all methods are reformulated or reinterpreted in a unified manner, allowing theoretical statistical comparisons. Furthermore, the practical consequences of some assumptions are discussed.

In order to perform the proposed analysis, a large set of distinct conditions, such as bridge span and road profile, is numerically simulated. The generated data set is employed for improving comparisons, since it enables relating the bridge properties and a method's performance. It may help users in choosing the most suitable method for their specific case. Furthermore, real-world signals are also evaluated, offering validation for the conclusions drawn based on the simulated model.

The main contribution of this paper is to conduct an investigation of B-WIM procedures, further explaining their theoretical considerations, the relation among methods and presenting a comparison of their performance. The methods are not only discussed as referenced, but are reinterpreted and reformulated in a unifying manner, employing a similar notation, which more easily enables comparisons. Thus, the analyses pursued here aim to lead to new insights for the development of novel methods or for choosing methods based on which cases they are likely to perform better.

The paper is organized as follows: Section 2 employs the same formulation framework on the analyzed methods, enabling not only numerical but conceptual, theoretical and implementation comparisons. Section 3 presents the numerical comparison of all methods using simulated data, and remarking some theoretically expected aspects. The results obtained from real field data are presented in Section 4. Finally, Section 5 addresses further comments and discussion while Section 6 presents concluding remarks.

2. Methods overview and discussion

General aspects regarding the methods that belong to the scope of this study are discussed in this section. Further analyses are conducted for five methods whose characteristics are relevant for the overall comparative context. For the remaining methods, the aspects that prevent their inclusion in the comparative study are remarked.

2.1. Matrix method

The study of O'Brien et al. [19] develops the matrix method for the acquisition of the influence line. It derives the equations for three-axle calibration trucks by minimizing a quadratic residual based on the predicted response of a pre-weighted vehicle. In their study, the influence line for two vehicles with three and seven axles are calibrated and the similarity between measured and predicted responses is shown. This work contributed with a systematic way of employing direct bridge measurements by first calibrating an influence line based on a pre-weighted truck, which is later used for weighing arbitrary vehicles. This contrasts with the use of the theoretical influence line proposed in Moses [4].

2.1.1. The Matrix approach itself

The Matrix method results in a least squares solution. The basic idea starts from Moses algorithm, where one must minimize an error function R, comprised of the sum of the squares of differences between the measured bending moment M^m and the theoretical bending moment M^t. The measured term, for an instant k is given by:

$$M_k^m = \sum_{g=1}^{G} E_g Z_g \varepsilon_g, \tag{1}$$

where G represent the number of girders and E_g, Z_g, and ε_g are the elastic modulus, section modulus and measured strain of the $g-th$ girder, respectively. The theoretical term introduces the influence line and reads:

$$M_k^t = \sum_{j=1}^{J} W_j IL_{(k-C_j)}, \tag{2}$$

where

$$C_j = \frac{d_j f}{v} \tag{3}$$

and J is the number of axles, W_j is the weight of the $j-th$ axle, $IL_{(k-C_j)}$ is the influence ordinate at the position of the $j-th$ axle, d_j is the distance between the first axle and $j-th$ axle, C_j is the number of scans corresponding to d_j, f is the sampling frequency and v is the vehicle velocity. It is worthwhile to point that if $k - C_j$ results in a index that does not match an influence line ordinate, it is attributed the value of zero for it.

The error term reads:

$$R = \sum_{k=1}^{K} (M_k^m - M_k^t)^2, \tag{4}$$

where k represent each scan and K the total number of scans. Differently from the work of Moses, where M^t was based on a theoretical influence line, in O'Brien et al. [19] it is now considered an unknown and solved with the minimization of Eq. (4) by calibrating with a vehicle with known weight. After the influence line based on direct measurements has been found, one can proceed as usual with Moses method for finding the unknown weights of vehicles passing over the bridge. That is:

$$W = (\Lambda^T \Lambda)^{-1} \Lambda^T M^m \tag{5}$$

where W is the vector of predicted axle weights and Λ is a matrix based on the influence line ordinates, shifted according to the axle spacing, defined as:

$$\Lambda^{kj} = IL_{k-C_j}. \tag{6}$$

Thus, Λ is a $K \times J$ matrix. Furthermore, the same observation done for IL_{k-C_j} in Eq. (2) holds here.

The error function in Eq. (4) defines a least square problem. By the statistical point of view, such an approach is the maximum likelihood estimator when errors are independent and normally distributed random variables [20]. These errors are related to each measured ordinate and at the same event of calibration. However, these underlying error assumptions may not be met when considering practical cases. For example, missing information on formulation could be seen as correlated errors [21]. Thus, if the model description in Eq. (2) is not accurate enough, a least squares estimate would not provide the best result. Nevertheless, one could still extract useful information from the method application, even though minor violations on the assumptions are present [22].

2.2. The maximum likelihood approach

The work of Ieng [8] proposes the utilization of maximum likelihood estimation (MLE) to obtain the influence line, based on calibration data. The main reason argued by the study is that the method proposed by O'Brien et al. [19] is not robust enough and the influence line derivation is done only with signals produced by a single pass of the calibration vehicle. Thus, the MLE approach intended to overcome these issues, applying an iid (independent and identically distributed) Gaussian random noise to the formulation.

Ieng [8] compared the MLE approach with the matrix method of O'Brien et al. [19], using the L_∞ norm (maximum value of error) between predicted and measured strains as the criterion. The presented results were derived using different traffic data from the one employed in calibration. Ieng [8] concluded that the applied methodology achieved smaller errors regarding to the measured strains.

2.2.1. The MLE approach itself

Measurements are always corrupted with some kind of noise, introducing uncertainties into the analysis. The maximum likelihood approach aims to include these uncertainties into the formulation in order to reach a more robust influence line.

It is assumed in MLE that the measurements are corrupted by ϵ, a zero mean multivariate normal random variable:

$$M^m = M^t + \epsilon, \tag{7}$$

where M^m and M^t are the vectors of measured and modeled moments, respectively, such that:

$$M^t = AIL, \tag{8}$$

where IL is a vector with the ordinates of the influence line and A is a Toeplitz matrix of the loads. The matrix A is based on the impulse load vector $\mathscr{W}$:

$$\mathscr{W}_i = \begin{cases} W_j, & \text{if } i = C_j + 1 \\ 0, & \text{otherwise} \end{cases}, \tag{9}$$

where W_j is the weight of the $j-th$ axle and C_j is analogous to that one defined in Eq. (3). This impulse vector represents the whole vehicle, with each axle load at their respective axle position. The A matrix is formed by shifting $\mathscr{W}$ in each line of A, which corresponds to a discrete convolution, where each line relates to a time step of the vehicle passing over the bridge:

$$A = \begin{bmatrix} W_1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ W_J & \dots & W_1 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & W_J & \dots & W_1 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & W_J \end{bmatrix}_{(K \times K - C_J)}, \tag{10}$$

where K and J are the total number of scans and axles, respectively.

Furthermore, several calibration trucks could be employed, passing N times over the bridge. Each passage of the vehicle could be counted as a realization of the random variable ϵ. Then, the notation may be modified to:

$$M_i^m = M_i^t + \epsilon_i, \tag{11}$$

where the index i ranges from 1 to N, the total number of signals collected. In order to merge data from distinct runs, one could suppose that all data collected is related to the same random variable. In this way, it becomes necessary to interpolate the data vectors to common ordinates.

Assuming that the realizations of this variable are independent among events, the likelihood can be written as the product of the individual probabilities:

$$L = \prod_{i=1}^{N} \text{pdf}(\epsilon_i | IL), \tag{12}$$

where pdf is the probability density function of ϵ and L is the likelihood. The principle of MLE is equivalent to minimizing the negative of its natural logarithm with respect to IL:

$$\text{argmax}_{IL}\left(\prod_{i=1}^{N} \text{pdf}(\epsilon_i | IL)\right) = \text{argmin}_{IL}\left(-\sum_{i=1}^{N} (\log(\text{pdf}(\epsilon_i | IL)))\right). \tag{13}$$

Inserting Eq. (11) into Eq. (13), it reads:

$$\text{argmin}_{IL}\left(-\sum_{i=1}^{N} (\log(\text{pdf}(M_i^m - M_i^t | IL)))\right). \tag{14}$$

As the variable ϵ follows a multivariate normal distribution, the expression of its pdf could be introduced into the formulation:

$$\text{argmin}_{IL}\left(-\sum_{i=1}^{N} \log\left(\frac{1}{|\Sigma|\sqrt{(2\pi)^K}} \exp^{-\frac{1}{2}((M_i^m - AIL)^T \Sigma^{-1} (M_i^m - AIL))}\right)\right), \tag{15}$$

where Σ is the covariance matrix of the random variable ϵ, $|.|$ is the determinant operator and K is the number of dimensions of the multivariate distribution, which corresponds to the number of scans in this case. Using the logarithm properties:

$$\text{argmin}_{IL}\left(\sum_{i=1}^{N} \left(\log(|\Sigma|\sqrt{(2\pi)^K}) + \frac{1}{2}(M_i^m - AIL)^T \Sigma^{-1} (M_i^m - AIL)\right)\right). \tag{16}$$

This expression can be minimized by setting to zero its first derivative with respect to IL. Supposing that the covariance matrix Σ is independent of the IL ordinates, it results that the derivative of the first term is zero. The derivative of the remaining expression can be written as:

$$\sum_{i=1}^{N} \frac{1}{2} \frac{\partial}{\partial IL_j} ((M_i^m - AIL)^T \Sigma^{-1} (M_i^m - AIL)) = 0, \tag{17}$$

which results in:

$$\sum_{i=1}^{N} ((M_i^m - AIL)^T \Sigma^{-1} A) = 0. \tag{18}$$

Recalling that $(AB)^T = B^T A^T$ and Σ is a symmetric matrix:

$$\sum_{i=1}^{N} (A^T \Sigma^{-1} (M_i^m - AIL))^T = 0. \tag{19}$$

Rearranging the expression and assuming that the covariance is a diagonal matrix, with equal variance:

$$\sum_{i=1}^{N} A^T AIL = \sum_{i=1}^{N} A^T M_i^m, \tag{20}$$

the expression proposed in the study is reached. To find the solution it is necessary for $\sum_{i=1}^{N} A_i^T A_i$ to be invertible, condition that is satisfied according to Ieng [8]. Similarly to the work of O'Brien et al. [19], the MLE method provides a way to find the influence line based on calibration data. Thus, the weighing procedure is also analogous to that stated by Moses.

