# State estimation and inverse modeling applied to noise pollution at a urban scale

Antoine Lesieur

Sorbonne Université

École doctorale : ED 386
Sciences Mathématiques de Paris Centre

Inria

**Thèse de Doctorat**
**en**
**Mathématiques appliquées**

présentée par

Antoine LESIEUR

# Estimation d'état et modélisation inverse appliquées à la pollution sonore en milieu urbain

Soutenue le 09/09/2021 devant un jury composé de :

| | | |
|---|---|---|
| M. Claudio Guarnaccia | Université de Salerne | rapporteur |
| M. Olivier Roustant | INSA Toulouse | rapporteur |
| M. Julien Salomon | INRIA | directeur de thèse |
| M. Vivien Mallet | INRIA | encadrant de thèse |
| M. Arnaud Can | Université Gustave Eiffel | encadrant de thèse |
| M. Régis Marchiano | Sorbonne Université | examinateur |
| Mme. Olga Mula | Université Paris Dauphine | examinatrice |
| M. Pierre Aumond | Université Gustave Eiffel | invité |

Le bruit fait peu de bien, le bien fait peu de bruit.

*Saint François de Sales*

# Exergue

"Durant le temps nécessaire à l'écriture de ce livre, le site d'observation américain de Mauna Loa [1], à Hawaii, indique que la concentration de $CO_2$ atmosphérique a franchi la barre des 400, puis des 410 ppm. Cette mesure, qui enregistre la transformation du climat d'origine humaine, prouve qu'à l'échelle d'une activité aussi minuscule que la rédaction d'un ouvrage de philosophie, la réalité écologique se dégrade silencieusement dans des proportions spectaculaires. Indiquons seulement que cette valeur était restée sous la barre des 300 ppm pendant l'intégralité de l'histoire humaine préindustrielle et que l'auteur de ces lignes est né à 340 ppm. Une étude allemande très médiatisée (Hallmann et al., 2017) a également montré que la biomasse d'insectes volants a été réduite de 76 % en vingt-sept ans : malgré les mesures de protection et la création de zones naturelles, trois quarts des insectes ont disparus en quelques décennies. Et cela n'est encore qu'un indice au milieu d'un vaste ensemble de recherches sur la dégradation des sols, des eaux, des fonctions de pollinisation et d'entretien des écosystèmes, qui indiquent que la transformation de la Terre se déroule désormais à un rythme commensurable avec la durée d'une vie, et même d'un simple projet d'écriture.

Lors de la même période de cinq ans, le paysage politique mondial a subi des transformations tout aussi sidérantes. L'accession au pouvoir de Donald Trump aux Etats-Unis en 2017, de Jair Bolsonaro au Brésil en 2019, mais aussi la victoire des partisans du Brexit dès juin 2016 sont les repères les plus nets dans une série d'événements souvent interprétés comme la désagrégation de l'ordre libéral. Un peu partout dans le monde, un mouvement de retour aux frontières et de conservatisme social fédère certains perdants du globalisme désespérément à la recherche de nouveaux protecteurs et des élites économiques décidées à entrainer les peuples dans le jeu de la rivalité entre nations pour préserver l'accumulation de capital. Un peu plus tôt pourtant, les accords de Paris signés dans l'enthousiasme général en décembre 2015 laissaient entrevoir l'émergence d'une diplomatie d'un nouveau genre, chargée de faire entrer dans l'ère climatique le concert des nations. En dépit des faiblesses constitutives de cet accord, c'est à cette articulation entre coopération diplomatique et politique climatique que se sont attaqués les nouveaux maitres du chaos : pas question en effet pour eux de fonder un ordre mondial sur la limitation de l'économie.

Durant cette même période encore, nous avons pu assister à la multiplication des fronts de contestation sociale qui mettent directement ou indirectement en question l'état de la Terre. Les dernières corrections apportées à ce livre se sont faites au rythme des mobilisations des Gilets jaunes en France, dont on ne peut oublier qu'elles ont été déclenchées par un projet de taxe sur les carburants. L'invention d'un nouveau rapport au territoire au sein de la ZAD de Notre-Dame-des-Landes, ou à l'occasion du conflit entre les habitants de la réserve amérindienne de Standing Rock et le projet de pipeline au Dakota, a débuté au moment ou je commençais, dans mes séminaires, à nouer des liens entre l'histoire de la pensée politique moderne et la question des ressources, de l'habitat et plus largement des conditions matérielles d'existence. L'actualité en somme, confirme et alimente sans cesse l'idée d'une réorientation des conflits sociaux autour des subsistances humaines. Mais à côté de tout cela, à côté des marches pour le climat,

1. gml.noaa.gov/ccgg/trends/

des discours de Greta Thunberg, des opérations de désobéissance menées par Extinction Rebellion à Londres, il y avait aussi Haïti, Porto Rico, Houston : L'intensification des ouragans tropicaux et la faillite des réponses gouvernementales ont fait de la vulnérabilité climatique le révélateur d'inégalités sociales de plus en plus politisées. La distribution des richesses, des risques, des mesures de protection oblige à comprendre dans un même geste la destinée des choses, des peuples, des lois et des machines qui les assemblent.

Cinq ans suffisent donc à enregistrer des mutations capitales. Cinq ans suffisent à regarder un passé pourtant proche comme un univers totalement différent de celui dans lequel on évolue désormais, et vers lequel on ne reviendra jamais. La rapidité de ces évolutions nous laisse aussi devant une question plus sombre : où en serons nous quand cinq ans de plus se seront écoulés ?"

**Pierre Charbonnier**,
Introduction de *Abondance et Liberté*, 2020.

# Résumé

La pollution sonore est un problème de santé publique bien identifié par les autorités sanitaires. Une exposition prolongée à des niveaux de bruit trop élevés peut entraîner différents types de complications, qu'elles soient de natures psychiques ou cardio-vasculaires. Dans une étude publiée en février 2019, Bruitparif, l'organisme de surveillance du bruit en Île-de-France, a conclu qu'à l'échelle de l'agglomération parisienne, ce sont 3 années de vie en bonne santé qui sont perdues à cause de la pollution sonore.

Afin d'établir l'exposition au bruit des populations, des cartes de bruit sont régulièrement générées à la demande des agglomérations et en conformité avec les réglementations de l'Union européenne. Ces cartes sont produites par des bureaux d'études, pour les sources de bruit principales que sont les trafics routier, ferré et aérien, ainsi que les principales industries. Pour le trafic routier, qui est reconnu comme étant la source de bruit principale en milieu urbain, elles sont le résultat de simulations qui estiment le niveau de bruit à partir des données de circulation, des données météorologiques, de la topographie, de la distribution des bâtiments et de la végétation. Les données qui varient dans le temps sont moyennées à l'échelle annuelle. Les cartes obtenues sont donc une estimation de la distribution spatiale des niveaux de bruit moyens sur la zone d'étude. Ces données sont en plus limitées dans l'espace, et il existe une incertitude qui empêche de connaître précisément le trafic moyen annuel sur l'ensemble des voies de la zone d'étude. Des estimations génériques complètent donc les mesures de trafic en remplissant les zones non mesurées avec des données trafic qui sont des fonctions de la nature de la voie. La précision des cartes de bruit est limitée aussi par la durée du temps de calcul qui impose d'avoir des modèles assez simple de la propagation acoustique. Il est aussi à noter que lors de comparaisons à des mesures sonométriques, il peut y avoir une différence importante par le fait que, dans le cadre de cette étude, seules les données de bruit routier sont prises en compte dans la simulation. Enfin, le fait d'avoir en général une carte annuelle interdit d'estimer l'évolution temporelle des niveaux de bruit.

En plus de cartes de bruit, les parties prenantes réalisent parfois des campagnes de mesures de niveaux de bruit à l'aide de sonomètres. Ils mesurent ainsi l'évolution temporelle du niveau de bruit à une série d'endroits donnés. Ces données donnent un reflet plus complet et réaliste du niveau sonore réel que les résultats des simulations de cartes de bruit mais elles sont très locales. Dans le cas de mesures fixes, elles permettent de ne connaître le niveau de bruit qu'à l'endroit où l'appareil de mesure est situé. Elles sont également coûteuses, ce qui interdit un quadrillage extensif d'une zone avec un réseau de capteurs.

Combiner les approches de la modélisation et des mesures permettrait d'augmenter la quantité de données utiles pour la production des cartes de bruit. Une carte de bruit qui combine les deux approches peut pallier les limitations de la simulation et des mesures, et fournir une cartographie dynamique et en temps réel des niveaux sonores.

L'objectif de cette thèse est de mettre en oeuvre des méthodes dites d'assimilation de données pour unir les bénéfices des deux approches, simulation et observation. Ces méthodes utilisent à la fois un simulateur de cartes de bruit et les données d'obser-

vation d'un réseau de capteurs de niveaux de bruit distribués sur l'ensemble d'une zone d'étude donnée. La première méthode propose d'ajouter une couche de correction aux simulations de cartes de bruit qui dépend des observations à un instant donné. La seconde méthode propose une paramétrisation des données d'entrée du simulateur qui utilise les niveaux de bruit observés. Afin de contourner la contrainte du temps de calcul lié à la simulation numérique, il est possible de générer des cartes de bruits en quelques millisecondes à partir de nouveaux paramètres à l'aide d'un méta-modèle, c'est-à-dire d'un algorithme qui reproduit fidèlement le comportement du simulateur sur une zone donnée grâce à des méthodes statistiques. Sa construction et l'étude de ses performances est l'objet du chapitre II. La mise en application des méthodes d'assimilation de données faisant usage du méta-modèle est l'objet des chapitres III et IV. Dans le chapitre III, l'étude porte sur une méthode de correction d'une carte, générée par le méta-modèle, en utilisant les niveaux de bruit observés. Cette méthode est appelée BLUE pour Best Linear Unbiased Estimator. Le chapitre IV quant à lui, s'intéresse à une méthode de modélisation, la modélisation inverse, dont le but est de générer les paramètres d'entrée du méta-modèle grâce auxquels la carte simulée obtenue approche au mieux les données fournies par les sonomètres. Ce chapitre étudie également une combinaison des méthodes de modélisation inverse et d'assimilation de données BLUE avec une méthode appelée JSPE pour Joint State Parameter Estimation. De plus , cette méthode permet, dans certaines conditions, de s'affranchir de l'acquisition préalable de données trafic et météorologiques en temps réel. Ce travail de thèse se termine par le chapitre V avec l'étude d'une méthode de diagnostic et d'estimation des paramètres d'une matrice de covariance. En effet, pour mener à bien l'assimilation de données, il est nécessaire d'introduire les matrices de covariance des erreurs d'ébauche et d'observation. Les composantes de ces matrices dépendent d'un vecteur de paramètres inconnu. Il convient donc de diagnostiquer des paramètres estimés *a priori*. Les ouvertures qu'amènent ces résultats sont mentionnés à la fin de la thèse :

— l'extension de l'application des méta-modèles à l'analyse de sensibilité ;
— l'enrichissement des méta-modèles avec un nombre de paramètres plus élévé afin d'analyser plus finement la distribution des niveaux de bruits ;
— l'étude de l'optimisation de la distribution des réseaux de mesures destinés à l'assimilation de données.

## Abstract

Noise pollution is a public health problem well identified by health authorities. Prolonged exposure to excessively high noise levels can lead to various types of complications, both psychological and cardiovascular. In a study published in February 2019, Bruitparif, the noise monitoring organization in the Paris region, concluded that in the Paris metropolitan area, 3 years of healthy life are lost due to noise pollution.

In order to establish the noise exposure of populations, noise maps are regularly generated at the request of the agglomerations and in compliance with European Union regulations. These maps are produced by engineering offices for the main sources of noise such as road, rail and air traffic, as well as the main industries. For road traffic, which is recognized as the main noise source in urban areas, they are the result of simulations that estimate the noise level from traffic data, meteorological data, topography, building distribution and vegetation. Data that vary over time are averaged on an annual basis. The resulting maps are therefore an estimate of the spatial distribution of average noise levels over the study area. These data are moreover spatially limited, there exists some uncertainties which prevent to know precisely the annual average traffic on all the roads in the study area. Generic estimates therefore complement the traffic measurements by filling in the unobserved areas with traffic data that are functions of the nature of the road. The accuracy of the noise maps is also limited by the length of the computation time, which requires fairly simple models of the acoustic propagation. It should also be noted that in comparisons with sonometric measurements, there can be an important difference by the fact that in this study, only road noise data are taken into account in the simulation. Finally, usually working on an annual map makes it impossible to study the temporal evolution of noise levels.

In addition to noise maps, stakeholders sometimes carry out noise level measurement campaigns using sound level meters. In this way, they measure the temporal evolution of noise levels at a series of given locations. These data give a more complete and realistic reflection of the actual noise level than the results of noise map simulations, but they are very local, in the case of fixed measurements they allow to know the noise level only at the level where the measuring device is placed. They are also expensive, which prohibits an extensive gridding of an area with a network of sensors.

Combining modeling and measurement approaches would increase the amount of data useful for the production of noise maps. A noise map that combines the two approaches can overcome the limitations of simulation and measurement and provide dynamic, real-time mapping of noise levels.

The objective of this thesis is to implement so-called data assimilation methods to unite the benefits of both approaches, simulation and observation. These methods use both a noise map simulator and observation data from a network of noise level sensors distributed over a given study area. The first method proposes to add a correction layer to noise map simulations that depends on the observations at a given time. The second method proposes a parameterization of the simulator input data using the observed noise levels. In order to circumvent the computation time constraint related to numerical simulation, it is possible to generate noise maps in a few milliseconds from new parameters using a meta-model, i.e. an algorithm that faithfully reproduces the

simulator's behavior over a given area using statistical methods. Its construction and the study of its performance is the subject of chapter II. The implementation of data assimilation methods using the meta-model is the subject of chapters III and IV. In the chapter III, the study deals with a method for correcting a map, generated by the meta-model, using the observed noise levels. This method is called BLUE for Best Linear Unbiased Estimator. As for the chapter IV, it deals with a modeling method, the inverse modeling, whose goal is to generate the input parameters of the meta-model, thanks to which the simulated map obtained approaches as well as possible the data provided by the sound level meters. This chapter also studies a combination of inverse modeling and BLUE data assimilation methods with a method called JSPE for Joint State Parameter Estimation. In addition, this method allows, under certain conditions, to dispense with the prior acquisition of traffic and meteorological data in real time. This thesis work ends with the chapter V, which proposes a diagnostic and estimation of the parameters of a covariance matrix. Indeed, in order to carry out data assimilation, it is necessary to introduce the covariance matrices of the background and observation errors. The components of these matrices depend on an unknown vector of parameters. It is therefore necessary to diagnose estimated parameters *a priori*. The openings brought by these results are mentioned at the end of the thesis:

— the extension of the application of meta-models to sensitivity analysis;
— the enrichment of the meta-models with a higher number of parameters in order to analyze more finely the distribution of noise levels;
— the study of the optimization of the distribution of measurement networks for data assimilation.

# Remerciements

Au moment de conclure ces presque quatre années de travail doctoral, je me dois d'exprimer ma gratitude envers ceux qui ont rendus possible la réalisation de ce manuscrit. Mes premières pensées vont à Vivien Mallet, mon principal encadrant dont la disponibilité, la clarté de ses explications et sa vision des problèmes ont grandement facilité mon travail. Il a su toujours être soucieux du bien-être de ses encadré.e.s quand bien même nous étions mis à distance par le confinement d'abord puis par l'océan Atlantique ensuite. Coté acoustique, je n'ai pas été en reste non plus, l'aide et le soutien que m'ont apporté Arnaud Can et Pierre Aumond pendant ce travail m'ont été infiniment précieux et je les en remercie grandement. Ils ont eu la patience de m'enseigner leur discipline, l'acoustique, qui m'était presque inconnue au début de ma thèse. Toujours disponibles pour résoudre mes bloquages, tant dans les domaines de la physique que de l'informatique. Grâce à eux, les problèmes techniques n'ont jamais été qu'une formalité. L'accueil qu'ils m'ont réservés sur le site de l'UMRAE a toujours été extrêmement chaleureux et les séjours à Nantes ainsi que les différents congrès effectués à Gand, Strasbourg et Aix-la-Chapelle font parti des bons souvenirs que je garderai de cette époque. Je remercie également Julien Salomon, mon directeur de thèse, qui a rendu notamment possible la publication du quatrième papier de ma thèse. Ces remerciements ne sauraient êtres complets sans remercier également Nicolas Fortin pour sa réactivité pour débuguer NoiseModelling, Jacques Sainte-Marie le directeur de l'équipe ANGE et Julien Guieu toujours disponible pour m'aider dans les tâches administratives. Enfin, tous les acteurs trop invisibles et qu'on oublie souvent mais qui font de l'INRIA un lieu où l'on fait de la recherche dans d'excellentes conditions : services généraux, services informatiques, secrétariat, personnel nettoyant, etc.

Je remercie l'ensemble des membres du jury d'avoir accepté d'évaluer ce travail de thèse. Je remercie particulièrement les rapporteurs Olivier Roustant et Claudio Guarnaccia pour leur lecture attentive du manuscrit et les commentaires constructifs qui en furent issus.

Cette thèse n'aurait pas non plus eu la même saveur sans les rencontres, discussions à la cantine, verres, footings, parties de foot et de baby foot effectués avec les doctorants et ingénieurs de l'équipe ANGE et du troisième étage du bâtiment A : Apolline, Bao, Fabien, Frédéric, Hussam, Janelle, Léa, Liudi, Matthias, Mathieu, Olivier, Raphaël, Thibault, Van-Thanh et tous les autres. Ils ont fait du centre un lieu on l'on se rend quotidiennement avec plaisir.

Mes dernières pensées vont à ma famille et mes proches qui m'ont aidé à me convaincre de me lancer dans cette entreprise, et qui ont alimenté tout au long de ces quatre années les réflexions sur mon travail, son rôle et le sens à lui donner.

# Foreword

There is a growing need for city planners to monitor the noise pollution across their administration area, in order to develop decision-making tools or to supervise the population exposed to this annoyance.

It is possible to harvest observed noise data with an array of sound level meters scattered across a study area. This method provides very accurate data about the noise levels at point locations. It can also be used for special events such as outdoor concerts or public works. However, the noise level can only be known in the vicinity of the sound level meter.

An other way to monitor the noise level distribution is to perform a noise map across the study area. This map is computed with a noise mapping reference software which takes as inputs traffic data, weather data, topographic data and sets as outputs the resulting noise level for each element of a grid of receptors. These noise maps suffer from a lack of accuracy due to the time and space approximations of the noise sources, and the approximations in the sound propagation calculation.

A new study area is the production of dynamic noise maps which allows to compute the noise distribution every hour to improve the noise exposure assessment. Dynamic noise mapping requires the generation of a large amount of noise maps and thus motivates the creation of techniques designed to reduce the necessary computation burden. Statistical methods and supervised machine learning are promising tools and will help to develop so-called meta-models which allow to quickly generate noise maps over a defined study area.

Data assimilation allows to merge computed noise maps with noise observations. The dynamic behavior of observed noise levels is very relevant and can provide useful information which can be integrated to a noise map. The combination of meta-model generated noise maps and observed noise levels give dynamic noise maps with a lower RMSE than the simulation only.

## Thesis Plan

Chapter I introduces the mathematical tools and concepts which will be used throughout the thesis, data assimilation principles, meta-modeling principles, and the regular statistical tools required for the study. In addition, some basic acoustics concepts are defined in this section, metrics, noise measures and simulation principles. This section has been written in order to make the document understandable for audiences with an acoustical background as well as people with a mathematical background.

Chapter II exposes the construction process of the meta-model, latin hypercube sampling, dimension reduction, interpolation (RBF and kriging). Once the meta-model is built its performance is tested by comparing its outputs with the outputs of the reference noise mapping software.

Chapter III presents the data assimilation method BLUE applied to the outputs of the meta-model when its inputs are the observed traffic and weather data. Its performance is evaluated with a leave-one-out cross-validation process and an improvement of 30 % of the root mean square error is observed between the noise map which uses data assimilation and the output of the meta-model.

Chapter IV presents an inverse modeling method which helps finding the most accurate input parameters of the parameters with respect to the observed noise levels. A method called JSPE merges the results of the inverse modeling with the data assimilation principles presented in Chapter III. These methods allow to reduce the root mean square error even more than the BLUE data assimilation used in chapter III. In addition, it can be performed in areas where no observation input is conducted.

Chapter V presents a new approch to perform a diagnosis of the parameterized covariance matrices of the error between the predicted and observed noise levels used in chapter III and IV. It is compared with the current diagnosis method called $\chi^2$ diagnosis. The extension of this method to parameters approximation is also presented.

Chapter VI concludes this study. It gives some perspectives on further uses of meta-models such as sensitivity analysis, further refinements of the meta-model and optimization of the distribution of the sound level meters applied to data assimilation.

# Contents

# I. Introduction

*We introduce in this chapter the notions of acoustics that will be used throughout this manuscript. In particular, the means of calculation and acquisition of noise levels will be presented, as well as the methods used for the elaboration of noise maps. The second part of this chapter formally introduces the elementary concepts of data assimilation. The third part presents the meta-model and its underlying mathematical principles.*

## Table of contents

# I.1. Environmental Acoustics

## I.1.1. Acoustic waves and sound levels

Acoustics is the science that studies sounds and their propagation as longitudinal pressure waves through elastic media. When an acoustic wave passes through a fluid environment, the pressure, speed, temperature, entropy and volumic mass fields are perturbed. The pressure field is written as follows:

$$P(\mathbf{X}, t) = P_{atm} + p_1(\mathbf{X}, t) \tag{I.1}$$

With $p_1 \ll P_{atm}$. The pressure of the fluid at rest is the atmospheric pressure $P_{atm}$, and $p_1(\mathbf{X}, t)$ represents the disturbance related to the passage of the wave. The sound level at a point $\mathbf{X}$ is defined relative to the evolution of the value of $p_1(\mathbf{X}, t)$ over time. The space dependency will now be implied and omitted to simplify the notation, and the perturbation will be noted $p(t)$. Several noise characterizations can be defined, the most commonly used being the SPL ("Sound Pressure Level"), expressed in decibels (dB):

$$L_p = 10 \log_{10} \left( \frac{p_{RMS}}{p_0} \right)^2 = 20 \log_{10} \left( \frac{p_{RMS}}{p_0} \right). \tag{I.2}$$

This level describes the power ratio between on the one hand $p$, and on the other hand a reference level set at $p_0 = 20\,\mu\text{Pa}$. This reference level $p_0$ is defined as the audibility threshold. In this equation, $p_{RMS}$ is the RMS value of the pressure signal, calculated through an RMS ("root-mean-square") average over a time interval of duration $T$:

$$p_{RMS} = \sqrt{\frac{1}{T} \int_0^T p^2(t)\, \mathrm{d}t}. \tag{I.3}$$

In practice, the mean RMS is estimated by measuring instruments through a discrete sum

$$p_{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} p_n^2}, \tag{I.4}$$

where $p_n$ are instantaneous pressure levels, measured periodically at a given sampling frequency.

The pressure levels associated with a $p$ acoustic disturbance are very low compared to the atmospheric pressure level. For example, for an overpressure of 1 Pa, corresponding to a (high) level of 94 dB, the ratio is

$$\frac{1}{1013.25 \cdot 10^2} \simeq 1 \cdot 10^{-5}. \tag{I.5}$$

The non-linear effects, associated with the products of the perturbation terms in the fluid mechanics equations, are therefore quite negligible.

#### I.1.1.1. Summing and averaging sound levels

The acoustic level $L_{tot}$ resulting from the combination of two signals $p_1(t)$ and $p_2(t)$, can be computed with equations (I.2) and (I.3). The RMS pressure $p_{RMS}$ of the signal during a time period $T$ is then:

$$\begin{aligned} p_{tot\,RMS} &= \sqrt{\frac{1}{T}\int_0^T (p_1+p_2)^2(t)\,\mathrm{d}t} \\ &= \sqrt{\frac{1}{T}\int_0^T p_1^2(t)+p_2^2(t)+2\,(p_1p_2)(t)\,\mathrm{d}t}. \end{aligned}$$

When $p_1$ and $p_2$ are not coherent signals, the last term of the sum is equal to zero, We get:

$$L_{p_{tot}} = 10\log_{10}\left(\frac{{p_1}^2_{RMS}+{p_2}^2_{RMS}}{p_0^2}\right). \tag{I.6}$$

We then define the sum of two acoustic levels from non coherent signals as:

$$L_1 \oplus L_2 = 10\log_{10}(10^{L_1/10}+10^{L_2/10}). \tag{I.7}$$

This equation can be generalised with $N$ non coherent signals $(p_i)_{i\in[\![1,\ N]\!]}$:

$$\mathrm{L}_{\sum p_i} = L_1 \oplus L_2 \oplus \ldots L_N = 10\log_{10}\left(\sum_{i=1}^{N} 10^{L_i/10}\right). \tag{I.8}$$

#### I.1.1.2. Noise measures

The function of a measuring device is to convert the measured acoustic energy into another form of energy (often electrical), as accurately as possible regarding the information conveyed. Ideally, the acquired signal $x$ is strictly proportional to the input signal $p$, regardless of the frequency content of $p$. The reference instrument for measuring the noise level is the sound level meter.

An audio aquisition device fulfills the following confitions:

— The frequency response, generally represented by the phase and amplitude transfer functions between $p(t)$ and $x(t)$;
— The dynamic, i.e. the difference between the maximum amplitude of a signal which can be recorded without being distorted and the background sound level produced by the recording device;
— The correctness, i.e. the uncertainty level related with the recording;
— The precision, i.e. the reproducibility of a recording.

The reference instrument for measuring sound levels is the sound level meter. There are international metrological standards for classifying industrial sound level meters according to the expected measurement quality. These standards define minimum accuracies that the sound level meter under evaluation must achieve when subjected to standardized stimuli (pure sounds at predefined frequencies).

### I.1.1.3. Perception and frequency weighting

A sound signal is characterized by its amplitude and frequency content. The dynamic range of the human ear extends over approximately 140 dB, with a generally accepted audible frequency range from 20 Hz to 20 kHz (Huber and Runstein, 2013). The sensitivity of the hearing system is highly dependent on frequency.

Measured sound levels are often corrected by means of frequency weightings to account for this dependence. Psychoacoustic studies (Fletcher and Munson (1933)) aimed at measuring *sound*, i.e. the loudness as perceived by listeners, have been used as a basis for defining different weightings. The curves representing the A, B, C, and D weightings, as described by ANSI S1.4 [1] are presented in figure I.1. These weightings are constructed on the basis of an estimation of the pitch of pure (sinusoidal) sounds. For this reason, other weightings, such as the one described in Geddes (1968), seem to be more suitable for a use of environmental nuisance measurement, composed of complex sounds, but their use has remained very limited in practice compared to the A-weighting. For this reason, we will use A-weighting throughout this manuscript, unless otherwise stated. Z-weighting (for "zero", or linear "dB", refers to a level calculated without frequency weighting.

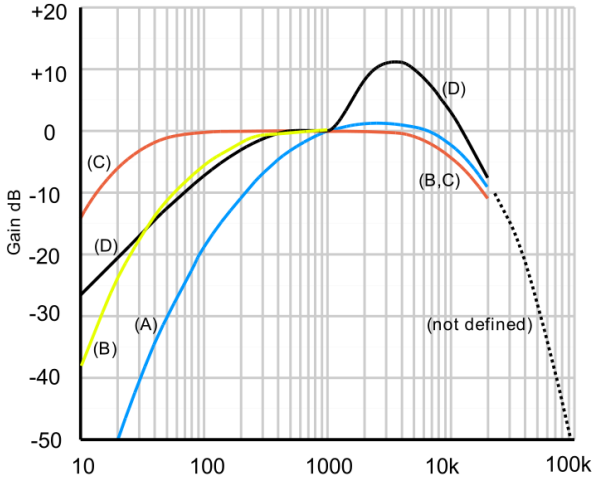


Figure I.1. – Gain (dB) of the usual weightings as a function of frequency.

### I.1.1.4. frequency windowing

In this case study, the noise data is stored in octave band noise level. An octave is a frequency interval $[f_1, f_2]$ logarithmically centered around a frequency $f_c$. The logarithm means that $\log_2(f_1) = \log_2(f_c) - \frac{1}{2}$ and $\log_2(f_2) = \log_2(f_c) + \frac{1}{2}$. Hence, the interval of an octave around $f_c$ is $[\frac{f_c}{\sqrt{2}}, \sqrt{2}f_c]$. In order to get the noise level at a given octave, the raw signal $x(t)$ is dicretized to $x(n)$ with a given timestep $\tau$ such that $x(n) = x(n\tau)$. It is then filtered by a band pass filter to give the filtered signal $y(n)$. The general expression of the filter is:

1. https://law.resource.org/pub/us/cfr/ibr/002/ansi.s1.4.1983.pdf last consulted february 8, 2021

$$y(n) = \sum_{i=1}^{\ell} b_i y(n-i) + \sum_{i=0}^{k} a_i x(n-i) \tag{I.9}$$

Its transfer function is given after a Z-transform as:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^{k} a_i z^{-i}}{1 - \sum_{i=1}^{\ell} b_i z^{-i}} \tag{I.10}$$

The simplest band pass filters are second order filter, i.e. $\ell = k = 2$. It is possible to study the harmonic behavior of these filters. The attenuation of the amplitude against the frequency of the signal is given in figure I.2.

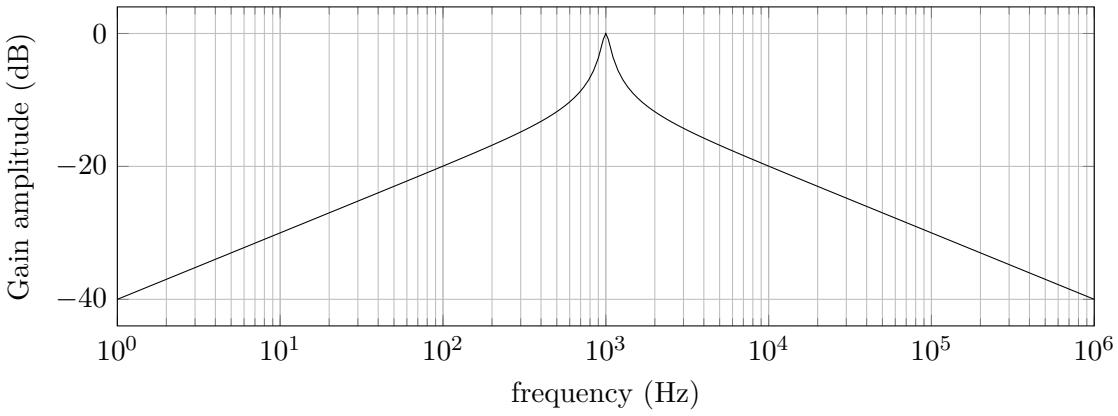


Figure I.2. – Gain of the amplitude of an octave band pass filter

For each octave $f_i$, we get a signal $y_{f_i}(n)$ obtained by a band pass filter $H_{f_i}(z)$.

#### I.1.1.5. time windowing

The sound level meter calculates a moving average in real time in order to provide the user with an overview of the evolution of the noise level. This is called the instantaneous noise level. The duration (or number of samples) to be considered for this average must be chosen with regard to perceptual aspects. For example, if the sounds measured are very brief but of significant amplitude (shocks, detonations), it is better to use a short duration window, in order to integrate the signal only during the period when it has a significant power level. The use of a longer window would also integrate the part of the signal of negligible power, and thus artificially reduce the value of the calculated noise level. The calculation of a level with a time window is performed through the following moving average:

$$L_{f_i,\tau}(t_i) = 20\ \log_{10}\left(\sqrt{\frac{1}{\tau}\int_{-\infty}^{t}\left(\frac{y_{f_i}(\xi)}{p_0}\right)^2 \exp\left(-\frac{t-\xi}{\tau}\right)\mathrm{d}\xi}\right), \tag{I.11}$$

where $y_{f_i}$ is the filtered signal for a given octave; $\tau$ is a time window. The time constant used in this study is called F (Fast) and lasts 125 ms. When no time windowing is applied, the signal power $y_{f_i}$ is integrated for the entire measurement interval $[0,\ T]$.

This is called the $L_{eq,T}$ level, or *equivalent* level: the level thus expressed is that of a signal of constant power that would contain the same amount of energy as $y_{f_i}$ during the measurement interval.

#### I.1.1.6. construction of the noise level

The combination of the frequency and the time windowing gives for each time window $Y$ a set of noise levels for $N$ octaves. It is possible to build the $LA_Y$ noise level by weighting the noise level of each frequency $f_i$ by its corresponding weight $w_{f_i}$. The noise levels are shown in table I.1.

| frequency $f_i$ (Hz) | 63 | 125 | 250 | 500 | 1000 | 2000 | 4000 | 8000 |
|---|---|---|---|---|---|---|---|---|
| weighting $w_{f_i}$ (dB) | -26.2 | -16.1 | -8.6 | -3.2 | 0 | 1.2 | 1.0 | -1.1 |

Table I.1. – attenuation table for the A weighting

The equivalent noise level at time $t$ is built as follows:

$$LA_\tau(t) = 10\log_{10}\left(\sum_{i=1}^{N} 10^{(L_{f_i,\tau}(t)+w_{f_i})/10}\right) \tag{I.12}$$

With $LA_Y(t_i)$, it is possible to build an average noise level for a larger time period $Y'$, at a given time $t_j$. with $T = \{t_i | t_j - \tau' \leq t_i \leq t_j\}$, we get :

$$LA_{\tau'}(t_j) = 10\log_{10}\left(\frac{1}{|T|}\sum_{t\in T} 10^{LA_\tau(t)/10}\right) \tag{I.13}$$

#### I.1.1.7. The $L_{den}$ noise index

This level is an indicator expressing an overall average level over the whole day ( day, evening, night). Together with the $L_{night}$ level, it is one of the indicators recommended in the diagnosis of population exposure by the European "Environmental Noise Directive 2002/49/EC" (END). Its expression is as follows:

$$L_{den} = 10\log_{10}\left(\frac{12}{24}10^{L_{day}/10} + \frac{4}{24}10^{(L_{evening}+5)/10} + \frac{8}{24}10^{(L_{night}+10)/10}\right), \tag{I.14}$$

where $L_{day}$, $L_{evening}$ and $L_{night}$ respectively represent an average level during the following 3 periods of the day:

**day** from 7am to 7pm;

**evening** from 7pm to 11pm;

**night** from 11pm to 7am.

Penalties of 5 dB and 10 dB are thus introduced to reflect the increased discomfort of local residents during the evening and night periods. The timetables are given as an indication, and can be defined at the discretion of the Member State.

### I.1.2. Acoustic models and noise maps

Modelling can be used in acoustics to study the impact of environmental noise. In particular, it is useful during studies prior to environmental modifications likely to change the exposure of populations. Such modifications are, for example: the installation of new infrastructures (buildings, noise barriers), changing the surface of a roadway, changing the speed limit of traffic lanes, etc. Modelling is also an essential complement to *in situ* measurement which, although potentially accurate, is extremely costly in terms of time and human resources when the area under study is significant or when its geometry or meteorological conditions result in an inhomogeneous acoustic field. The product of a simulation is typically a *noise map*, representing an acoustic level, for a given duration, over a certain geographical area.

Different modeling strategies can be adopted. Depending on the desired level of precision and the technical and material constraints, the methods employed may use simple heuristics, ray tracing type methods, or solve the acoustic equations on the studied domain using finite element type numerical methods. Whichever method is considered, the first stage of modelling consists in defining the sources of noise and the environment in which its propagation takes place.