2.2.2. Comparing MLE and matrix method

The comparative aspect between MLE and matrix method was already assessed in the work of Ieng [8]. The matrix method was derived based on data of only one calibration run, while the MLE approach could be seen as a generalization of the matrix method for cases where more calibration runs were performed. Hence, the matrix method needs some kind of assembly strategy, such as performing the mean of resulting influence lines from several runs. It is worthwhile mentioning

that, as previously discussed for the matrix method, the MLE approach will only be the maximum likelihood estimator when errors are independent and normally distributed random variables, with equal standard deviation.

2.3. pBWIM

The pBWIM approach was proposed by O'Brien et al. [10], which aimed to directly incorporate probabilistic information about several passages of calibration vehicles to construct the influence line. This formulation assumed that each measured influence line ordinate follows a normal distribution, obtained from field testing. The matrix method is employed to derive the parameters of these distributions, based on a single influence line for each event. Finally, the estimated axle weights are those with the highest probability of occurrence among all possible combinations.

In order to derive the pBWIM results, two different levels of information were applied to generate the influence line: the whole calibration data or only a subset of events. The goal was to reproduce a situation closer to the real case, whose small number of calibration events are available. The results showed that with less information, the pBWIM achieved better results in comparison with the traditional approach.

2.3.1. The pBWIM approach itself

Considering that the response M^t is the sum of products of axle weights and influence line ordinates, the response related to scan k, M_k^t, could be written as:

$$M_k^t = \sum_{j=1}^{J} W_j IL_{k-C_j}, \tag{21}$$

where J is the number of axles, W is the vector of axle weights, IL_{k-C_j} is the influence line value at ordinate $k - C_j$ and C_j is the offset distance between the ordinates of the influence line related to axle j. If the value of $k - C_j$ results in an index that does not correspond to an influence line ordinate, the value zero is attributed to it, since it reflects a situation where the axle is out of the bridge.

Since each influence line ordinate follows a known normal distribution and the response M_k^t is a linear combination of normal variables, the response M_k^t itself also follows a normal distribution. Thus, each ordinate has a corresponding normal distribution defined by the parameters:

$$\mu_k^M = \sum_{j=1}^{J} W_j \mu_{k-C_j} \tag{22}$$

$$\sigma_k^M = \sqrt{\sum_{j=1}^{J} \left(W_j \sigma_{k-C_j}\right)^2 + \tau^2}, \tag{23}$$

where μ_k^M and σ_k^M are the mean and standard deviation of the random variable related to the predicted moment of each bridge ordinate k, respectively. Furthermore, τ and σ_{k-C_j} are the standard deviation of the measurement noise and influence line, respectively.

Given the measured moments M_k^m, related to a specific bridge ordinate, the probability of a set of weights being responsible for generating such measures could be formulated as (highlighting those terms that rely on the weights W):

$$P_k(W) = \frac{1}{\sigma_k^M(W)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{M_k^m - \mu_k^M(W)}{\sigma_k^M(W)}\right)^2\right). \tag{24}$$

Assuming that all random variables related to the measured moments on each ordinate are independent, an expression for finding the weights that most likely have generated such data could be formulated as:

$$\text{argmax}_W\left(\prod_{k=1}^{K} P_k(W)\right), \tag{25}$$

where K is the total number of scans in the measured response. Therefore, it is necessary to find the weights that maximize this expression. The procedure described by O'Brien et al. [10] to obtain the axle weights is based on a grid search, whose bounds and increments are arbitrary parameters. Such study has applied 0.8 and 1.2 times the weights predicted by the matrix method as bounds and 0.1 kN as increment. It is worth pointing that such a procedure is quite time-consuming, especially when increasing the number of axles.

2.3.2. The assumption of independence

Although pBWIM method provided reasonable results, an assumption made in the derivation of the procedure seems to be violated. In order to better understand it, one can write the response M^t in matrix formulation, as already discussed in Eq. (8):

$$M^t = AIL, \tag{26}$$

where A is the Toeplitz matrix for a given set of axle loads and IL is the influence line vector, assumed to be comprised by independent Gaussian random variables by O'Brien et al. [10]. As IL is supposed to be Gaussian, it is straightforward to calculate the covariance matrix of the resulting moments random vector M^t, dealing with A as a linear transformation applied to the IL random variables:

$$\Sigma_{M^t} = A^T \Sigma_{IL} A, \tag{27}$$

where Σ_{M^t} and Σ_{IL} are the covariance matrix of the predicted moments and influence line, respectively. Recalling that Σ_{IL} is diagonal, since the influence line ordinates are defined as independent random variables, and writing the matrix multiplication with index notation, results in an expression for each term of the Σ_{M^t} matrix:

$$\Sigma_{M^t}^{ij} = \sum_{k=1}^{K} A_{ik} \Sigma_{IL}^{kk} A_{jk}, \tag{28}$$

where K is the total number of scans. As already mentioned, for the predicted moments to be independent random variables, it is necessary that Σ_{M^t} be diagonal. Matematically, for all i and j, with $i \neq j$, one must have $\Sigma_{M^t}^{ij} = 0$. Thus, for showing that such independence does not occur, it is enough that there are i and j, with $i \neq j$, such that $\Sigma_{M^t}^{ij} \neq 0$.

Observing Eq. (28), it could be noticed that A_{ik}, A_{jk} and Σ_{IL}^{kk} will always be non-negative. The first two because of the structure of matrix A, which is comprised only by axles weights and zeros, as seen in Eq. (10). The last one is strictly greater than zero, since it is the diagonal of a covariance matrix. Therefore, just one term of the summation greater than zero is sufficient to ensure a non-diagonal covariance. In other words, it is enough that, for any k, there are i and j, with $i \neq j$, such that:

$$A_{ik} A_{jk} > 0. \tag{29}$$

One could see that the condition of Eq. (29) is always met when, at least, one column in A has more than one non-negative value. For the Toeplitz matrix A, it occurs when the passing vehicle has more than one axle, being the difference between the corresponding i and j equal to the number of scans separating both axles. Thus, for the general case, Σ_{M^t} is not a diagonal matrix and predicted moments are not independent random variables, violating this assumption. It is worth to mention that if the matrix A is diagonal, which corresponds to a vehicle with only one axle passing over the bridge, the independence holds. However, such a case is clearly unrepresentative for B-WIM applications.

2.3.3. Comparing pBWIM and MLE approaches

In the definition of the pBWIM approach, O'Brien et al. [10] account for a zero mean error in the measurement. In this way, the stated formulation could be seen as a total least squares, with the form:

$$(IL + \epsilon)W = M^m + \tau, \tag{30}$$

where IL now is a matrix containing the influence line ordinates in each column, shifted by the corresponding axle spacing in the rows and W is the vector of weights. However, as it became necessary to have some estimate for the standard deviation of the measurement noise, O'Brien et al. [10] set this value to zero. Therefore, resulting in a problem such that:

$$(IL + \epsilon)W = M. \tag{31}$$

Such a formulation remarks the difference between pBWIM and MLE approaches. The former accounts for error in the independent variable while the latter sums it to the dependent variable. While the formulation of MLE results in a closed-form solution, the approach of pBWIM requires some form of optimization procedure in order to find the most likely weights. Furthermore, pBWIM does not make any assumption regarding errors with equal standard deviation, which is the case for MLE.

2.4. Tikhonov regularization

O'Brien et al. [11] applied Tikhonov regularization to the matrix method equations. The reason is that the final system of equations used to solve the axle weights has an ill-conditioned or ill-posed nature. With this approach, the authors intended to achieve better results mainly for the weight by axle, which is acknowledged to have worse prediction precision than total vehicle weight. The author used the well known L curve method [23] to define the regularization parameter, evaluating parameters ranging from 10^{-90} to 600,000. It is worth to mention that the solution is unique for each parameter.

The method was theoretically tested using dynamic simulations of a series of moving forces on a bridge. The author concluded that the regularized solution performed better than the matrix method. However, as the vehicle dynamics increased, the convergence of the regularized solution was not as accurate.

2.4.1. The Tikhonov regularization approach itself

Tikhonov or ridge regression is a regularization technique that uses the least squares framework with the addition of another term that depends on a regularization parameter λ. This term can be viewed as a penalization aimed at improving the conditioning of the system. One can employ the colon notation to denote the Frobenius inner product and Frobenius norm, respectively:

$$A\colon B = \mathrm{tr}(A^T B), \tag{32}$$

$$\|A\|_F^2 = \mathrm{tr}(A^T A) = A\colon A, \tag{33}$$

and consider the error function of the full matrix case:

$$R = M^m - TW, \tag{34}$$

where M^m represents measured moments, W are axle weights and T is a matrix used to perform the convolution procedure between influence line and weights.

In Tikhonov regularization, one should minimize a function f, comprising the error norm as well as the regularized solution norm, with respect to the matrix W. Therefore, writing function f as:

$$f = \|R\|_F^2 + \lambda\|W\|_F^2 = R\colon R + \lambda W\colon W, \tag{35}$$

it is straightforward to compute the differential and gradient as follows:

$$\begin{aligned} df &= 2R\colon dR + 2\lambda W\colon dW \\ &= 2(M - TW)\colon -T dW + 2\lambda W\colon dW \\ &= 2T^T(TW - M)\colon dW + 2\lambda W\colon dW \\ &= (2T^T(TW - M) + 2\lambda W)\colon dW \end{aligned} \tag{36}$$

$$\frac{\partial f}{\partial W} = 2T^T(TW - M) + 2\lambda W. \tag{37}$$

Setting this gradient to zero, one can find an expression for the optimal weight matrix:

$$\begin{aligned} 2T^T(TW - M) + 2\lambda W &= 0 \\ T^T TW - T^T M + \lambda W &= 0 \\ (T^T T + \lambda I)W &= T^T M \\ W &= (T^T T + \lambda I)^{-1} T^T M, \end{aligned} \tag{38}$$

where the matrix I represents the identity matrix.

Considering that each λ parameter defines a unique regularized solution W_λ, the nontrivial task is to obtain the optimal regularization parameter for the problem solution. Numerous methods exist for this task such as cross-validation [24], ridge trace [25], and the L-curve method [26]. In their study, O'Brien et al. [11] opts for the L-curve method where two norms are defined. The first one is the residual norm of the error for each specific regularization parameter, given by:

$$E_{norm} = \sqrt{(M - TW_\lambda)^T (M - TW_\lambda)}. \tag{39}$$

The second, is the norm of the solution for each regularization parameter, given by:

$$F_{norm} = \sqrt{W_\lambda^T W_\lambda}. \tag{40}$$

According to the method, the optimal λ is located at the corner of the curve constructed by plotting F_{norm} and E_{norm} on a log-log scale. The corner, represents a trade-off between bias and variance on the system approximation. Fig. 1 illustrates the usual shape of the L-curve. O'Brien et al. [11] does not details the process of optimal lambda selection. Nevertheless, other regularization parameters could be obtained given different approaches for finding the L-curve corner such as the Spline-based Curvature Method [27], Triangle Method [28] or Adaptive Pruning [29]. Considering that the numerous algorithms focused only on the L-curve approach exist, which often returns different "optimal" points, the task of finding the optimal regularization parameter is seen as complex and subjective [10].

2.4.2. Statistical aspects behind the regularization approach

In order to understand the statistical interpretation of the formulation employed by O'Brien et al. [11], we make use of the analogy of Bayesian methods and regularization where the maximum a priori estimate of a normal prior with normal likelihood results in the same estimation as a Tikhonov regularization (for demonstration and proof the reader is referred to Aster et al. [20]). That is, by the Bayesian perspective, initial probability statements are updated, providing a posterior distribution that combines both prior knowledge and the data

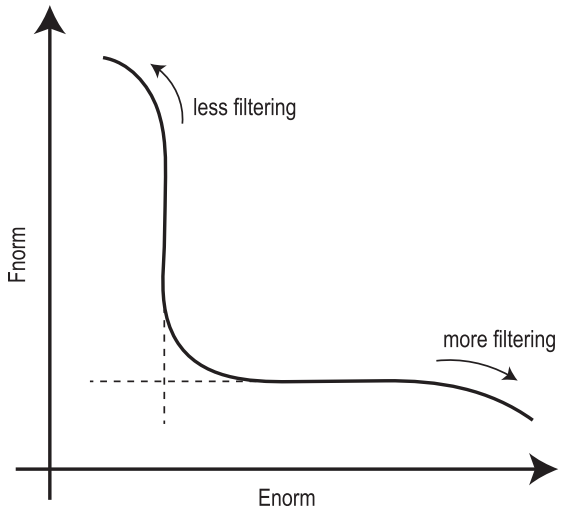


Fig. 1. The generic form of the L-curve plotted in double-logarithmic scale, adapted from: [23].

at hand [30]. Thus, instead of a single estimate as output, Bayesian statistics provide a probability distribution for the model parameters. It is worthwhile mentioning that it is possible to derive a single estimate from such posterior as, for instance, its maximum *a posteriori* (MAP). The prior distribution is obtained independently of the results of measurements [31]. It could represent in a B-WIM application, for example, the knowledge about the non-negativity of predicted axle weights.

The regularization procedure could be interpreted from a Bayesian perspective. The MAP solution obtained by using a prior with independent and normally distributed model parameters is precisely the Tikhonov regularized solution, as demonstrated in Aster et al. [20]. Aligning this interpretation with the already discussed aspect of a trade-off between bias and variance, the regularization parameter arises as the ratio between noise and prior variances [32]. Thus, a lower regularization parameter means a less informative prior is adopted. For the extreme case of zero in the regularization parameter, the least square solution is obtained.

However, Tikhonov regularization and Bayesian approach are not entirely equivalent, since Bayesian solution is a probability distribution, while the Tikhonov solution is a single set of parameters [20]. Hence, the regularization procedure could be seen as a bridge between non-Bayesian and Bayesian estimation problems [32].

When employing Bayesian or regularization approaches, some bias is introduced to the formulation. Accordingly to the adopted method, straightforward ways for incorporating each kind of prior knowledge could be reached. However, the formulation developed by O'Brien et al. [11] has not made use of any prior information as, for example, the already discussed non-negativity of axle weights. Thus, making use of such knowledge could result in improved weigh procedures, for both regularization and Bayesian approaches.

2.5. Including transverse position into formulation

The work of Zhao et al. [9] accounts for the transverse distribution of axle loads on each girder in the formulation. By using the calculated influence line of each girder as a reference, a modified 2D Matrix method was derived to identify axle weights of moving vehicles. Zhao et al. [9] emphasized that, although there are theoretical approximations for the transverse load distribution, the measured data was employed to generate one transverse distribution coefficient for each girder.

The results presented are based on two calibration vehicles, passing 10 times on each lane. Therefore, the influence line of each girder is the mean of each event. Zhao et al. [9] concluded that the method is suitable for simply supported concrete slab-girder bridges. It is noteworthy that the study did not compare the results with the matrix method, or any other method. Finally, some limitations are cited at the end of the paper, such as that the methodology is not suitable for box-girder bridges or other long-span bridge forms.

2.5.1. The modified 2D Matrix method itself

An additional parameter Q_g is introduced into the model of the 2D modified method. It is supposed to represent the transverse distribution of the vehicle loads on each girder. Zhao et al. [9] calculated it based on the 50 largest strains, as the ratio between the strain of a given girder and the total value:

$$Q_g = \frac{\varepsilon_{g,k}}{\sum_{g=1}^{G} \varepsilon_{g,k}}, \tag{41}$$

where k is a scan related to one of the 50 largest strains and $\varepsilon_{g,k}$ is the strain measured for scan k and girder g, of a total of G girders. Thus, this measure can be seen as the percentage of total strain that distributes on each girder. Although the study recognizes that along the driving direction the transverse load distribution factor of each girder is position-dependent, a constant value for each girder was applied. The reason is that most slab-girder bridges have equally distributed lateral connectivity in the vicinity of the mid-span. Furthermore, it supposes that, after calculating the contribution of each girder, each one is responsible for their own load only, independent of the others.

The algorithm for influence line calculation is analogous to the Matrix method, however resulting in one influence line per girder. It is based on a least-squares minimization of the difference between the measured and predicted strains at mid-span:

$$R_g = \sum_{k=1}^{K} (\varepsilon_{g,k}^m - \varepsilon_{g,k}^t)^2, \tag{42}$$

where $\varepsilon_{g,k}^m$ and $\varepsilon_{g,k}^t$ are the measured and predicted strains at scan k and girder g. Furthermore, R_g is the squared error, calculated individually for each girder g. The predicted strain could be written as:

$$\varepsilon_{g,k}^t = \frac{1}{E_g Z_g} \sum_{j=1}^{J} W_j Q_g I_{g,(k-C_j)} \tag{43}$$

where $W_j Q_g$ is the weight contribution of the $j-th$ axle on each girder.

Using the matrix notation applied for all methods analyzed here, the influence line for each girder could be calculated as:

$$IL_g = ((AQ_g)^T (AQ_g))^{-1} (AQ_g)^T E_g Z_g \varepsilon_g^m, \tag{44}$$

where IL_g is the vector of influence line ordinates for the girder g, A is Toeplitz matrix of loads described in Eq. (10).

The weighing procedure is based on a least square minimization between predicted and measured moments, considering the axle weights as variables of interest. In order to maintain consistency with the previous formulation, the weighing procedure may be written as:

$$W = ((\Lambda_Q)^T \Lambda_Q)^{-1} \Lambda_Q \sum_{g=1}^{G} (E_g Z_g \varepsilon_g^m) \tag{45}$$

$$= ((\Lambda_Q)^T \Lambda_Q)^{-1} \Lambda_Q M^m, \tag{46}$$

where Λ_Q is a matrix of influence line ordinates defined by:

$$\Lambda_Q = IL_G \times_n Q^T, \tag{47}$$

where Q is the vector grouping the transverse distribution parameters of all girders, $\times_n$ is the n-mode tensor product [33] of a third-order tensor IL_G with the vector Q, with IL_G defined as:

$$IL_G^{kjg} = \Lambda_g^{kj} \tag{48}$$

with Λ_g defined by Eq. (6) for every girder g.

2.5.2. Comparing Matrix method and its 2D modification

In order to better understand the relation of the method proposed by Zhao et al. [9] and the Matrix Method in O'Brien et al. [19], one should take into account the formulation differences. Recalling Eq. (44), it is thus, possible to extract the constant Q as in:

$$\begin{aligned} IL_g &= ((AQ_g)^T (AQ_g))^{-1} (AQ_g)^T E_g Z_g \varepsilon_g^m \\ &= Q_g^{-2} (A^T A)^{-1} Q_g A^T E_g Z_g \varepsilon_g^m \\ &= Q_g^{-1} (A^T A)^{-1} A^T E_g Z_g \varepsilon_g^m. \end{aligned} \tag{49}$$

For the special case where the transverse distribution is constant along the bridge, this expression can be further simplified by writing:

$$Q_g = \frac{\varepsilon_g^m}{\varepsilon^m}, \tag{50}$$

where ε^m is the total strain over all girders obtained during one calibration event. Merging Eqs. (49) and (50):

$$IL_g = \left(\frac{\varepsilon_g^m}{\sum_{g=1}^{G} \varepsilon_g^m} \right)^{-1} (A^T A)^{-1} A^T E_g Z_g \varepsilon_g^m = (A^T A)^{-1} A^T E_g Z_g \sum_{g=1}^{G} \varepsilon_g^m. \quad (51)$$

Therefore, if all girders have the same properties, namely E_g and Z_g, the Modified 2D method results in equal influence lines for each girder. Still under this condition, IL_g recovers the expression that would be obtained by employing the Matrix method. Thus, under such assumptions, both methods are equivalent in calibration.