The sources are defined by their location, directivity, power and frequency profile over time. The profile of the noise radiated by a source depends strongly on its nature. For this reason, specific computation methods have been established for different source categories: road traffic (Besnard et al., 2009; Abbaléa et al., 2009), rail traffic (van Leeuwen, 2000), air traffic (CEAC, 1997), and industrial noise (ISO 8297 standard). As road traffic is one of the main sources of environmental noise in urban areas, its estimation has a major influence on the results of a model. Traffic is sometimes estimated by numerical simulation (Mai, 2006).

The environment corresponds to the propagation medium and to all the physical elements that will be on an acoustic trajectory connecting a source to a *receiver*, i.e. a point in space where a noise value is to be calculated. The geometry of the terrain, the reflective or absorbent character of the materials constituting this terrain, the meteorological conditions (wind direction and speed, temperature) are all parameters that have a strong influence on the results. The propagation of acoustic energy can take place directly between source and receiver, but it is necessary to consider the phenomena of specular reflection or absorption by surfaces, diffraction at the boundaries of objects, and atmospheric absorption.

The calculation of the noise level at a given receiver is performed by summing the energy contributions (equation (I.8)) from the surrounding sources. The consideration of interference phenomena, i.e. the coherent interaction of signals from the same source, is generally limited to the rays propagating in the same vertical plane. In order to limit the complexity of the calculations, the spatial extent considered around a source, as well as the number of reflections and diffractions to be considered on the path of a ray coming from this source are often limited.

#### I.1.2.1. Noise mapping reference software: NoiseModelling

The noise mapping software used all along this study is called NoiseModelling (Bocher et al., 2019). This software follows the Common NOise aSSessements MethodS in

Europe (CNOSSOS-EU) (European Commission, 2015) which is the standard reference in the European Union. In order to compute a noise map, the software proceeds in two steps:

**Noise emission** The only emission source considered in this thesis is the road traffic source. It distinguishes light vehicles from heavy vehicles. The inputs are road sections which are described as line sources. This step consists in breaking down these sources into a series of point sources. The sound emission level of the points are determined for a given frequency band according to the acoustic power level per metre of a line source.

**Noise propagation** The sound propagation from a point source to a receiver is submitted to various phenomena such as the geometrical spreading due to the expansion of the wavefront and the atmospheric absorption that results from the molecular relaxation effect. The propagation can be modelled by separating the contributions of each phenomenon, which leads to consider the sound field received at a given location as a combination of:

— the direct fields, which corresponds to the sound waves propagating directly from the source to the receiver;
— the diffracted field, related to the diffraction of the sound waves around and over the buildings;
— the reflected field, associated with the reflections on the ground and on the buildings facades along the propagation path that can also include absorption by these elements.

The reflected fields on vertical surfaces are modelled by introducing an order of reflection which correspond to the number of reflections that are taken into account from the source to the receiver. Diffraction is also taken into account. An graphical exemple of the paths computed between a punctual source and the receivers is shown in figure I.3

Figure I.3. – 3D example of the computed rays of a single source

#### I.1.2.2. Exposure and qualitative aspects

During the modeling, receptors are placed on the most exposed facades of residential buildings (Kephalopoulos et al., 2012). By establishing a correspondence between the number of inhabitants (previously estimated) of each building and the noise levels that

are calculated at the receptors located on this building, an estimate of the overall exposure of the urban population is obtained. Other associations are possible, for example, Havard et al. (2011) calculates the exposure of the study participants by considering the neighbourhood around their place of residence. This neighborhood is defined by a circle of 250 m around the place of residence. The question of how the neighbourhood is defined and the size of the neighbourhood to be considered has an important influence on the estimated exposure levels (Tenailleau et al., 2015).

This measure of exposure is therefore purely quantitative and does not include perceptual aspects. However, it is clear that annoyance is intrinsically linked to the nature of the noise, beyond the energy it contains: a concert of music can be enjoyed when the noise level is very high, while neighbourhood noise is the source of many complaints, even though the level is often low. Soundscape studies, such as Lavandier and Defréville (2006), seek to determine the links that may exist between objective (calculated) acoustic indicators and the subjective judgements expressed by participants exposed to a certain sound environment.

As an indication, the exposure levels not to be exceeded, as recommended by the World Health Organization (WHO) are 53 dB(A) during the day (in $L_{day}$) and 40 dB(A) in $L_n$ equivalent level at night (Hurtley, 2009). Indeed, sensitivity and potential annoyance vary according to the activity under consideration, and therefore the time of day. In this perspective, it would seem beneficial to have a representation of temporal noise at a finer scale than the 3 periods defined in section I.1.1.7. Average levels such as $L_{day}$, $L_{evening}$ and $L_{night}$ do not allow local dynamics to be detected. One of the objectives of this thesis is to try to overcome this limitation through methods that generate dynamic maps. This question is dealt with in the chapters III and IV of this manuscript.

## I.2. Why do environmental acoustics need data assimilation?

Environmental acoustics consists in studying the sound environment of a given area, in this particular case study, in urban areas. Two common approaches have been presented with differents advantages and drawbacks.

— The noise mapping approach: for a given area, with a given set of input data such as traffic, weather, topographic data, etc. a noise mapping reference software computes the resulting for the study area, i.e. for every reception point, noise level in dB can be computed. However, the accuracy of the data may be flawed and the computation time only allows to perform a limited amount of noise maps. In other words, this method is useful when there is a need for a global result but no need for studying the dynamics of the noise ditribution.

— The sound level meter approach: For a local point, the sound level meter gives a precise and dynamical evolution of the noise level in dB.This approach is complementary to the previous one since it is both local and dynamic.

At this point, we have two approaches, the first is global and static and the second is local and dynamic. One may wonder if it is possible to combine the two antagonic benefits of these methods, in other words, is it possible to get a method which gives both global and dynamic distribution of the noise levels?

Implementing data assimilation principles in the solution presented in this work to

merge the benefits of the two approaches and to get a dynamic noise map. The need for dynamic noise maps is needed by urban planners to find the best traffic regulation policies and for epidemiologists, to infer the influence of the noise pollution on some health issues. Data assimilation principles are presented in section I.3.

## I.3. Data Assimilation Principles

Data assimilation methods allow the products of a numerical model to be combined with measurements. They were developed as early as the 1950s in the field of meteorology (Bergthörsson and Döös, 1955). Their objective is to estimate at each point of the model, in the most realistic way possible, the value of one or more variables, represented as a state vector. A review of classical data assimilation methods can be found in Bouttier and Courtier (2002). The application of such methods results in the computation of an *analysis*, which combines the theoretical knowledge of the studied system and the knowledge derived from its observation. The section I.3.1 gives a general formalization of the problem, the notion of error being detailed in section I.3.2. We then present assimilation methods used in practice in sections I.3.3 and I.3.4.

### I.3.1. Formalisation

**State vector** In the general case, the studied variable $\mathbf{x}$ is a vector field and belongs to a space of infinite dimension $\mathcal{B}$. Since the simulation calculations are based on computer tools, we give a discrete representation of the state vector $\mathbf{x}^t \in \mathbb{R}^n$ (for "true") of dimension $n$. Formally, this results in the following operation:

$$\mathbf{x}^t = \mathbf{\Pi}(\mathbf{x}), \tag{I.15}$$

where $\mathbf{\Pi}\colon \mathcal{B} \to \mathbb{R}^n$ is an operator to obtain a representation of $\mathbf{x}$ in $\mathbb{R}^n$. Typically, $\mathbf{\Pi}$ can operate the discretization of the studied spatial domain into a grid of regularly spaced points.

**Background** The theoretical knowledge *a priori* of the system, also called *background*, is noted $\mathbf{x}^b \in \mathbb{R}^n$. It is a direct estimate of $\mathbf{x}^t$. This estimation is performed using the output of the model.

**Observations** The *observations* of the system, represented by the vector $\mathbf{y} \in \mathbb{R}^p$, reduce the error of the model. The comparison of the observations and the state vector is not immediate:

— observations are not made at the model grid points;
— the observed quantities may not correspond to the variables making up the state vector, but rather be more or less complex functions of these variables.

An *observation operator* $H$ is therefore used, linking the space of the state vector to that of the observations, thus allowing comparison:

$$H\colon \mathbb{R}^n \to \mathbb{R}^p \tag{I.16}$$

$$\mathbf{x}^t \mapsto H(\mathbf{x}^t). \tag{I.17}$$

Usually, only a small portion of the system state variables are accessed through observation, and $p \ll n$.

**Probabilistic interpretation** As it will be developed in the next section, the state vector representations provided by the background and the observations are tainted by unknown errors, which we represent by random vectors. The quality of the estimation of the distributions associated with these random vectors conditions the relevance of the analysis offered by data assimilation, which therefore inevitably takes on a statistical character, and is naturally formulated within the framework of Bayesian inference (Cohn, 1997). Probability distributions $\mathbb{P}(\mathbf{x})$ and $\mathbb{P}(\mathbf{y})$ are associated with $\mathbf{x}$ and $\mathbf{y}$. According to Bayes' formula, we have:

$$.\mathbb{P}(\mathbf{x} \mid \mathbf{y}) = \frac{\mathbb{P}(\mathbf{x})\,\mathbb{P}(\mathbf{y} \mid \mathbf{x})}{\mathbb{P}(\mathbf{y})} \propto \mathbb{P}(\mathbf{x})\,\mathbb{P}(\mathbf{y} \mid \mathbf{x}). \tag{I.18}$$

In the previous equation, $\mathbb{P}(\mathbf{x})$ represents the *a priori* knowledge that we have of the state vector, $\mathbb{P}(\mathbf{y} \mid \mathbf{x})$ concerns the error committed during the comparison between the observations and this state vector. The analysis phase corresponds to calculating $\mathbf{x}^a$ as an estimator of $\mathbf{x}^t$ based on the conditional distribution $\mathbb{P}(\mathbf{x} \mid \mathbf{y})$. The choice of the estimator is not unique: it can be the conditional expectation $\mathbf{x}^a = \mathbb{E}(\mathbf{x} \mid \mathbf{y})$ of the distribution, but other choices are possible, such as the *a posteriori* maximum :

$$\mathbf{x}^a = \max_{\mathbf{x}} \mathbb{P}(\mathbf{x} \mid \mathbf{y}) = \max_{\mathbf{x}} \frac{\mathbb{P}(\mathbf{y} \mid \mathbf{x})\mathbb{P}(\mathbf{x})}{\int_{\boldsymbol{\zeta}} \mathbb{P}(\mathbf{y} \mid \boldsymbol{\zeta})\,\mathrm{d}\boldsymbol{\zeta}} = \max_{\mathbf{x}} \mathbb{P}(\mathbf{y} \mid \mathbf{x})\mathbb{P}(\mathbf{x}). \tag{I.19}$$

## I.3.2. Errors

Various approximations and assumptions are made in order to estimate $\mathbf{x}$, which translates into errors associated with the variables previously described. In this section, we describe the nature of the errors associated with each of the components of the assimilation system, as well as the statistical properties expected for the production of a $\mathbf{x}^a$ analysis.

### I.3.2.1. Errors categorisation

**Observation errors** There are several kinds of errors associated with observations. In the previous section, the $H$ operator was introduced to compare elements in the space of the discretized state vector $\mathbf{x}^t$ and those in the observations space. One may in fact wish to compare $\mathbf{y}$ to the continuous state $\mathbf{x}$, and not $\mathbf{x}^t$, through an operator $h\colon \mathcal{B}. \to \mathbb{R}^p$. On the other hand, any observation is associated with an *instrumental* error$\mathbf{e}^i$, reflecting the metrological qualities of the measuring instrument used to make this observation. Thus, we have:

$$\mathbf{y} = h(\mathbf{x}) + \mathbf{e}^i. \tag{I.20}$$

Going from $\mathbf{x}$ to its discrete representation $\mathbf{x}^t$ carries an additional error called *representativity* error:

$$\mathbf{e}^r = h(\mathbf{x}) - H(\mathbf{x}^t) = h(\mathbf{x}) - H(\boldsymbol{\Pi}(\mathbf{x})). \tag{I.21}$$

Finally, we get an observation error $\mathbf{e}^o$ wich is the sum of the latter errors:

$$\mathbf{y} = h(\mathbf{x}) + \mathbf{e}^i = H(\mathbf{x}^t) + \mathbf{e}^r + \mathbf{e}^i \tag{I.22}$$
$$= H(\mathbf{x}^t) + \mathbf{e}^o, \tag{I.23}$$

In other terms:

$$\mathbf{e}^o = \mathbf{y} - H(\mathbf{x}^t). \tag{I.24}$$

**Background and analysis errors** Background and analysis errors respectively represent the deviation of the *a priori* and *a posteriori* estimates to $\mathbf{x}^t$

$$\mathbf{e}^b = \mathbf{x}^b - \mathbf{x}^t, \tag{I.25}$$
$$\mathbf{e}^a = \mathbf{x}^a - \mathbf{x}^t. \tag{I.26}$$

#### I.3.2.2. Errors statistics

Several methods can be used to obtain an analysis, depending on the characteristics of the problem to be solved: its size, its dynamics, the available computer resources, etc. Independently of the method used, the quality of the analysis will depend on the ability to estimate the different errors and their statistical properties.

It is assumed that each error has a second order moment , i.e. a covariance matrix. This is defined as follows:

— $\mathbf{R} = \mathrm{Var}(\mathbf{e}^o) = \mathbb{E}[\mathbf{e}^o \mathbf{e}^{oT}] \in \mathbb{R}^{p \times p}$ The observation error covariance matrix. This matrix is generally supposed to be diagonal, i.e. the observation errors are independant from one another. This hypothesis is genrally condiered to be reasonable, especially when different devices conduct the observation in different places as it is the case in our studies.

— $\mathbf{B} = \mathrm{Var}(\mathbf{e}^b) = \mathbb{E}[\mathbf{e}^b \mathbf{e}^{bT}] \in \mathbb{R}^{n \times n}$ The background error covariance matrix. The simulated noise level distribution error obeys a spatial covariance structure. The extra diagonal term $\mathbf{B}_{ij} = \mathrm{Cov}(e_i^b, e_j^b)$ with $i \neq j$ describes the covariance between the background error at the points $i$ and $j$ of the state vector. The estimate of $\mathbf{B}$ allows us to account for the spatial covariance of the error over the whole domain, and thus conditions the correction brought by the analysis.

These matrices are supposed to be positive-definite. Hence every component of $\mathbf{x}^b$ of $\mathbf{y}$ is considered to be flawed.

### I.3.3. BLUE analysis

This section presents the general principles of a static data assimilation technique used in this study which aims at computing the state of the system at a given moment combining both simulated and observed levels. this presentation relies on the following works (Bouttier and Courtier, 2002; Gelb, 1974).

The method computes an analysis which is a observations based correction of the background, it is a statistical interpolation. We want this interpolation to be optimal, meaning that $\mathrm{tr}(\mathrm{Cov}(\mathbf{e}^a))$ is minimized. Here, we make the following assumptions:

— $H$ is a linear operator noted $\mathbf{H}$.

— The background and observation errors are unbiased: $\mathbb{E}[\mathbf{e}^b] = 0$ and $\mathbb{E}[\mathbf{e}^o] = 0$

— The background and observation errors are uncorrelated: $\mathbb{E}[\mathbf{e}^b\mathbf{e}^o] = 0$

We aim at building an analysis $\mathbf{x}^a$ which has the following properties:

— A linear combination of the background and the observations:

$$\mathbf{x}^a = \mathbf{L}\mathbf{x}^b + \mathbf{K}\mathbf{y} \tag{I.27}$$

— unbiased:

$$\mathbb{E}[\mathbf{x}^a] = 0 \tag{I.28}$$

Since:

$$\begin{aligned}\mathbf{e}^a &= \mathbf{L}\mathbf{x}^b + \mathbf{K}\mathbf{y} - \mathbf{x}^t \\ &= \mathbf{L}\mathbf{x}^b - \mathbf{L}\mathbf{x}^t + \mathbf{L}\mathbf{x}^t + \mathbf{K}\mathbf{y} - \mathbf{K}\mathbf{H}\mathbf{x}^t + \mathbf{K}\mathbf{H}\mathbf{x}^t - \mathbf{x}^t \\ &= \mathbf{L}\mathbf{e}^b + \mathbf{L}\mathbf{x}^t + \mathbf{K}\mathbf{e}^o + \mathbf{K}\mathbf{H}\mathbf{x}^t - \mathbf{x}^t,\end{aligned}$$

and $\mathbf{x}^b$ and $\mathbf{y}$ are unbiased:

$$\mathbb{E}(\mathbf{e}^a) = 0 \iff (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})\mathbb{E}(\mathbf{x}^t) = 0. \tag{I.29}$$

This leads to $\mathbf{L} = \mathbf{I} - \mathbf{K}\mathbf{H}$, Hence:

$$\begin{aligned}\mathbf{x}^a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x}^b + \mathbf{K}\mathbf{y} \\ \mathbf{x}^a &= \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b).\end{aligned} \tag{I.30}$$

— Optimal, i.e. $\mathrm{tr}(\mathrm{Cov}(\mathbf{e}^a))$ is minimal. Starting with $\mathbf{e}^a$ expressed as a linear combination of the background and observation errors:

$$\begin{aligned}\mathbf{e}^a &= \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b) - \mathbf{x}^t \\ &= \mathbf{e}^b + \mathbf{K}(\mathbf{e}^o - \mathbf{H}\mathbf{e}^b) \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{e}^b + \mathbf{K}\mathbf{e}^o\end{aligned}$$

$$\mathbf{e}^a = \mathbf{L}\mathbf{e}^b + \mathbf{K}\mathbf{e}^o. \tag{I.31}$$

The analysis error covariance matrix $\mathbf{A}_\mathbf{K}$ satisfies:

$$\begin{aligned}\mathbf{A}_\mathbf{K} &= \mathbb{E}[\mathbf{e}^a\mathbf{e}^{aT}] = \mathbb{E}\left[(\mathbf{L}\mathbf{e}^b + \mathbf{K}\mathbf{e}^o)(\mathbf{L}\mathbf{e}^b + \mathbf{K}\mathbf{e}^o)^T\right] \\ &= \mathbb{E}\left[\mathbf{L}\mathbf{e}^b\mathbf{e}^{bT}\mathbf{L}^T + \mathbf{K}\mathbf{e}^o\mathbf{e}^{oT}\mathbf{K}^T\right] \\ &= \mathbf{L}\mathbf{B}\mathbf{L}^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T\end{aligned}$$

Using $\mathbb{E}[\mathbf{e}^b\mathbf{e}^o] = 0$ Let $\mathbf{V}$ be the variation of $\mathbf{A}_\mathbf{K}$ under the infinitesimal variation $\delta\mathbf{K}$:

$$\mathbf{V} = \mathbf{A}_{\mathbf{K}+\delta\mathbf{K}} - \mathbf{A}_\mathbf{K}. \tag{I.32}$$

The optimum value of $\mathbf{K}$ must satisfy:

$$\mathrm{tr}(\mathbf{A}_{\mathbf{K}+\delta\mathbf{K}}) - \mathrm{tr}(\mathbf{A}_\mathbf{K}) = \mathrm{tr}(\mathbf{V}) = 0, \tag{I.33}$$

Hence, we aim at finding the optimal matrix $\mathbf{K}^\star$ wich fulfills this condition:

$$\mathbf{V} = -\delta\mathbf{K}\mathbf{H}\mathbf{B} + \delta\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T - \mathbf{B}\mathbf{H}^T\delta\mathbf{K}^T + \mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\delta\mathbf{K}^T - \delta\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\delta\mathbf{K}^T + \mathbf{K}\mathbf{R}\delta\mathbf{K}^T + \delta\mathbf{K}\mathbf{R}\delta\mathbf{K}^T + \delta\mathbf{K}^T\mathbf{R}\mathbf{K}.$$

Neglecting the second order terms, with $\mathrm{tr}(\mathbf{M}) = \mathrm{tr}(\mathbf{M}^T)$ and $\mathbf{B}$ and $\mathbf{R}$ symmetrical:

$$\mathrm{tr}(\mathbf{V}) = 2\,\mathrm{tr}\left((-\mathbf{B}\mathbf{H}^T + \mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{K}\mathbf{R})\delta\mathbf{K}^T\right), \tag{I.34}$$

Hence:

$$\mathrm{tr}(\mathbf{V}) = 0 \iff \mathbf{K}^\star\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{K}^\star\mathbf{R} = \mathbf{B}\mathbf{H}^T, \tag{I.35}$$

In other terms

$$\mathbf{K}^\star = \mathbf{B}\mathbf{H}^T\left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}. \tag{I.36}$$

Such estimator is called *BLUE* as " Best Linear Unbiased Estimator". The $\mathbf{K}$ matrix is called gain matrix or Kalman matrix. Its terms depend on the guessed uncertainties concerning the background and the observations, and their corresponding covariance structures. In eqauation I.30, the term $\mathbf{d}_b^o = \mathbf{y} - \mathbf{H}\mathbf{x}^b$ is called the *innovation* vector and represents the differences between the observation and the background at the observation points. The new vector $\mathbf{d}_b^a = \mathbf{x}^a - \mathbf{x}^b$ between the background and the analysis is the output of the Kalman matrix applied to the innovation vector. this vector depends of covariance matrices structures, the spatial distribution of the observations and level of confidence given to the background and the observations.

Consider the following toy model. Define a state vector of dimension 2 and an observation of the second component of the state vector:

$$\mathbf{x}^b = \begin{pmatrix} x_1^b \\ x_2^b \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 0 & 1 \end{pmatrix}. \tag{I.37}$$

Let $\sigma_b^2$ be the background error variance, $\sigma_o^2$ the observation error variance and $\rho \in [-1, 1]$ the correlation between the two component of the state vector $\mathbf{x}^b$. Hence, the error matrices are expressed as follows:

$$\mathbf{B} = \begin{pmatrix} \sigma_b^2 & \rho\sigma_b^2 \\ \rho\sigma_b^2 & \sigma_b^2 \end{pmatrix} = \sigma_b^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \mathbf{R} = \sigma_o^2\mathbf{I}_1 = \sigma_o^2. \tag{I.38}$$

Hence the Kalman matrix is expressed as:

$$\begin{aligned} \mathbf{K}^\star &= \sigma_b^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \left( \begin{pmatrix} 0 & 1 \end{pmatrix} \sigma_b^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \sigma_o^2 \right)^{-1} \\ &= \sigma_b^2 \begin{pmatrix} \rho \\ 1 \end{pmatrix} (\sigma_b^2 + \sigma_o^2)^{-1} \end{aligned}$$

The analysis vector $\mathbf{x}^a$ is then expressed as:

$$\begin{aligned} \mathbf{x}^a &= \mathbf{x}^b + \mathbf{K}^*(y - \mathbf{H}\mathbf{x}^b) \\ &= \mathbf{x}^b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2} \begin{pmatrix} \rho \\ 1 \end{pmatrix} (y - x_2^b). \end{aligned}$$

Hence, the value of $x_2^a$ is a fonction of the innovation, weighted by the quantity $\frac{\sigma_b^2}{\sigma_b^2+\sigma_o^2}$. Let consider two extreme scenarios:

— The level of confidence to the observation is high, i.e. $\sigma_o^2 \ll \sigma_b^2$ then:

$$x_2^a \simeq x_2^b + (y - x_2^b) = y; \tag{I.39}$$

— The level of confidence to the observation is low, i.e. $\sigma_o^2 \gg \sigma_b^2$ then:

$$x_2^a \simeq x_2^b + 0 \times (y - x_2^b) = x_2^b, \tag{I.40}$$

This consideration is also valid for the first component of $\mathbf{x}^a$ but the increment is weighted by $\rho$.

### I.3.4. Variational approaches

The results shown in section I.3.3 can be also proven in a different approach which involves the optimization of some cost function. This approach will be a key tool for the study shown in chapter IV.

Using the notation of section I.3.3, define the following cost function:

$$\mathcal{J}(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}^b\|^2_{\mathbf{B}^{-1}} + \|\mathbf{y} - H(\mathbf{x})\|^2_{\mathbf{R}^{-1}} \tag{I.41}$$

where $\|\mathbf{x}\|_{\mathbf{M}} = \sqrt{\mathbf{x}^T\mathbf{M}\mathbf{x}}$. The gradient formula is

$$\nabla\mathcal{J}(\mathbf{x}) = 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - 2\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x})) \tag{I.42}$$

With $\mathbf{H}$ being the Jacobian matrix of $H$ at $\mathbf{x}$. If $H$ is a linear function, then $\mathbf{H}$ et $H$ and $\mathcal{J}$ is convex since $\mathbf{B}$ and $\mathbf{R}$ are positive-definite. Hence, there is a uniquely defined minimal value $\mathbf{x}^\star$ such that $\mathcal{J}(\mathbf{x}^\star) = 0$ which is:

$$\mathbf{x}^\star = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}) \tag{I.43}$$

**Lemma 1.** *Using the Sherman-Morrisson-Woodbury identity (Sherman and Morrison, 1950), we find:*

$$(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{I.44}$$

*Proof.* The Sherman-Morrisson-Woodbury identity reads, for matrices with ad-hoc dimensions:

$$(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{V}\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}\mathbf{A}^{-1} \tag{I.45}$$

Hence, using the matrices above

$$\begin{aligned}
&\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} - (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} \\
=&\mathbf{B}\mathbf{H}^T(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{H}(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}) - (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} \\
=&\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1} - \mathbf{B}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} - (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} \\
=&\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1} - (\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{I})(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} \\
=&\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1} - \mathbf{B}\underbrace{(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}}_{\mathbf{I}}\mathbf{H}^T\mathbf{R}^{-1} \\
=&\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1} - \mathbf{B}\mathbf{H}^T\mathbf{R}^{-1} = 0
\end{aligned} \tag{I.46}$$

Hence, we have $\mathbf{B}\mathbf{H}^T(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} = (\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$ then the matrices product found in (I.43) corresponds to the same Kalman matrix as described in section I.3.3: $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$. □

With

$$\mathbf{x}^\star = \mathbf{x}^b + \mathbf{K}(\mathbf{y}-\mathbf{H}\mathbf{x}) \tag{I.47}$$

We find the same result as in the BLUE estimation used before. If the cost function has non linear components, an optimal value can be found with Newton-like optimization algorithms as shown in chapter IV. The algorithm wich consists in minimizing the cost function is called 3D-VAR, in a linear framework, output is the same as BLUE.

## I.4. Meta-modeling: General principles

### I.4.1. Latin hypercube sampling

When sampling a high-dimensional set, a regular grid includes too many points for the simulations to be carried out. For example, if the input set is $[a,b]^k$, a regular grid with $r$ discretization points along each axis leads to a sample of size $r^k$. Since the number of simulations should be limited to about $2{\cdot}10^3$, with $k=7$ (in our case), it is impossible to have a regular grid. Even with only 4 points along each axis, the number of simulations is too high ($4^7 = 16384 \gg 2000$). A randomly chosen sample with a uniform distribution is also intractable since this method leads to some clustering effects which prevent us to efficiently span the input space. In order to emulate a relatively smooth function, an effective sampling method with a limited sample size should guarantee a uniform distribution along each axis. The LHS is appropriate for this task. The latin hypercube sampling method (LHS) (McKay et al., 1979) is a sampling method that applies to an hypercubic input interval, say $I_{LHS} = [0,1]^k$. In our case, the parameters are proportionality coefficients or physical values $I = [\min(p_1)-\delta_1, \max(p_1)+\delta_1]\times\cdots\times[\min(p_k)-\delta_k, \max(p_k)+\delta_k]$, with a margin $\delta_i$ that is used to avoid edge effects at the interpolation step of the meta-model. It is trivial to go from $I$ to $I_{LHS}$ with an affine transformation. To get a latin hypercube sample of size $r$, we split $I$ into $r^k$ hypercubes of equal size $\left[\frac{i_1}{r}, \frac{i_1+1}{r}\right]\times\cdots\times\left[\frac{i_k}{r}, \frac{i_k+1}{r}\right]$, with $\{i_1,\ldots,\ i_k\} \in [\![0,\ r-1]\!]^k$, and then iteratively choose $r$ points with the following rules:

— The first sample vector $\mathbf{p}^{(1)}$ is randomly chosen in $I_1 = I_{LHS}$

— The i-th sample vector $\mathbf{p}^{(i)}$ is randomly chosen in $I_i$ which is recursively defined as follows. $\mathbf{p}^{(i)}$ is chosen in

$$H = \left[\frac{j_{1H}}{r}, \frac{j_{1H}+1}{r}\right]\times\cdots\times\left[\frac{j_{kH}}{r}, \frac{j_{kH}+1}{r}\right] \in \mathcal{H}'_i \tag{I.48}$$

with

$$\mathcal{H} = \left\{\left[\frac{i_1}{r}, \frac{i_1+1}{r}\right]\times\cdots\times\left[\frac{i_k}{r}, \frac{i_k+1}{r}\right], \{i_1,\cdots,i_k\} \in [\![0, r-1]\!]^k\right\} \tag{I.49}$$

and

$$\mathcal{H}'_i = \{x, x\in\mathcal{H}\wedge x\subset I_i\} \tag{I.50}$$

— $I_{i+1}$ is the remaining part of $I_i$ after we removed $H$ and all the hypercubes with one interval in common with $H$, hence all the hypercubes in:

$$E_i = \left\{ \left[\frac{\ell_1}{r}, \frac{\ell_1+1}{r}\right] \times \cdots \times \left[\frac{\ell_k}{r}, \frac{\ell_k+1}{r}\right] \subset I_i, \; \{\ell_1, \cdots, \ell_k\} \in [\![0, r-1]\!]^k \middle| \; \exists!\, m \in [\![1,k]\!],\ \ell_m \neq j_{m_H} \right\} \tag{I.51}$$

$$I_{i+1} = I_i \setminus (E_i \cup H_i) \tag{I.52}$$

With this strategy, we get a sample of values uniformly distributed along each axis. An example of a LHS of size 10 in $[0,1]^2$ is shown in figure I.4.

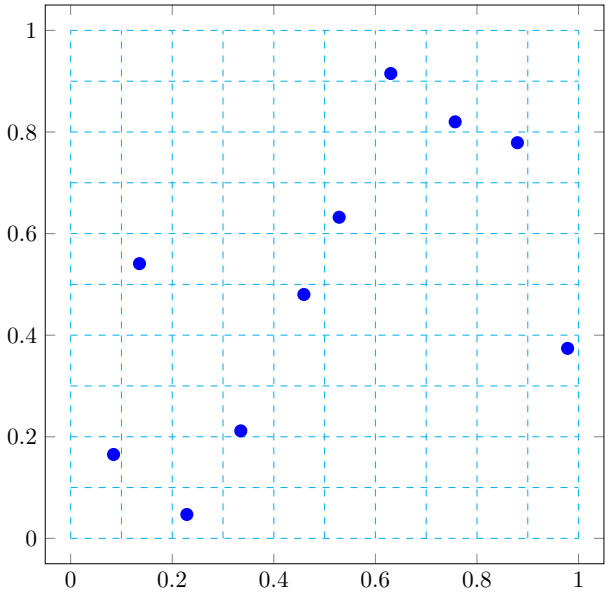


Figure I.4. – Example of a latin hypercube sampling (LHS) in dimension 2 and with 10 samples.

## I.4.2. Dimension Reduction

Before the emulation step, which applies to scalar functions, we need to represent the model outputs with a few scalars. Indeed, a direct application of the emulation to the simulator output would imply to design an emulator for each receptor. With tens of thousands of receptors, the overall computation time would be too high. The strategy is to project $\mathbf{x}$ onto a subspace spanned by a reduced basis $(\mathbf{\Psi}_i)_{i\in[\![1,d]\!]}$ with $d \ll n$. If $\mathbf{\Psi} = [\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_d] \in \mathbb{R}^{n\times d}$ and $\bar{\mathbf{x}} = \frac{1}{r}\sum_{i=1}^{r} \mathbf{x}^{(i)}$, we would expect that

$$\mathbf{x} \simeq \bar{\mathbf{x}} + \mathbf{\Psi}\mathbf{\Psi}^T(\mathbf{x} - \bar{\mathbf{x}}) \tag{I.53}$$

Since $\bar{\mathbf{x}}$ and $\mathbf{\Psi}$ are known, one would only have to generate $d$ emulators for $\mathbf{\Psi}^T\mathcal{M}(\mathbf{p}) \in \mathbb{R}^d$ instead of $n$ emulators for $\mathcal{M}(\mathbf{p}) \in \mathbb{R}^n$.

The reduced basis is chosen so as to represent the variability of the noise level fields. It is computed with the training set $\mathbf{X} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)}]$ obtained with the LHS strategy, and it is determined by a principal component analysis (PCA). The components $\mathbf{\Psi}_\ell$ ($\ell \in [\![1, n]\!]$) are sorted so that the i-th component accounts for the i-th largest variability $\lambda_i$ of the data. The total quadratic error of the approximations of the training set is

$$\sum_{h=1}^{r} \|\mathbf{x}^{(h)} - \bar{\mathbf{x}} - \sum_{j=1}^{d}((\mathbf{x}^{(h)} - \bar{\mathbf{x}})^T \mathbf{\Psi}_j)\mathbf{\Psi}_j\|_2^2 = \sum_{\ell=d+1}^{n} \lambda_\ell \tag{I.54}$$

$(\mathbf{x}^{(h)} - \bar{\mathbf{x}})^T \mathbf{\Psi}_j$ is the projection of $\mathbf{x} - \bar{\mathbf{x}}$ in the direction $\mathbf{\Psi}_j$.

If we call $\bar{\mathbf{X}}$ the centered training set $\bar{\mathbf{X}} = [\mathbf{x}^{(1)} - \bar{\mathbf{x}}, \ldots, \mathbf{x}^{(r)} - \bar{\mathbf{x}}]$, then for any $\ell \in [\![1, n]\!]$, the variability of the $\ell$-th components is defined with $\bar{\mathbf{X}}\bar{\mathbf{X}}^T \mathbf{\Psi}_\ell = \lambda_\ell \mathbf{\Psi}_\ell$. The part of the variance (of the training set) explained by the first $d$ components is

$$\frac{\sum_{\ell=1}^{d} \lambda_\ell}{\sum_{\ell=1}^{n} \lambda_\ell} \tag{I.55}$$

A high explained variance, like 98 %, can be achieved with only a few vectors in the reduced basis (about four, i.e., $d = 4$).

## I.4.3. Emulation

For any $\mathbf{p} \in I$, we now wish to emulate $\mathbf{\Psi}^T \mathcal{M}(\mathbf{p}) \in \mathbb{R}^d$. We denote $\alpha_i(\mathbf{p}) = \mathbf{\Psi}_i^T \mathcal{M}(\mathbf{p}) \in \mathbb{R}$, and $\hat{\alpha}_i$ its emulator which should satisfy $\forall \mathbf{p} \in I,\ \hat{\alpha}_i(\mathbf{p}) \simeq \alpha_i(\mathbf{p})$. If we denote $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_d)$ then

$$\forall \mathbf{p} \in I,\ \hat{\boldsymbol{\alpha}}(\mathbf{p}) \simeq \mathbf{\Psi}^T \mathcal{M}(\mathbf{p}) \tag{I.56}$$

The following sections will describe two methods usually used to generate an emulator for a function with multidimensional inputs and scalar outputs:

— Radial basis functions;

— Kriging.

### I.4.3.1. Radial Basis Function Interpolation

Once we computed the projections $\left\{\mathbf{\Psi}_i^T \mathcal{M}(\mathbf{p}^{(1)}), \ldots, \mathbf{\Psi}_i^T \mathcal{M}(\mathbf{p}^{(r)})\right\}$ of the simulations along each axis $i$, the next step is to find an approximation of $\mathbf{\Psi}_i^T \mathcal{M}(\mathbf{p})$ for every $\mathbf{p}$ in $I$, with an interpolation function. The radial basis function interpolator (RBF) (Broomhead and Lowe, 1988) is an interpolator which demands very low computational power. The idea is to define a distance $\Lambda(\cdot, \cdot)$ over the input space and to choose a function $\phi : \mathbb{R}_+ \to \mathbb{R}$. The resulting interpolator is a linear combination of the values of the function applied to the distance between the input parameter $\mathbf{p}$ and the training set inputs:

$$\begin{array}{rcccc} \hat{\alpha}_i & : & I & \to & \mathbb{R} \\ & & \mathbf{p} & \mapsto & \displaystyle\sum_{j=1}^{r} \gamma_{i,j} \phi(\Lambda(\mathbf{p}, \mathbf{p}^{(j)})) \end{array} \tag{I.57}$$

Note that the distance $\Lambda$ and the radial function $\phi$ could be different for different components $i$, but in our case, we used the same for all $d$ components. The weights are chosen so that the predictions at the training points match the data, hence

$$\forall j \in [\![1, r]\!],\ \hat{\alpha}_i(\mathbf{p}^{(j)}) = \alpha_i(\mathbf{p}^{(j)}) = \boldsymbol{\Psi}_i^T \mathcal{M}(\mathbf{p}^{(j)}) \tag{I.58}$$

If we denote

— $\boldsymbol{\Phi} = \left(\phi\left(\Lambda\left(\mathbf{p}^{(i)}, \mathbf{p}^{(j)}\right)\right)\right)_{(j,k)\in[\![1,r]\!]^2} \in \mathbb{R}^{r\times r}$;

— $\boldsymbol{\gamma}_i = (\gamma_{i,1} \cdots \gamma_{i,r})^T \in \mathbb{R}^r$;

— $\boldsymbol{b}_i = \left(\boldsymbol{\Psi}_i^T \mathcal{M}(\mathbf{p}^{(1)}) \cdots \boldsymbol{\Psi}_i^T \mathcal{M}(\mathbf{p}^{(r)})\right)^T \in \mathbb{R}^r$;

The computation of the weights amounts to solving the linear system

$$\boldsymbol{\Phi}\boldsymbol{\gamma}_i = \boldsymbol{b}_i \tag{I.59}$$

The offline computation of the emulator consists in solving $d$ (one for each $\hat{\alpha}_i$) linear systems of size $r$. Usually, and as well in our case study, the distance chosen is the Euclidean distance $\Lambda(\cdot,\cdot) = \|\cdot - \cdot\|_2$ with scaling factors, and the radial function is the cubic function $\phi : r \to r^3$. The figure I.5 shows a visual example in 1d of the RBF interpolation of the sine function

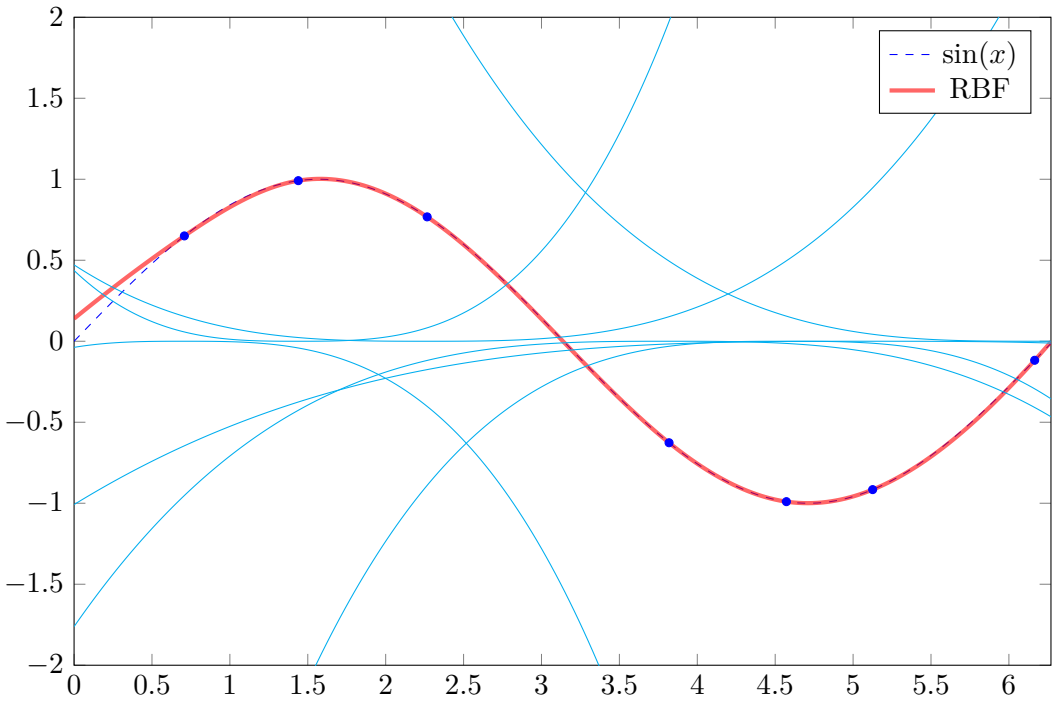


Figure I.5. – RBF interpolation in 1D, the red curve is the weighted sum of the blue curves (the cubic radial basis functions of the example)

### I.4.3.2. Kriging

The Kriging emulator (Matheron, 1962) is a statistical interpolator which demands a little more computational power than the RBF interpolator. Since it relies on a

assumption on the statistical behavior of the function $\alpha_i$, the validity of this assumption could improve the performance of the emulation, compared to the RBF approach. The founding assumption is that we consider $\alpha_i$ as a centered stationary stochastic process, which follows a Gaussian distribution with a fixed variance $\sigma_i^2$:

$$\forall \mathbf{p} \in I,\ \alpha_i(\mathbf{p}) \sim \mathcal{N}(0, \sigma_i) \tag{I.60}$$

and

$$\begin{aligned} &\forall (\mathbf{p}_1 \cdots \mathbf{p}_q) \in I^q, \exists\ \mathbf{\Gamma} \in \mathbb{R}^{q \times q}, \\ &(\alpha_i(\mathbf{p}_1) \cdots \alpha_i(\mathbf{p}_q))^T \sim \mathcal{N}(0, \mathbf{\Gamma}) \end{aligned} \tag{I.61}$$

With this assumption, it is possible to build a unique estimator (Rasmussen and Williams, 2005) with the following properties:

— linear—it is a linear combination of the training data;
— unbiased—the prediction matches the data at the training points;
— optimal—in the sense that it minimizes the variance of the prediction error $\mathbb{E}\left[(\hat{\alpha}_i(\mathbf{p}) - \alpha_i(\mathbf{p}))^2\right]$.

This interpolation method is known as universal Kriging. The Kriging interpolator depends on the covariance function of the Gaussian process. More precisely, if we denote

— $\mathbf{\Gamma} = \left(Cov\left(\alpha_i(\mathbf{p}^{(j)}), \alpha_i(\mathbf{p}^{(k)})\right)\right)_{(j,k)\in[\![1,r]\!]^2} \in \mathbb{R}^{r\times r}$,
— $\boldsymbol{\gamma}(\mathbf{p}) = \left(Cov\left(\alpha_i(\mathbf{p}), \alpha_i(\mathbf{p}^{(j)})\right)\right)_{j\in[\![1,r]\!]} \in \mathbb{R}^{r}$,
— $\boldsymbol{\lambda} = \left(\alpha_i(\mathbf{p}^{(j)})(\omega)\right)_{j\in[\![1,r]\!]} \in \mathbb{R}^{r}$,

with $\omega$ being a random selection in $\Omega$ (the events space), $\boldsymbol{\lambda}$ being a random variable, whereas $\mathbf{\Gamma}$ and $\boldsymbol{\gamma}(\mathbf{p})$ being deterministic.

The Kriging interpolator estimates the expected output, it is formulated as:

$$\begin{aligned} \hat{\alpha_i}(\mathbf{p}) &= \mathbb{E}[\alpha_i(\mathbf{p})|\boldsymbol{\lambda}] \\ &= \boldsymbol{\lambda}^T \mathbf{\Gamma}^{-1} \left[\boldsymbol{\gamma}(\mathbf{p}) + \mathbf{1}_r \frac{1 - \mathbf{1}_r^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}(\mathbf{p})}{\mathbf{1}_r^T \mathbf{\Gamma}^{-1} \mathbf{1}_r}\right] \end{aligned} \tag{I.62}$$

A key offline computation of the emulator consists in estimating the covariance kernel of the Gaussian process. We assume the second-order stationarity of the Gaussian process, which implies the covariance between two points only depends on the distance separating them. In addition, for computing considerations, we choose a separable kernel, i.e., a tensor product of $k$ univariate functions, each depending on only one parameter.

The regularity of the Gaussian process is a determining factor for the selection of the covariance function.The Matérn function $\phi_\nu(r, \theta)$ is a function whose parameter $\nu$ represents the level of roughness of the process, and $\theta$ represents the parameters to be tuned. The parameter $\nu$ goes from $\frac{1}{2}$ for a non differentiable continuous process ($\phi_{\frac{1}{2}}$ is an exponentially decreasing function) to $+\infty$ for an indefinitely differentiable process ($\phi_{+\infty}$ is a Gaussian function). $\phi_{\frac{3}{2}}$ and $\phi_{\frac{5}{2}}$ respectively describe once and twice differentiable processes. $\theta = (\sigma\ \ell)^T$ where $\sigma^2$ represents the variance of the process and $\ell$ its characteristic distance.

The covariance kernel has the following form

$$\forall(\mathbf{p}_1, \mathbf{p}_2) \in I^2,\ Cov(\alpha_i(\mathbf{p}_1), \alpha_i(\mathbf{p}_2)) = \prod_{i=1}^{k} \phi_\nu(|p_{1_i} - p_{2_i}|, \theta_i) \tag{I.63}$$

The covariance kernel is differentiable with respect to $\theta$, hence it is possible to infer $\theta_{opt}$ under the maximum likelihood estimation with a gradient descent algorithm for

$$\left(\alpha_i(\mathbf{p}^{(1)}) \cdots \alpha_i(\mathbf{p}^{(r)})\right)^T \sim \mathcal{N}(0, \mathbf{\Gamma}) \tag{I.64}$$

The figure I.6 shows a visual example in 1d of the Kriging interpolation of the sine function

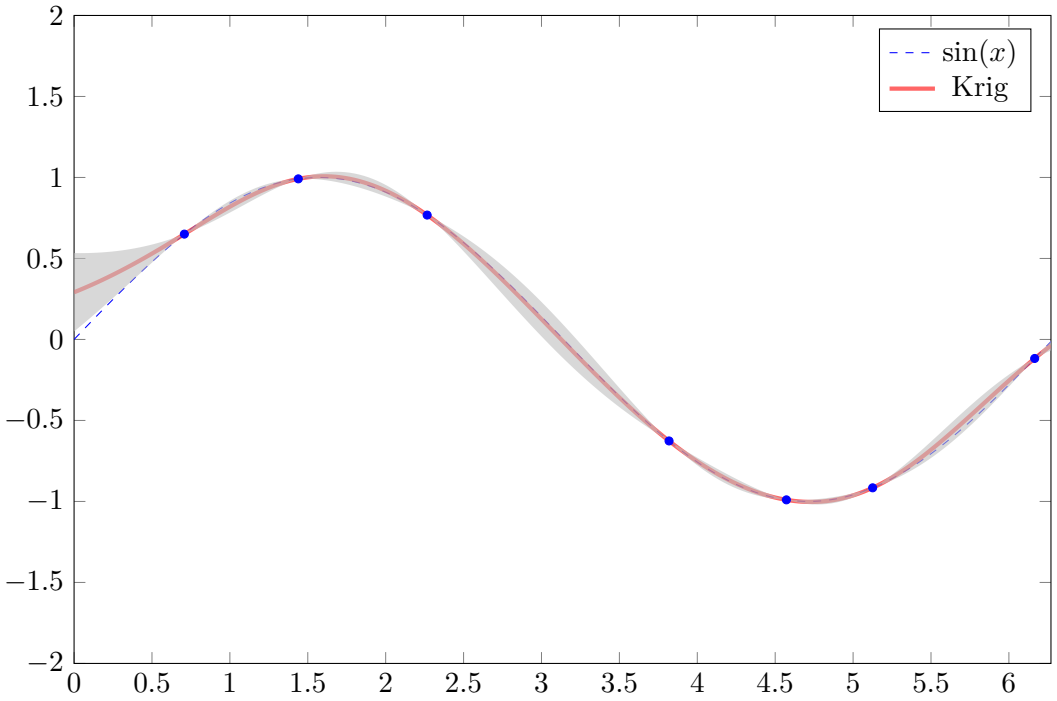


Figure I.6. – Kriging interpolation in 1D

## I.5. Some useful statistical tools

All along this study several concepts related to the statistics are used. A brief overview is presented in this section

### I.5.1. The $\chi^2$ law

The $\chi^2_n$ law with $n$ degrees of freedom is a random distribution followed by $\sum_{i=1}^{n} X_i^2$ where $(X_i)_i$ are random variables with a standard normal distribution. If $Y = \sum_{i=1}^{n} X_i^2$, then $\mathbb{E}[Y] = p$. Different $\chi^2$ distributions are shown in figure I.7.

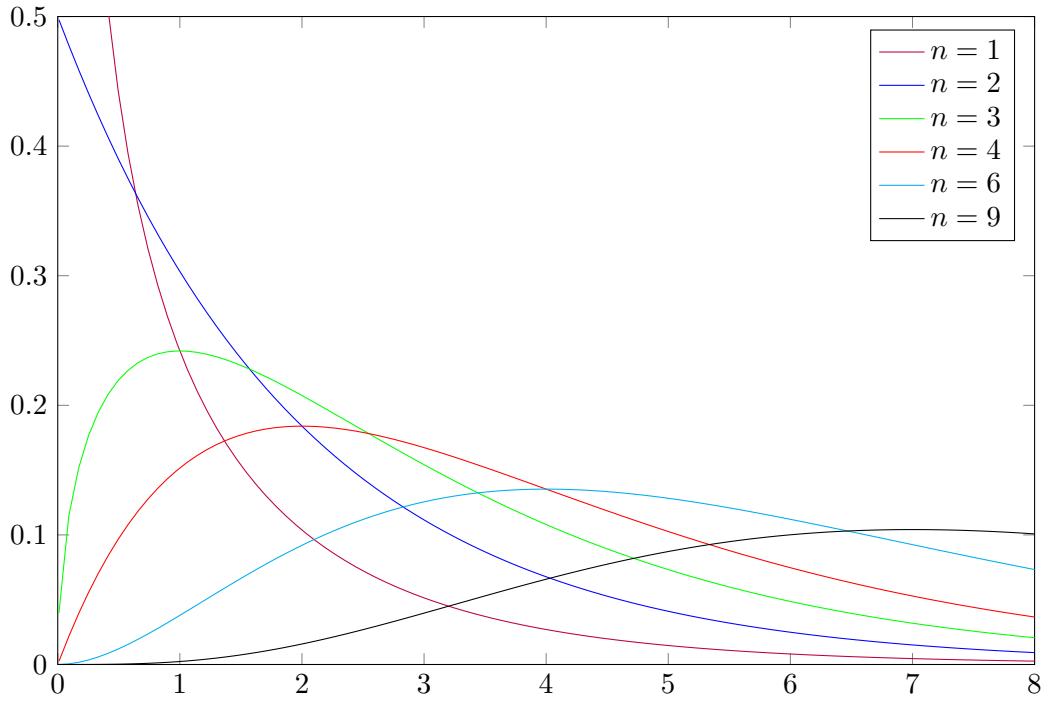


Figure I.7. – probability density function of different $\chi^2$ laws

If $\boldsymbol{\nu}$ is a random variable of dimension $n$ such that $\boldsymbol{\nu} \sim \mathcal{N}(0, \boldsymbol{\Gamma})$ then, following the definition, $\boldsymbol{\nu}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\nu} \sim \chi^2_n$. Then, we find the following propositions used in chapters III and V

— $\mathbb{E}[\boldsymbol{\nu}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\nu}] = n$

— If $z_\alpha$ is a quantile of order $1 - \alpha$ of the law $\chi^2_n$, then

$$\mathbb{P}(\boldsymbol{\nu}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\nu} \leq z_\alpha) = \int_0^{z_\alpha} f_{\chi^2}(x) \mathrm{d}x = 1 - \alpha \tag{I.65}$$

## I.5.2. Maximum likelihood

### I.5.2.1. Definition

The maximum likelihood estimator is a method used to estimate which parameters $\boldsymbol{\theta}$ of a distribution $p$ maximizes the probability for a given sample $\boldsymbol{\nu}$ to occur, i.e. where the density $p(\boldsymbol{\nu}, \boldsymbol{\theta})$ is maximum.

The real function $\boldsymbol{\theta} \to p_n((\boldsymbol{\nu}^{(k)})_k, \boldsymbol{\theta})$ where $p_n((\boldsymbol{\nu}^{(k)})_k, \boldsymbol{\theta}) = \prod_{i=1}^k p(\boldsymbol{\nu}^{(i)}, \boldsymbol{\theta})$ and $(\boldsymbol{\nu}^{(k)})_k$ fixed, is called the likelihood of the sample $(\boldsymbol{\nu}^{(k)})_k$. There exists a unique $\hat{\boldsymbol{\theta}}_n$ such that the likelihood is maximal. The random variable $\hat{\boldsymbol{\theta}}((\boldsymbol{\nu}^{(k)})_k)$ is called the maximum likelihood estimator.

Define the log-likelihood as the logarithm of the likelihood:

$$\boldsymbol{\theta} \to L_n((\boldsymbol{\nu}^{(k)})_k, \boldsymbol{\theta}) = \log(p_n(\boldsymbol{\nu}^{(k)})_k, \boldsymbol{\theta}) = \sum_{i=1}^k \log(p(\boldsymbol{\nu}^{(i)}, \boldsymbol{\theta})) \tag{I.66}$$

#### I.5.2.2. Application to a normal distribution

Suppose now that for all $i$, $\boldsymbol{\nu}^{(i)}$ follow a Gaussian distribution, $\boldsymbol{\nu}^{(i)} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$. Hence, its probability density function is:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\sqrt{|\boldsymbol{\Gamma}|}} \exp(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}(\mathbf{x}-\boldsymbol{\mu})) \tag{I.67}$$

Hence, the log-likelihood of $(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ for a sample $\boldsymbol{\nu}$ of size 1 is:

$$L(\boldsymbol{\nu}, \boldsymbol{\mu}, \boldsymbol{\Gamma}) = -\frac{p}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}(\boldsymbol{\nu}-\boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{\nu}-\boldsymbol{\mu}) \tag{I.68}$$

In the case studied in chapter V, $\boldsymbol{\Gamma}$ is fixed and $\boldsymbol{\mu}$ depends on a set of parameters $\boldsymbol{\theta}$ and is expressed as $\boldsymbol{\mu}(\boldsymbol{\theta})$, hence the maximum likelihood estimator minimizes the cost function $\mathcal{J}$ defined as follow:

$$\mathcal{J}(\boldsymbol{\theta}) = (\boldsymbol{\nu}-\boldsymbol{\mu}(\boldsymbol{\theta}))^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{\nu}-\boldsymbol{\mu}(\boldsymbol{\theta})) \tag{I.69}$$

### I.5.3. Matérn function

The Matérn class of covariance functions is given by:

$$k(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}r}{\ell}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}r}{\ell}\right) \tag{I.70}$$

with positive parameters $\nu$ and $\ell$, where $K_\nu$ is a modified Bessel function (Abramowitz and Stegun, 1964). Note that the scaling is chosen so that for $\nu \to \infty$ we obtain the squared exponential covariance function $e^{-r^2/2\ell}$. For the Matérn class, the process is $k$-times mean square differentiable (Soo, 1973) if and only if $\nu > k$. The Matérn covariance functions become especially simple when $\nu$ is half-integer: $\nu = p + 1/2$, where p is an integer. In this case the covariance function is a product of an exponential and a polynomial of order $p$, of general expression

$$k_{\nu=p+1/2}(r) = \exp\left(-\frac{\sqrt{2\nu}r}{\ell}\right)\frac{\Gamma(p+1)}{\Gamma(2p+1)}\sum_{i=0}^{p}\frac{(p+i)!}{i!(p-i)!}\left(\frac{\sqrt{8\nu}r}{\ell}\right)^{p-i} \tag{I.71}$$

The cases used in this study are $\nu = 3/2$ and $\nu = 5/2$ for which:

$$k_{\nu=3/2}(r) = \left(1+\frac{\sqrt{3}r}{\ell}\right)\exp\left(-\frac{\sqrt{3}r}{\ell}\right) \tag{I.72}$$

$$k_{\nu=5/2}(r) = \left(1+\frac{\sqrt{5}r}{\ell}+\frac{5r^2}{3\ell^2}\right)\exp\left(-\frac{\sqrt{3}r}{\ell}\right) \tag{I.73}$$

When setting $\nu = 1/2$, the the Matérn class gives the exponential covariance function $k_{\nu=1/2} = \exp -r/\ell$.

Figure I.8 shows the different covariance functions and figure I.9 shows an exemple of their corresponding Gaussian processes for $\ell = 1$.

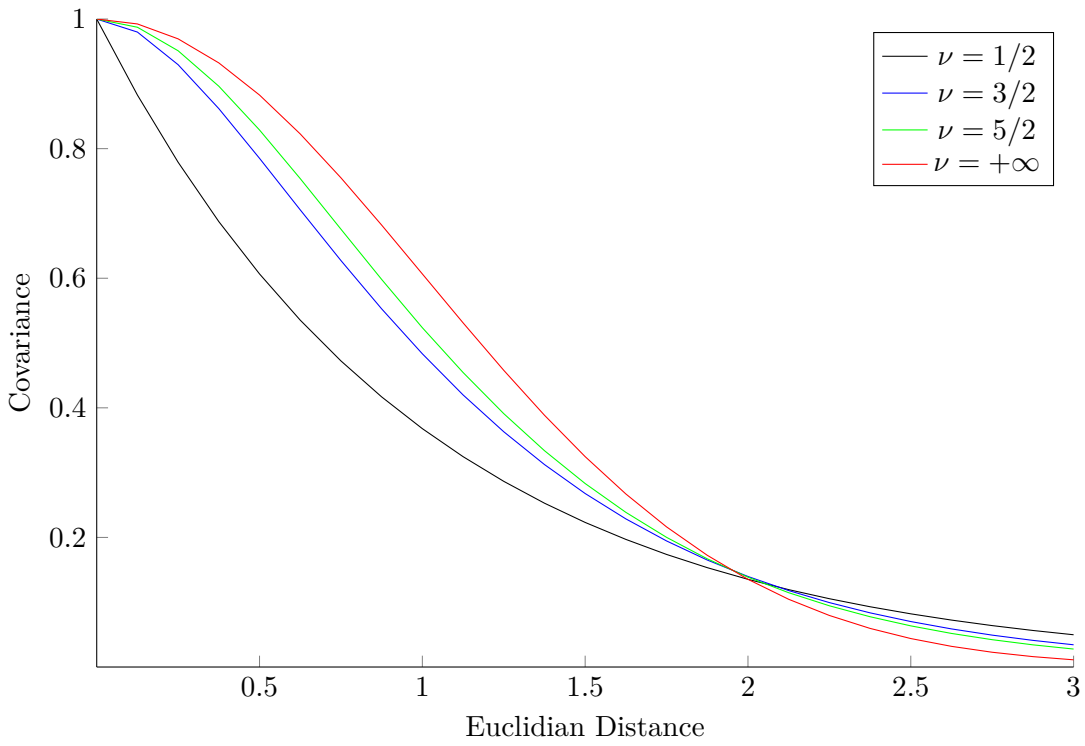


Figure I.8. – Representation of the Matérn functions

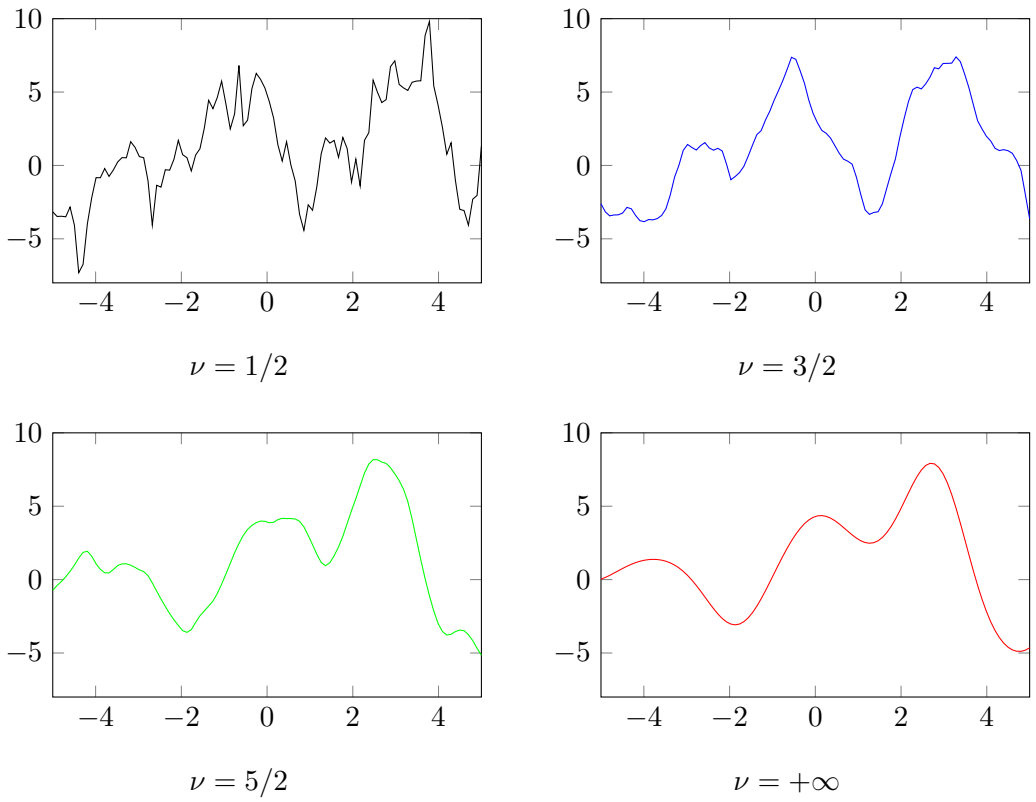


Figure I.9. – Representation of Gaussian processes with Matérn covariance kernels with different $\nu$ coefficients

# II. Meta-modeling for urban noise mapping

*This chapter introduces the construction of a meta-model for urban noise mapping. After explaining the principles of a meta-model, it justifies the need for such a device for the noise mapping of urban areas and applies it to the city of Lorient, France. Then, it evaluates its performance when compared to the output of NoiseModelling, the noise mapping software it is based on.*

## Table of contents

This chapter is a transcription of the following article: Antoine Lesieur, Pierre Aumond, Vivien Mallet, and Arnaud Can. Meta-modeling for urban noise mapping. *The Journal of the Acoustical Society of America*, 148(6):3671–3681, 2020. doi: 10.1121/10.0002866. URL https://doi.org/10.1121/10.0002866.

## Abstract

Urban noise mapping generally consists of simulating the emission and attenuation of noise in an area by following rules such as Common NOise aSSessment methOdS (CNOSSOS-EU). The computational cost makes these models unsuitable for applications such as uncertainty quantification, where thousands of simulations may be required. One solution is to replace the model with a meta-model that reproduces the expected noise levels with highly reduced computational costs. The strategy is to generate the meta-model in three steps. The first step is to generate a training sample exploring the large dimension model's inputs set. The second step is to reduce the dimension of the outputs. In the third step, statistical interpolators are defined between the projected values of the training sample over the reduced space of the outputs. Radial basis functions or kriging are used as interpolators. The meta-model was built using the open source software NoiseModelling. This study compares the proximity of the meta-model outputs to the model outputs against the reduced basis, the class of the kriging covariance function and the training sample size. Simulations using the meta-model are more than 10 000 times faster than the model, while maintaining the main behavior.

## II.1. Introduction

Noise pollution is associated with multiple psychological and physical disorders such as sleep disturbance, nervousness, cognitive impairment or hypertension (Basner et al., 2014). It is a major challenge to tackle for urban areas. The World Health Organization (WHO) recommends in its guideline, an outdoor night level of 40 dB(A) and a $L_{den}$ of 65 dB(A) (Hurtley, 2009). However, 8 million European adults suffer sleep disturbance due to environmental noise (European Environment Agency, 2014).

To diagnose sound environments, European cities with more than 100 000 inhabitants are required to produce noise maps at least every five years. These maps should present the distribution of $L_{den}$ and $L_{night}$ indicators over the studied area, in such a way that the exposure of dwellers to environmental noise can be estimated. These generic approaches may help to prevent or reduce the harmful effects of noise.

A noise map is a representation of the equivalent average noise level field over an urban area. A noise map which relies on annually averaged traffic and weather data is usually represented in regulatory noise maps. In Europe, the attenuation rules obey the European standard CNOSSOS (Kephalopoulos et al., 2012) which has become the reference since the application of the directive 2015/996 (European Commission, 2015). When this method is applied at the city level, it becomes computationally expensive due to the high number of sources and receivers. The computation time makes this process unsuitable for the generation of hourly or even daily noise maps.

Recent studies (Sevillano et al., 2016), have endeavored to enrich noise maps with data from sensor networks. Even so, computational time constraints have been problematic. Too much computation time leads to the intractability of large Monte Carlo simulations and prevents applications of advanced methods for uncertainty quantification, data assimilation, inverse modeling or network design which can require tens of thousands of calls to the model. These applications are nonetheless crucial to evaluate

the range of validity of the model, or to carry out real-time generation of observation-based noise maps.

Statistical models have been proposed to link noise levels to road traffic and urban morphology variables, in order to reduce computation times to their minimum as in Silva et al. (2013). Genaro et al. (2010) construct the estimated noise maps based on a neural network and a set of 24 variables describing the sources (number of light vehicles, etc.) and propagation conditions (building height, street width, etc.). However, although useful for understanding the traffic and urban morphology variables that influence noise levels, these works are based on measurements and so cover only a small area of the input parameters space.

A meta-model (also sometimes called surrogate model) is a tool used in several fields to statistically reproduce the behavior of a reference simulation software. Iooss et al. (2010) has proven the efficiency of a meta-model generation method for a wide class of models that rely on Gaussian processes, Janon et al. (2014) has shown that the meta-models are suitable to quickly perform sensitivity analysis on complex systems. In applied sciences, meta-models reproduce the behavior of several systems to perform various studies. For instance Iooss and Marrel (2019) has performed uncertainty quantification for a nuclear reactor using a meta-model by performing a substantial amount of calls to the model. Mallet et al. (2018) has compared the accuracy of air quality predictions with observations by generating a high amount of air quality maps computed with a meta-model. Other studies have shown the efficiency of the use of a meta-model for predicting noise level distribution, as Hart et al. (2016), which shows the performance of a meta-model for blasts sounds in local areas. In this paper, the authors propose the development and use of a meta-model for environmental acoustics to predict the urban distribution of noise level indices such as $LA_{eq,\ 1h}$.

A meta-model for noise mapping shall help to reproduce the main features of a noise model while requiring negligible computational resources. On a given area with given input data, a meta-model is a way to quickly generate a new noise map obtained by slightly modifying the input data. This approach enables to determine the impact of input data modifications on the noise map, e.g., if the traffic or weather conditions change by a certain amount.

In this paper, the authors explain the construction of such meta-model based on (1) dimension reduction and (2) interpolation methods applied to each projection coefficient on the basis vectors of the reduced output space. The interpolation methods are kriging and an interpolation based on Radial Basis Functions (RBF). Explanations about the reduction and interpolation strategy are provided in section II.2. In section II.3, the case study is described, results are given in section II.4. The paper concludes in sectionII.5 with several applications made possible by the substitution of the regular simulator by a meta-model.

## II.2. Generation of the noise mapping meta-model: Training set, Reduction and interpolation

### II.2.1. Vocabulary

In order to avoid any confusion, several terms used in this paper are described below:

— A simulator: An algorithm that computes an output from inputs according to a mathematical model of a physical phenomenon (here, noise emission and propagation);
— A meta-model (also sometimes called surrogate model): An algorithm that computes an output (in this study, a noise map) from inputs (in this study, traffic and weather data) in order to mimic a simulator with much lower computational requirements;
— A simulation: an output of the simulator for a given input;
— An interpolation: An output of the kriging or RBF method.

### II.2.2. Purpose of the meta-model

In the noise context, a meta-model can be generated with a training sample of a given amount of noise maps computed with the noise mapping software.

A meta-model $\widehat{\mathcal{M}}$ generates the same type of outputs as the noise mapping model $\mathcal{M}$. Once the meta-model is built, each call with new inputs takes a very small amount of time, because it only consists in an interpolation between the projection of the outputs of a training sample onto the vectors of a reduced basis. These reduced base vectors are constructed to capture the behavior of the reference noise mapping software in spite of a lower dimension. For every input $\mathbf{p}$ within the boundaries of the case study,

$$\widehat{\mathcal{M}}(\mathbf{p}) \simeq \mathcal{M}(\mathbf{p}) \tag{II.1}$$

### II.2.3. Description of the framework of the meta-modeling process

Noise mapping models use various input data:
— Road traffic distribution: flow rate and mean speed in road sections;
— Buildings distribution;
— Building absorption coefficient;
— Ground properties;
— General topography;
— Weather data.

A simulation is also characterized by a set of propagation parameters:
— propagation distance;
— number of reflections;
— whether or not to consider diffractions.

For a given set of parameters $\mathbf{p}$ let $\mathbf{x} = \mathcal{M}(\mathbf{p})$ where $\mathcal{M}$ is a noise mapping model that relates the model inputs $\mathbf{p}$ to the predicted noise levels $\mathbf{x}$ on a grid of receivers. If each building and road section is considered separately, the noise mapping model can have a very large number of inputs, the dimension of the input space is then very large. A common way to describe the input data as a low dimension vector of $k$ parameters $\mathbf{p} = (p_1 \ \dots \ p_k)^T$ for a given area is to apply the same multiplicative coefficient on a given set of inputs. For instance, $p_1$ can be a multiplicative coefficient applied to the yearly averaged traffic flow rate in all roads of the study area, so that the total traffic changes with the variations of $p_1$—but the traffic cannot be changed in a single specific road. By doing so, the degrees of freedom are reduced and so is the dimension of the input parameters. $\mathbf{p}$ contains all the inputs that vary, e.g., all uncertain inputs

if uncertainty quantification is carried out. See section II.3.3 for the list of parameters $p_i$ considered in this paper.

A call to $\widehat{\mathcal{M}}$ for a new set of parameters $\mathbf{p}$, only implies the computation of an interpolation function $\hat{\boldsymbol{\alpha}}(\mathbf{p})$. It consists in a short sequence of elementary operations. Hence, a rather cheap computation power is involved.

## II.2.4. Meta-model Generation

The present study relies on previous studies that have been done for urban air quality models (Mallet et al., 2018) dealing with spatial distribution of pollutants such as $NO_2$ and $PM_{10}$.

### II.2.4.1. General Formula

The training sample is composed of a set of input parameters and their corresponding simulation outputs. As a toy example, one can imagine a really basic meta-model whose output for a given input $\mathbf{p}$ would be the output corresponding to the nearest neighbor to $\mathbf{p}$ in the training sample.