In order to extend the analysis for the weighing procedure, the same assumption made in Eq. (50) could be incorporated into the theoretical bending moments expression, for any given scan:

$$\sum_{g=1}^{G} E_g Z_g \varepsilon_g^t = \sum_{g=1}^{G} W Q_g \Lambda_g. \quad (52)$$

Supposing that E_g and Z_g are equal for all girders, namely E and Z, it results that the influence lines IL_g would be the same, as previously discussed. Thus, Λ_g will also be the same, since it is a function of the influence lines of each girder. Calling Λ such a matrix of influence lines:

$$EZ \sum_{g=1}^{G} \varepsilon_g^t = W \Lambda \sum_{g=1}^{G} \left(\frac{\varepsilon_g^m}{\varepsilon^m} \right) \quad (53)$$

$$\sum_{g=1}^{G} E_g Z_g \varepsilon_g^t = W \Lambda. \quad (54)$$

Thus, the predicted moments for Matrix method and modified 2D Moses are the same, in this case. Therefore, one can conclude that, under the hypotheses of constant weight distribution and same mechanical properties of each girder, the weighing procedure also recovers the same solution found when applying the Matrix method, becoming independent of the distribution parameters.

2.6. Other approaches found in the literature

Regarding the methods that are focused in this paper, it is worth to point out that there are some proposed methodologies that address the same problem, however with characteristics that prevent a suitable comparison. In what follows, some of them are summarized and such aspects are discussed.

An approach to construct a more realistic influence line was proposed by Zhao et al. [34]. Their research intended to overcome the limitations of the theoretical influence line used by Moses in his first work. The bridge was modeled with semi-rigid connections and horizontal springs as boundary conditions. Moreover, it was included in the formulation the transverse load distribution, as proposed by Zhao et al. [9]. To completely define the model, it was necessary to estimate the values of some stiffness coefficients. In order to find these values, the author performed a trial and error procedure comparing measured and modeled values. Two algorithms were proposed: semi-rigid approach, adjusting end moments and semi-rigid approach, using moments of the whole bridge. Such methods were compared with Moses, employing the theoretical influence line. The experimental procedure used two trucks to calibrate and validate the method. The results showed that both proposed approaches achieved better results, specially the second method, which reached the lowest error among all. However, the procedure for finding the stiffness parameters that define the model does not have a clear definition. Thus, as such analysis could be user-dependent, performed comparisons may be inconclusive.

Kim et al. [18] proposed an approach to weigh vehicles using deformation measures as inputs to neural networks, obtaining the weights of each axle as the output. The error for both gross vehicle weight (GVW) and weight by axle were considerably low, indicating that the proposed approach could be applied in real situations. Nevertheless, the training process needed numerous training examples, preferably containing vehicles with distinct number of axles, which is not available in most practical cases. The advantage of this method was that it could be applied in cases where the traditional approach have some difficulties, which is not the focus of the present work.

In the work of Helmi et al. [35] three weigh methods were compared utilizing data of a real bridge in Canada. The first two methods were developed by the authors and consisted in the creation of an equivalent uniform distributed load to represent the axle loads, considering the influence line of a simply supported beam. The authors tried to find the fraction of the bridge span, corresponding to the length of the equivalent distributed load, which causes the maximum moment in the bridge. Thus, GVW could be calculated as the ratio between the maximum moment and this length. Nevertheless, both of the authors' proposed methods performed worse than the third alternative tested, namely the Beta method from Ojio and Yamada [36]. This method used the area under the moment or strain curve to calculate GVW, where errors of less than 5% were observed. However, none of such methods is able to distinguish the weight contribution of each axle, which is a parameter of comparison in the present paper.

The work of Frøseth et al. [3] intended to overcome issues related to implementation complexity and computational cost through the realization that the response of the structure is the convolution of the influence line and the loading. Thus, instead using the well-established matrix method, the author suggested that the convolution could be performed in frequency domain, since the convolution integral transforms into an element-wise multiplication operation, which is very efficiently handled. The reported gains in computational time were, in general, of one order of magnitude, at least. Another advantage of viewing the problem under the proposed aspect, was that the matrices utilized in the least-squares approach in O'Brien et al. [11] were straightforward to generalize for arbitrary number of axles. It is worth to mention that it may be necessary to apply a Tikhonov regularization in order to perform the deconvolution. The reason is that the system solution could result in an ill-posed problem in the frequency domain, for example, when the passing vehicle has two axles with identical loads. The author concluded that the obtained influence line provides virtually identical results in comparison with the matrix method. Therefore, the main practical utility of this approach was not precision, but computational gains. As computational complexity is not addressed in the present paper, no further analyses are performed for this method.

3. Numerical investigation

In addition to the theoretical argument presented in Section 2, this study aims to evaluate the numerical performance of discussed methods. This is justified given the usual absence of comparison among methods in literature. Table 1 illustrates this matter on studies that shall be evaluated. In this table, each row presents the method name, the comparison form employed, the type of data gathered and the method's characteristic of obtaining either influence line or weight. It becomes clear that recent work on the field has not been taken into account.

In order to compare the methods in a set of different conditions, a parametric investigation is conducted. The main factors that may

Table 1
Literature method comparisons.

Method	Comparison	Data	IL	Weight
Matrix [19]	Measured/Predicted	Real	✓	–
MLE [8]	Matrix method	Real	✓	–
PBWIM [10]	Matrix method	Real	–	✓
Regularization [11]	Matrix method	Synthetic	–	✓
Modified 2D Moses [9]	Measured/Predicted	Real	✓	✓

influence the performance of B-WIM algorithms are simulated numerically and the trend in gross vehicle weight error is measured and presented as result.

3.1. Numerical simulations

Recalling, the methods evaluated are the Matrix method [19], maximum likelihood (MLE) [8], pBWIM [10], Regularization [11] and Modified 2D Moses [9]. The procedure employed for finding the corner of the L curve for the regularization approach consisted on minimizing the Euclidean distance between adjacent points, as suggested in [37]. This is done utilizing all the pairs (E_{norm}, F_{norm}) employed for plotting the L-curve.

Regarding computational effort, almost all methods show similar values, being negligible for simulation purposes. The only exception is the pBWIM method, since this method depends on an optimization procedure for finding the most likely weights. Thus, while other methods can perform the computation directly by solving a single linear system, pBWIM may need several similar solution steps. The computational effort is shown to be increased by some orders of magnitude, mainly when the number of axles increases. This issue is partially handled in this study by employing a modified optimization procedure than that proposed by O'Brien et al. [10], which employed a grid search. Nevertheless, the computational effort still remains as a drawback for this method.

The model applied to artificially simulate the bridge strains is based on approximating the bridge behavior by a simply supported Euler-Bernoulli beam model, employing systems of sprung masses to represent each axle. The whole description of numerical procedures employed is further discussed in Appendix A.

The influence of two main aspects are evaluated on gross vehicle weight prediction, namely, road pavement profile and bridge span. Three distinct road profiles are evaluated, with Power Spectral Density (PSD) amplitudes of zero (no roughness), 4 and 16, where roughness increases with the amplitude. Further details on the description of this model are given in the Appendix A. For the bridge span parameter, three cases are considered, with spans of 10, 20 and 30 meters. Thus, the combination of every case of road profile and bridge span results in 9 distinct cases. Furthermore, a Gaussian noise with signal to noise ratio of 20 is added to every signal, aiming to incorporate measurement errors due to other sources than pavement roughness. It is worth to mention that the values adopted for all these cases intended to reflect the recommendations of Jacob [38] for B-WIM sites, hence, reproducing practical cases of interest.

In order to better approximate simulations with the real in-service operation behavior, a total of 200 vehicles, with number of axles ranging from 2 to 9, are simulated. The procedure for generating vehicles is random, where the absolute value of a normally distributed random variable ($\theta \sim \mathcal{N}(0, 1)$), is sampled and applied to Eq. (55).

$$J = \min(1 + \lceil 3.5(\mathrm{abs}(\theta)) \rceil, 9). \tag{55}$$

where $\lceil . \rceil$ is the ceil function, abs(.) is absolute value and J is the number of axles of the generated vehicle.

The histogram of vehicles sampled by this procedure is presented in Fig. 2 to illustrate that lower weight vehicles are more frequently created, an approach that intended to simulate a usual real scenario.

After the definition of the number of axles of the vehicle, a vehicle type is randomly chosen, which defines the bounds on axle spacing, damping, stiffness and maximum allowable weight on each axle. A total of 16 vehicle types are applied in this study, which are defined in Appendix B, based on a classification often used in Brazil. Axles spacing, damping, stiffness and weights are uniformly sampled based on the previously defined limits. Since axles weights does not have a minimum defined, a value of 80% of the maximum allowable weight is adopted. Furthermore, the velocity is uniformly sampled from a random variable with 10 m/s and 25 m/s as lower and upper bounds,

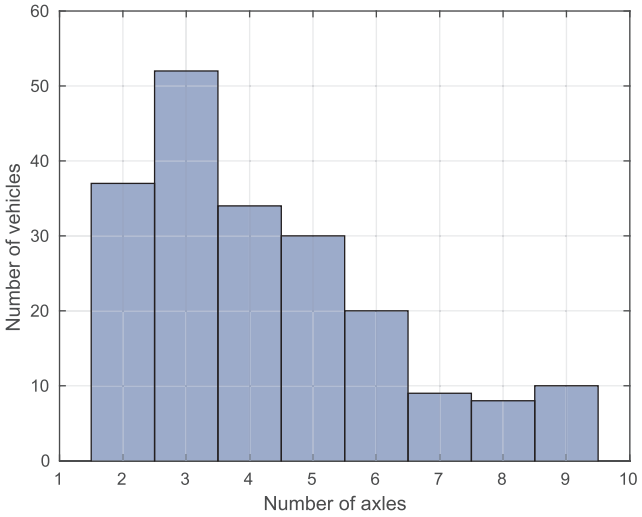


Fig. 2. Histogram of number of axles of vehicles created.