In this study, the generation of the meta-model is composed of three steps:

— The training sample is strategically generated with a LHS. From a random seed, it generates a set of $N$ input parameters $(\mathbf{p}^{(i)})_{i\in[\![1,\ N]\!]}$ for the model $\mathcal{M}(\mathbf{p})$. The simulations $\left(\mathcal{M}(\mathbf{p}^{(i)})\right)_{i\in[\![1,\ N]\!]}$ of a reference simulation software are carried out, its outputs $\mathbf{x}^{(i)} = \mathcal{M}(\mathbf{p}^{(i)})$ are the noise levels in dB(A) in a grid of receivers whose input parameters correspond to an average for a given time interval. For instance, if the input parameters correspond to the average traffic and weather data at a given hour, the output will be a noise map of the corresponding $LA_{eq,\ 1h}$ distribution;

— A dimension reduction algorithm is applied to select the output subspace that gathers most of the variance of the outputs. From the outputs $\mathbf{x}^{(i)} = \mathcal{M}(\mathbf{p}^{(i)})$, the resulting output is $\boldsymbol{\Psi}$, the projection matrix to the reduced subspace. This matrix is computed with a Principal Component Analysis (PCA) (Jolliffe, 1986) as explained in section II.2.4.3;

— An interpolation function is generated for each basis vector of the reduced subspace. The interpolation function $\hat{\boldsymbol{\alpha}}(\mathbf{p})$, whose input is $\mathbf{p}$, interpolates its output, the projected values of the outputs of the simulator onto the vectors of the reduced basis. It is designed by an RBF algorithm or a kriging algorithm, as detailed in section II.2.4.4.

Let $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}^{(i)}$ be the mean value of the simulated noise maps of the $N$ LHS generated inputs.

The complete meta-model $\widehat{\mathcal{M}}$ reads (Mallet et al., 2018):

$$\widehat{\mathcal{M}}(\mathbf{p}) = \bar{\mathbf{x}} + \boldsymbol{\Psi}(\hat{\boldsymbol{\alpha}}(\mathbf{p}) - \boldsymbol{\Psi}^T\bar{\mathbf{x}}) \tag{II.2}$$

### II.2.4.2. Training sample

A training sample is generated using Latin Hypercube Sampling (LHS) McKay (1992) in the input space and computing the corresponding simulations at the sample points.

Since the dimension of the input space is relatively large, a Cartesian regular grid would grow too rapidly with the step size. If the number of input parameters (i.e. the dimension of the input space) is $k = 15$ as in this study, a Cartesian regular grid which splits each axis in two parts would lead to a training sample of size $2^{16} = 65536$. This is not compatible with the computation time allocated for this study which corresponds to a sample size of $\sim 10^3$. The latin hypercube sampling provides an adequate sampling size that allows the implementation of the adapted interpolation strategies.

### II.2.4.3. Dimension Reduction

Instead of considering the output of the model as a vector of $n$ independent receivers, it is represented as a linear combination of $d$ vectors of size $n$, with $d \ll n$. Doing such a reduction allows to design only a small number of interpolation functions. It significantly reduces the complexity of the meta-model since the following section II.2.4.4 is only applied to $d$ vectors instead of $n$. The computation time, which is proportional to the number of interpolators, is significantly reduced. The strategy is to project $\mathbf{x}$ onto a subspace spanned by a reduced basis $(\mathbf{\Psi}_i)_{i\in[\![1,d]\!]}$. If $\mathbf{\Psi} = [\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_d] \in \mathbb{R}^{n\times d}$, the following result is expected

$$\mathbf{x} \simeq \bar{\mathbf{x}} + \mathbf{\Psi}\mathbf{\Psi}^T(\mathbf{x} - \bar{\mathbf{x}}) \tag{II.3}$$

Since $\bar{\mathbf{x}}$ and $\mathbf{\Psi}$ are known, one would only have to generate $d$ interpolating functions based on $\mathbf{p}$ as input and $\mathbf{\Psi}^T\mathcal{M}(\mathbf{p}) \in \mathbb{R}^d$ as output—using the low-dimensional training set $(\mathbf{\Psi}^T\mathcal{M}(\mathbf{p}^{(i)}))_i$—instead of $n$ interpolators with $\mathbf{p}$ as input and $\mathcal{M}(\mathbf{p}) \in \mathbb{R}^n$ as output—using the high-dimensional $(\mathcal{M}(\mathbf{p}^{(i)}))_i$. In other words, the interpolation is applied to basis vectors – i.e. noise maps of norm 1 – instead of individual receivers.

The reduced basis is chosen so as to represent the variability of sound levels at receivers over the outputs of the training sample. It is computed with the training sample $\mathbf{X} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(r)}]$ obtained with the LHS strategy, and it is determined by a principal component analysis (PCA) (Jolliffe, 1986). The components $\mathbf{\Psi}_\ell$ ($\ell \in [\![1, n]\!]$) are computed so that the first component accounts for the axis along which the largest variability of the data is observed, the second for the largest variability in the orthogonal complement, and so on. Note this PCA takes place in a $n$-dimensional vector space where $n$ is the number of receivers. The PCA is applied on the outputs of the model (i.e., noise maps), each noise map being organized in a vector whose $n$th component is the noise level at receiver $n$.

### II.2.4.4. Interpolation

For any new set of parameters $\mathbf{p} \in I$, the value $\mathbf{\Psi}^T\mathcal{M}(\mathbf{p}) \in \mathbb{R}^d$ needs to be interpolated. With the notation $\alpha_i(\mathbf{p}) = \mathbf{\Psi}_i^T\mathcal{M}(\mathbf{p}) \in \mathbb{R}$, and $\hat{\alpha}_i$ its interpolator that approximates $\alpha_i(p)$, $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_d)$:

$$\hat{\boldsymbol{\alpha}}(\mathbf{p}) \simeq \mathbf{\Psi}^T\mathcal{M}(\mathbf{p}) \tag{II.4}$$

Two common methods for an interpolation function with multidimensional inputs and scalar outputs are used in this study:

— Radial Basis Functions interpolation (Broomhead and Lowe, 1988), a linear combination of radial functions centered around the sample points;

— Kriging (Matheron, 1962), a statistical interpolator which gives an expected value based on a covariance function computed between the sample points.

The compared performance of these two methods is evaluated in section II.4.3.

The reader might be familiar with kriging in its classical geostatistical background as it has been applied in some studies (Baume et al., 2009). However, contrary to the usual kriging used in several acoustics applications using geographical data, this paper applies kriging to the abstract vector space of the input parameters which has no relation with the geographical 2D space.

The interpolation algorithms used in this paper, kriging and RBF, have been selected because they fit with some characteristics which are specific to the case study, i.e., a large dimension of the input space and a small training sample, which leads to a sparse distribution of the training sample in the input space, as opposed to Cartesian regular grids.

## II.3. Meta-modeling of NoiseModelling applied to Lorient (France)

### II.3.1. Case study

This study is a part of a national research project called CENSE whose goal is to characterize urban sound environments combining numerical predictions and local observations. The case study of this article is the city of Lorient, France, which has been selected by the CENSE project and for which we have access to a complete input data set.

The receiver grid is a regular grid with a space step of 10 m where the points that are inside the buildings have been removed. They are located 4 m above the ground, which is approximately the height of the physical sensors. The total number of remaining points is 9780.

### II.3.2. Noise mapping reference software

NoiseModelling [1] (Aumond et al., 2020a) is an open source software designed to produce noise maps for evaluating the noise impact on urban mobility plans. It uses traffic, topographic and meteorological data to generate noise maps. The noise calculation method implemented within NoiseModelling is based on the standard European method CNOSSOS, as a reference method to be used under the Directive 2015/996 related to the assessment and management of environmental noise (Kephalopoulos et al., 2012). The version used for this paper is sightly different than NoiseModelling v3.0 as it has been especially designed to run multiple simulations.

The NoiseModelling software implements the CNOSSOS specifications. The CNOSSOS method is described in the reference document (Kephalopoulos et al., 2012). In order to comply with the CNOSSOS specification, the process can be summed up by the following steps: the input traffic data along each road is discretized into a set of punctual noise sources, and the noise level field is represented as Cartesian regular

1. See http://noise-planet.org/noisemodelling.html

grid. Noisemodelling computes the emissions. Then Noisemodelling uses the path finding method designed by CNOSSOS which computes the transmission paths between the sources and the receivers, taking into account the distance propagation, the number of reflections and diffactions. The software then computes the attenuation along each path for a selected range of octave bands. The sound level at a receiver is finally calculated from the contributions of all sources. Its output is a grid of receivers with an attributed sound level value computed in dB.

The authors have chosen the CNOSSOS model over other models such as HARMONOISE (Salomons et al., 2011) or NORD2000 (Jónsson and Jacobsen, 2008). Regardless of the compared performance of each model, CNOSSOS is a reference across the European Union since the application of the directive 2015/996. The case study of this paper takes place within the EU area where the directive is applied. However, this approach can be applied in practice for every type of model.

### II.3.3. Input parameters

Table III.1 shows the list of the $k = 15$ coefficients in $\mathbf{p}$.

Table II.1. – Input parameters and their input ranges for the meta-model. They are all dimensionless multiplicative coefficients applied to the reference values (i.e. the annually averaged traffic data of each road section), except the temperature and absorption coefficient which are the actual values.

| Parameter | Minimum | Maximum |
|---|---|---|
| Vehicle speed | 0.1 | 2 |
| Light vehicle flow in major roads | 0.1 | 2 |
| Medium vehicle flow in major roads | 0.1 | 2 |
| Two-wheeled vehicle flow in major roads | 0.1 | 2 |
| Heavy vehicle flow in major roads | 0.1 | 2 |
| Light vehicle flow in medium roads | 0.1 | 2 |
| Medium vehicle flow in medium roads | 0.1 | 2 |
| Two-wheeled vehicle flow in medium roads | 0.1 | 2 |
| Heavy vehicle flow in medium roads | 0.1 | 2 |
| Light vehicle flow in small roads | 0.1 | 2 |
| Medium vehicle flow in small roads | 0.1 | 2 |
| Two-wheeled vehicle flow in small roads | 0.1 | 2 |
| Heavy vehicle flow in small roads | 0.1 | 2 |
| Temperature (°C) | -5 | 30 |
| Absorption coefficient | 0 | 1 |

In addition, a few numerical parameters were kept fixed in the noise mapping calculation, although they were candidate parameters: the maximum propagation distance (set to 250 m), the distance after reflection (set to 100 m), the number of reflections (set to 1). The diffractions have been taken into account. These parameters have been fixed, since they are usual propagation parameters for dense urban areas such as the one studied in this paper. Other parameters have remained fixed whereas in future research it might be interesting to tune them: the uncertainties in topography, the

ground reflection, the height of the buildings and the height of the receivers.

The traffic data varies from street to street, but a strong correlation exists along the road network, the traffic usually goes up or down at the same rate across the entire area according to the time of the day. A proportionality coefficient can then be applied globally to the traffic data (Wei et al., 2016). Two pieces of data are given for traffic information, speed and flow rate. More variability can be captured if the flows of each categories of vehicles are segregated. Studies suggest that in urban areas, roads can be segregated into different categories with different traffic time evolutions (Barrigón Morillas et al., 2005). The study segregates three categories of roads:

— Major roads with a flow rate $> 1000\,\text{vehicles}\,\text{h}^{-1}$;
— Medium roads with a flow rate between $300\,\text{vehicles}\,\text{h}^{-1}$ and $1000\,\text{vehicles}\,\text{h}^{-1}$;
— Small roads with a flow rate $< 300\,\text{vehicles}\,\text{h}^{-1}$.

The impact of temperature in noise emission and noise propagation is not negligible, the minimum and maximum values of temperature were chosen to be typical of the variation over a complete year at this site.

Some data used as inputs for noise mapping are sometimes roughly approximated and may alter the noise map prediction. A degree of freedom has been added to monitor the impact of these uncertainties on the reference noise mapping software output. For this study, one parameter has been selected: the absorption coefficient of the walls.

For computational efficiency, parameters that have an impact on the matrix of rays have been excluded, like the buildings height, the receivers height or the type of junction. The selected parameters only changed the emission or the attenuation, hence there was no need to recompute the matrix of rays.

### II.3.4. Reference Noise Map

The traffic inputs are derived from the daily average traffic data gathered by the government department in charge of the application of noise directives and the city of Lorient traffic department, the French national study center for territory development, for the year 2016. Soil, topography and buildings data come from the French national geographical institute (IGN [2]) and the city of Lorient. The reference $L_{day}$ noise map of the city of Lorient is represented in figure II.1, as computed with the reference noise mapping software and with the unperturbed input data (yearly averaged daily road traffic and speed, average year temperature (15 °C) and an absorption coefficient of the walls of 0.2).

### II.3.5. Interpolation tools

The authors rely on the R package **Dicekriging 1.5.5** (Roustant et al., 2012) to fit the kriging covariance model and on the Python library **SciPy 0.17.0** (Virtanen et al., 2020) to compute the interpolation based on Radial Basis Functions. The radial basis functions studied in this paper are called monomial functions $\phi : r \to r^k$, $k$ being an integer. The kriging interpolation studied in this paper uses a covariance function called the Matérn function of parameter $\nu$ (Rasmussen and Williams, 2005). If $\nu = \frac{1}{2}$, the Matérn function is a decreasing exponential function, if $\nu = \infty$, the Matérn function is a Gaussian function.

2. See http://professionnels.ign.fr/

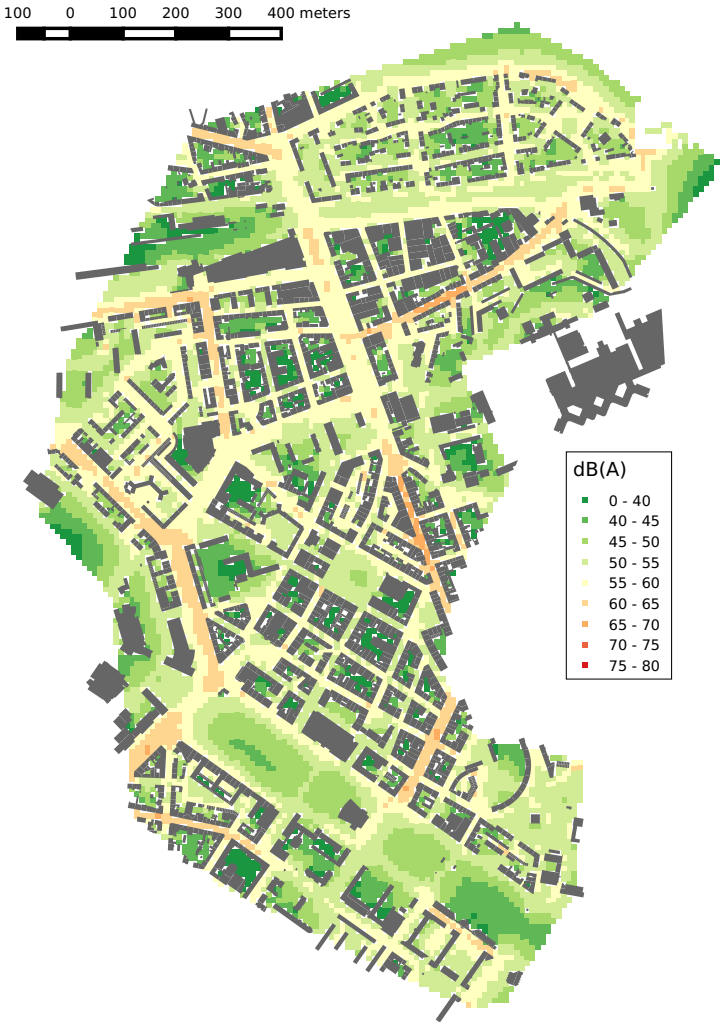


Figure II.1. – Simulation of the noise map of Lorient computed with the reference noise mapping software and the unperturbed input data.

## II.4. Results

### II.4.1. Testing setup: Scores and testing sample

This section aims at evaluating the meta-model performance, by comparing ouputs of the meta-model with outputs of the reference noise mapping software with the same set of inputs. 90 noise maps have been generated with a new set of parameters called testing sample. It depends on the original LHS, which has allowed to build the training sample, and is a complementary sample whose points are chosen to be the farthest from the original LHS. This strong constraint guarantees the validity of the produced results. The testing part consists in comparing the meta-model output with the output of the reference noise mapping software, and to observe its performance through the following scores described in table IV.3:

— Bias (dB);

— Correlation coefficient (dimensionless quantity);
— Root Mean Square Error (RMSE) (dB).

Table II.2. – Scores for the performance evaluation of a model. $(c_i)_i$ is the simulated sequence. $(o_i)_i$ is the corresponding reference sequence. $n$ is the total number of elements in the sequence. $\bar{c}$ and $\bar{o}$ are respectively the mean of $(c_i)_i$ and $(o_i)_i$.

| Score | Formula |
|---|---|
| Bias | $\frac{1}{n}\sum_{i=1}^{n}(c_i - o_i)$ |
| Correlation | $\frac{\sum_{i=1}^{n}(c_i - \bar{c})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{n}(c_i - \bar{c})^2}\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2}}$ |
| RMSE | $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(c_i - o_i)^2}$ |

The influences of the sample size (II.4.3.1), the dimension reduction (II.4.3.2), the class of the covariance kernel (for kriging) (II.4.3.3) and radial basis function (II.4.3.4) on the meta-model have been evaluated. Then the spatial distribution of the scores are displayed to identify the area with a loss of performance (II.4.3.5). Finally, the complete distributions of the scores will be analyzed (II.4.3.6).

## II.4.2. Dimension Reduction

As explained in section II.2.4.3, dimension reduction is a key step to reduce the complexity of the meta-model and to generate a small set of interpolating functions. The PCA is applied to a training sample whose size is $2{\cdot}10^3$ (this size is shown to be optimal in section II.4.3.1). Figure II.2 shows the remaining unexplained variance in the training sample against the number of selected principal components. With 4 principal components, the remaining unexplained variance goes below 3 %. There is little interest in selecting a higher number of principal components. Indeed, if a vector $\mathbf{\Psi}_i$ conveys too little explained variance, the function $\alpha_i : \mathbf{p} \to \mathbf{\Psi}_i^T \mathcal{M}(\mathbf{p})$ would be too random and meaningless, and also the interpolation function might not perform well.

The projection of a simulation over a basis composed of the 4 principal vectors of the PCA efficiently reflects the actual spatial distribution of the noise map. Table II.3 displays the RMSE of the projection against the real simulations of the testing sample under the number of selected principal components. The RMSE decreases quickly with the first vectors, and slowly goes to zero as the number of principal components increases. Figure II.3 clearly shows the huge decrease of the RMSE whether 3 or 4 vectors are used in the projection basis and the slow decrease of RMSE error as the size of the basis increases (here with a basis of size 12). 4 vectors are kept in the following.

If the principal components $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ are displayed as spatial maps, some features can be intuitively identified, as shown in figure II.4. High traffic areas stand out in the first principal component as they are the main cause of variability. As the number of

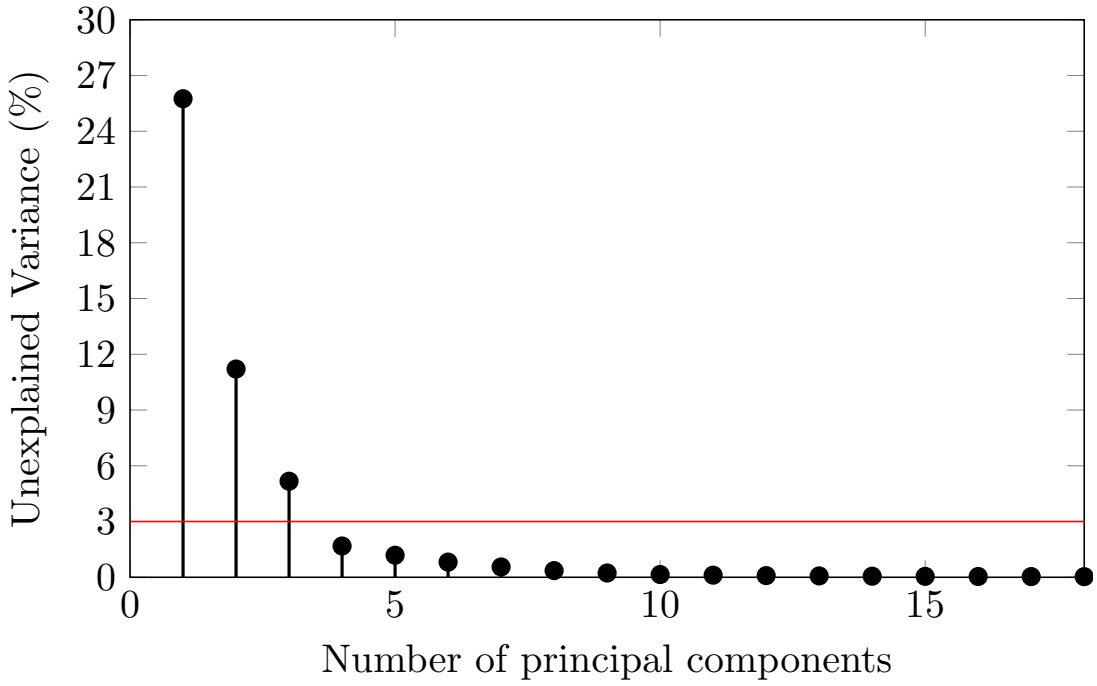


Figure II.2. – The unexplained variance (%) of the training sample against the size of the projection basis $d$: over 97 % of the variance is explained with 4 principal components.

Table II.3. – Scores of the projection against the size of the basis. The simulation is compared to its projection on the testing sample, at all the receivers.

| Basis size | 3 | 4 | 12 | 20 | 50 | 100 |
|---|---|---|---|---|---|---|
| RMSE (dB) | 0.75 | 0.40 | 0.11 | $5.6{\cdot}10^{-2}$ | $1.5{\cdot}10^{-2}$ | $5.9{\cdot}10^{-3}$ |

projection vectors increases, it is noticeable that traffic remains the principal cause of variation, nevertheless different zones in the map are highlighted, hence representing the spatial variability due to part of the traffic or local propagation effects. The first vector explains the noise distribution caused by raw traffic data. For the second component, the variability remains related to the traffic, but occurs in the quieter areas with less traffic.

### II.4.3. Comparison of the meta-model with the reference simulation software

#### II.4.3.1. Performance against the size of the training sample

The performance of the resulting meta-models is compared against the size of the training sample. Since the interpolation functions need all the elements of the training sample to compute their outputs, the larger the training sample is, the longer the meta-model takes to generate a noise map. The performance of 4 meta-models generated with different sample sizes is compared in table II.4. They have been generated on a

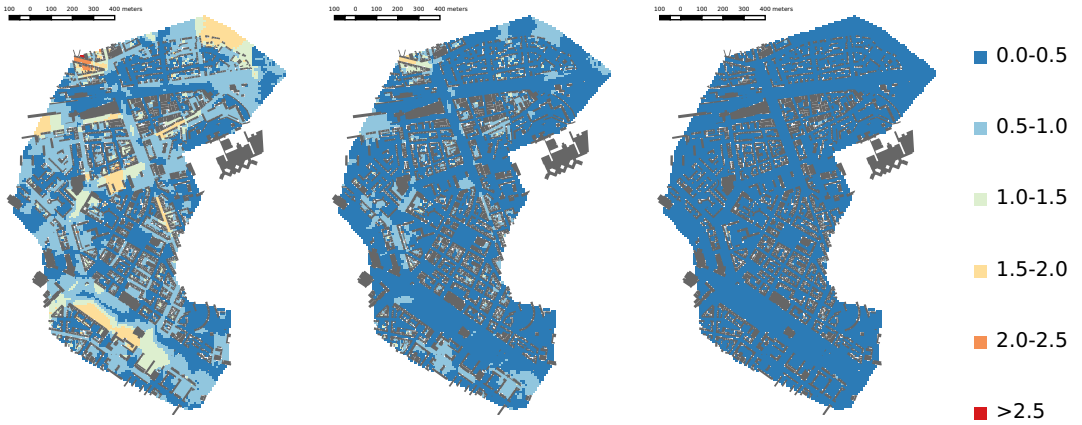


Figure II.3. – RMSE representation in dB for the simulation of the whole test sample. The simulation is compared to its projection on basis of size 3, 4 and 12.

reduced subspace of size $d = 4$ as defined in section II.2.4.3, with a Matérn $\frac{5}{2}$ function as the kriging covariance kernel and a cubic function for the RBF.

Table II.4. – Scores of the meta-model, with kriging and RBF, against the size of the sample with a reduced basis of size 4. The reference software simulations are compared to the meta-model results over the whole complementary LHS sample, at all receivers.

| sample size | $10^3$ | $2{\cdot}10^3$ | $3{\cdot}10^3$ | $10^4$ |
|---|---|---|---|---|
| **kriging** | | | | |
| Bias (dB) | $-3.5{\cdot}10^{-2}$ | $9.2{\cdot}10^{-3}$ | $4.6{\cdot}10^{-3}$ | $-5.3{\cdot}10^{-3}$ |
| RMSE (dB) | 0.94 | 0.70 | 0.76 | 0.61 |
| **RBF** | | | | |
| Bias (dB) | $9.2{\cdot}10^{-2}$ | $16{\cdot}10^{-2}$ | $7.8{\cdot}10^{-2}$ | $4.8{\cdot}10^{-2}$ |
| RMSE (dB) | 2.17 | 1.81 | 1.63 | 1.65 |

As expected, the RMSE decreases with the size of the training sample, whether kriging or RBF interpolator is used. The RMSE decreases substantially from a training size of $10^3$ simulations to a training size of $2{\cdot}10^3$ simulations. The use of more than $2{\cdot}10^3$ simulations does not seem to improve the results significantly, hence a training sample of size $2{\cdot}10^3$ will be selected for the rest of the study.

### II.4.3.2. Performance against the size of the reduced basis

The performance of the kriging and RBF approaches are now compared on the complete testing sample and over all the receivers against the size of the reduced basis. See table II.5.

The performance slightly increases with the size of the basis. However, the increase

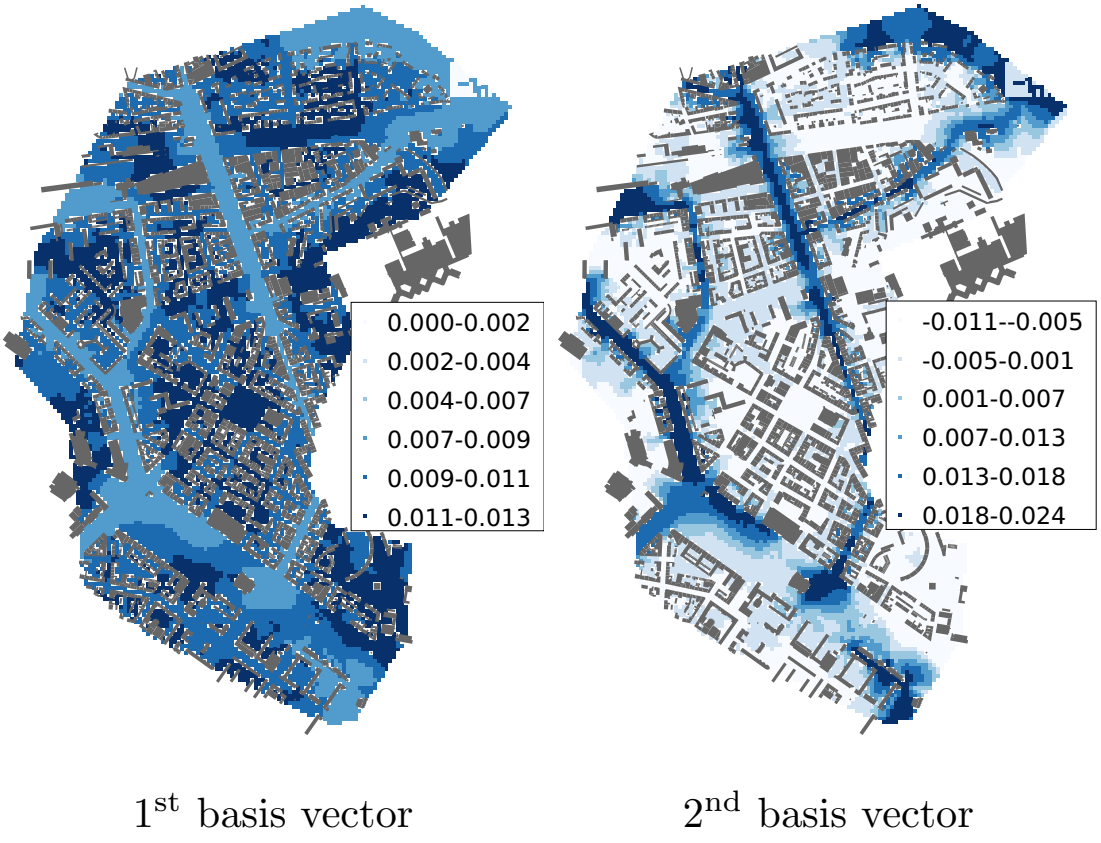


Figure II.4. – This figure represents two of the selected basis vectors. The values represented on the figure are in dB but they have no physical meaning, here the noise maps expressed as a vector are normalized so that their norm is 1. Every noise map generated by the meta-model with a physical meaning is the noise map $\bar{\mathbf{x}}$ plus a linear combination of these maps.

Table II.5. – Scores of the meta-model, with kriging and RBF, against the size of the basis. The reference software simulations outputs are compared to the meta-model results over the whole complementary LHS sample, at all receivers.

| Basis size | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| **kriging** | | | | | | |
| Bias (dB) | 0.090 | 0.093 | 0.094 | 0.093 | 0.091 | 0.092 |
| RMSE (dB) | 0.94 | 0.70 | 0.68 | 0.65 | 0.63 | 0.61 |
| **RBF** | | | | | | |
| Bias (dB) | 0.16 | 0.16 | 0.16 | 0.16 | 0.16 | 0.16 |
| RMSE (dB) | 1.91 | 1.81 | 1.82 | 1.81 | 1.81 | 1.80 |

of performance is rather negligible. The performance is much more sensitive to the interpolation method (kriging or RBF). In any case, the bias is very low: the kriging

overestimates the mean noise level by $9.3{\cdot}10^{-2}$ dB, and the RBF by 0.16 dB. The size of the basis has virtually no impact on the bias since the PCA is computed on centered data, the mean value of the simulation is retrieved from the output values.

The rest of the study will be carried out with a meta-model constructed with the four first principal components vectors.

#### II.4.3.3. Performance against the covariance function for the kriging interpolator

Four classes of kernel covariance functions have been compared over the testing sample. The test has been performed with a reduced basis of size 4 and a training sample of size $2{\cdot}10^3$. The performance of the resulting meta-models is summed up in table II.6.

Table II.6. – Scores of the kriging meta-model, against the class of covariance function, with a reduced basis of size 4. The reference software simulations outputs are compared to the meta-model results over the whole complementary LHS sample, at all receivers.

| Kernel | Exponential | Matérn $\frac{3}{2}$ | Matérn $\frac{5}{2}$ | Gaussian |
|---|---|---|---|---|
| $\nu$ | $\frac{1}{2}$ | $\frac{3}{2}$ | $\frac{5}{2}$ | $+\infty$ |
| **kriging** | | | | |
| Bias (dB) | $2.1{\cdot}10^{-2}$ | $1.1{\cdot}10^{-2}$ | $9.2{\cdot}10^{-3}$ | $2.6{\cdot}10^{-3}$ |
| RMSE (dB) | 1.82 | 0.72 | 0.70 | 0.75 |

The Matérn $\frac{5}{2}$ kernel gives the best results with a global RMSE of 0.70 dB and a resulting bias of $9.2{\cdot}10^{-3}$ dB. Nonetheless, all kernels show low errors except the exponential one ($\nu = \frac{1}{2}$).

#### II.4.3.4. Performance against the basis function for the RBF interpolator

Three monomial functions have been compared over the testing sample. The test has been performed with a reduced basis of size 4 and a training sample of size $2{\cdot}10^3$. The performance of the resulting meta-models is summed up in table II.7.

Table II.7. – Scores of the RBF meta-model, against the chosen monomial function, with a reduced basis of size 4. The reference software simulations outputs are compared to the meta-model results over the whole complementary LHS sample, at all receivers.

| $\phi(r)$ | $r$ | $r^3$ | $r^5$ |
|---|---|---|---|
| **RBF** | | | |
| Bias (dB) | 0.59 | 0.16 | 0.13 |
| RMSE (dB) | 1.96 | 1.81 | 1.99 |

The $r^3$ function gives the best results with a global RMSE of 1.18 dB and a resulting bias of 0.15 dB. Nonetheless, all kernels show similar error rates higher than the kriging ones.

### II.4.3.5. Spatialized scores

Figures II.5 and II.6 show the spatial distributions of the correlation and the RMSE scores. The results show that the meta-models perform best in areas with low building density where the influence of reflections and diffractions is lower. Hence in the regions where the influence of the reflections and diffractions is higher, the fraction of acoustical energy received from a direct field is lower, which leads to a higher variability. The intensity of the noise sources seems to have little impact on the performance. The spatial representation also shows that the kriging meta-model gives better results than the RBF meta-model.


Figure II.5. – Maps of the correlation between the meta-models results and the reference noise mapping software simulation.

### II.4.3.6. Scores distribution and local evolution

The distributions, over all the simulations of the testing sample, of the scores (computed on an entire map for each simulation of the testing sample) are shown in figure II.7. It reflects the distribution of bias, correlation and RMSE over the input parameters space. The bias histogram shows that the majority of the error is centered around 0 dB and it has a very narrow error variation range. More than 90 % of the meta-models outputs with kriging have a mean spatial bias lower than 1 dB, and 95 % of them have a spatial RMSE below 1 dB. The average RMSE across all simulations is 0.51 dB.

In figure II.8, the reference noise mapping software outputs are compared with their kriging meta-model equivalent, for a single simulation of the testing sample. This simulation was chosen so that the RMSE score is similar to the average RMSE score

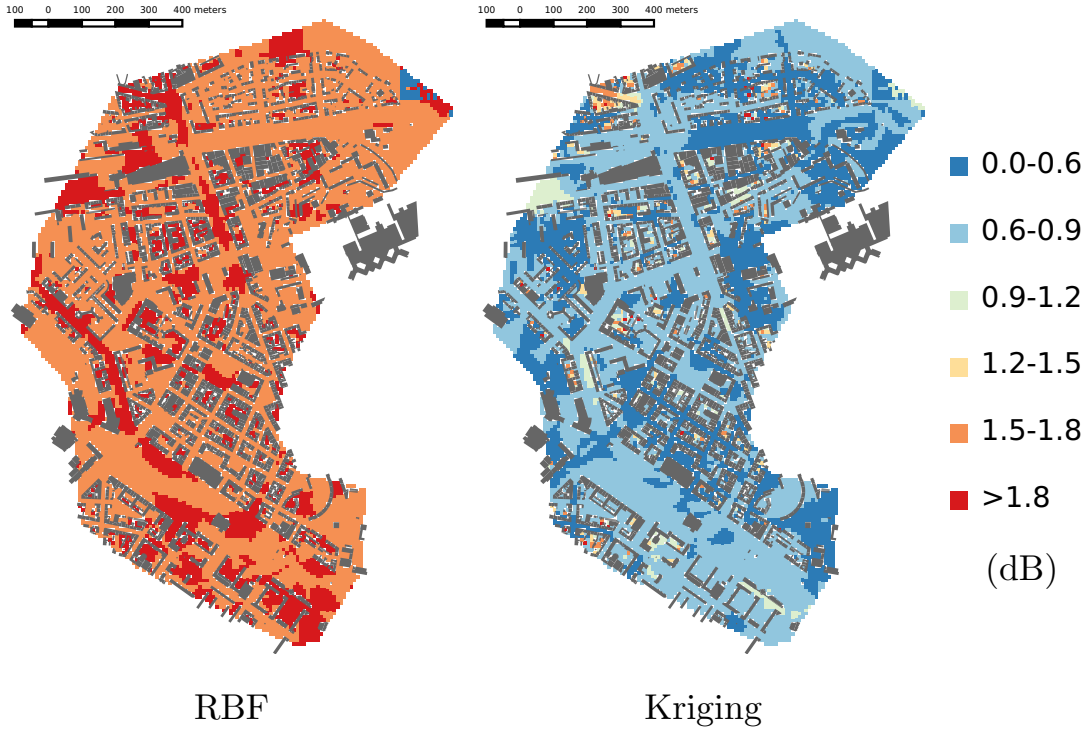


Figure II.6. – Maps of the RMSE between the meta-models results and the reference noise mapping software simulation.

over the entire testing sample. The difference between the two maps is noticeable only in a few quiet areas, as shown in figure II.9. The main patterns are preserved.

### II.4.4. Computational costs: building and applying the meta-model

The $2{\cdot}10^3$ simulations have been fully run in parallel during 2 h on 192 cores. This generated data allowed us to generate the reduced basis $\mathbf{\Psi}$ and the interpolation parameters for both RBF and kriging interpolators. The offline computation consists in the building of the meta-model (mostly the computation of the simulations, hence 2 h). The online computation consists in a call to the meta-model, which takes 35 ms with RBF and 65 ms with kriging. This corresponds to $4.6{\cdot}10^4$ and $2.5{\cdot}10^4$ speedups ratio compared to the reference noise mapping software respectively.

## II.5. Conclusions

The meta-model results remain very close to the results of the reference noise mapping software with an average RMSE of 0.96 dB. The meta-model has been carried out in three steps: generation of a training sample, dimension reduction and interpolation. The training sample has been carried out with a LHS, this sample sweeps the whole input space with a limited size, even in high-dimensional spaces. With a training sample of $2{\cdot}10^3$ simulations, satisfactory results have been obtained with respect to the bias, correlation and RMSE scores when comparing the meta-model with the reference noise

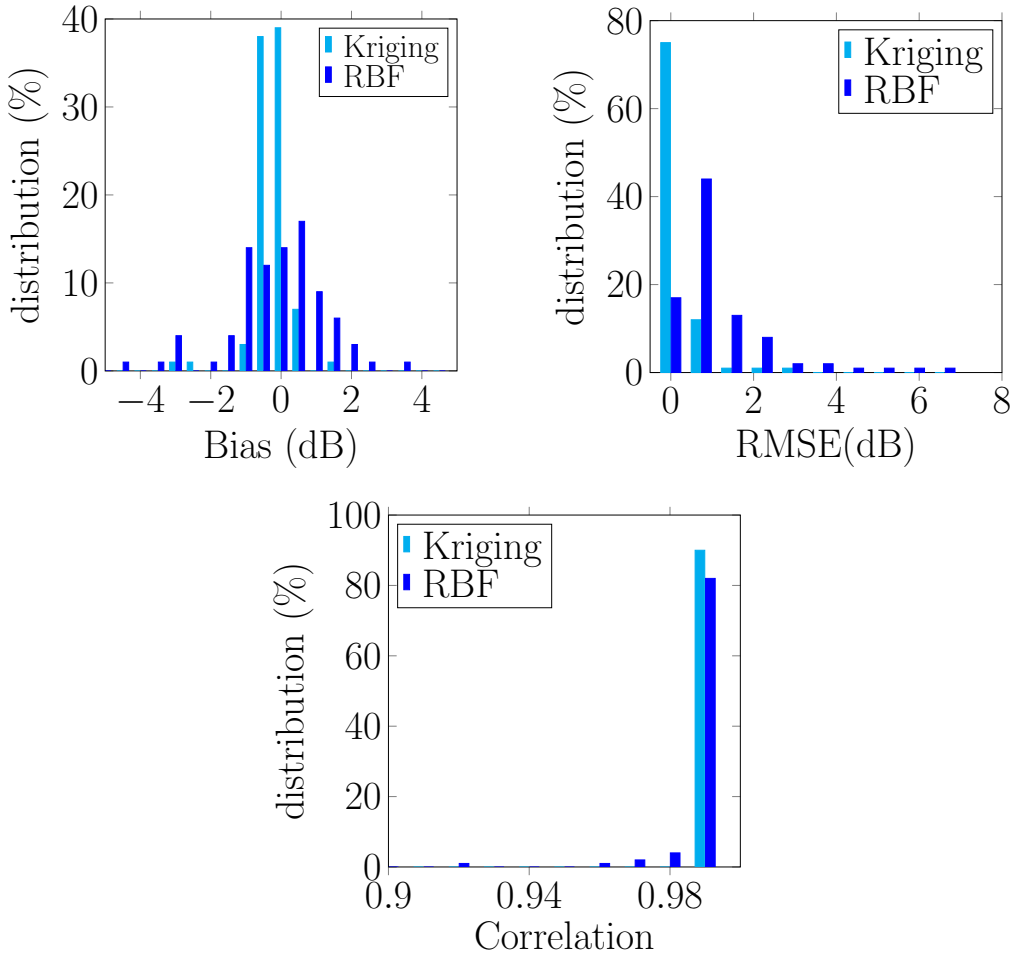


Figure II.7. – Distribution of the scores (per simulation of the testing sample) of the meta-models. RBF is in blue and kriging is in cyan.

mapping software on a testing sample. With this training data, a reduced basis of size 4 has been generated, which explains 98 % of the variance. Then, the projections of the noise maps have been interpolated on each basis vector.

The computational cost of the meta-model to generate one noise map is 65 ms with kriging and 35 ms with RBF, compared to 27 min for the reference noise mapping software, which corresponds to a speed-up of $2.5{\cdot}10^4$ to $4.6{\cdot}10^4$. The computation time saving is the main benefit of the meta-model. With an offline computation cost of $2{\cdot}10^3$ simulations and an acceptable loss of performance compared to the reference noise mapping software, a very high amount of computation time has been gained in online computations, which opens the way to many applications. In the context of uncertainty quantification, the propagation of the uncertainties in the model can require huge numbers of simulations in order to determine the sensitivity of the outputs to the parameters i.e. the variability of the ouputs regarding the variability of an input. Another useful application of the meta-model is for variance-based sensitivity analysis such as Sobol procedures (Sobol, 1990). A very large amount of calls to the simulator

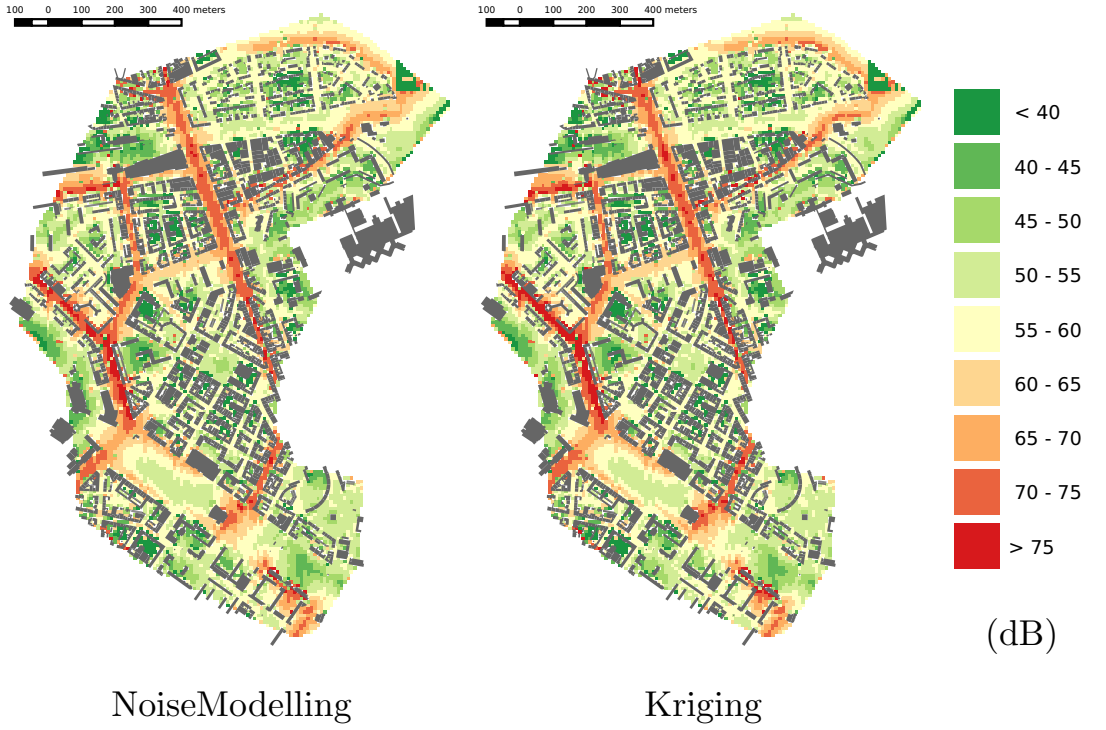


Figure II.8. – Noise Maps of the city of Lorient. The noise levels are computed by the reference noise mapping software on the left and by the kriging meta-model on the right. This simulation was chosen in the testing sample, so that the RMSE performance of the meta-model for that simulation is similar to the performance over the entire testing sample.

is required to compute estimators of such Sobol indices ($10^5$ simulations or more). In this case, a noise mapping software can advantageously be replaced by its meta-model to conduct the evaluation of the indices. In the context of data assimilation, one may need one or several maps every hour before correcting them with the observations collected for the given time period, and around $10^4$ noise maps to compute the so-called background error covariance matrix. For data assimilation processes that use inverse modeling, 40 to 50 calls to the meta-model can be required at each time step in an optimization algorithm to find the optimal input parameters, hence this process is impossible with a noise mapping software only. The performance of the meta-model against observed values from local sensors in the same area should be carried out in future work. The CENSE project, in which this study has been done, will provide these data since a network of sensors is already installed in the area considered in this paper.

The meta-model has its least conclusive results in densely built-up areas. The hypothesis formulated is the greater variability of outputs in these areas. It would be interesting to test the construction of a metamodel on a training set with more advanced calculation parameters, to test whether this improves the model's performance in areas with high building density.

Finally, further studies shall investigate other machine learning algorithms applied to reproduce the behavior of noise maps, such as random forests or neural networks,

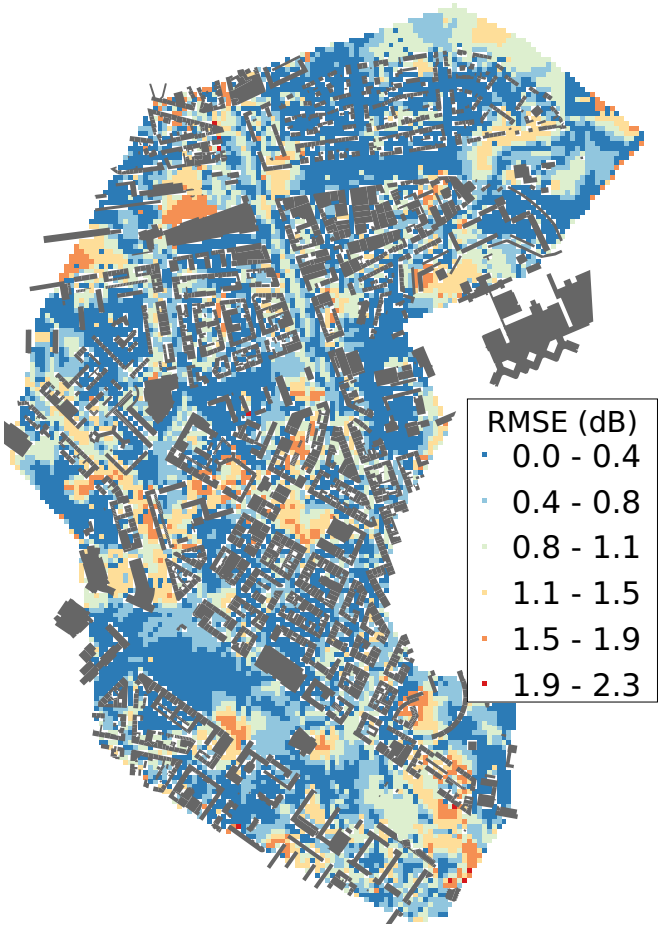


Figure II.9. – Difference between the simulated noise map and the meta-model output.

whether on a reduced subspace or directly on the computed noise maps.

A video published online [3] shows a real time application of the meta-model, where the user manually tunes the parameters of the meta-model.

## Acknowledgments

This study has been funded in the context of the ANR project CENSE (ref. ANR-16-CE22-0012). The authors woud like to thank the CENSE project team for initiating this project, the Agence Nationale de la Recherche (ANR) for their financial support, the city of Lorient for their interest in this project and Nicolas Fortin of the UMRAE department at Université Gustave Eiffel for his technical support on the NoiseModelling software. The authors would also like to express their grateful aknowledgements to Judicaël Picaut and Gráinne Duffy for their careful reading of the article which has greatly improved it.

3. See https://youtu.be/orc5ZbN2dlY

# III. Data assimilation for urban noise mapping with a meta-model

*This chapter is dedicated to the explanation of noise data assimilation principles using a metamodel. After recalling the principles of meta-modeling noise mapping simulator explained in chapter II, the construction of the input parameters based on traffic and weather observation is explained. Then the BLUE data assimilation process is explained and applied to the noise mapping process. Finally the performance of the method is tested with a leave-one-out in the results section, and shows that the observations allow to display dynamic noise maps with a low RMSE.*

## Table of contents

This chapter is a transcription of the following article: Antoine Lesieur, Vivien Mallet, Pierre Aumond, and Arnaud Can. Data assimilation for urban noise mapping with a meta-model. *Applied Acoustics*, 178:107938, 2021. ISSN 0003-682X. doi: https://doi.org/10.1016/j.apacoust.2021.107938. URL https://www.sciencedirect.com/science/article/pii/S0003682X21000311.

## Abstract

Accurately predicting dynamic noise levels in urban environments is non-trivial. This study aims to optimally combine both simulated and empirical data. Acoustic data from microphone arrays, traffic and weather data was merged with a simulated noise map, created with a statistical emulator tool (meta-model). Each hour, a noise map is generated by the meta-model with the measured traffic and weather data. This map is algorithmically merged with the measured readings to form a new composite map. The resulting analyzed map is the best linear unbiased estimator under certain assumptions. The performance is evaluated with leave-one-out cross-validation. The performance of the method depends on the accuracy of the meta-model, the input parameters of the meta-model and the structure of the error covariances between the simulated noise level errors. With 16 microphones over an area of $3\,\mathrm{km}^2$, this new method achieves a reduction of 30 % of the root-mean-square error when compared to a meta-model only.

## III.1. Introduction

A regulatory noise map is a representation of the equivalent average noise level field over an urban area. These maps are useful to evaluate scenarios for predictive applications, and are produced by many cities around the world. Within Europe, every agglomeration with more than 100 000 inhabitants is required to produce a noise map every 5 years. The standard reference for producing noise maps in the EU is CNOSSOS-EU (Kephalopoulos et al., 2012), it is currently applied following the directive 2015/996 (European Commission, 2015).

Most urban scale noise maps use models similar to CNOSSOS-EU. They proceed as follows, first, the emission algorithm is based on a discretization of the road. Each road section is described by a set of point sources whose intensities depend on the road, traffic (flow rate and speed) and weather data. Secondly, the propagation algorithm determines the sound attenuation between each point source and each receiver based on geometry and weather data, after a ray path calculation. When this method is applied at a city scale, the high number of sources and receivers makes this method computationally expensive. It is not adapted to the generation of hourly or even daily noise maps: the computation time would be too high. This approach is limited by the quality of the input data used to generate the noise map. Traffic data may only be measured on a subset of roads, which may not provide an accurate representation of wider situation.

Observation-based approaches have been developed to build maps solely from noise measurements. (Wei et al., 2016) built an interpolating model based on observations and a method to propagate the observed sound level. (Aumond et al., 2018) statistically interpolated noise measurements with a kriging approach; this approach gave satisfactory results in the vicinity of the measurement devices but suffered from the lack of sensors when estimating the noise level over a wider range.

Several attempts have been made to enrich noise maps with observational data. (Ventura et al., 2018) assimilated observations from mobile phone microphones in order to improve a pre-calculated noise map. However this process is useful only when a large number of mobile observations can be collected for this task. The DYNAMAP project

(Zambon et al., 2016) updates noise maps by scaling the noise levels of pre-calculated noise maps using the differences observed between measured and calculated original grid data for different classes of roads. (De Coensel et al., 2015) speculated about combining a noise mapping model with mobile noise measurements to increase the resolution of the noise map and fixed measurements to allow a dynamic noise mapping. These methods however purely rely on statistical methods (Least Mean Squares) and do not take into account either error propagation or traffic dynamics models. (Eerden et al., 2014) finally infers the noise mapping software inputs from noise observation by estimating the emission sources from the observed noise measures in the vicinity of the roads. This method uses a forecasting sequential approach rather than a spatial approach.

The approach proposed in this paper offers (1) to use a dynamic noise mapping, quickly generated by a meta-model (also called surrogate model) and based on real-time weather and traffic input data, and (2) to spatially enrich the noise map with noise observation data thanks to a data assimilation process which includes a simulation error model.

The objective of this paper is to develop a strategy to display accurate hourly noise maps, starting with:

— An open source noise mapping software,
— Annually averaged input data for the simulation,
— Hourly traffic and weather data from local sensors,
— A network of microphones spread across the study area.

The approach presented in this paper uses both meta-model outputs and observational data to allow a dynamic approach where 1-hour noise levels can be evaluated. This provides a view on the city noise dynamics and a finer analysis of the exposure to noise, at night and during peak hours and it is very fast to compute. This paper proposes a three-step approach:

1. Meta-modeling: the noise map produced by a classical noise model is quickly approximated using a meta-model. A meta-model is a replacement for a noise model when only a limited number of selected input parameters are allowed to change. The meta-model is trained with a sample of simulated noise maps, from which it is possible to statistically represent the relation between the selected input parameters and the corresponding noise maps.

2. Data assimilation design: the simulation error covariance matrix is computed with some training observations.

3. Data assimilation process: the simulation is corrected anywhere in space by the observation data.

A performance evaluation step will be executed with a leave-one-out cross validation method.

This paper explains in section III.2 the framework where the study lies (section III.2.1), the software and data used for the study (sections III.2.2 and IV.2), how the meta-model is built (section III.2.4), how the data assimilation process works (section III.2.5) and the method conducted to validate the process (section III.2.6). In section IV.4, the performance of the data assimilation is evaluated.

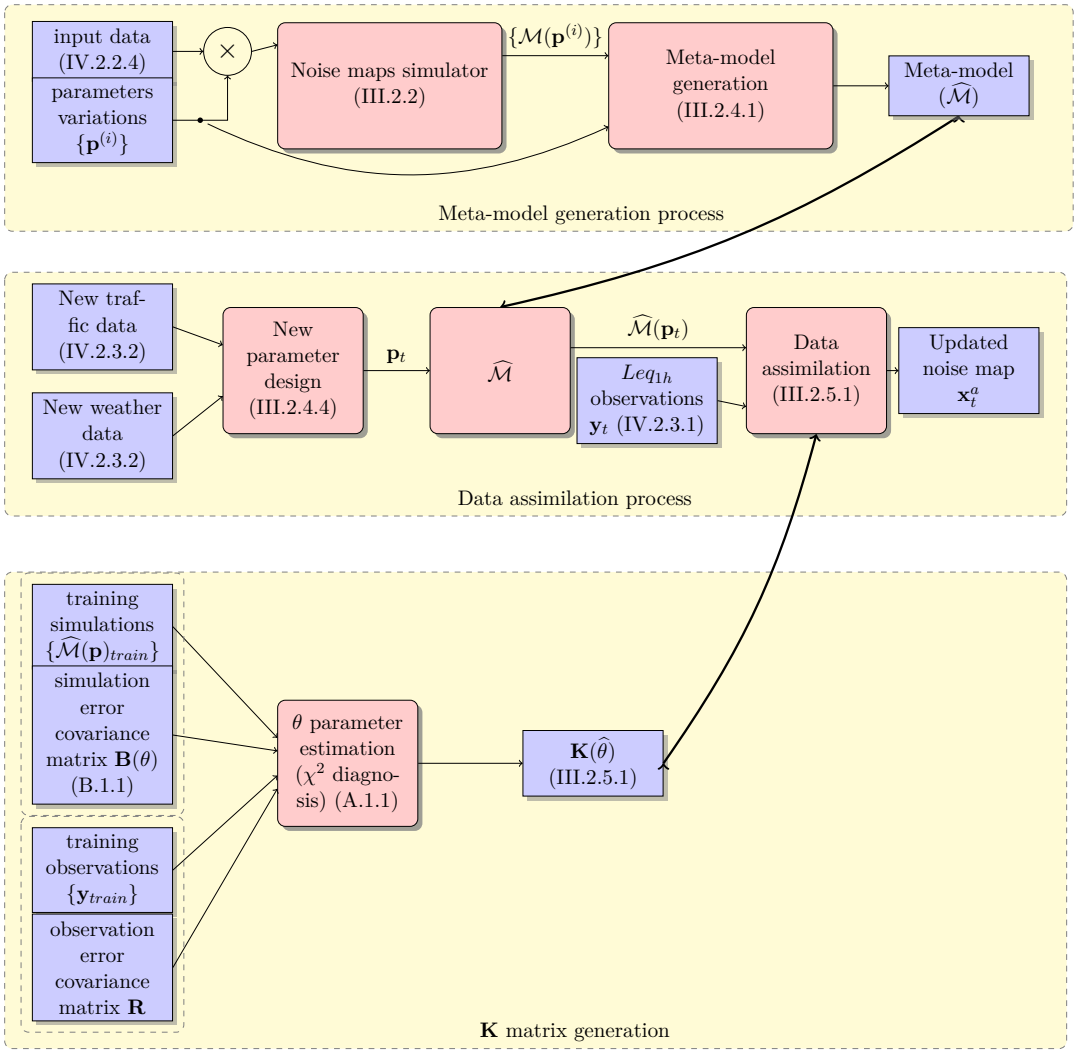


Figure III.1. – Functional diagram of the data assimilation process.

# III.2. Methods

## III.2.1. Description of the framework

The proposed approach consists in relying on a metamodel which reproduces the outputs of a reference software in a very short computing time, and hourly traffic and acoustic data, to produce hourly noise maps. This approach is inspired by a study done for air quality simulation (Tilloy et al., 2013). All the steps of the data assimilation process are summarized in figure III.1.

The approach has three steps:

— Build a meta-model from reference simulations (section III.2.4);
— Build a gain matrix $\mathbf{K}$ used in the data assimilation algorithm (appendix A);
— Build a data assimilation algorithm which corrects the results based on a error covariance matrix (section III.2.5).

## III.2.2. Reference noise mapping software

### III.2.2.1. Presentation

The simulator used in this study is NoiseModelling v3.0 (Bocher et al., 2019), an open source software designed to produce noise maps for evaluating the noise impact on urban mobility plans. Its input are traffic data, topographic data (buildings distribution, altitude, surface types, ...) and meteorological data (temperature, hygrometry, probability of occurrence of favorable atmospheric conditions ...). The version used in this paper is especially designed to run multiple simulations (Aumond et al., 2020a). A first call to the simulator computes a matrix of rays between the source/receiver pairs, which is the longest step for a noise mapping software. Each call to the reference noise mapping software with a new set of parameters is then much faster since it only requires to compute the attenuation along each pre-computed path.

The input traffic data along each road is discretized into a set of point noise sources, an emission value is assigned which depends on the speed, the flow rate and other traffic parameters (type of road, type of vehicle, ...). The output noise level field is represented by a regular grid of receivers. NoiseModelling uses a path finding method similar to the one proposed by CNOSSOS-EU which computes the path taken by the acoustic waves between the sources and the receivers, taking into account the propagation distance, the number of reflections and whether the diffractions are taken into account or not. The software then computes the attenuation along each path for a selected range of octave bands (63 Hz, 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz and 8 kHz) in accordance with the CNOSSOS-EU method. The sound level at a receiver is the sum of the contributions of each source. The final output is a grid of receivers with an attributed sound level value computed in dB(A).

### III.2.2.2. Characteristics of the simulation

The simulated area is an area located in the 13th district of Paris, France. Its surface is about $3\,\mathrm{km}^2$. The receivers, where noise levels will be calculated, are located on a regular grid of step $\Delta x = 15\,\mathrm{m}$. Receivers have also been located at the location of the observation points. The total number of receivers is $n = 8456$. All these receivers are positioned at a fixed height of 4 m. Note that the actual height of the microphones does not always match the height of the simulation. The only available value for the height of the microphone is the floor level where they were positioned, which is too vague to be precisely set as an input value. The propagation parameters are: 250 m for the propagation distance, 100 m for the distance after reflexion, one order of reflection and one order of diffraction.

### III.2.2.3. Standard noise map using annually averaged data

The nominal $L_{day}$ field distribution given by the reference noise mapping software with this input data is shown in figure IV.2. The nominal $L_{evening}$ and $L_{night}$ have also been computed in the study.


Figure III.2. – $L_{day}$ level distribution over the study area

## III.2.3. Observation data

### III.2.3.1. Noise data

This study uses data collected in the 13th district of Paris, France during a measurement campaign conducted in 2015 by (Lavandier et al., 2016). A network of 25 microphones has been deployed across the study area. 16 of them have been selected for the present study, the ones that had simultaneously an exploitable set of data during the entire time period. They have recorded the noise level $L_{f_i}$ in dB for every one-third octave $f_i$ from 63 Hz to 20 kHz with a time step of $\Delta t = 125$ ms, from January 1, 2015 to September 30, 2015. The indicator used in this study is $LAeq_{1h}$

The spatial distribution of the sensors is shown in figure IV.1.

### III.2.3.2. Annually averaged input data

The input files of the reference noise mapping software are the buildings distribution and the annually averaged road traffic map during the daytime (between 06:00 and 18:00, local time), the evening (between 18:00 and 22:00) and the night (between 22:00 and 06:00) computed for the year 2014. These data are respectively given by the IGN , the French national geographical institute, and the city of Paris. Data like ground surface or altitude have been ignored in this study.

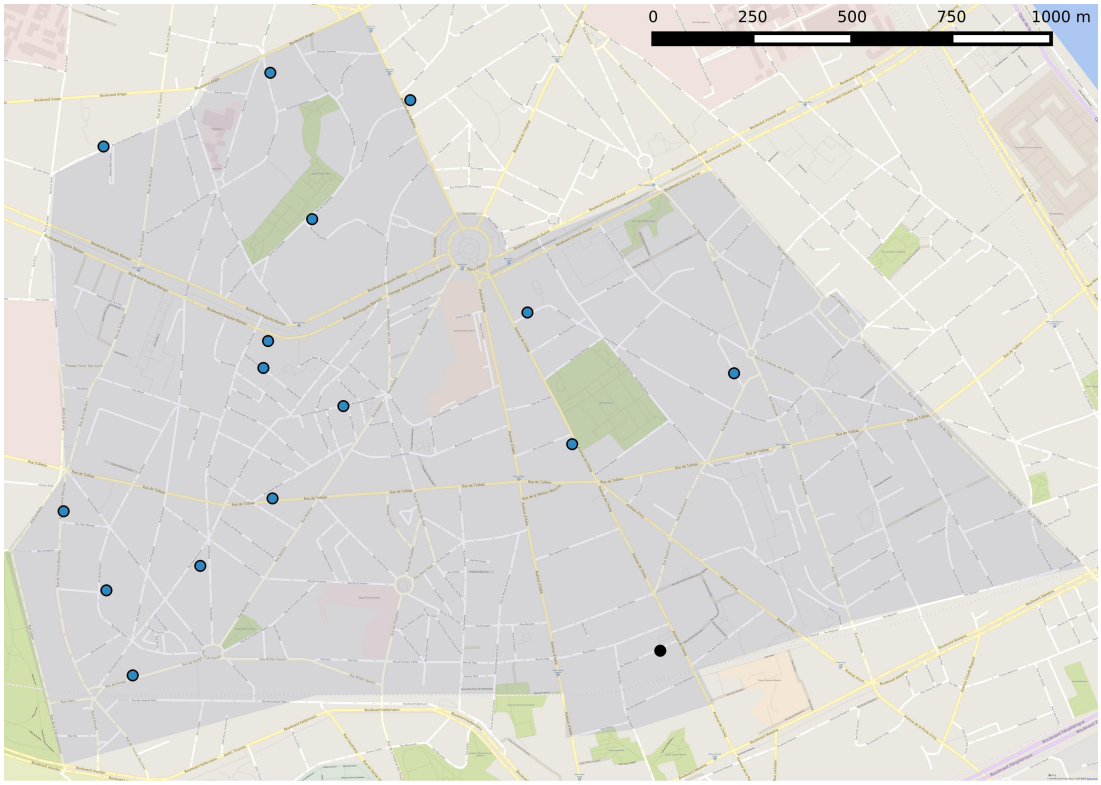


Figure III.3. – Distribution of the sensors across the study area, the observation point colored in black is the one which is displayed in figure III.4

#### III.2.3.3. Real-time 1 h traffic and weather data

The real-time 1 h traffic data on the main roadways of Paris during the case study time period is available online on the city of Paris open data platform. For every measured track section, at every hour, the hourly average flow $F(t)$ (vehicles/h) and the occupation rate $R(t)$ (ratio of time during which the sensors detect an occupation) are given. Given the track section length $\ell_t$ and assuming an average vehicle length of $\ell_v = 4\,\mathrm{m}$, it is possible to estimate the average vehicle speed $S(t)$ on the track section in $\mathrm{km\,h^{-1}}$:

$$S(t) = 3.6 \cdot \frac{F(t) \cdot \ell_v}{R(t) \cdot \ell_t} \tag{III.1}$$

The real-time weather data has been taken from an open access source at the Montsouris weather station near the study area. The weather station updates its data every hour in real time. The reference noise mapping software only takes into account the temperature but upcoming versions might also take into account the wind intensity and direction, the temperature gradient, and more generally, the probability of occurrence of favourable atmospheric conditions.

## III.2.4. Meta-model

This section introduces the tools built during the study and their justification: the meta-model and the design of its parameters (III.2.4.4).

### III.2.4.1. Construction of the meta-model

When a noise map is computed with annually averaged input data, one may wonder how this noise map changes with a slight modification of the parameters, e.g, a multiplicative coefficient applied to the traffic flow or a temperature variation in °C.

$\mathbf{p} \in \mathbb{R}^k$ denotes the input vector of parameters and $\mathcal{M}(\mathbf{p}) \in \mathbb{R}^n$ is the output of the reference noise mapping software. In order to speed up the computation of the hourly maps, a meta-model $\widehat{\mathcal{M}}(\mathbf{p})$ which is very similar to $\mathcal{M}(\mathbf{p})$ but with a computational time much lower (in this case, $\sim 100\,\text{ms}$ versus $\sim 1\,\text{h}$) is built. $\widehat{\mathcal{M}}$ is generated using a training sample of size $r = 2{\cdot}10^3$, $\{\mathbf{p}^{(1)}, \cdots, \mathbf{p}^{(r)}\}$ and the corresponding simulations $\{\mathcal{M}(\mathbf{p}^{(1)}), \cdots, \mathcal{M}(\mathbf{p}^{(r)})\}$. $\widehat{\mathcal{M}}$ is a weighted sum of a small number of maps obtained with a dimension reduction algorithm described in section III.2.4.2, the weights are computed by a kriging interpolation function based on the weights of the training set described in section III.2.4.3.

### III.2.4.2. Dimension reduction

A PCA (Principal Components Analysis) (Jolliffe, 1986) algorithm determines which of the principal components explain most of the variance for the training sample. The structure of the noise mapping behavior allows us to explain more than 97 % of the variance with a short sequence of vectors $\{\mathbf{map}_i\}_{i\in[\![1,k]\!]}$ ($k \simeq 3$) as previously shown in (Lesieur et al., 2020):

$$\mathcal{M}(\mathbf{p}) = \bar{\mathbf{x}}_b + \sum_{i=1}^{k} w_i(\mathbf{p}) \times \boldsymbol{map}_i + \boldsymbol{\epsilon}, \tag{III.2}$$

where $\bar{\mathbf{x}}$ is the average of the training set outputs:

$$\bar{\mathbf{x}}_b = \frac{1}{r} \sum_{i=1}^{r} \mathcal{M}(\mathbf{p}^{(i)}), \tag{III.3}$$

$w_i(\mathbf{p})$ is the projection coefficient of the centered $\mathcal{M}(\mathbf{p})$ on $\boldsymbol{map}_i$,
$\boldsymbol{map}_i^T(\mathcal{M}(\mathbf{p}) - \bar{\mathbf{x}}_b)$, and $\boldsymbol{\epsilon}$ is the residual error.

### III.2.4.3. Kriging

The reader might be familiar with kriging in its classical geostatistical background as it has been applied in some studies (Baume et al., 2009). However, contrary to the usual kriging employed in several acoustics applications using geographical data, this paper applies kriging to the abstract vector space of the input parameters which has no relation with the geographical 2D space.

The goal of the second step is to obtain for each map $i$ selected in section III.2.4.2 a fast approximation of the weight function:

$$\begin{array}{rccc} w_i & : & \mathbb{R}^k & \rightarrow & \mathbb{R} \\ & & \mathbf{p} & \mapsto & w_i(\mathbf{p}) \end{array} \tag{III.4}$$

where $w_i(\mathbf{p}^{(j)})$ is known for every $\mathbf{p}^{(j)}$ in the training sample. Kriging Matheron (1962) is a technique which allows to determine an approximation $\widehat{w}_i(\mathbf{p})$ of $w_i(\mathbf{p})$ by interpolating from the known values $w_i(\mathbf{p}^{(j)})$.

Once the kriging interpolators are computed for every projection vector, the meta-model $\widehat{\mathcal{M}}$ is defined as follows:

$$\widehat{\mathcal{M}}(\mathbf{p}) = \bar{\mathbf{x}}_b + \sum_{i=1}^{k} \widehat{w}_i(\mathbf{p}) \times \boldsymbol{map}_i \tag{III.5}$$

$\widehat{\mathcal{M}}(\mathbf{p})$ is a short sequence of elementary operations, hence its computation is much lower than the reference noise mapping software and then allows:

— A fast real-time computation of a noise map with the real-time input data;

— The generation of the statistical error covariance matrix $\mathbf{B}_{stat}$ with a Monte Carlo method, which will be used in data assimilation (See III.2.5.2).

### III.2.4.4. Design of the parameters

The parameters selected as inputs of the meta-model are shown in Table III.1 with their respective amplitudes. For each input, the amplitude corresponds to the range of values observed in real-time.

Table III.1. – Input parameters and their input ranges for the meta-model. They are all dimensionless multiplicative coefficients applied to the reference values (the annually averaged daily values) except the temperature which is the actual value in °C.

| Parameter | Minimum value | Maximum value |
|---|---|---|
| Light vehicle speed | 0.1 | 2 |
| Heavy vehicle speed | 0.1 | 2 |
| Light vehicle flow | 0.1 | 2 |
| Heavy vehicle flow | 0.1 | 2 |
| Temperature (°C) | -5 | 30 |

The goal of this section is to infer a value for the input coefficient from the real-time traffic data. As seen in the beginning of section III.2.4.1, the input which represents traffic variations in the meta-model is a multiplicative coefficient applied to the annually averaged traffic data. It therefore models the variation in time of the traffic, and assumes a constant spatial distribution. Regarding the flow rate and the speed, the input data $p_i$ is the ratio between the average quantity on the observed roads at a given time and the average quantity on the observed roads over one year.