Table 2
Constant parameters.

Propertie	Value (Units)
Bridge modulus of elasticity	10^{10} (Pa)
Bridge damping coefficient	0.05 (–)
Bridge moment of inertia	0.5 (m^4)
Bridge mass per unit length	10^4 (kg/m)

respectively. The remaining parameters related to the bridge are taken as constants and are defined in Table 2. The same set of vehicles are applied to each one of the 9 cases.

A total of 40 runs of the vehicles 2C and 2S2, 20 for each of them, are applied to calibrate the system, simulating the real scenario where only a limited amount of calibration vehicles is available. For the calibration runs, vehicles speed are adopted accordingly with suggestions of Jacob [38]. Thus, 12 runs are executed with a mean velocity, adopted as 20 m/s here, 4 runs with 16 m/s and the remaining 4 runs with 24 m/s. After the calibration, all methods are applied to a test set comprised of the 200 runs which were previously generated.

3.2. Numerical results

The algorithms are evaluated in a set of different scenarios for assessing performance aspects and sensitivity of the methods. The comparison criterion consists of the mean absolute percentage error related to the known GVW. The results are shown in Fig. 3, as a function of roughness amplitude and bridge length. In order to facilitate the distinction among the performance of the methods, these values are also presented in Table 3, where some differences appear more evidently.

The simultaneous presentation of performance evolution with respect to both parameters, in Fig. 3 and Table 3, allow for quite interesting remarks. Firstly, all methods showed similar performance, mainly Matrix method, Regularization and Modified 2D Moses. The MLE method also shared the same trend, however, its performance surpassed the other methods when bridge span increases. This fact is observed independently of the roughness amplitude applied. Thus, although the difference is not so remarkable, the MLE method can be argued as the most accurate method for this data set.

On the other hand, pBWIM method showed similar performance for lower roughness amplitude, becoming worse than the others with the increase in this parameter. It is worth mentioning that the performance loss occurs in conjunction with a more computationally expensive prediction procedure.

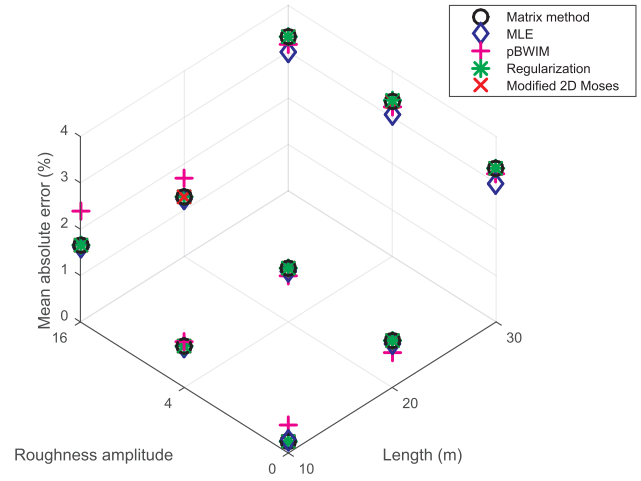


Fig. 3. Mean absolute error of each method as a function of bridge length and roughness amplitude.

Table 3
Mean absolute error for each method and case.

Amplitude	Length (m)	Matrix	MLE	pBWIM	Regularization	Modified 2D Moses
0	10	0.26	0.26	0.61	0.26	0.26
4	10	0.88	0.88	0.99	0.88	0.88
16	10	1.65	1.62	2.39	1.65	1.65
0	20	1.02	0.93	0.75	1.02	1.02
4	20	1.17	1.09	0.98	1.17	1.17
16	20	1.28	1.24	1.69	1.28	1.28
0	30	3.31	2.99	3.19	3.31	3.31
4	30	3.35	3.06	3.21	3.36	3.35
16	30	3.32	3.00	3.15	3.32	3.32

Table 4
Max of absolute difference between each method and the Matrix method.

Amplitude	Length (m)	MLE	pBWIM	Regularization	Modified 2D Moses
0	10	0.12	1.32	0.02	0.00
4	10	0.30	5.44	0.02	0.00
16	10	0.53	18.26	0.03	0.00
0	20	0.52	2.43	0.01	0.00
4	20	0.93	6.30	0.01	0.00
16	20	1.54	15.21	0.01	0.00
0	30	1.22	5.24	0.01	0.00
4	30	1.30	6.15	0.01	0.00
16	30	1.38	5.35	0.01	0.00

Since the whole bridge is simulated as a unique beam, it is already expected for Modified 2D Moses and Matrix method results to be numerically identical. Indeed, in all evaluated cases and runs, the results are exactly the same. However, Regularization also shows similar behavior, presenting only slight deviations from the trends in the Matrix method. Such aspects can be seen in Table 4, where the maximum deviation of each method with respect to the Matrix method is presented. One can conclude that, in the conditions of this analysis, the difference in performance among Matrix method and Regularization is not significant. The reason is that the parameters found were usually close to zero, which promotes solutions very close to the least squares in the Matrix method. Nevertheless, it does not mean that the regularization approach is not useful, however, in the examples, the application of the L-curve corner lead to similar results. Therefore, it did not have a large influence in the cases analyzed here.

In order to correctly analyze the method presented by Zhao et al. [9], it is necessary to take into account multiple beams in the bridge, which is addressed in the following sections.

When comparing the evolution of mean absolute error regarding bridge length and roughness amplitude separately, one can conclude that the first is more problematic. The difficulty of static methods in dealing with long span bridges is a fact already well known on B-WIM literature [38,18]. However, the comparison of variation of bridge length and roughness amplitude shows an interesting aspect. Analyzing the results, it is clear that bridge length has a higher impact on the overall result than the roughness amplitude, since the cases of smooth profile and 30 m bridge span easily surpass the error of the 10 m bridge span and roughness amplitude of 16.

It is interesting to notice that the increase in bridge length spreads the effect of roughness in the prediction accuracy. Thus, the result is more sensible to the road pavement profile for short span bridges. The opposite is not true, in other words, independently of the roughness amplitude, the increase in bridge length decreases the accuracy of weight prediction.

3.3. Multiple beams analysis

Given that Modified 2D Moses and Matrix method resulted in rigorously the same predictions in previous analyses, it becomes necessary simulating cases where the bridge response is modeled considering multiple beams. In this section, the bridge structure is comprised of 3 distinct beams, where the simulation details are also referred to Appendix A. Moreover, the same vehicles and properties previously defined are adopted in this section. Three distinct cases of transverse distribution (Q) are considered, defined by Table 5.

Since pBWIM, MLE and Regularization methods do not make any assumption regarding transverse distribution of loads, their evaluation for this new case should not bring any new insight. Thus, only as a comparison criterion, the Matrix method is jointly evaluated with Modified 2D Moses. As a result of this analysis, Figs. 4–6 present the influence of roughness amplitude and transverse distribution case for each bridge span.

In all cases, Modified 2D Moses performed equally or worse than Matrix method. In the case of low roughness amplitude, both methods could be argued as similar. When roughness amplitude increases, otherwise, there is a trend for the Modified 2D Moses method to present higher errors, for all analyzed bridge spans. Such performance gap is specially remarkable in the 10 m bridge. This last statement is in agreement with the previous section, where the effect of road profile showed higher impact for short span bridge cases.

One possible explanation for the performance gap can be seen in Fig. 7. It shows the influence of the error in approximating Q (measured as the mean of Euclidean distance among predicted and real Q values) and the percentage difference between the results of the two methods. This figure makes clear that the accuracy in Q estimate is directly related to the difference in performance when compared with the Matrix method. Therefore, if transverse distribution factors can be accurately predicted, Modified 2D Moses approximates Matrix method performance. This fact is in accordance with Section 2.5.2, since Q values are kept constant in the simulation. On the other hand, the increase in difference of such factors also increases the likelihood that weight prediction is corrupted.

Table 5
Load distribution factors for each beam (%).

Case	Beam 1	Beam 2	Beam 3
Q1	33.3	33.3	33.3
Q2	20.0	40.0	40.0
Q3	25.0	50.0	25.0

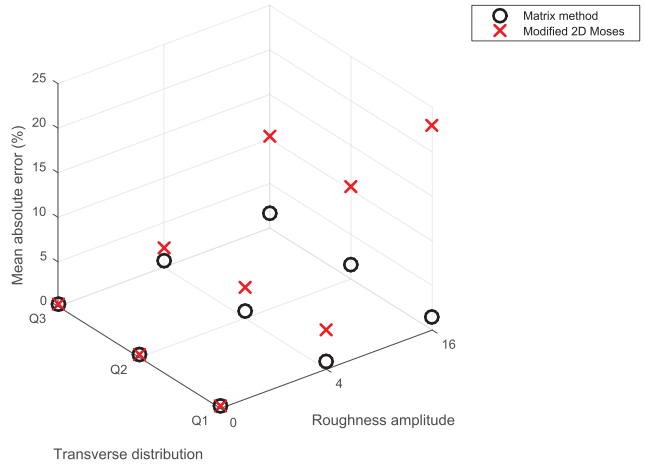


Fig. 4. Comparison between Matrix method and Zhao for bridge span of 10 m.

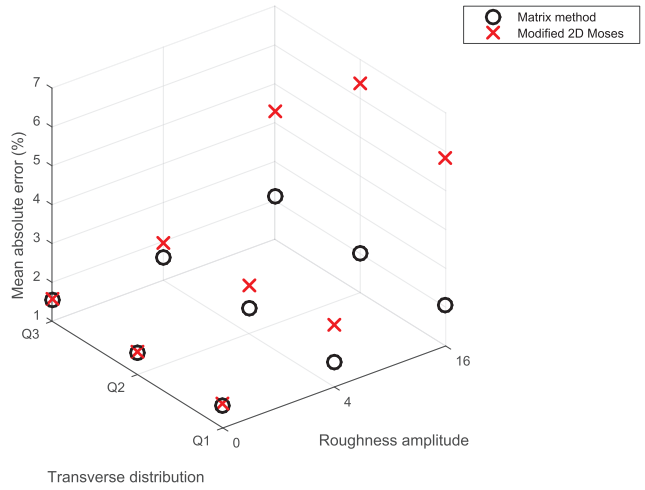


Fig. 5. Comparison between Matrix method and Modified 2D Moses for bridge span of 20 m.

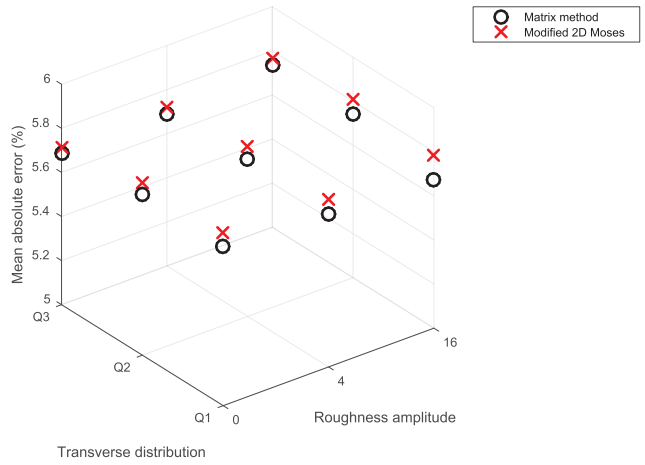


Fig. 6. Comparison between Matrix method and Modified 2D Moses for bridge span of 30 m.

4. Field-testing

In order to compare the different methods presented in Section 3.2 on a real-world setting, data from a bridge located in the city of Uruaçu, Brazil, is employed. The main aspects regarding such bridge and the calibration procedure are described in what follows.

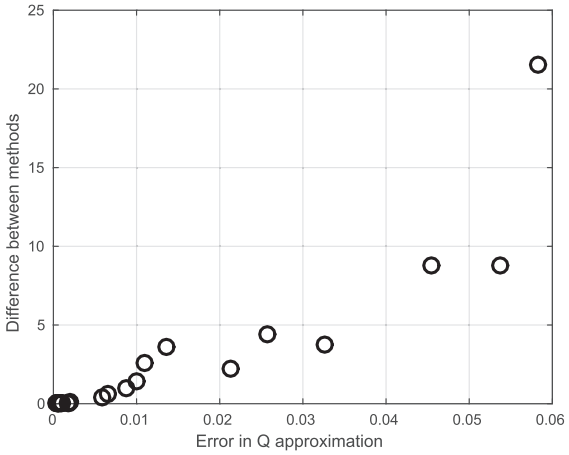


Fig. 7. Influence of error in transverse distribution factor on the difference between methods.

Fig. 8. Itinguijada bridge.

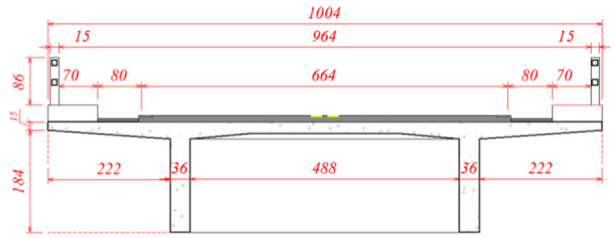


Fig. 9. Mid-span cross section dimensions.

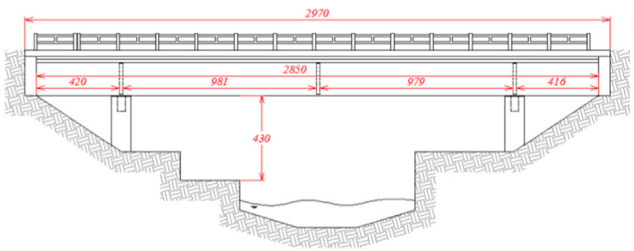


Fig. 10. Lateral view dimensions.

Table 6
Axle weights and spacing for Itinguijada bridge.

	Axle weight (kN)					Axle spacing (m)			
	Axle 1	Axle 2	Axle 3	Axle 4	Axle 5	d_{12}	d_{23}	d_{34}	d_{45}
3 axle vehicle	67.7	146.2	125.5	–	–	4.78	1.28	–	–
5 axle vehicle	73.6	138.3	130.4	108.9	90.3	3.57	5.59	1.26	1.23

Table 7
Mean absolute error.

	Matrix method	MLE	pBWIM	Regularization	Modified 2D Moses
GVW (%)	4.47	4.08	4.91	4.40	4.41
Single axle (%)	17.79	15.26	36.46	15.35	16.91
Group of axles (%)	6.62	5.38	5.78	6.22	6.52

Table 8
Difference between each approach and the Matrix method.

	MLE	pBWIM	Regularization	Modified 2D Moses
Mean (%)	1.2101	4.2554	0.3860	0.1526
Std (%)	0.6351	3.6975	0.2901	0.1153
Maximum (%)	2.5673	16.6928	1.2900	0.6864

4.1. Bridge and vehicle description

The Itinguijada bridge, shown in Fig. 8, is comprised by two girders and five cross beams, with a total length of 29.0 m. Figs. 9 and 10 show the main dimensions of the cross section and the lateral view, respectively. One FAD sensor is installed in the mid-span of the bridge, while the other is spaced 4 meters longitudinally from the first. Two strain sensors were attached to each girder, at the mid-span.

Two trucks, with three and five axles, are used for calibrating and evaluating the system performance. Several runs with each of these vehicles are conducted, totaling at least 10 runs per truck and lane. The axle spacing and weight distribution for the calibration vehicles are shown in Table 6.

4.2. Numerical results

In this section, following what is done in practice, as reported in Lydon [5] and Yu et al. [1], one influence line is constructed for each lane. The analyses are based on prediction for three distinct quantities, defined by Jacob [38]: GVW, single axle and group of axles. The mean absolute error of such quantities for each method on the whole data set is presented in Table 7.

All methods showed higher errors for single axle prediction when compared to the group counterparts, which is in accordance with most studies in this subject, as in Zhao et al. [9] and O'Brien et al. [11]. The prediction for GVW and group of axles weight showed reasonable results, with mean absolute errors always smaller than 7%. However, single axle prediction did not present the same level of performance, achieving values as high as 36% for pBWIM method.

MLE reached the best results, independently of the quantity being measured. However, for almost all methods and quantities measured, the mean absolute errors reported remained at a quite similar level. Since all methods disregarded dynamic effects, it is expected that the higher such effects are, the lower the suitability of all approaches are. As in this bridge the dynamic behavior is not remarkable, the results for MLE suggests that this method should perform better in cases where the model seems to be more suitable.

The similarity of measured errors previously discussed justify a more detailed analysis on it. Table 8 focus on this statement, based on the absolute difference in GVW, for each event, between all methods and the Matrix method, taken as reference here. The parameters presented are the mean, standard deviation and maximum value of the absolute difference.

From the five analyzed methods, two can be seen as almost identical to the Matrix method. The first of them is the Regularization method, whose maximum difference in GVW for all events does not surpass 1.29 %. This lack of difference, as already discussed in Section 3, is caused by small regularization parameters obtained from the application of L-curve method. The second method quite similar to the Matrix method is Modified 2D Moses, in which maximum difference did not reached 1%. As already discussed, all girders have the same mechanical properties, which remain constant along the span. Thus, it is possible that the contribution factor Q_g resulting for each girder and event be approximately a constant value. As presented in Section 2.5.2, in such a case it is already expected for results of Matrix method and modified 2D Moses to be similar.

The remaining two methods, MLE and pBWIM, showed more distinct values, rendering them as alternative approaches for the Matrix method. Observing these three methods by a probabilistic point of view, these differences are remarked. The Matrix method is based on least squares, which assumes that errors are uncorrelated, normally distributed and with the same variance, using all the measurements of one event as the realization of only one normal random variable. The MLE method also applies the least squares approach, however now considering a multivariate normal distribution whose variables are moments measured in each ordinate, independently of the other ordinates. The pBWIM relaxes the assumption of equal variance, allowing for each ordinate to have a standard deviation estimated by the value of each influence line previously calculated from calibration events. Therefore, in some cases, namely when the data from both distinct events and the whole bridge have the same normal distribution, all methods should

Table 9
Vehicle types employed in the study.

Vehicle Class	Axles
2C	2
3C	3
4CD, 2S2	4
3I2, 2S3, 3S2	5
2R4, 3S3, 3D3, 3N3	6
3D4, 3N4	7
3D5	8
3M6, 3Q6	9

Table 10
Mean values of stiffness and damping coefficients for each kind of axle

Axle type	Stiffness (N/m)	Damping (Ns/m)
Rear	1,000,000	10,000
Front	400,000	10,000
Semi trailer	750,000	10,000

present virtually the same results. However, considering the real-world examples presented, this is not the case in practical applications.

One could say that pBWIM should achieve the best result, since this method has the least restrictive formulation. However, one must realize that this can only occur if the assumptions are met, and this is not the case for the pBWIM formulation, as demonstrated in Section 2.3.2. It could be observed that this method assumes uncorrelated error moments. Yet, the formulation based on uncertainties in influence line will necessarily result in banded correlation of the resulting moments random variable, violating this assumption.

5. Further remarks

Among five tested methods, three of them showed almost identical results for all events evaluated, namely, the Matrix method, Regularization and Modified 2D Moses. Although the pBWIM method could be argued as an alternative approach for weight prediction, both theoretical and practical results are not promising. Comparing MLE and Matrix method, the results for the simulated runs were very similar, however with MLE being slightly better. Furthermore, when field-testing was conducted, MLE approach showed even better performance. These points, together with its more suitable statistical background, indicates that MLE is the most promising method evaluated in the present study. Some suggestions for improving these methods are included in the following.