Since the observed traffic data do not discriminate the heavy vehicles from the light vehicles, the same coefficient is applied to the heavy and light vehicle input — in other words, the fraction of heavy vehicle is assumed to be constant.

Figure III.4 compares the time evolution of the measured noise level at an observation point and the corresponding output value of the meta-model with the input parameters built from the traffic and weather data. It shows that the meta-model fairly reproduces the hourly noise level distribution.

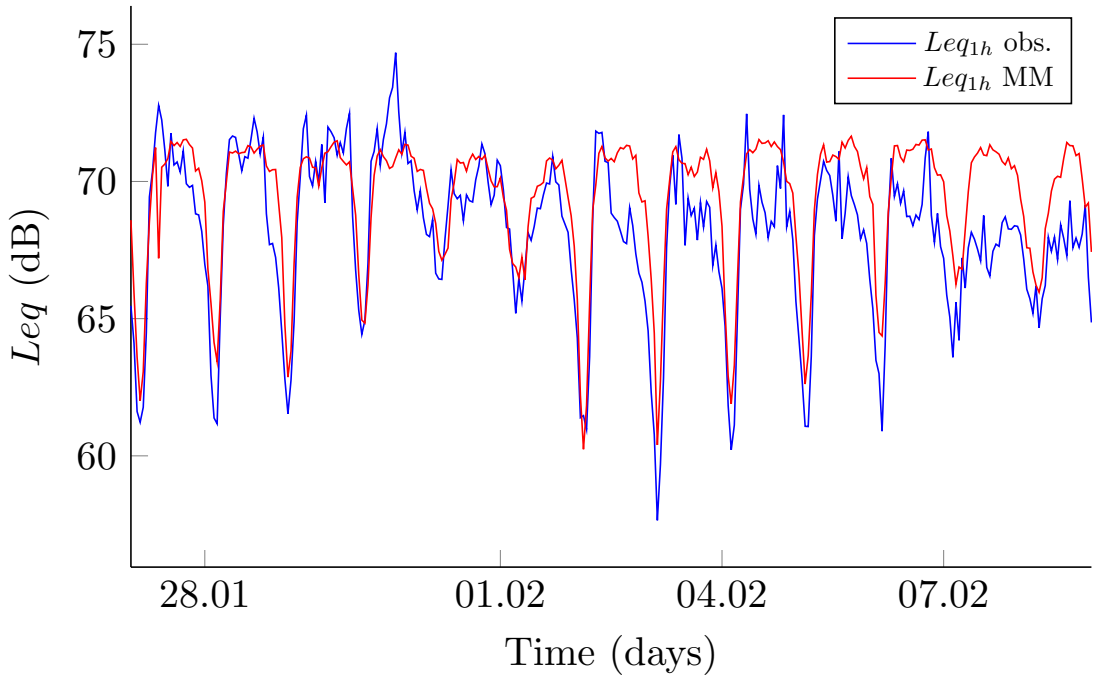


Figure III.4. – time evolution of the observed $LA_{eq,1h}$ and the meta-model output for the sensor point colored in black in figure IV.1

### III.2.5. Data assimilation

#### III.2.5.1. Presentation

The data assimilation algorithm which will be used in this paper is called BLUE, for Best Linear Unbiased Estimator (Bouttier and Courtier, 2002). It makes use of the output of the meta-model and applies a correction which depends on $\mathbf{B}$, the covariance matrix of the simulation error, $\mathbf{R}$, the covariance matrix of the observational error, and the discrepancy between observations and simulations at the observation location. Matrices $\mathbf{B}$ and $\mathbf{R}$ characterize the spatial behavior of the model error and the observation error respectively.

At each timestep, a nominal noise map $\mathbf{x}^b = \widehat{\mathcal{M}}(\mathbf{p})$ is computed. It is an estimation of the real, exact state noise level distribution field $\mathbf{x}^t$. It is impossible to know the exact vector $\mathbf{x}^t$, either from the erroneous simulation or the incomplete (and slightly erroneous) observations. It would be the noise level field observed if each receiver had an ideal microphone on it. The set of observations is denoted $\mathbf{y} \in \mathbb{R}^p$. Let:

— $\mathbf{H} \in \mathbb{R}^{p \times n}$ be the observation matrix which maps a state vector, e.g., $\mathbf{x}^t$, to the observed values $\mathbf{y}$, so that $\mathbf{y}$ can be compared to $\mathbf{H}\mathbf{x}^t$. In this case, each observation point has a simulated counterpart, $\mathbf{H}$ is therefore a matrix in which

each row is full of zeros except for a one in the column that corresponds to the receiver located at the observation point. The elements $[\mathbf{H}]_{ij}$ are equal to one if and only if the j$^{th}$ receiver corresponds to the i$^{th}$ observation point;
— $\mathbf{e}^b \in \mathbb{R}^n$ be the simulation error $\mathbf{x}^b - \mathbf{x}^t$, a random vector, which is assumed to be unbiased $\mathbb{E}[\mathbf{e}^b] = 0$;
— $\mathbf{e}^o \in \mathbb{R}^p$ be the observation error $\mathbf{y} - \mathbf{H}\mathbf{x}^t$, a random vector, which is assumed to be unbiased $\mathbb{E}[\mathbf{e}^o] = 0$;
— $\mathbf{B} = \mathbb{E}[\mathbf{e}^b \mathbf{e}^{bT}] \in \mathbb{R}^{n \times n}$ be the covariance matrix of the simulation error,
— $\mathbf{R} = \mathbb{E}[\mathbf{e}^o \mathbf{e}^{oT}] \in \mathbb{R}^{p \times p}$ be the covariance matrix of the observation error.

Based on $\mathbf{x}^b, \mathbf{B}, \mathbf{y}, \mathbf{R}$ and $\mathbf{H}$, an analysis state vector $\mathbf{x}^a$ is computed as the "Best Linear Unbiased Estimator" which is linearly dependent on $\mathbf{x}^b$ and $\mathbf{y}$, has unbiased error $\mathbf{e}^a = \mathbf{x}^a - \mathbf{x}^t$, and has an error variance with minimum trace (Bouttier and Courtier, 2002). This estimator is uniquely defined as follows, under the assumption that the simulation and observation errors are uncorrelated:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b) \tag{III.6}$$

where

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{III.7}$$

$\mathbf{K}$ is called the gain matrix or the Kalman matrix. The observation errors at two different receivers are not correlated. The observation error essentially depends on the accuracy of the microphone which is described by its standard deviation $\sigma_r$, therefore:

$$\mathbf{R} = \sigma_r^2 \mathbf{I}_p \tag{III.8}$$

with $\mathbf{I}_p$ the identity matrix of dimension $p$. In the rest of the study, $\sigma_r$ =1 dB$^2$, which corresponds to the industrial accuracy of the sensor.

#### III.2.5.2. Computation of the simulation error covariance matrix B

A good estimation of the covariance matrix $\mathbf{B}$ is crucial to perform an efficient data assimilation. There are two approaches to compute the simulation error covariance matrix: analytical and statistical. On the one hand, the analytical approach gives a covariance formula based on the assumption that where there is an error on the noise estimation at a given location, the traffic estimation is subject to similar errors in the surrounding streets. On the other hand, the statistical approach generates an error covariance matrix between the meta-model output and the ideal meta-model output, with a large random selection of noise maps made possible by the short computation time of the meta-model in a Monte Carlo algorithm.

**Analytical covariance matrix $\mathbf{B}_{ana}$** A parametrized simulation error covariance function has been successfully used in a previous work (Ventura et al., 2018). The simulation error cross correlations between two receivers $i$ and $j$ are expressed in terms of:
— $d_{ij}$, the length of the shortest path along the road network between the projection of receivers $i$ and $j$ on the closest road computed with a Johnson algorithm (Johnson, 1977) and shown for a specific observation point in figure III.5,

— the nominal noise field itself, through a correlation function $\rho(x_i, x_j)$, where $x_i$ and $x_j$ are the values of $\mathbf{x}^b$ at points $i$ and $j$ as proposed by (Riishøjgaard, 2002). The corresponding simulation error covariance between points $i$ and $j$ reads:

$$[\mathbf{B}_{ana}]_{ij} = \sigma_b^2 \exp\left(-\frac{d_{ij}}{L_d}\right) \exp\left(-\frac{|x_i - x_j|}{L_b}\right) \tag{III.9}$$

where $\sigma_b^2$ is the characteristic variance of each point, $L_d$ is a characteristic distance in m and $L_b$ is a characteristic noise level value in dB.

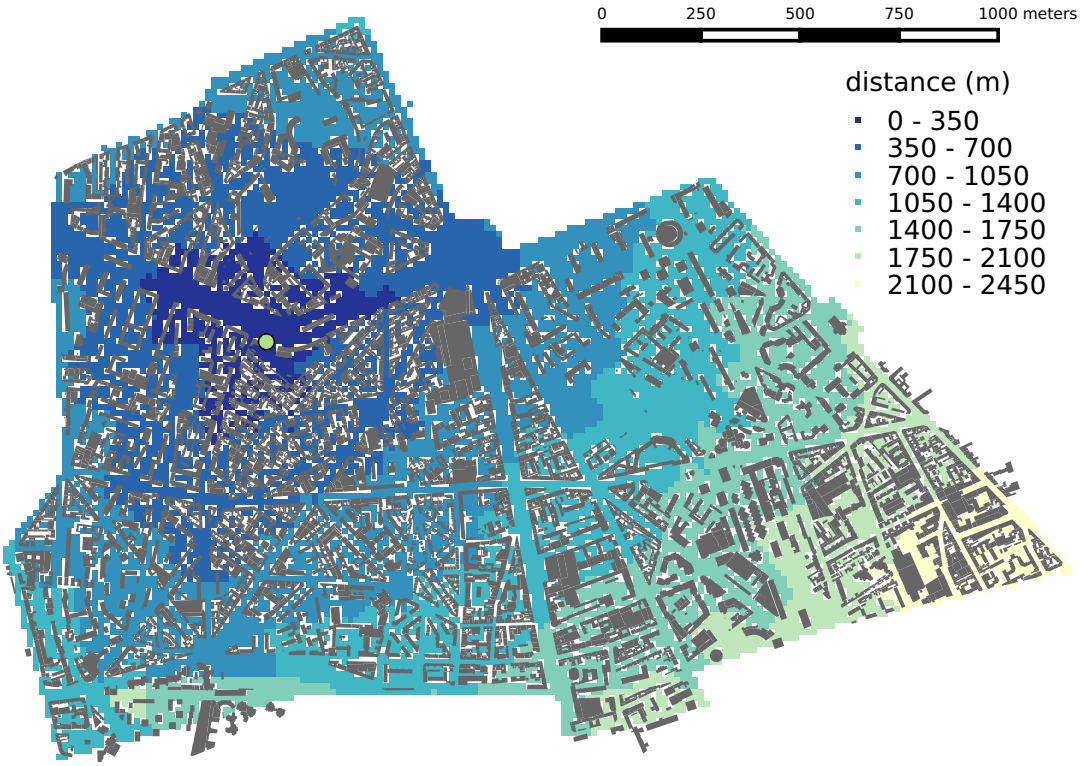


Figure III.5. – This map displays the distance along the road network between the highlighted point and the other points of the simulated map

Figure III.6 shows the covariance between the same specific point and the other points, using the analytic covariance function with $\sigma_b^2 = 67\,\mathrm{dB}^2(\mathrm{A})$, $L_d = 500\,\mathrm{m}$ and $L_b = 6\,\mathrm{dB}$ ($A = 1.03$). These values were validated by a method called $\chi^2$ diagnosis which checks whether the observations are consistent with the formulation of their error covariance (see appendix A.1 for further details). These values are consistent with other studies that computed the correlation length of the noise error distribution such as (Aumond et al., 2018). It points out that the correlation is mostly significant for the points located in the neighborhood of the selected point.

**Statistical covariance matrix $\mathbf{B}_{stat}$** The meta-model allows to run a large number of simulations. It is possible to compute the error covariance matrix between the simulated noise map and the ideal meta-model output, within the subspace $\{\boldsymbol{map}_i\}_i$

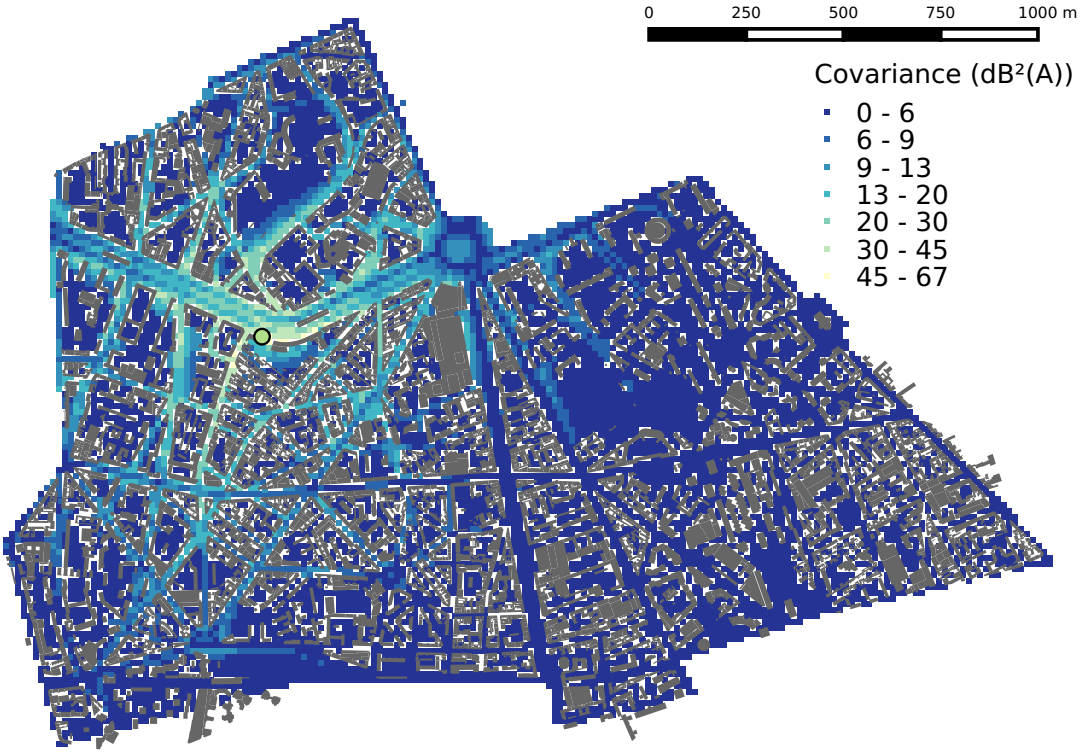


Figure III.6. – This map displays the analytical covariance between the highlighted point and the other points of the simulated map

of the meta-model $\mathbf{B}_{stat}$ with a Monte Carlo algorithm (Fishman, 1996), sampling the uncertainties. Let $\{\mathbf{p}^{(1)}, \cdots, \mathbf{p}^{(N)}\}$ be $N$ random draws of a uniform distribution over the input interval (in this case study, $N = 3{\cdot}10^4$). Then, with:

$$\bar{\mathbf{x}}^b = \frac{1}{N} \sum_{k=1}^{N} \widehat{\mathcal{M}}(\mathbf{p}^{(k)}) \tag{III.10}$$

$\mathbf{B}_{stat}$ is defined by:

$$\mathbf{B}_{stat} = \frac{1}{N-1} \sum_{k=1}^{N} \left(\widehat{\mathcal{M}}(\mathbf{p}^{(k)}) - \bar{\mathbf{x}}^b\right) \left(\widehat{\mathcal{M}}(\mathbf{p}^{(k)}) - \bar{\mathbf{x}}^b\right)^T \tag{III.11}$$

Figure III.7 shows, for the same point as in figure III.6, the covariance function computed with the Monte Carlo algorithm. Since the dimension of the subspace is very low, it has a much more global distribution, the low ranking implies that an error covariance at some receiver is likely to be correlated with an error covariance of any receiver elsewhere in the study area. The point is not only correlated to its neighborhood but with all the points of the map located near a roadway.

**Sum of the analytical and statistical covariance matrices** The covariance matrix $\mathbf{B}_{ana}$ built as in section III.2.5.2 has shown good results in studies where the data

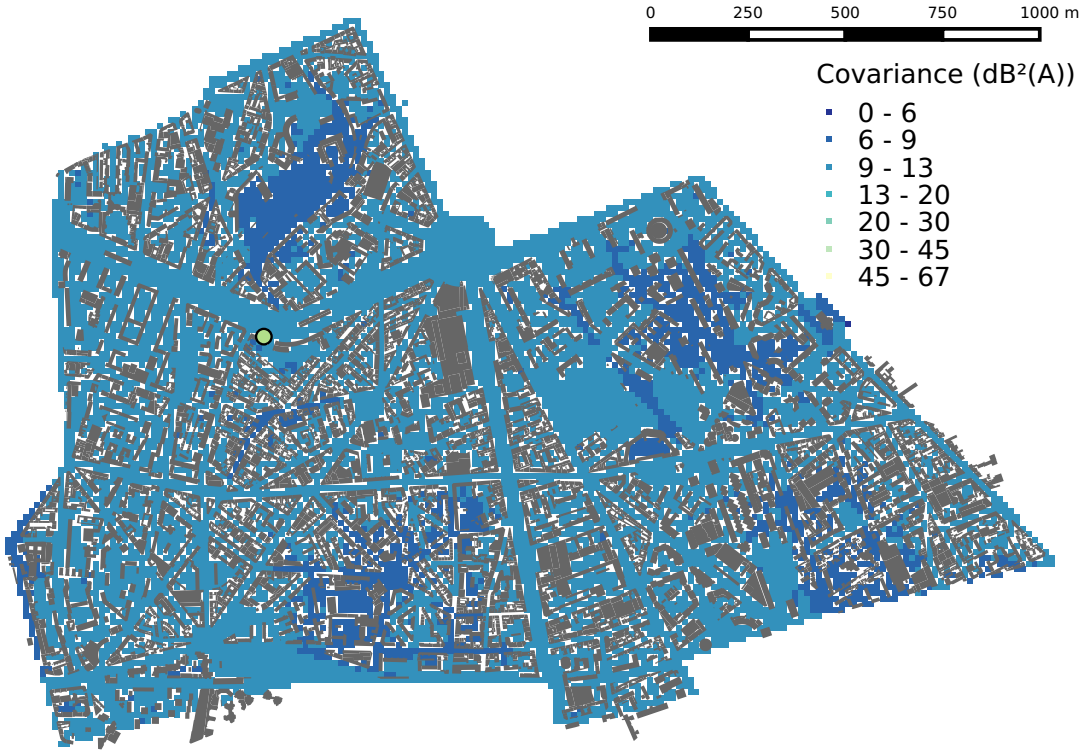


Figure III.7. – This map displays the statistical covariance of each point with the highlighted point

assimilation is computed on one fixed nominal noise map (Ventura et al., 2018). On the other hand, the statistical matrix $\mathbf{B}_{stat}$ cannot be used alone for data assimilation since it only applies on the reduced subspace and thus cannot be used to compute the covariance of $\mathbf{x}^b - \mathbf{x}^t$. In the appendix A.2, the authors have shown that a good error covariance matrix for a data assimilation process, performed on a noise map generated with a meta-model, is the sum of the two previous covariance matrices: $\mathbf{B}_{ana} + \mathbf{B}_{stat}$.

### III.2.6. Validation method

The leave-one-out cross-validation consists in removing the observations of a given microphone from the data assimilation process. Only the observations from the other microphones are used to correct the noise level map. This procedure is carried out for all microphones, one by one — only one microphone is removed at a time. At the location of the removed microphone, the meta-model performance is compared to the performance after assimilation of the observations of the other microphones. This allows to check whether the assimilation properly propagates in space the corrections that originate from the observed locations. The cross-validation evaluates the effects of the data assimilation at locations without any observations.

The bias and RMSE indicators defined in table IV.3 are used to quantify the performance of the simulations for the $LAeq_{1h}$.

Table III.2. – Scores for the performance evaluation of a model. $(c_i)_i$ is the simulated sample. $(o_i)_i$ is the corresponding observed sample. $n$ is the total number of elements in the sample. $\bar{c}$ and $\bar{o}$ are respectively the mean of $(c_i)_i$ and $(o_i)_i$. RMSE stands for Root Mean Square Error.

| Score | Formula |
|---|---|
| Bias | $\frac{1}{n}\sum_{i=1}^{n}(c_i - o_i)$ |
| RMSE | $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(c_i - o_i)^2}$ |

## III.3. Results

### III.3.1. Cross validation

#### III.3.1.1. Dispersion

Figure IV.3 shows the error dispersion for all the indicators at every step of the data assimilation process (meta-modeling, data assimilation with $\mathbf{B} = \mathbf{B}_{ana}$, data assimilation with $\mathbf{B} = \mathbf{B}_{ana} + \mathbf{B}_{stat}$). The distribution gets narrower as the precision of the data assimilation process increases. This evolution is also presented in table IV.4 where the evolution of the absolute bias is showed at each level of correction.

Table III.3. – Dispersion in dB(A) of the error between the observation and the simulation or the analysis in leave-on-out cross-validation for all the microphones during all the measurement campaign

| **error amplitude** (dB) | $[0,1]$ | $[1,3]$ | $[3,5]$ | $>5$ |
|---|---|---|---|---|
| meta-model | 18 % | 33 % | 26 % | 22 % |
| assimilation with $\mathbf{B} = \mathbf{B}_{ana}$ | 26 % | 41 % | 22 % | 10 % |
| assimilation with $\mathbf{B} = \mathbf{B}_{ana} + \mathbf{B}_{stat}$ | 30 % | 44 % | 18 % | 8 % |

In addition, Figure III.9 shows the time evolution of the observed and assimilated sound levels at the same reception point and during the same time period as in figure III.4. The temporal evolution of sound levels is finely reproduced by the data assimilation process.

#### III.3.1.2. RMSE

The root mean squared error of the $LA_{eq,1h}$ is displayed for each month in figure IV.4 and for each hour of the day in figure III.11 at every step of the data assimilation process. The reduction of the error covariance is noticeable. At each step, the RMSE is reduced and goes down to an average of 2.9 dB(A) at the final stage. The assimilation process mostly improves the estimation during the night time when the noise level is less correlated to the traffic flow rate.

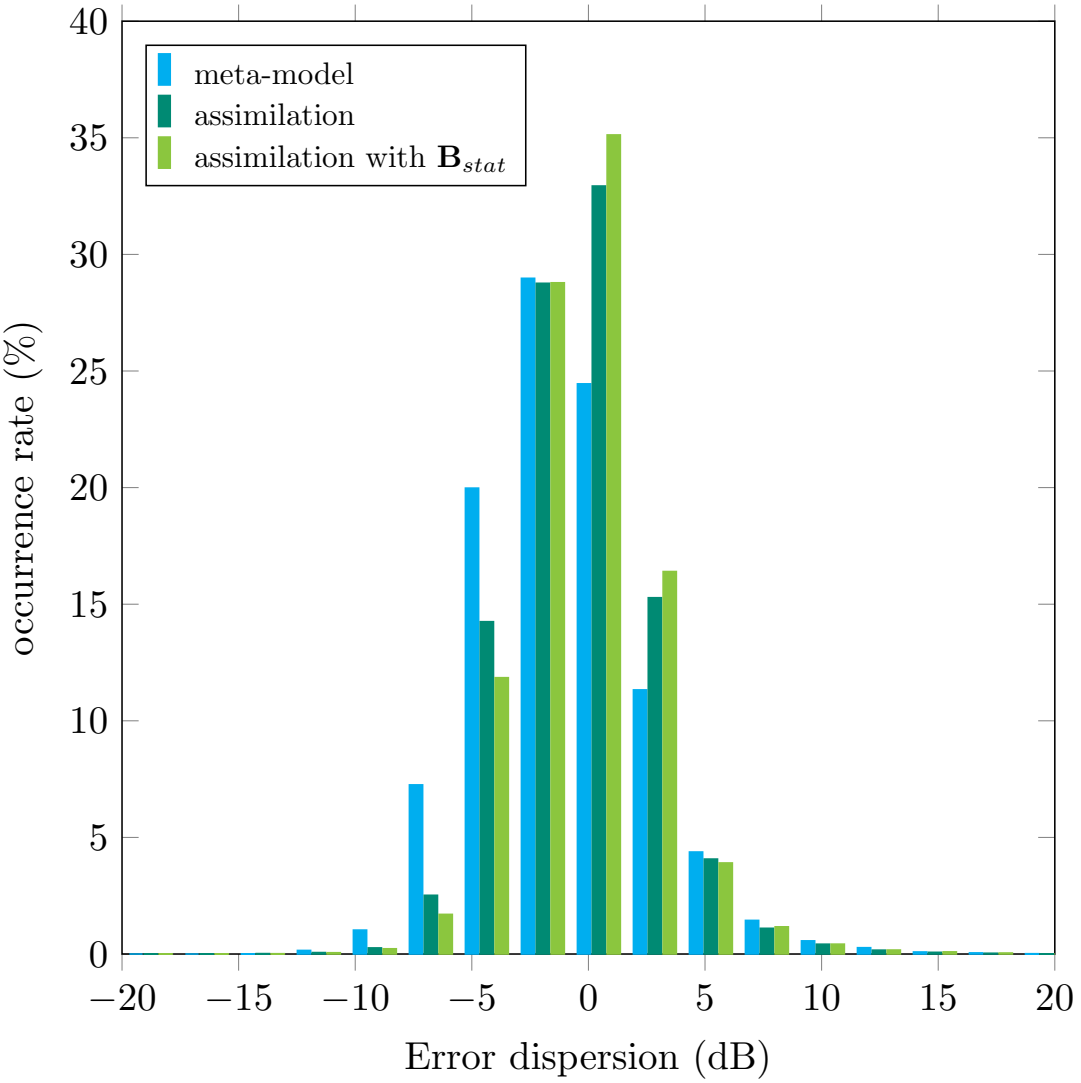


Figure III.8. – Dispersion in dB(A) of the error between the observation and the simulation (blue) or the assimilation in leave-on-out cross-validation (green) for all the microphones during all the measurement campaign.

As expected in section III.2.5.2, the matrix $\mathbf{B} = \mathbf{B}_{ana} + \mathbf{B}_{stat}$ better characterize the error covariance behavior and gives better results than $\mathbf{B}_{ana}$ alone for both dispersion and RMSE.

#### III.3.1.3. Distance to the network

This section proposes to connect the individual performance of the receiver with its position in the road network. Table III.4 shows the performance of the data assimilation process at each receiver.

Figure III.12 shows the RMSE reduction percentage $\rho$ (equation (III.12)) in the leave-one-out cross validation context with respect to the distance of the observation point

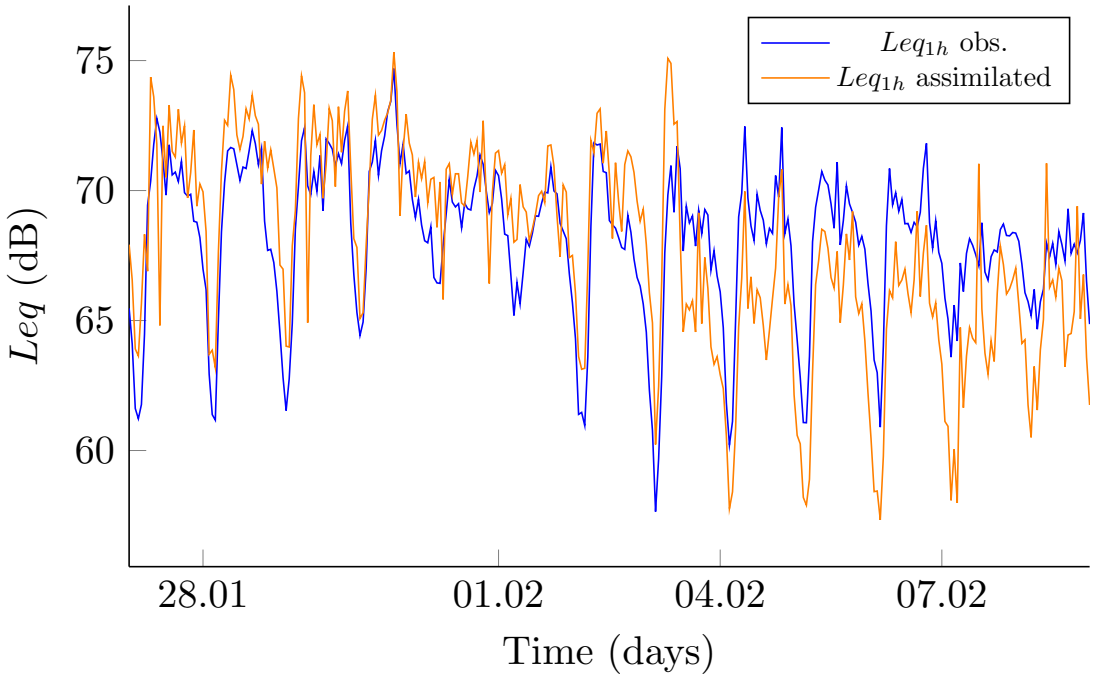


Figure III.9. – time evolution of the observed $LA_{eq,1h}$ and the leave-one-out data assimilation process output for the sensor point colored in black in figure IV.1

Table III.4. – Performance of the noise map and computed with the meta-model and the assimilation regarding the observations at each microphone point, the values are given in dB(A)

| Node | Bias meta-model | Bias assimilation | RMSE |
|---|---|---|---|
| 251 | -2.19 | -0.16 | 1.97 |
| 252 | -3.41 | -1.79 | 3.72 |
| 254 | -0.12 | 0.13 | 2.20 |
| 255 | 0.88 | 0.92 | 2.65 |
| 256 | 0.42 | 0.86 | 2.28 |
| 259 | 1.72 | 1.42 | 3.71 |
| 261 | -2.65 | -1.30 | 3.07 |
| 262 | -0.37 | 0.67 | 2.07 |
| 264 | 3.00 | 0.93 | 3.17 |
| 265 | 1.71 | 0.72 | 3.22 |
| 266 | -2.02 | -0.59 | 1.75 |
| 268 | -5.22 | -2.31 | 4.08 |
| 270 | -2.93 | -0.46 | 1.87 |
| 272 | -2.81 | -1.04 | 2.62 |
| 274 | -4.35 | -1.60 | 2.90 |
| 275 | -0.35 | -0.05 | 3.25 |

to the network (i.e., the distance to the nearest microphone).

$$\rho = 100 \cdot \frac{\mathrm{RMSE}_{meta-model} - \mathrm{RMSE}_{assimilation}}{\mathrm{RMSE}_{assimilation}} \tag{III.12}$$

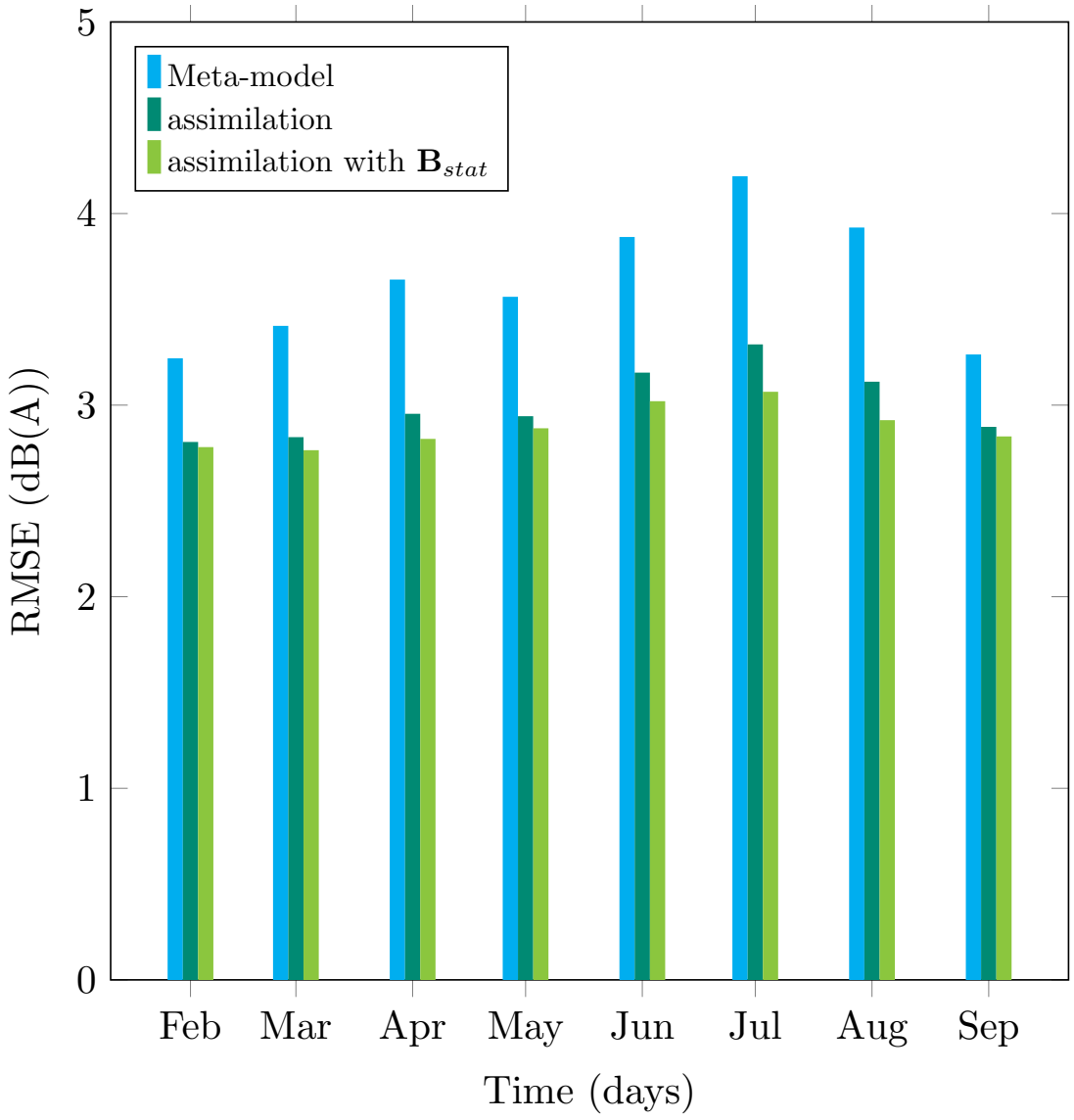


Figure III.10. – RMSE in dB(A) of the error between the observation and the nominal noise level values (blue) or the analysis in leave-on-out cross-validation (green) for all the microphones, for each month of the measurement campaign. It is noticeable to observe that the summer months have a higher error rate, which is strongly corrected by the data assimilation.

The dependency on the distance to the network is mainly unnoticeable. This is explained by the distribution of the Monte Carlo covariance matrix which is mainly non-local. The RMSE reduction shows that an error covariance between two points of the network depends on other factors than the distance between them or the difference between their noise levels. Some future studies shall work on identifying these factors, such as for instance, the nature of the roads linked to the receivers.

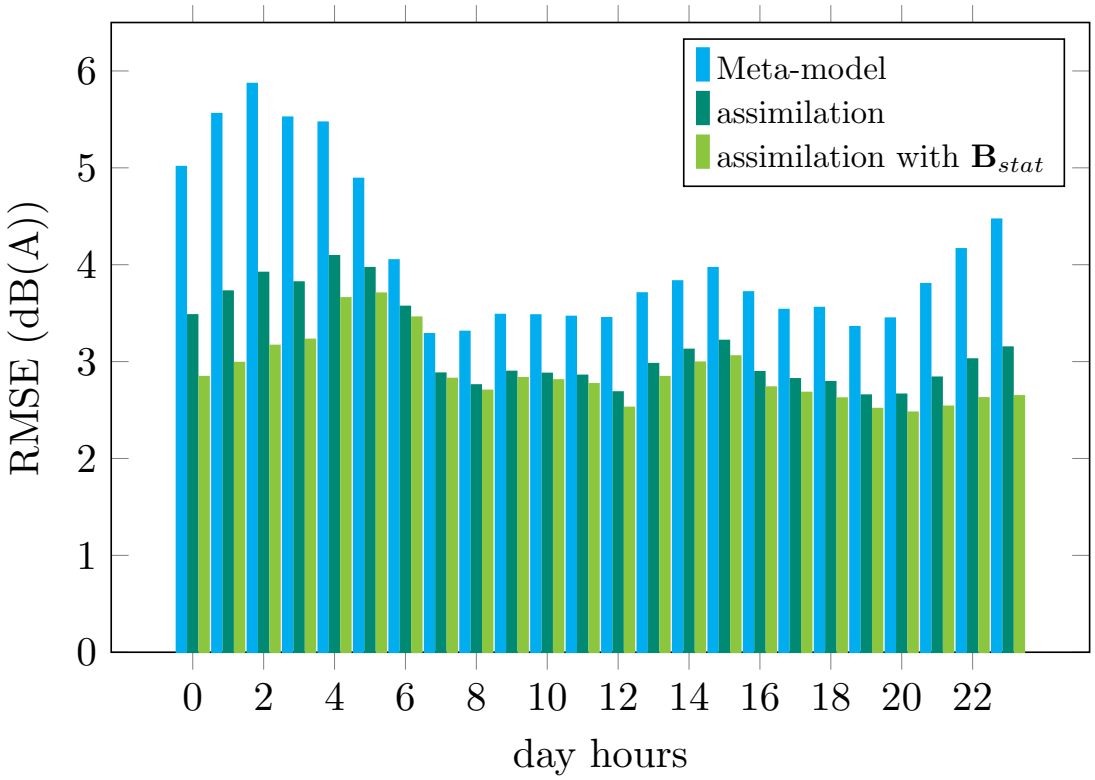


Figure III.11. – RMSE in dB(A) of the error between the observation and the nominal noise level values (blue) or the analysis in leave-on-out cross-validation (green) for all the microphones, for each hour of the day.

### III.3.2. Spatial analysis

Figure IV.6 shows how the correction due to the observations propagates across the study area when $\mathbf{B} = \mathbf{B}_{ana} + \mathbf{B}_{stat}$. It follows the intuition that for every observation point, the correction its neighborhood receives grows with the underestimation of its sound level.

## III.4. Discussion

The global RMSE reduction of all the points is noticeable and proves the efficiency of this approach. Since $\mathbf{B}$ has been built partly under the assumption that the noise errors are correlated along the road network, the data assimilation procedure shall apply in so far as the traffic is the predominant source of noise pollution at the observation points, and the bias remains low.

Amongst all the parameters shown in table III.1, the parameter “wall absorption coefficient” is not taken into account in the data assimilation process. It has in fact been set to 0.2 during the whole study. Another value would give better results on some areas of the case study and worse results on others, since there is a variability of the absorption coefficient along the buildings. The meta-model has been built this

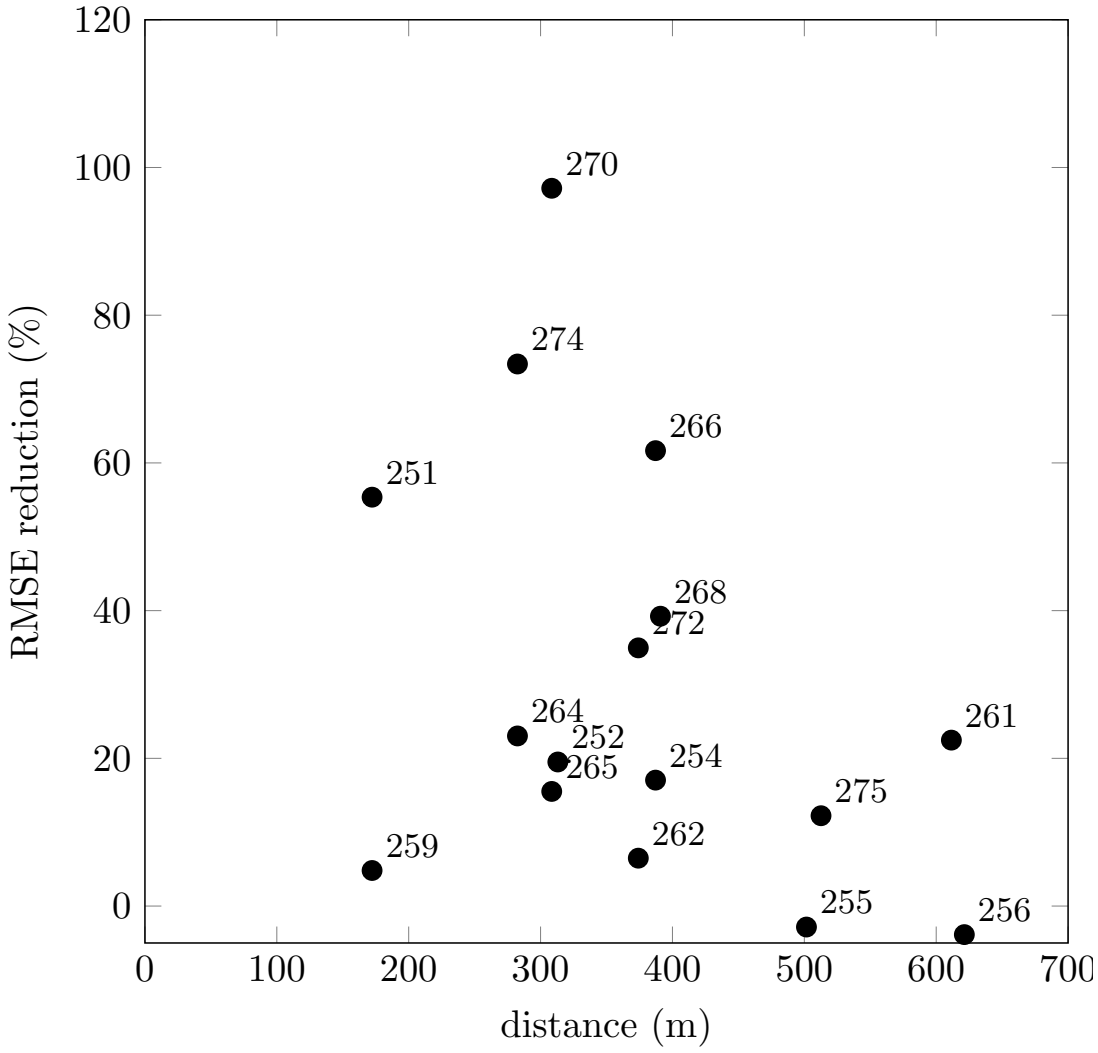


Figure III.12. – RMSE reduction rate versus the distance to the network for data assimilation with the statistical covariance matrix $\mathbf{B}_{stat}$

way for other analyses such as sensitivity analysis or inverse modeling, where more parameters are necessary in the meta-model than the condition parameters (Le Gratiet et al., 2017).

The number of parameters chosen for this study is restricted, some meta-models segregate road categories into minor, medium and major roads due to their different traffic dynamics (Barrigón Morillas et al., 2005). With a finer granularity, it is possible to better seize the noise dynamics. However, in this study, the lack of traffic data out of the major axes will preclude achieving higher levels of precision.

The observational data in this study have not been specifically collected for data assimilation purposes. An interesting approach would be to explore how the spatial distribution and the number of receivers affect the results of the data assimilation procedure. The CENSE project aims to study this question by displaying a large

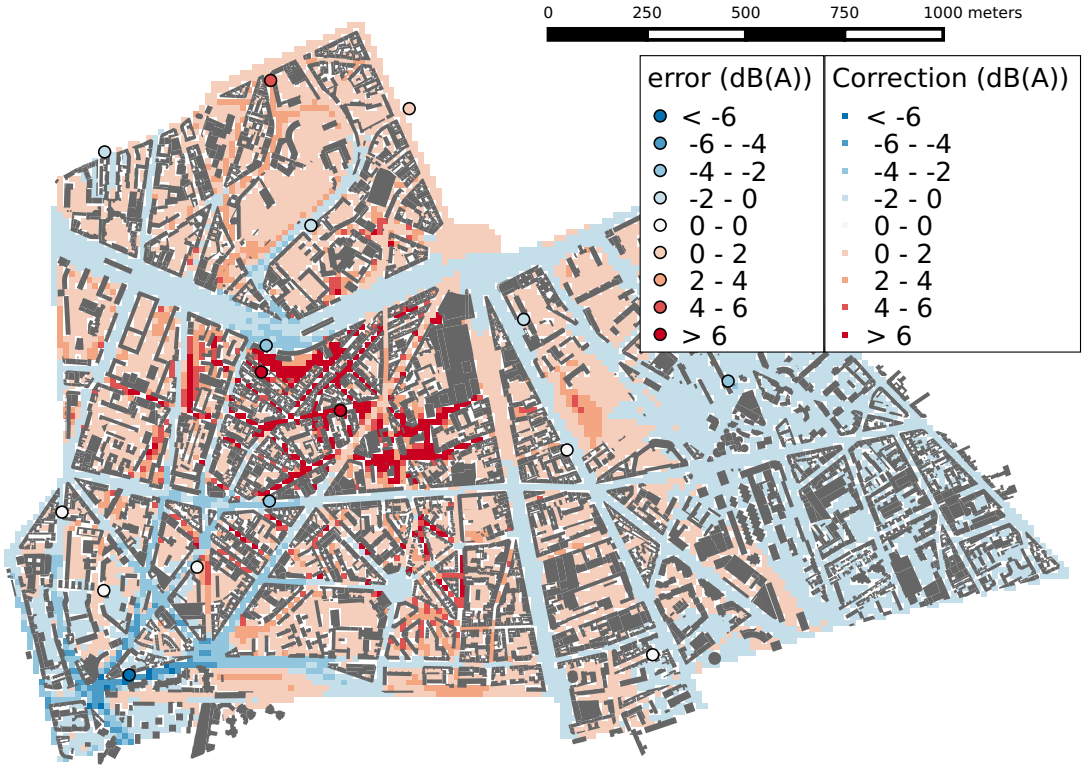


Figure III.13. – This map displays an example of the correction induced by the assimilation of the observations at a given time. The error disks are the $\mathbf{e}^b$ at the observation points, and the map represents the data assimilation correction over the map. The chosen date is June, 21 at 22:00, during a popular music festival (fête de la musique), hence the observation error and the correction are rather high. The positively corrected area is a lively neighborhood with a high density of bars and restaurants (Butte-aux-Cailles).

amount of microphones across a study area in order to optimize the data assimilation reference procedure for model and observation-based noise mapping design.

## III.5. Conclusion

The data assimilation applied to noise maps generated through a meta-model improves the performance of the generated noise map, from an average RMSE of 3.9 dB(A) to 2.8 dB(A), and better grasps the hourly evolution of the noise level distribution across the study area than a meta-model alone.

The process was carried out in three steps, the collection of data for the road traffic, the generation of the meta-model (section III.2.4.1), and then the data assimilation itself (section III.2.5.1)

This gain matrix $\mathbf{K}$ has been computed with the simulation error covariance matrix $\mathbf{B}$ and the observation error covariance matrix $\mathbf{R}$. $\mathbf{B}$ is the sum of two matrices $\mathbf{B}_{stat}$

and $\mathbf{B}_{ana}$.

— On the one hand, $\mathbf{B}_{ana}$ is called the analytical covariance matrix and relies on a simulation error propagation analytical model whose parameters have been designed with a so-called $\chi^2$ diagnosis (section B.1.1). It takes into account the physical behavior of a simulation error on a road network.

— On the other hand, $\mathbf{B}_{stat}$ has been statistically generated with a large number of calls to the meta-model using a Monte Carlo algorithm (section III.2.5.2). It describes the behavior of the error between the meta-model output given by the traffic and weather observations and the ideal meta-model output.

New data assimilation techniques relying on a meta-model are yet to be explored. The first among them is inverse modeling which aims at approaching the true input parameters of the meta-model with the help of the noise observations. This method requires an enrichment of the meta-model with more uncertain input data (height of the building, ground absorption, topography uncertainty, etc.) in order to increase the number of degrees of freedom. The distribution of the sensors across the network also needs to be optimized, and further investigation will be dedicated to the optimization of the network of sensors.

## Acknowledgments

The authors would like to thank the CENSE project (ref. ANR-16-CE22-0012) team for initiating this project, the Agence Nationale de la Recherche (ANR) for their financial support, the city of Paris and BRUITPARIF for granting access to their data, GRAFIC project for the measurement campaign whose noise data were used for the data assimilation and Nicolas Fortin of the UMRAE department at Université Gustave Eiffel for his technical support on the NoiseModelling software. The authors would also like to express their grateful aknowledgements to Dr. Isaac Jones and Dr. Ian Jones for their careful reading of the article which has greatly improved it.

# IV. Meta-model aided inverse modelling and Joint State Parameter Estimation for noise data assimilation

*This chapter is dedicated to the presentation of data assimilation techniques called Inverse Modeling (IM) and Joint State Parameter Estiamtion (JSPE). Thse methods allow to assimilate noise observation into meta-model generated noise maps in order to perform a dynamic noise mapping without the need of the traffic and weather observation as opposed as chapter III. In this chapter, The new methods are derived, then their accuracy is compared to the meta-model output and to the BLUE data assimilation method developed in chapter III with a leave-one-out testing method.*

**Table of contents**

This chapter is a transcription of the following article: Antoine Lesieur, Pierre Aumond, Arnaud Can, and Vivien Mallet. Meta-model aided inverse modelling and joint state parameter estimation for noise data assimilation (submitted). *The Journal of the Acoustical Society of America*, a.

## Abstract

This sudy aims to produce dynamic noise maps based on a noise model and acoustic measurements. To do so, inverse modeling and joint state parameters methods are proposed. These methods estimate the input parameters that optimize a given cost function calculated with the resulting noise map and the noise observations. The accuracy of these two methods is compared with a noise map generated with a meta-model and with a classical data assimilation method called Best Linear Unbiased Estimator. The accuracy of the data assimilation processes is evaluated using a "leave-one-out" cross-validation method. The most accurate noise map is generated computing a Joint State Parameter Estimation algorithm without *a priori* knowledge about traffic and weather and shows a reduction of approximately 26 % in the root mean square error from 3.5 dB to 2.6 dB compared to the reference meta-model noise map with 16 microphones over an area of 3 $km^2$.

## IV.1. Introduction

The growing awareness of the health effects of noise has made it a first-rate nuisance that needs to be characterized and mitigated (Miedema et al., 2011). Many countries have implemented policies and strategies to reduce noise pollution, of which noise mapping is a part. Directive 2002/49/EC requires for instance every European agglomeration with more than 100 000 inhabitants to produce a noise map every 5 years, as a starting point for the implementation of action plans to reduce noise pollution, but also as a tool for communicating between the different stakeholders (European Commission, 2002). As a response, model-based numerical engineering methods have been developed and are currently widely used to make road traffic urban sound maps: CNOSSOS-EU in Europe (European Commission, 2015), FHWA in USA (Fleming et al., 1995), ASJ-RTN in Japan (Yamamoto, 2010) amongst others.

For instance, CNOSSOS-EU obeys the following method, which is in two steps. First, road sections are discretized into a set of point noise sources whose emission parameters depend on the given traffic, road pavement, weather data, etc. Secondly, propagation paths are computed between each couple point source / receiver, on which the sound attenuation are calculated based on atmospheric input data such as the temperature or humidity, and physical input data such as the buildings layout and buildings absorption coefficients. This method is known to provide a good compromise between accuracy and computation time, and has made it possible to quantify the number of people exposed in cities to noise levels associated with health impacts. It is however criticized on three points. First, the quality of the input data and the accuracy of the model parameters strongly affect the quality of the noise maps produced. Secondly, it focuses on noise sources that are considered annoying, and does not reflect the diversity in sources of the perceived sound mixtures, although research in soundscapes provides evidence on the importance to characterize this diversity. Thirdly, this method is computationally expensive; thus, it only allows for the elaboration of maps based on long-term indicators and is not adapted to the production of time–evolving noise maps. The production of such maps would be a strong asset to better characterize noise exposures and the impacts of punctually noisy activities.

To address these three shortcomings, noise observatories have been developed. They enable to monitor sound levels evolution in urban areas with a fixed network of low cost noise level meters. The possibility of relying on these observatories to build noise maps recently emerged. (Aumond et al., 2018) statistically interpolated noise measurements with a kriging approach; this approach gave satisfactory results in the vicinity of the measurement devices but suffered from the lack of sensors when estimating the noise level over a wider scope. An other study, (Segura Garcia et al., 2016), showed that using similarly a kriging approach the residuals are spatially correlated except for the [7:00 P.M.- 10 P.M.] period, probably because it entails specific noise behaviors (leisure noise activities, etc.). Also, (Can et al., 2014) showed that interpolation methods were defective when the spacing between sensors was too large. Finally, (Rey Gozalo and Morillas, 2016) showed that a stratification of roads based on their functionality was helpful before interpolating sound levels and (Wei et al., 2016) built an interpolating model based on observations and a method to propagate the observed sound level.

More recently, studies propose to merge the benefits of the two approaches that is combining the spatial information from the model-based approaches and the temporal information from the observation-based approaches. (Ventura et al., 2018) assimilated observations from mobile phone microphones in order to improve a pre-calculated noise map. However, this process is useful only when a large number of mobile observations can be collected for this task. (Eerden et al., 2014) inferred the noise mapping software inputs from noise observation by estimating the emission sources from the observed noise measurements in the vicinity of the roads by using a forecasting sequential approach rather than a more spatial approach. These methods can be adapted to estimate either long-term or dynamic 1-hour maps. The Life project "Dynamap" updated 1-hour noise maps by scaling the noise levels of pre-calculated noise maps using the observed differences between measured and calculated original grid data for different classes of roads (Zambon et al., 2016; Sevillano et al., 2016). However, the production of dynamic maps requires that noise or road traffic data are available continuously, what can be tedious.

The approach presented in this paper uses both model and observation data and then allows a much more dynamic approach where 1-hour noise levels can be evaluated in real time. The method aims at optimizing several cost functions designed to improve the accuracy of the noise maps using the observations. Two classes of methods will be discussed in this paper:

— Inverse modeling (IM) (Tarantola, 2004), a method which aims at finding the most accurate noise map parameters that can explain the observed noise level at the observed locations.
— Joint state–parameter estimation (JSPE) (Zayane et al., 2010), a method which aims at finding both the most accurate noise map parameters and the corrected noise map regarding the observations. This type of method corrects at the same time the state (the output) and the parameters (the input) of the model.

Because of the large amount of calls to the model in the optimization algorithm and the large time period where the inverse modeling needs to be processed, this method is only made tractable after the substitution of the noise mapping software by a meta-model, i.e., a statistical emulation of the noise mapping software which can compute similar maps at a much lower computation cost. The construction of such meta-model

is shown in (Lesieur et al., 2020)

Since the goal of the optimization algorithm is to compute input parameters for the meta-model. The authors define *a priori* parameters as the input parameters obtained by computing observed traffic and weather data as opposed to *a posteriori* parameters which are the output of the optimisation algorithms.

The performance of these noise maps, generated using IM and JSPE, will be compared with the performance of (1) the noise maps generated by the meta-model with the given *a priori* parameters and (2) the noise maps corrected with a widely used data assimilation technique called Best Linear Unbiased Estimator (BLUE). Some of the optimization algorithms do not require knowing the *a priori* parameters values. If the performance of IM and JSPE with and without *a priori* parameters are similar, it could alleviate the work of noise maps designers to gather input observations such as traffic and weather data to build their real time noise maps, and only rely instead on noise observations.

The performance evaluation will be carried out with a leave-one-out cross validation procedure, with measurements gathered over 8-months in Paris (Aumond et al., 2017) during the ADEME project GRAFIC. Section IV.2 presents the material used in the study, the available data, the measurement network and the metamodel, section IV.3 presents the inverse modeling methods and section IV.4 presents the results using the cross validation method.

## IV.2. Materials

### IV.2.1. Case study

The study area is located in the 13th district of Paris, France and covers an area of $3\,km^2$. During the ADEME project GRAFIC, an array of 25 microphones has been deployed across the study area in 2015 (Aumond et al., 2017), from which 16 have been selected for the present study. The spatial distribution of the 16 sensors is shown in Figure IV.1.

### IV.2.2. Meta-model

#### IV.2.2.1. NoiseModelling and CNOSSOS-EU

The noise mapping software used in this study by which the meta-model is generated is NoiseModelling [1] (Aumond et al., 2020a). It is an open source software designed to produce noise maps. It uses traffic, topographic and meteorological data to generate noise maps. The process can be summed up by the following steps: the input traffic data along each road is discretized into a set of punctual noise sources. Noisemodelling computes the emissions. Then Noisemodelling computes the transmission paths between the sources and the receivers, taking into account the distance propagation, the number of reflections and diffractions. The software then computes the attenuation along each path. The sound level at a receiver is finally calculated from the contributions of all sources. The emission and propagation laws are computed following the CNOSSOS-EU guidelines (European Commission, 2015).

1. See http://noise-planet.org/noisemodelling.html

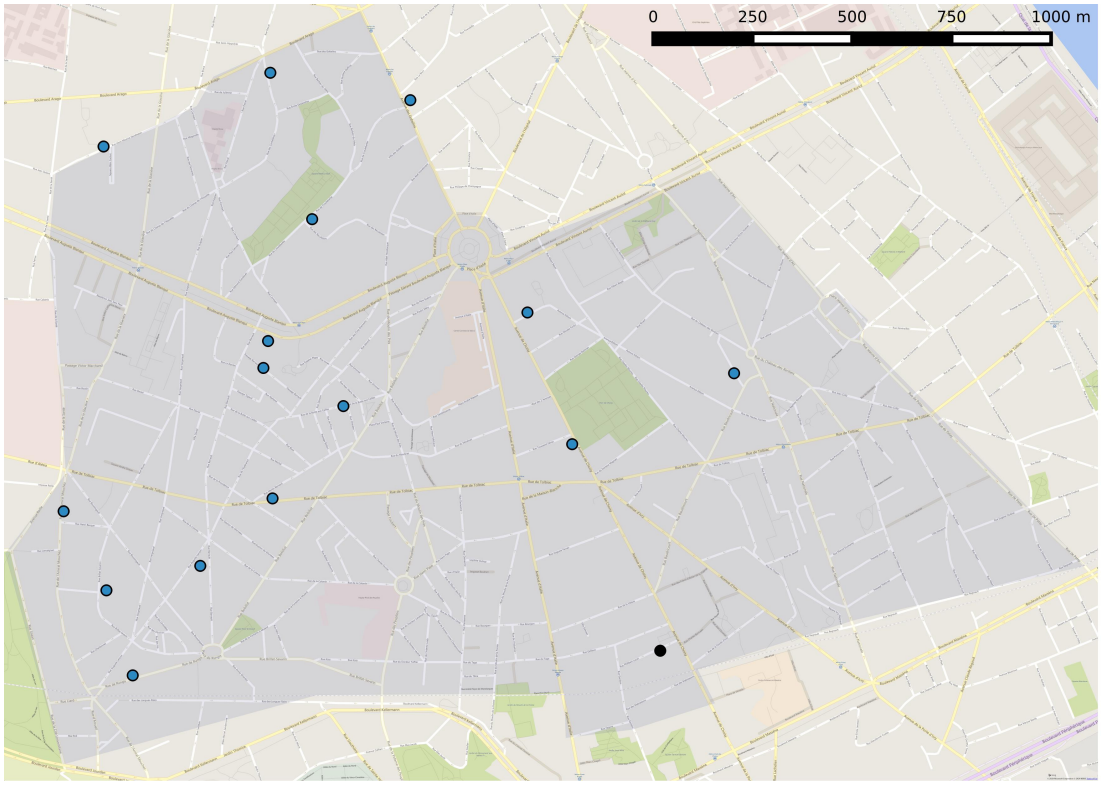


Figure IV.1. – distribution of the sensors across the study area

### IV.2.2.2. Meta-model method

Updating noise maps based on input data on an hourly basis is not tractable with classical noise mapping software, which imply too high calculation times. To circumvent this shortcoming, the proposed solution in (Lesieur et al., 2020) is to build a metamodel, which generates the same type of outputs as the reference noise mapping software, which is in this study Noisemodelling (Aumond et al., 2020b). The meta-model is built in three steps:

— The generation of the training sample;
— The dimension reduction, through a PCA, aims to select a reduced amount of maps that explain most of the variance of the noise maps of the training sample;
— The kriging, wich interpolates with a statistical method the projected values of the training sample onto the basis vectors of the reduced subspace.

Once the meta-model is built, each call with new inputs takes a very small amount of time (100 ms versus $\sim$ 1 h). For more details about the metamodel construction, the reader can refer to (Lesieur et al., 2020).

The parameters selected as inputs of the meta-model are shown in table IV.1 with their respective amplitudes. For each input, the amplitude corresponds to the range of values observed on a given time period. These parameters have been selected because the noise map software is mostly sensitive to these parameters (Aumond et al., 2020b). Regarding the flow rate and the speed, the input data $p_i$ is the ratio between the average quantity on the observed roads at a given time and the average quantity on

the observed roads over one year.

Table IV.1. – Input parameters and their input ranges for the meta-model. They are all dimensionless multiplicative coefficients applied to the reference values (the annually averaged daily values) except the temperature which is the actual value in °C.

| Parameter | Minimum value | Maximum value |
|---|---|---|
| Light vehicle speed | 0.1 | 2 |
| Heavy vehicle speed | 0.1 | 2 |
| Light vehicle flow | 0.1 | 2 |
| Heavy vehicle flow | 0.1 | 2 |
| Temperature (°C) | -5 | 30 |

#### IV.2.2.3. Configuration of the meta-model

The simulated reception points is the union of a regular grid of step $\Delta x = 15\,\text{m}$ and the set of the geographical position of the acoustical noise measurement points. They are all positioned at a fixed height of 4 m. The total amount of simulated points is $n = 8456$. The propagation parameters are set to 250 m for the propagation distance and one order of reflection.

#### IV.2.2.4. Input data for the meta-model

The input files of the noise mapping reference software are the buildings distribution and the annually averaged road traffic map during the daytime (between 06:00 and 18:00), the evening (between 18:00 and 22:00) and the night (between 22:00 and 05:00). These data are respectively given by the IGN [2], the French national geographical institute, and the city of Paris.

The computation of the noise maps with the original traffic input data, which is the annually averaged hourly traffic count for daytime period give access to the so-called $L_{day}$ indicator. It is the A-weighted, equivalent continuous sound level, over the 12-hour day period (06:00-18:00).

The nominal $L_{day}$ field distribution given by NoiseModelling with this input data is shown in figure IV.2. The nominal $L_{evening}$ and $L_{night}$ have also been computed in the study.

### IV.2.3. Real time observations

#### IV.2.3.1. Acoustic measurements

The microphones scattered across the study area have recorded the noise level $L_{f_i}$ in dB for every one-third octave $f_i$ from 63 Hz to 20 kHz with a time step of $\Delta t = 125\,\text{ms}$, from January 1, 2015 to September 30, 2015. In this study, all the noise data is given in global value and in dB(A) because the meta-model is designed to give an output in dB(A).

2. See http://professionnels.ign.fr/

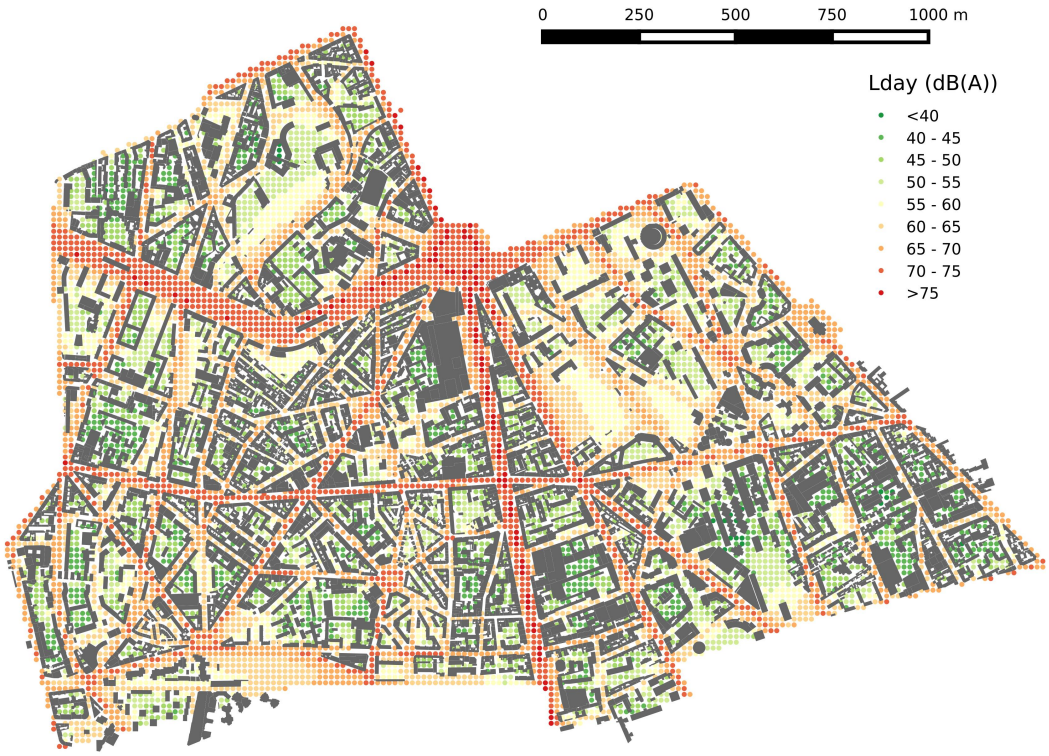


Figure IV.2. – $L_{day}$ level distribution over the study area

Although the receptors were not all simultaneously active, this does not affect the study because the small lack of data did not alter significantly the performance of the chosen data assimilation algorithm.

### IV.2.3.2. Real time 1-hour traffic and weather data

The real-time 1-hour traffic data on the main roadways of Paris is available online on the city of Paris open data platform [3]. For every measured track section, at every hour, the hourly average flow $F(t)$ (vehicles/h) and the occupation rate $R(t)$ (ratio of time where the sensors detect an occupation) are given. Given the track section length $\ell_t$ and assuming an average vehicle length value $\ell_v = 4\,\mathrm{m}$. It is possible to estimate the average vehicle speed $S(t)$ on the track section in $\mathrm{km\,h^{-1}}$

$$S(t) = 3.6 \cdot \frac{F(t) \cdot \ell_v}{R(t) \cdot \ell_t} \tag{IV.1}$$

The real-time weather data has been taken from the Montsouris weather station near the study area [4], it is also on open access online. The weather station updates its data every hour in real time, the noise mapping software only takes into account the temperature but upcoming versions might also take into account the probability of occurence of favorable conditions for each direction of propagation.

3. See https://opendata.paris.fr/
4. See https://www.infoclimat.fr/

# IV.3. Data assimilation methods

## IV.3.1. Purpose of the data assimilation

When noise measures are available, noise map designers wish to take this additional data into account in the computation of the noise map. For instance, a BLUE (Best Linear Unbiased Estimator) might be computed in order to add a correction layer to a meta-model generated with *a priori* parameters as in section IV.2.2. The BLUE method has already given staisfying results in a previous work conducted by the authors of this paper (Lesieur et al., 2021). Its principle will be quickly reminded in section IV.3.2.1 with a variational approach because the IM and JSPE algorithm are derived from this approach. The input parameters of the meta-model may be flawed, especially when they aim at estimating the traffic in minor roads. Then, data assimilation methods which allow a modification of the input parameters are proposed. Two methods will then be discussed in this section. Section IV.3.2.2 will present the inverse modeling technique. It consists in finding the parameters of the meta-model which give a noise map which best approaches the observed values at the observation point the sense of least squares. Section IV.3.2.3 will present the Joint state parameters approach. It is a mixed approach of BLUE and Inverse modeling. It consists in finding the parameters which gives a noise map which best fit the observed values at the observation points when the BLUE correction is applied. *A priori* parameters might be available, it is possible to take them into account in the process and to give new *a posteriori* parameters. The functions associated with these 5 methods (BLUE, IM and JSPE with and without *a priori* parameters) can be found in section IV.3.3.1

## IV.3.2. Data assimilation methods

### IV.3.2.1. Common data assimilation using the BLUE algorithm

This section shows a variational approach to find the BLUE for a fixed noise map $\mathbf{x}^b$ called background. The IM and JSPE algorithms respectively described in sections IV.3.2.2 and IV.3.2.3 wil be derived from this algorithm. At each timestep, a nominal noise map $\mathbf{x}^b = \widehat{\mathcal{M}}(\mathbf{p})$ is computed. It is an estimation of the real, exact state noise level distribution field $\mathbf{x}^t$. The set of observations is denoted $\mathbf{y} \in \mathbb{R}^p$.

It is impossible to know the exact vector $\mathbf{x}^t$, either from the erroneous simulation or the incomplete (and slightly erroneous) observations. $\mathbf{x}^t$ is then a random vector. The set of observations is denoted $\mathbf{y} \in \mathbb{R}^p$. Let:

— $\mathbf{H} \in \mathbb{R}^{p \times n}$ be the observation matrix which maps a state vector, e.g., $\mathbf{x}^t$, to the observed values $\mathbf{y}$, so that $\mathbf{y}$ can be compared to $\mathbf{H}\mathbf{x}^t$. In this case, each observation point has a simulated counterpart, $\mathbf{H}$ is therefore a matrix in which each row is full of zeros except for a one in the column that corresponds to the receiver located at the observation point. The elements $[\mathbf{H}]_{ij}$ are equal to one if and only if the $\mathrm{j}^{th}$ receiver corresponds to the $\mathrm{i}^{th}$ observation point.

— $\mathbf{e}^b \in \mathbb{R}^n$ be the simulation error $\mathbf{x}^b - \mathbf{x}^t$, a random vector, which is assumed to be unbiased $\mathbb{E}[\mathbf{e}^b] = 0$,

— $\mathbf{e}^o \in \mathbb{R}^p$ be the observation error $\mathbf{y} - \mathbf{H}\mathbf{x}^t$, a random vector, which is assumed to be unbiased $\mathbb{E}[\mathbf{e}^o] = 0$,

— $\mathbf{B} = \mathbb{E}[\mathbf{e}^b \mathbf{e}^{bT}] \in \mathbb{R}^{n \times n}$ be the covariance matrix of the simulation error,

— $\mathbf{R} = \mathbb{E}[\mathbf{e}^o \mathbf{e}^{oT}] \in \mathbb{R}^{p \times p}$ be the covariance matrix of the observation error.

The BLUE output is the so-called analysis noise map $\mathbf{x}^a$ which is a linear combination of $\mathbf{y}$ and $\mathbf{x}^b$ and minimizes the variance of the error $\mathbf{e}^a = \mathbf{x}^a - \mathbf{x}^t$. It can equivalently be described with a variational approach, i.e. as an optimization problem over a functional (Le Dimet and Talagrand, 1986) as shown below.