The formulation employed for Regularization leads to a solution that recovers only weights associated with each axle. However, it is possible that the regularization approach can have improved results when trying to recover the whole impulse vector, as more information about its characteristics may be incorporated and enforced by the regularization procedure. Furthermore, it may be worth to employ different regularization procedures instead of Tikhonov, such as Lasso, which uses the L1 norm instead of the usual L2 [39]. As a further step, the non-negativity of the vector can be enforced as in the work of Hummelsheim [40].

Discussing specifically the pBWIM method, it could be concluded that an assumption made in the approach formulation is conceptually violated. Thus, it is useful to change such a methodology to account for a covariance matrix that is not diagonal. Furthermore, the procedure suggested to find the most likely weights in the original paper is quite-time consuming. Thus, applying an optimization procedure is necessary for achieving a reasonable computational cost.

The procedures applied are clearly divided in two steps: building an influence line and weighing the vehicles. From this perspective, one could easily apply distinct methods for each step. For instance, MLE could address the former and regularization the latter. Since it was not the focus of this work, coupling methods was not evaluated. However, it has potential to improve the accuracy of predicted axle weights.

Overall, it becomes clear that the main point that should be addressed is the incorporation of additional knowledge to the model. One aspect is related to dynamic effects. The results could be improved by a more suitable consideration on the dynamic behavior of the vehicle-bridge system model. A second aspect is related to the form of the solution. The knowledge that a vehicle is modeled by point loads on each axle imposes constraints on the form of the impulse vector and the sign of the resulting weights. Nevertheless, relaxing assumptions that over-constrain the model, such as the correctness of axle spacing measurement, may improve flexibility and robustness to the methods.

For all evaluated methods, the weighing procedure disregarded some prior information that could lead to more reliable estimates. As examples, one could cite the non-negativity of weights and the trend for axle weights having values at the same order of magnitude. As discussed previously, a Bayesian approach could be employed in such a case, using these prior information to create improved estimators for axle weights.

When observing methods by the statistical point of view, the assumptions implicitly made in the development of such formulations arose in a more clear fashion. Thus, the possibilities of relaxing some of those assumptions could be seen as a good initial point for the development of new methods. Furthermore, by knowing in advance which statistical assumption was made, it becomes possible to easily predict for which real cases the new methods could perform better. As an example, allowing errors to be correlated should perform better in problems where the model fits poorly, since missing information on formulation could be seen as correlated errors. Thus, enabling the model to allow correlated error is an alternative to address the dynamic effects.

The comparison criterion of the evaluated methods was done without a clear performance threshold. Thus, although it is possible to verify which method is best in the tested cases, nothing can be said about robustness or suitability of overall results. Therefore, it motivates the development of a methodology for extracting more useful information from such results, mainly regarding robustness in practical applications.

6. Concluding remarks

This work presented a comparative study regarding some methods applied in B-WIM systems to derive the bridge influence line and predict vehicle weights. In order to fairly compare the analyzed methods, only those which do not explicitly address the dynamic behavior of the vehicle-bridge system were considered. Furthermore, the methods were interpreted from a statistical point of view, where their assumptions were highlighted, remarking their theoretical differences.

In addition to the theoretical comparisons, the methods were implemented and numerically compared, addressing the lack of comparisons found in literature. In order to assess the similarity or the advantage of a method compared to the others, synthetically generated and real data were considered. Furthermore, the influence of some main factors that affect B-WIM systems were evaluated in simulated analyses, such as road profile and bridge length.

The analysis using simulated data showed that the increase in bridge length had more impact in corrupting the overall results than the roughness amplitude. The influence of road profile increased when decreasing the bridge span. When analyzing the performance of the methods, results indicated that pBWIM and Modified 2D Moses were surpassed by the other approaches. Indeed, pBWIM usually provided worse weight estimation performance, together with a considerably higher computational cost. Modified 2D Moses was, for all analyzed cases, equal or worse than Matrix method. The remaining methods, MLE, Matrix method and Regularization, presented very similar results, with MLE being slightly better. The results regarding real data followed the trend already observed in numerical simulation, however with MLE showing superior performance. Thus, from both theoretical and

practical perspectives, the MLE could be argued as the most promising method evaluated in this study.

Overall, the reformulation and reinterpretation of methods under a common statistical point of view allowed a better understanding of the underlying assumptions and theoretical hypothesis. This fact enabled a better comparison among methods and generated an opportunity for further improvements. For example, developing novel methods by relaxing the model statistical assumptions or applying a Bayesian approach to better account for prior information.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001 and the Transport Infrastructure National Department (DNIT).

Appendix A. Sprung-mass simulation

In the present study, the simulations are based on the works of Biggs [41], Yang et al. [42] and Yang and Lin [43]. The resulting dataset is available online [44]. The dynamic behavior of the bridge is modeled supposing a simply supported Euler Bernoulli beam model, under a set of moving sprung-masses. In this approach, each sprung-mass system represents one vehicle axle. Modal decomposition is performed for the equation of motion of the bridge, resulting in Eq. (56):

$$\ddot{q}_i + 2\xi_i\omega_i\dot{q}_i + \omega_i^2 q_i = \sum_{j=1}^{N} \frac{2P_j}{\bar{m}L}\sin\frac{i\pi Vt}{L}, \tag{56}$$

where N is the number of axles, q_i, ω_i and ξ_i are the modal coordinates, natural frequency and damping of the i mode, respectively. Furthermore, dotted variables represent the derivative with respect to time, P_j is the load of the j axle, V the speed of the vehicle, L is the length of the bridge and $\bar{m}$ is the mass per unit length of the bridge.

The equations of motion for each sprung-mass system is given by Eq. (57):

$$M_{vj}\ddot{z}_j + C_{vj}\dot{z}_j + K_{vj}z_j = K_{vj}(u + y)|_{x=Vt} + C_{vj}(\dot{u} + Vy')|_{x=Vt}, \tag{57}$$

where M_{vj}, C_{vj} and K_{vj} are the mass, damping and stiffness of axle j. Furthermore, u represents the bridge vertical displacement, z_j is the vertical position of the j axle, y is the pavement elevation ordinate and a prime denotes the derivative with respect to x.

Both bridge displacement and its first derivative could be computed directly with the modal coordinates and mode shapes of the simply supported beams:

$$u = \sum_{j=1}^{m} \sin\left(\frac{j\pi x}{L}\right)q_i, \tag{58}$$

$$\dot{u} = \sum_{j=1}^{m} \sin\left(\frac{j\pi x}{L}\right)\dot{q}_i, \tag{59}$$

where m is the number of modes applied, which is equal to 5 in all simulations.

The simulation of pavement roughness applied here is a common approach in studies in the same subject [45,46,16]. The ordinates of the pavement irregularities are modeled as a random process, with a specific power spectral density function (PSD):

$$\Phi(\Omega) = \Phi(\Omega_0)\left(\frac{\Omega}{\Omega_0}\right)^{-2}, \tag{60}$$

where $\Phi(\Omega_0)$ is the amplitude coefficient, analyzed in previous sections, measured in (10^{-6} m^3/cycle) and Ω_0 is a reference spatial frequency. Thus, the road profile is generated by sampling from this PSD, using the method of superposition of harmonics [47]:

$$y(x) = \sum_{i=1}^{n\Omega} \sqrt{2\Delta\Omega\Phi(\omega_i)}\cos(2\pi\Omega_i x - \phi_i) \tag{61}$$

where $y(x)$ is the generated road vertical profile, ϕ_i is a random uniform phase angle between 0 and 2π, $\Delta\Omega$ is a constant increment, $n\Omega$ is the total number of frequency increments in the interval and Ω_i is a frequency uniformly distributed in the range of Ω_{min} and Ω_{max}. The parameters adopted are $\Omega_0 = 0.01$ cycle/m, $\Omega_{min} = 0.001$ cycle/m, $\Omega_{max} = 4$ cycle/m and $\Delta\Omega = 0.001$ cycle/m.

It is worth to mention that for each run a distinct road profile is generated, since in the practical scenario lateral deviations occur. Moreover, a moving average with total size of 30 cm is employed to approximate the real contact between tire and pavement [46].

The interaction between bridge and vehicle clearly appears in Eq. (57), by means of the displacement term related to the beam at the current axle position. Moreover, such interaction also occurs in Eq. (56), due to the P_j term. This relation is remarked in Eq. (62):

$$P_j = p_j\delta(x - Vt)\left(H(t - t_j) - H\left(t - t_j - \frac{L}{V}\right)\right), \tag{62}$$

$$p_j = -M_{vj}g + K_{vj}(z - (u + y)|_{x=Vt}) + C_{vj}(\dot{z} - (\dot{u} + Vy')|_{x=Vt}), \tag{63}$$

with $\delta()$ and $H()$ representing the Dirac delta and Heaviside functions, respectively. Moreover, g is the acceleration of gravity and t_j is the time that the j axle arrives the bridge.

The equations of motion are solved independently, by a decoupled approach. Both bridge and vehicle equations are solved numerically by applying the Newmark-β method, with 1400 time steps. The time window begins when the first vehicle axle enters the bridge and ends when the last axle leaves it. The problem is solved iteratively, since the interaction force in the bridge-vehicle system changes with the displacement of both beam and sprung mass. An initial guess of interaction force is given to the beam equations, where the beam displacement is calculated. Such a displacement is then enforced to the vehicle model and a new interaction force is calculated. This procedure continues until the change in the

interaction force reaches a small tolerance (10^{-5}). Usually such procedure converges rapidly, within 5 iterations. The midspan strains (s), which are the main output of simulation, are also updated with this interaction. Adopting, without loss of generality, a unit vertical distance from the neutral axis, and utilizing the fact that strains are related with the second derivative of displacement with respect to x, the strains can be written as:

$$s = -\sum_{i=1}^{m} \left(\frac{i\pi}{L}\right)^2 q_i \sin\left(\frac{i\pi}{2}\right). \tag{64}$$

For the case of multiple beams, a transverse distribution factor (Q) is provided and applied to divide the axle loads for each beam. Thus, each beam is simulated independently.

In order to simulate the inherent imperfections of the measured signal due to all possible aspects in the measurement field, noise is applied to the simulated response. The noise applied consists of a white Gaussian random noise, with a constant signal to noise ratio of 20.

Appendix B. Vehicle types

The vehicles employed are based on the brazilian's traffic and infrastructure department report [48]. As some classes have only a lower bound for axle spacing, the upper bound for such cases is defined as 5 meters, since this is necessary for generating vehicles in a uniform distribution. Table 9 presents all types of vehicle employed as well as their number of axles. In order to allow a more concise presentation, axle spacing and weight of each employed truck are omitted. For assessing such values, the reader is referred to DNIT IPR 723 [48].

On the other hand, damping and stiffness coefficients still need to be defined. In this study, each vehicle have a specific coefficient related to each axle, which is a random variable uniformly distributed around a mean value, presented in Table 10. This distribution have bounds of 0.5 and 1.5 times this mean value, which is adopted based on Fancher [49] and Nosseir et al. [50]. It is worth to cite that such mean values differ accordingly with the type of axle, namely: rear, front or semi trailer axle.

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.engstruct.2019.109463.

References

[1] Yu Yang, Cai CS, Deng Lu. Nothing-on-road bridge weigh-in-motion considering the transverse position of the vehicle. Struct Infrastruct Eng 2018;14(8):1108–22.
[2] Lansdell Andrew, Song Wei, Dixon Brandon. Development and testing of a bridge weigh-in-motion method considering nonconstant vehicle speed. Eng Struct 2017;152:709–26.
[3] Frøseth Gunnstein T, Rønnquist Anders, Cantero Daniel, Øiseth Ole. Influence line extraction by deconvolution in the frequency domain. Comput Struct 2017;189:21–30. https://doi.org/10.1016/j.compstruc.2017.04.014. ISSN 0045-7949.
[4] Moses Fred. Weigh-in-motion system using instrumented bridges. J Transp Eng 1979;105(3).
[5] Lydon Myra, Taylor Su E, Robinson Desmond, Mufti A, Brien EJO. Recent developments in bridge weigh in motion (b-wim). J Civ Struct Health Monit 2016;6(1):69–81.
[6] Yu Yang, Cai CS, Deng Lu. State-of-the-art review on bridge weigh-in-motion technology. Adv Struct Eng 2016;19(9):1514–30.
[7] Žnidarič Aleš, Kalin Jan, Kreslin Maja. Improved accuracy and robustness of bridge weigh-in-motion systems. Struct Infrastruct Eng 2018;14(4):412–24.
[8] Ieng Sio-Song. Bridge influence line estimation for bridge weigh-in-motion system. J Comput Civ Eng 2015;29(1):06014006. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000384.
[9] Zhao Hua, Uddin Nasim, O'Brien Eugene J, Shao Xudong, Zhu Ping. Identification of vehicular axle weights with a bridge weigh-in-motion system considering transverse distribution of wheel loads. J Bridge Eng 2014;19(3):04013008. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000533.
[10] O'Brien Eugene J, Zhang Longwei, Zhao Hua, Hajializadeh Donya. Probabilistic bridge weigh-in-motion. Can J Civ Eng 2018;45(8):667–75. https://doi.org/10.1139/cjce-2017-0508.
[11] O'Brien Eugene J, Rowley Cillian W, Gonzalez Arturo, Green Mark F. A regularised solution to the bridge weigh-in-motion equations. Int J Heavy Veh Syst 2009;16(3):310–27. https://doi.org/10.1504/IJHVS.2009.027135.
[12] Wang Ning-Bo, He Li-Xiang, Ren Wei-Xin, Huang Tian-Li. Extraction of influence line through a fitting method from bridge dynamic response induced by a passing vehicle. Eng Struct 2017;151:648–64.
[13] Wang Yi, Wei-Lian Qu. Moving train loads identification on a continuous steel truss girder by using dynamic displacement influence line method. Int J Steel Struct 2011;11(2):109–15.
[14] Dowling Jason OBrien Eugene J, González Arturo. Adaptation of cross entropy optimisation to a dynamic bridge wim calibration problem. Eng Struct 2012;44:13–22. doi: https://doi.org/10.1016/j.engstruct.2012.05.047. http://www.sciencedirect.com/science/article/pii/S0141029612002908. ISSN 0141-0296.
[15] Deng Lu, Cai CS. Identification of dynamic vehicular axle loads: theory and simulations. J Vib Control 2010;16(14):2167–94.
[16] Wang Haoqi, Nagayama Tomonori, Zhao Boyu, Di Su. Identification of moving vehicle parameters using bridge responses and estimated bridge pavement roughness. Eng Struct 2017;153:57–70.
[17] Richardson Jim, Jones Steven, Brown Alan, O'Brien Eugene J, Hajializadeh Donya. On the use of bridge weigh-in-motion for overweight truck enforcement. Int J Heavy Veh Syst 2014;21(2):83–104.
[18] Kim Sungkon, Lee Jungwhee, Park Min-Seok, Jo Byung-Wan. Vehicle signal analysis using artificial neural networks for a bridge weigh-in-motion system. Sensors (Basel) 2009;9(10).
[19] O'Brien Eugene J, Quilligan Michael, Karoumi Raid. Calculating an influence line from direct measurements. Bridge Eng Proc Inst Civ Eng 2006;159(BE1):31–4.
[20] Aster Richard C, Borchers Brian, Thurber Clifford H. Parameter estimation and inverse problems vol. 90. Academic Press; 2011.
[21] Washington Simon P, Karlaftis Matthew G, Mannering Fred. Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC; 2010.
[22] Chatterjee Samprit, Hadi Ali S. Regression analysis by example. John Wiley & Sons; 2015.
[23] Hansen P. Rank-deficient and discrete Ill-posed problems. Soc Ind Appl Math 1998. https://doi.org/10.1137/1.9780898719697.
[24] Golub Gene H, Heath Michael, Wahba Grace. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 1979;21(2):215–23.
[25] Hoerl Arthur E, Kennard Robert W. Ridge regression: biased estimation for non-orthogonal problems. Technometrics 1970;12(1):55–67.
[26] Hansen Per Christian. Analysis of discrete ill-posed problems by means of the l-curve. SIAM Rev 1992;34(4):561–80.
[27] Hansen Per Christian, O'Leary Dianne Prost. The use of the l-curve in the regularization of discrete ill-posed problems. SIAM J Sci Comput 1993;14(6):1487–503.
[28] Longina Castellanos J, Gómez Susana, Guerra Valia. The triangle method for finding the corner of the l-curve. Appl Numer Math 2002;43(4):359–73.
[29] Hansen Per Christian. Toke Koldborg Jensen, and Giuseppe Rodriguez. An adaptive pruning algorithm for the discrete l-curve criterion. J Comput Appl Math 2007;198(2):483–92.
[30] Congdon Peter. Bayesian statistical modelling vol. 704. John Wiley & Sons; 2007.
[31] Tarantola Albert. Inverse problem theory and methods for model parameter estimation vol. 89. Siam; 2005.
[32] Kaipio Jari, Somersalo Erkki. Statistical and computational inverse problems vol. 160. Springer Science & Business Media; 2005.
[33] Kolda Tamara G, Bader Brett W. Tensor decompositions and applications. SIAM Rev 2009;51(3):455–500.
[34] Zhao Zhisong, Uddin Nasim, O'Brien Eugene J. Bridge weigh-in-motion algorithms based on the field calibrated simulation model. J Infrastruct Syst 2017;23(1):04016021. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000308.
[35] Helmi Karim, Bakht Baidar, Mufti Aftab. Accurate measurements of gross vehicle weight through bridge weigh-in-motion: a case study. J Civ Struct Health Monit 2014;4(3):195–208.
[36] Ojio T, Yamada K. Bridge weigh-in-motion systems using stringers of plate girder bridges. Third international conference on weigh-in-motion (ICWIM3). Ames: Iowa State University; 2002.
[37] Frasso Gianluca, Eilers PH. Smoothing parameter selection using the l-curve. In: 27th international workshop on statistical modelling, Prague, Czech Republic, 2012, Proceedings; 2012.
[38] Jacob B. Cost 323 weigh in motion of road vehicles. Final report, appendix 1

European wim specification; 1999.
[39] Tibshirani Robert. Regression shrinkage and selection via the lasso. J Roy Stat Soc Ser B (Methodological) 1996:267–88.
[40] Hummelsheim S. Lasso and equivalent quadratic penalized models. ArXiv e-prints; 2014.
[41] Biggs John M. Introduction to structural dynamics. McGraw-Hill College; 1964.
[42] Yeong-Bin Yang JD, Yau Zhongda Yao, Wu YS. Vehicle-bridge interaction dynamics: with applications to high-speed railways. World Scientific; 2004.
[43] Yang YB, Lin CW. Vehicle–bridge interaction dynamics and potential applications. J Sound Vib 2005;284(1–2):205–26.
[44] Gonçalves Matheus Silva, Carraro Felipe, Lopez Rafael Holdorf. Vehicle-bridge dynamics simulation. Mendeley Data; 2019. URL https://doi.org/10.17632/kt48wf5vjz.1.
[45] Múčka Peter. Simulated road profiles according to iso 8608 in vibration analysis. J Test Eval 2017;46(1):1–14.
[46] Miguel Leandro FF, Lopez Rafael H, Torii André J, Miguel Letícia FF, Beck André T. Robust design optimization of tmds in vehicle–bridge coupled vibration problems. Eng Struct 2016;126:703–11.
[47] Dharankar Chandrashekhar S, Hada Mahesh Kumar, Chandel Sunil. Numerical generation of road profile through spectral description for simulation of vehicle suspension. J Braz Soc Mech Sci Eng 2017;39(6):1957–67.
[48] DNIT IPR 723. Manual de estudos de tráfego. Departamento Nacional de Infraestrutura e Transportes, Rio de Janeiro, Brazil; 2006.
[49] Fancher Paul S. A factbook of the mechanical properties of the components for single-unit and articulated heavy trucks. Phase i. Final report; 1986.
[50] Nosseir TA, El-Gindy M, El-Saied FA. Tire radial properties. J Periodica Polytech Transp Eng 1982:21–8.