Let $\mathcal{J}$ be the following cost functional:

$$\begin{aligned} \mathcal{J}(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) + (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}) \\ &= \left\| \mathbf{x} - \mathbf{x}^b \right\|^2_{\mathbf{B}^{-1}} + \left\| \mathbf{y} - \mathbf{H}\mathbf{x} \right\|^2_{\mathbf{R}^{-1}} \end{aligned} \tag{IV.2}$$

With

$$\begin{array}{rccc} \|\cdot\|_{\mathbf{A}} & : & \mathbb{R}^n & \to & \mathbb{R} \\ & & \mathbf{x} & \mapsto & \sqrt{\mathbf{x}^T \mathbf{A} \mathbf{x}} \end{array} \tag{IV.3}$$

where $\mathbf{A}$ is a positive definite matrix.

The purpose is to find $\mathbf{x}^a$ such that

$$\mathcal{J}(\mathbf{x}^a) = \min_{\mathbf{x} \in \mathbb{R}^n} \{\mathcal{J}(\mathbf{x})\} \tag{IV.4}$$

$\mathbf{H}$, $\mathbf{B}$ and $\mathbf{R}$ are respectively the observation operator (matrix), the simulation error covariance matrix and the observation error covariance matrix. They are described more thoroughly in the appendix section B.1.

The expression of $\mathbf{x}^a$ is found by solving $\nabla \mathcal{J}(\mathbf{x}^a) = 0$:

$$\begin{aligned} \mathbf{x}^a &= \mathbf{x}^b + \underbrace{\mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}}_{\mathbf{K}} (\mathbf{y} - \mathbf{H}\mathbf{x}^b) \\ &= \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b) \end{aligned} \tag{IV.5}$$

This is the BLUE found with a minimal variance approach (Bouttier and Courtier, 2002).

#### IV.3.2.2. Inverse Modeling

Starting with this method, the idea of the inverse modeling is to find a cost function which is minimal when its inputs, the parameters set $\mathbf{p}^\star$, minimize the RMSE of the meta-model output, i.e. when they give the best $\mathbf{x}^b = \widehat{\mathcal{M}}(\mathbf{p}^\star)$ with no data assimilation process.

$$\mathcal{J}(\mathbf{p}) = \left\| \mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}) \right\|^2_{\mathbf{R}^{-1}} \tag{IV.6}$$

In this case study where $\mathbf{R}$ is proportional to the identity matrix, this technique corresponds to the Least Mean Square (LMS) estimation applied with a gradient method, $\mathcal{J}$ being differentiable.

The method selected in this case study is the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm as implemented in the **Scipy** library v1.2.2 in the function "scipy.optimize.minimize" with the method "L-BFGS-B" (Virtanen et al., 2020)

#### IV.3.2.3. Joint State Parameter Estimation

Starting with the BLUE $\mathbf{x}^a = \mathbf{x}^b + \underbrace{\mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^b)}_{\boldsymbol{\epsilon}}$, the goal of this method is to optimize at the same time the correction layer $\boldsymbol{\epsilon} = \mathbf{x}^a - \mathbf{x}^b$, and the set of input parameters $\mathbf{p}^\star$ so that $\mathbf{x}^a = \widehat{\mathcal{M}}(\mathbf{p}^\star) + \boldsymbol{\epsilon}^\star$.

The idea of the Joint State Parameter Estimation is to find a cost function which is minimal when the parameters $\mathbf{p}^\star$ and the correction layer $\boldsymbol{\epsilon}^\star$ minimize the RMSE of $\mathbf{x}^a$, under the constraint that $\boldsymbol{\epsilon}^\star$ has a covariance matrix $\mathbf{B}$.

This is equivalent to optimizing the following cost function:

$$\mathcal{J}(\mathbf{p}, \boldsymbol{\epsilon}) = \left\| \mathbf{y} \ - \mathbf{H}(\widehat{\mathcal{M}}(\mathbf{p}) + \boldsymbol{\epsilon}) \right\|^2_{\mathbf{R}^{-1}} + \|\boldsymbol{\epsilon}\|^2_{\mathbf{B}^{-1}} \tag{IV.7}$$

Note that equation IV.7 is equivalent to equation IV.2, only the input parameters change.

Define the following cost function

$$\mathcal{J}'(\mathbf{p}) = \left\| \mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}) \right\|^2_{(\mathbf{I}-\mathbf{H}\mathbf{K})^T\mathbf{R}^{-1}} \tag{IV.8}$$

with the same $\mathbf{K}$ as defined in section IV.3.2.1

Appendix B.1.3 shows that minimizing this new cost function is equivalent to minimizing $\mathcal{J}(\mathbf{p}, \boldsymbol{\epsilon})$ considering that $\boldsymbol{\epsilon} = \mathbf{K}(\mathbf{y} - \widehat{\mathcal{M}}(\mathbf{p}))$ is a necessary condition for $\mathcal{J}$ to be optimal. Due to the structure of the meta-model, $\mathcal{J}'$ is differentiable. This optimization problem can then be solved with a gradient method.

The method selected in this case study is the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm as implemented in the **Scipy** library v1.2.2 in the function "scipy.optimize.minimize" with the method "L-BFGS-B" (Virtanen et al., 2020)

#### IV.3.2.4. Consideration of the *a priori* about $\mathbf{p}$

The ground observation of traffic and weather data gives an *a priori* about $\mathbf{p}$ called $\mathbf{p}_{\text{prior}}$ as shown in section IV.2.2. The *a priori* can be taken into account in the cost function by adding the quantity $\|\mathbf{p} - \mathbf{p}_{prior}\|^2_{\mathbf{Q}}$, which is analogous to a restoring force that links $\mathbf{p}$ to its prior $\mathbf{p}_{prior}$.

For inverse modeling:

$$\mathcal{J}(\mathbf{p}) = \left\| \mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}) \right\|^2_{\mathbf{R}^{-1}} + \|\mathbf{p} - \mathbf{p}_{\text{prior}}\|^2_{\mathbf{Q}^{-1}} \tag{IV.9}$$

For Joint State Parameter Estimation:

$$\mathcal{J}(\mathbf{p}, \boldsymbol{\epsilon}) = \left\| \mathbf{y} \ - \mathbf{H}(\widehat{\mathcal{M}}(\mathbf{p}) + \boldsymbol{\epsilon}) \right\|^2_{\mathbf{R}^{-1}} + \|\boldsymbol{\epsilon}\|^2_{\mathbf{B}^{-1}} + \|\mathbf{p} - \mathbf{p}_{\text{prior}}\|_{\mathbf{Q}^{-1}} \tag{IV.10}$$

For equation (IV.10), since this new term does not modify the reasoning which lead to equation (IV.8), the new final cost function reads:

$$\mathcal{J}'(\mathbf{p}) = \left\| \mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}) \right\|^2_{(\mathbf{I}-\mathbf{H}\mathbf{K})^T\mathbf{R}^{-1}} + \|\mathbf{p} - \mathbf{p}_{\text{prior}}\|_{\mathbf{Q}^{-1}} \tag{IV.11}$$

The $\mathbf{Q}$ matrix reflects the level of confidence we have in the *a priori* parameter. In this study $\mathbf{Q} = \text{diag}(\sigma^2_{\mathbf{Q},1}, \cdots, \sigma^2_{\mathbf{Q},k})$ with $\sigma_{\mathbf{Q},i}$ being the standard deviation of the difference of two selections of the $i^{th}$ parameter assuming that they follow a uniform distribution over its range defined in table IV.1. If the confidence on the value of $\mathbf{p}_{prior}$ is high, the value of $\sigma_{\mathbf{Q},i}$ is low and the quantity which appears in the cost function $\sigma^{-2}_{\mathbf{Q},i}(p_i - p_{prior,i})^2$ weights more in the cost function, hence a significative difference between $p_i$ and $p_{prior,i}$ has more weight in the cost function. Whether we have an *a priori* or not about the input parameter changes the purpose of the application. On the one hand, if input data are available, there is a need to compare the performance of the joint state parameter estimation with regular data assimilation as studied in a previous work to tell how relevant is this method regarding the previous works. On the other hand, if good results are obtained with only sound level observations, this would really ease the use of data assimilation in areas with no traffic observation data and expand the area of noise level data assimilation.

## IV.3.3. Compared interests of the different methods

### IV.3.3.1. Sum up of the data assimilation algorithms

The following list sums up the different data assimilation algorithms studied in this paper:

- The classical data assimilation: $\widehat{\mathcal{M}}(\mathbf{p}_{\text{prior}}) + \mathbf{K}\left(\mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}_{\text{prior}})\right)$;
- The inverse modeling: $\widehat{\mathcal{M}}(\mathbf{p}^\star)$ with $\mathbf{p}^\star$ solution of
  - $\min_{\mathbf{p}} \left\{ \mathcal{J}_1(\mathbf{p}) = \left\| \mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}) \right\|^2_{\mathbf{R}^{-1}} \right\}$;
  - $\min_{\mathbf{p}} \left\{ \mathcal{J}_2(\mathbf{p}) = \left\| \mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}) \right\|^2_{\mathbf{R}^{-1}} + \|\mathbf{p} - \mathbf{p}_{\text{prior}}\|^2_{\mathbf{Q}^{-1}} \right\}$;
- The joint state parameter estimation: $\widehat{\mathcal{M}}(\mathbf{p}^\star) + \mathbf{K}(\mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}^\star))$ with $\mathbf{p}^\star$ solution of
  - $\min_{\mathbf{p}} \left\{ \mathcal{J}_3(\mathbf{p}) = \left\| \mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}) \right\|^2_{(\mathbf{I}-\mathbf{HK})^T\mathbf{R}^{-1}} \right\}$;
  - $\min_{\mathbf{p}} \left\{ \mathcal{J}_4(\mathbf{p}) = \left\| \mathbf{y} - \mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}) \right\|^2_{(\mathbf{I}-\mathbf{HK})^T\mathbf{R}^{-1}} + \|\mathbf{p} - \mathbf{p}_{\text{prior}}\|_{\mathbf{Q}^{-1}} \right\}$;

### IV.3.3.2. Comparison of the different methods

Both methods, IM and JSPE, can work with or without prior knowledge of the input parameters, four optimization algorithms are therefore analyzed. Table IV.2 sums up the data necessary to perform the optimization algorithm for each cost function. If satisfying results are obtained with optimization algorithms that do not need *a priori* parameters, then it is possible to expand the inverse modeling data assimilation processes to a wider range of urban areas where the input parameters values are not known. The JSPE may outperform the inverse modeling method thanks to the additional piece of information given by the background error covariance matrix $\mathbf{B}$. As shown in (Lesieur et al., 2021) the computation of $\mathbf{B}$ requires an additional amount of computation time in order to compute the parameters of the matrix with a so-called $\chi^2$ diagnosis and the computation of the statistical covariance which requires an order of magnitude of $10^5$ calls to the meta-model. Both methods, IM and JSPE, can work with

or without prior knowledge of the input parameters, four optimization algorithms are therefore analyzed. Table IV.2 sums up the data necessary to perform the optimization algorithm for each cost function. If satisfying results are obtained with optimization algorithms that do not need *a priori* parameters, then it is possible to expand the inverse modeling data assimilation processes to a wider range of urban areas where the input parameters values are not known. The JSPE may outperform the inverse modeling method thanks to the additional piece of information given by the background error covariance matrix $\mathbf{B}$. As shown in (Lesieur et al., 2021) the computation of $\mathbf{B}$ requires an additional amount of computation time in order to compute the parameters of the matrix with a so-called $\chi^2$ diagnosis and the computation of the statistical covariance which requires an order of magnitude of $10^5$ calls to the meta-model.

Table IV.2. – Synthesis table of the data necessary to compute each cost function and thus to perform the corresponding optimization algorithm.

| **optimization algorithms** | *a priori* parameters $\mathbf{p}$ | covariance matrices $\mathbf{B}$ and $\mathbf{R}$ |
|---|---|---|
| Inverse modeling | ✗ | ✗ |
| Inverse modeling with *a priori* parameters | ✓ | ✗ |
| JSPE | ✗ | ✓ |
| JSPE with *a priori* | ✓ | ✓ |

## IV.4. Results

### IV.4.1. Validation Methods

#### IV.4.1.1. Metrics

The indicators defined in table IV.3 are used to quantify the performance of the outputs of the simulations, data assimilations, inverse modeling and JSPE.

Table IV.3. – Scores for the performance evaluation of a model. $(c_i)_i$ is the simulated sequence. $(o_i)_i$ is the corresponding observed sequence. $n$ is the total number of elements in the sequence. $\bar{c}$ and $\bar{o}$ are respectively the mean of $(c_i)_i$ and $(o_i)_i$. RMSE stands for Root Mean Square Error.

| Score | Formula |
|---|---|
| Bias | $\frac{1}{n}\sum_{i=1}^{n}(c_i - o_i)$ |
| RMSE | $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(c_i - o_i)^2}$ |

#### IV.4.1.2. Leave-one-out cross validation

The leave-one-out cross-validation consists in removing the observations of a given microphone from the data assimilation process. Only the observations from the other

microphones are used to correct the noise level distribution. This procedure is carried out for all microphones, one by one, where only one microphone is removed at a time. At the removed station, the meta-model performance is compared to the performance after assimilation of the observations of the other microphones. This enables to check whether the assimilation properly distributes in space the corrections that originate from the observed locations. The cross-validation evaluates the effects of the data assimilation method at locations without any observations. It was carried out on the data of the first 7 weeks of the year 2015.

In addition to the performance of these noise maps generated with a data assimilation algorithm, the performance of the noise map generated by the meta-model only with the *a priori* parameters $\widehat{\mathcal{M}}(\mathbf{p}_{\text{prior}})$ will be tested. The evaluation of the performance of the meta-model is not a leave-one-out cross validation, it simply consists in comparing the meta-model output at an observation point at a given time and the corresponding observation.

## IV.4.2. Comparison of the methods

### IV.4.2.1. Bias dispersion

Figure IV.3 shows the bias dispersion for all the receptors for every optimization algorithm shown in section IV.3.3.1. The distribution of the bias is narrower for the inverse modeling than for the the meta-model only. It is even narrower when the JSPE is used. This evolution is also presented in table IV.4 where the evolution of the absolute bias is showed at each level of correction. The table shows that the JSPE with $\mathbf{p}_{\text{prior}}$ gives the best results with 35 % of the 1-hour estimates within a margin error of 1 dB and up to 79 % within a margin error of 3 dB. All the optimized cost function based algorithms show better results than the regular BLUE data assimilation procedure for both error margins 1 dB and 3 dB whether we consider the approach with or without taking the $\mathbf{p}_{\text{prior}}$ parameter.

Table IV.4. – Dispersion in dB(A) of the error between the observation and the simulation or the analysis in leave-on-out cross-validation for all the microphones during all the measurement campaign

| **error amplitude** (dB) | $[0,1]$ | $]1,3]$ | $]3,5]$ | $>5$ |
|---|---|---|---|---|
| meta-model | 21 % | 39 % | 24 % | 14 % |
| assimilation | 30 % | 45 % | 17 % | 7 % |
| Inverse modeling | 31 % | 44 % | 17 % | 7 % |
| Inverse modeling with $\mathbf{p}_{\text{prior}}$ | 31 % | 44 % | 17 % | 7 % |
| JSPE | 33 % | 45 % | 15 % | 6 % |
| JSPE with $\mathbf{p}_{\text{prior}}$ | 35 % | 44 % | 15 % | 6 % |

### IV.4.2.2. RMSE

In figure IV.4, the root mean squared error is displayed for each week, for each optimization algorithm shown in section IV.3.3.1 and for the regular meta-model computation and data assimilation. The reduction of the error variance is noticeable. The

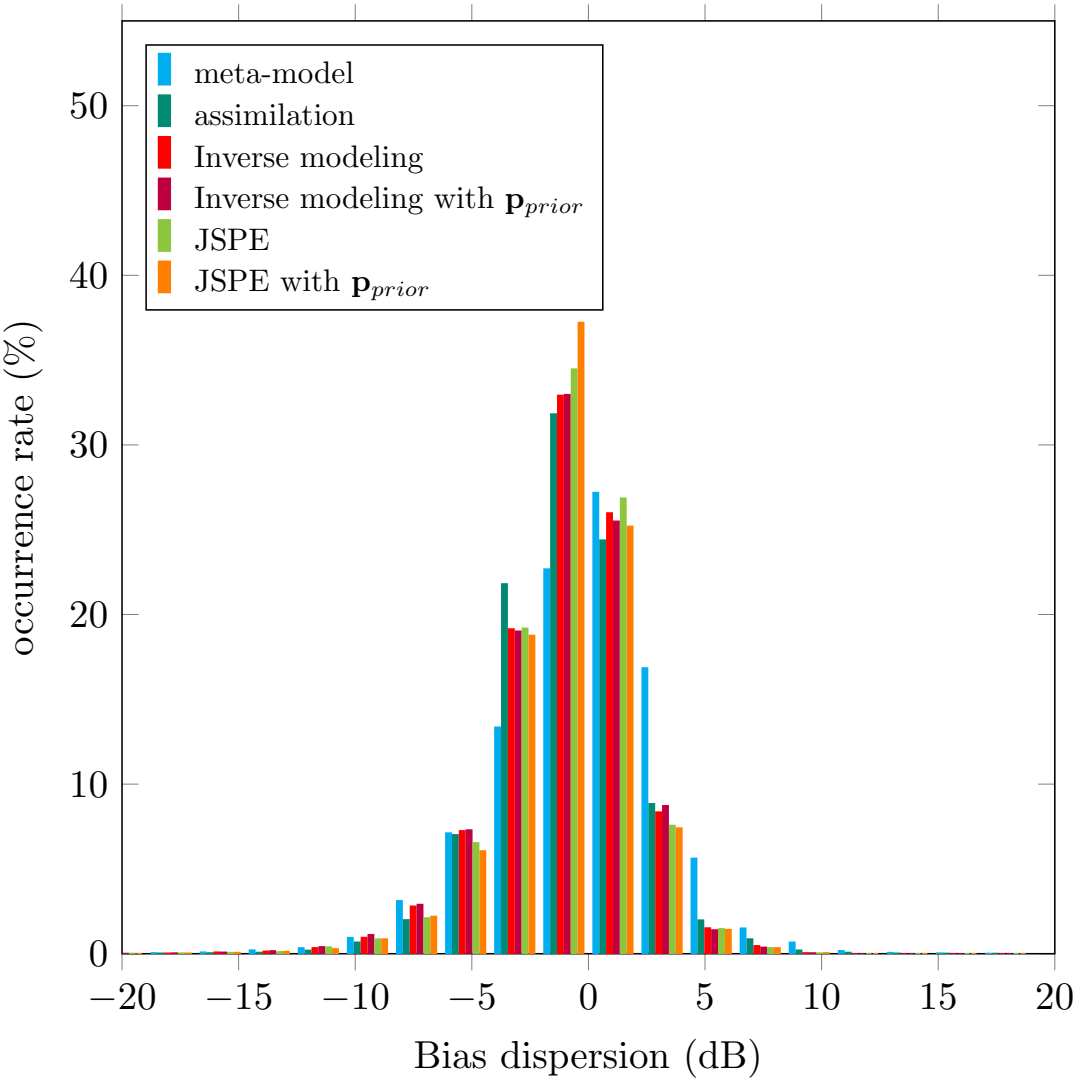


Figure IV.3. – Dispersion in dB(A) of the error between the observation and the simulation or the noise maps corrected by the data assimilation processes for all the microphones during the first 7 weeks of the measurement campaign.

JSPE with *a priori* parameters shows the best results as it goes down from an average of 3.3 dB(A) to an average of 2.5 dB(A) for the first week. It represents an improvement of the performance of 25 % compared to the meta-model prediction. The inverse modeling shows an improvement of 21 % with a RMSE of 2.7 dB compared to the meta-model prediction. These results are on average more efficient than the regular data assimilation procedure. As explained in section IV.3.3, the better performance of the JSPE is explained by the taking into account of the background error covariance matrix **B**.

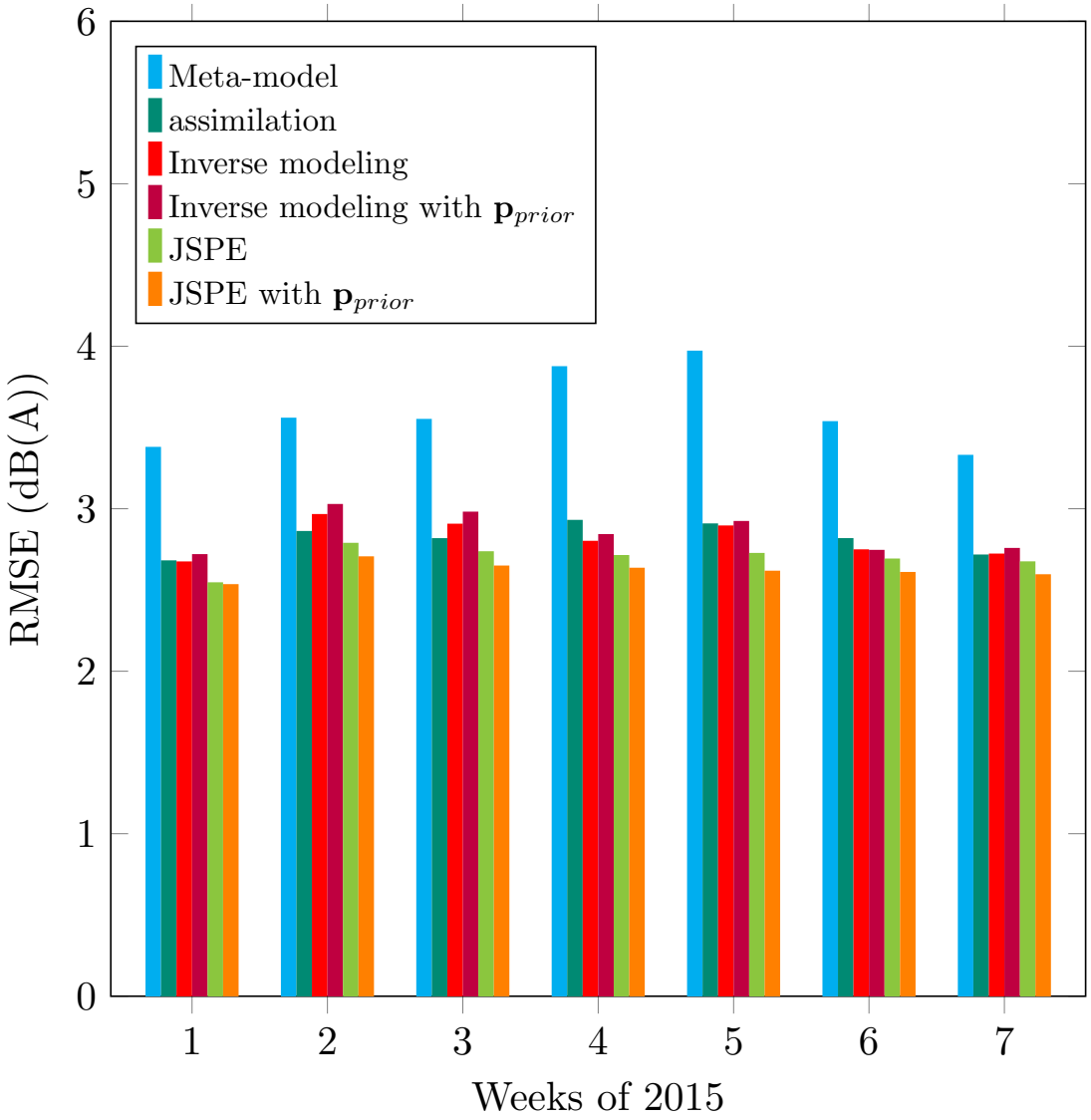


Figure IV.4. – RMSE in dB(A) of the error between the observation and the different outputs of the nominal and corrected noise maps for all the microphones, for the first 7 weeks of th measurement campaign.

## IV.4.3. Spatial analysis

### IV.4.3.1. Visualization of the noise and correction maps

An example of the correction induced by the JSPE based on the observations at a given time is diplayed in figures IV.5, IV.6 and IV.7. The emulation with the parameters found by the JSPE $\widehat{\mathcal{M}}(\mathbf{p}^\star)$ is represented in figure IV.5, figure IV.6 represents the correction found by the JSPE $\boldsymbol{\epsilon}^\star$ and finally figure IV.7 represents the resulting analysis map $\widehat{\mathcal{M}}(\mathbf{p}^\star) + \boldsymbol{\epsilon}^\star$. The chosen date is June, 21 at 22:00, during a popular street music festival (fête de la musique), hence the observation error and the correction are rather high. The positively corrected area is a lively neighborhood with a high density of bars and restaurants (Butte-aux-Cailles).

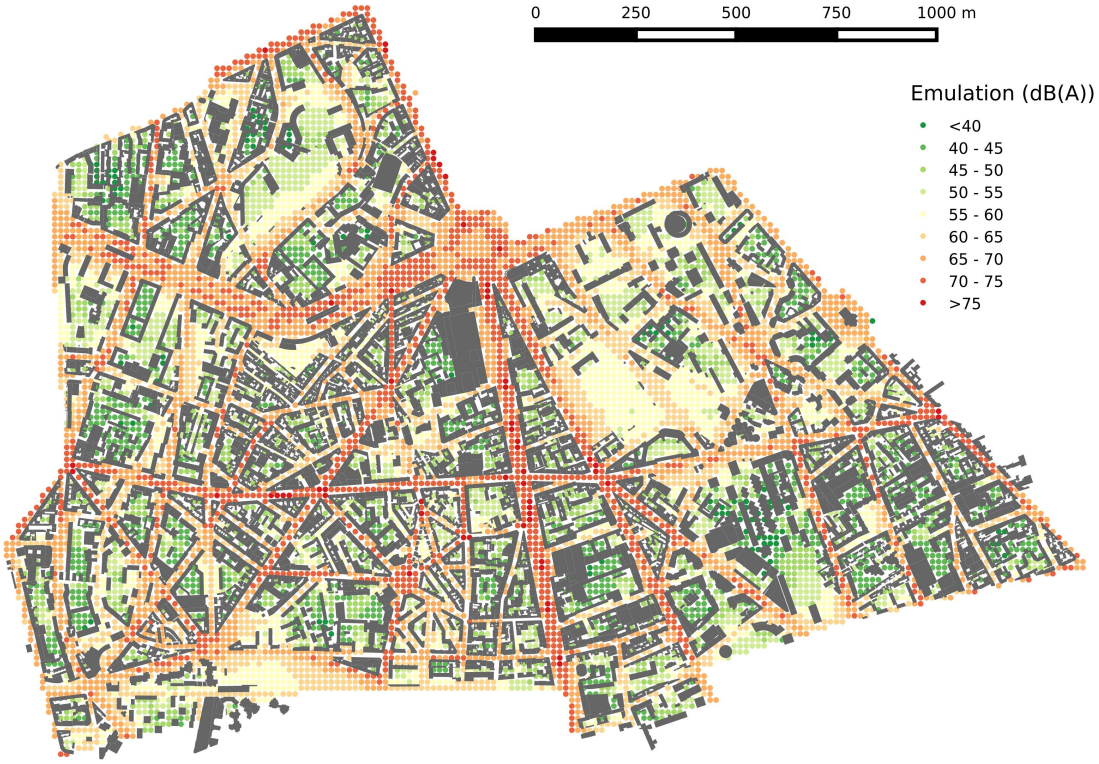


Figure IV.5. – Emulation map: $\widehat{\mathcal{M}}(\mathbf{p}^\star)$ computed on June 21 at 10 P.M. during a popular street music festival (fête de la musique)

### IV.4.3.2. Compared performance at the observation points

The performance of the leave-one-out cross-validation at the reception points for the inverse modeling and the JSPE without *a priori* parameters is shown in figure IV.8. The figure shows that the performance of the observation points is not uniformally distributed. A qualitative categorization of the roads has been done, the minor roads observation points have been selected manually as observation points nearby the roads visually classified as streets and the major roads are the remaining ones visually classified as roads. The performance seems to be connected to the nature of the road section

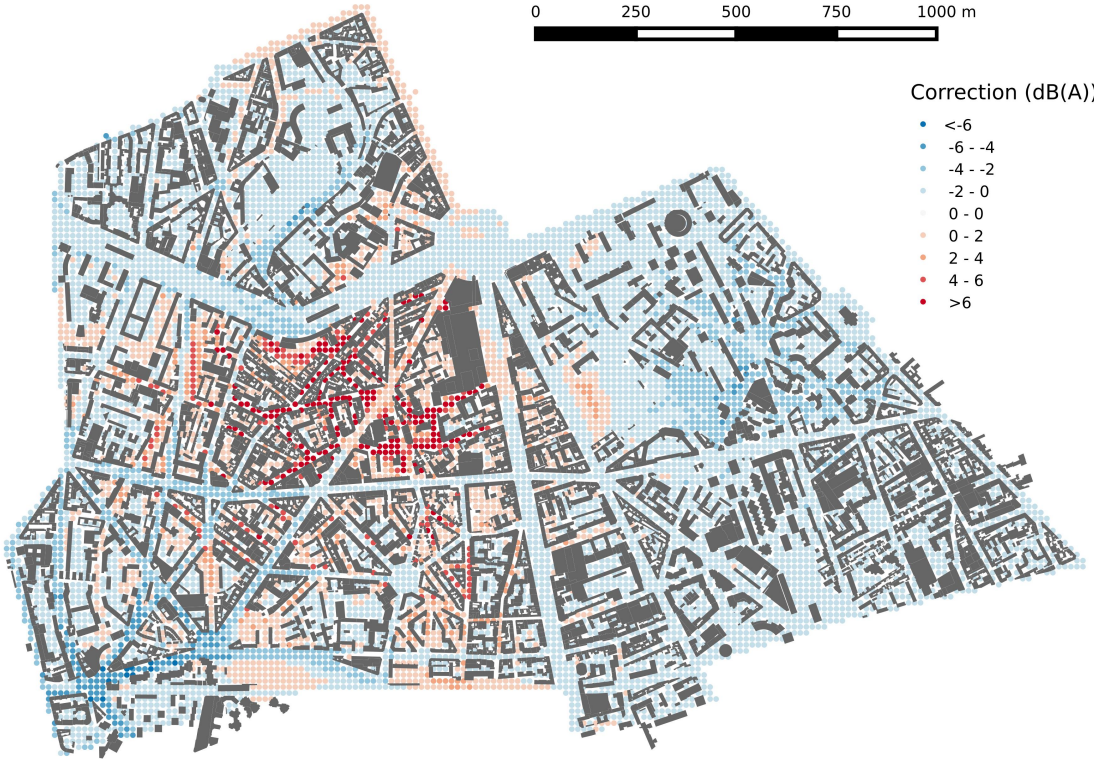


Figure IV.6. – Correction map: $\epsilon^\star$ computed on June 21 at 10 P.M. during a popular street music festival (fête de la musique)

associated to the observation point. It is shown in figure IV.8 and in table IV.5 that minor roads perform better than major roads. These figures only show the results for JSPE and IM without *a priori* parameters algorithms because we just aim at showing a trend, the other algorithms with *a priori* perform in the same fashion. This is explained by the higher variability of the time evolution of the traffic and thus the noise dynamic in these areas. In addition, in minor roads areas, since the traffic noise is not predominant, other noise sources (public works, shoutings, shops and restaurants, etc.) may increase the discrepancy between the simulated noise maps which only takes into account traffic noise and the observed sound level.

One may wonder if the distance of the observation points to the sensors network affects the performance of the RMSE with such distance between the observation points. The scatter plots displayed in figures IV.9 and IV.10 show that there is no trend that show a clear correlation between the distance to the network and the performance of the observed points. Note that these trends are valid for both inverse modeling and JSPE algorithms, not all reception points perform better with the JSPE rather than with the inverse modeling but no external factor has been identified to explain the change of performance. This lack of trend might be explained by the relatively large distance between each reception points (the majority of the points are more than 250 m apart). A tighter array of receptor might show a trend for reception points closer together.

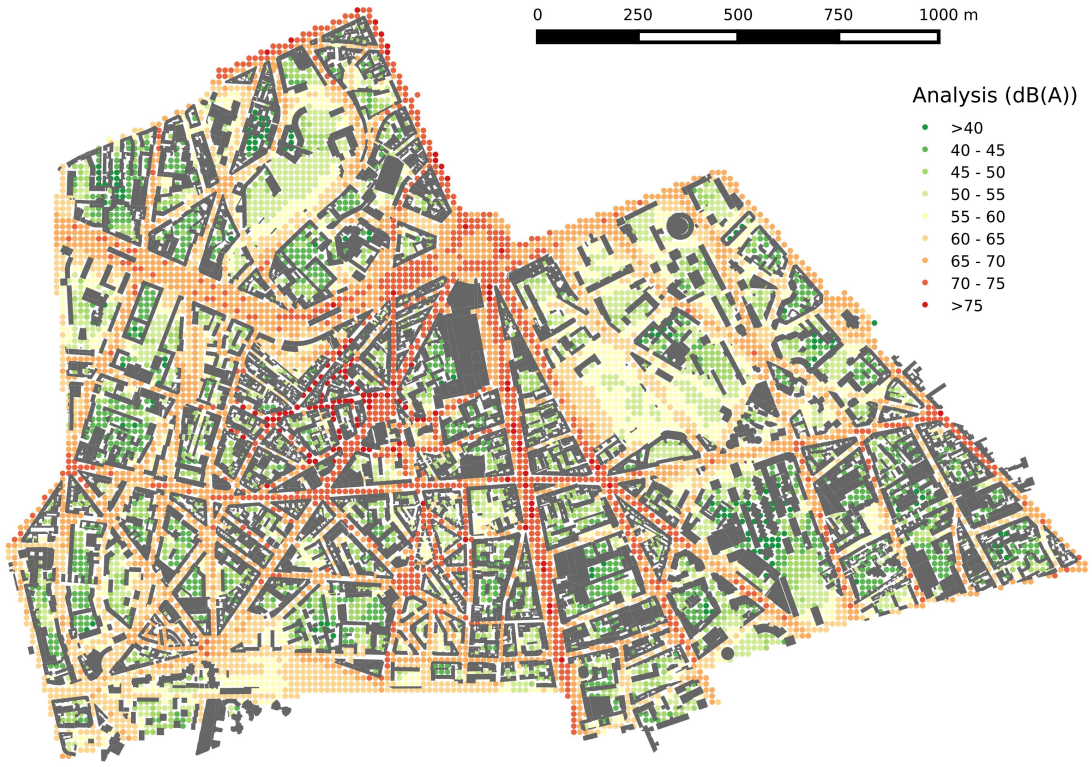


Figure IV.7. – Resulting analysis map: $\widehat{\mathcal{M}}(\mathbf{p}^\star) + \boldsymbol{\epsilon}^\star$ computed on June 21 at 10 P.M. during a popular street music festival (fête de la musique)

### IV.4.4. Comparison of the *a priori* and *a posteriori* parameters

The *a priori* input parameters of the meta-model and the input parameters computed with the JSPE algorithm are shown in figure IV.11. The two graphs follow the same daily trend, however, the variability of the computed parameters is higher due to the higher uncertainty of the observed sound levels. As these input parameters might seem too bumpy an unrealistic, recall that they are designed to fit with the observed noise data instead of reflecting the global trend of traffic intensity over the study area. This is why here, only the trend matters to check if it complies with the order of magnitude.

## IV.5. Discussion

### IV.5.1. Implementation of the algorithms and computation time

As explained in table IV.4, The JSPE algorithm requires to computes a background error covariance matrix $\mathbf{B}$ and thus requires additional knowledge of the state of the studied area. However, when all the required data is gathered, the computation time to process the optimization algorithm is the same for all the presented algorithm and is on average 70 s. The number of iteration fluctuates around an average value of 30 iteration to reach the optimal value, and so does the computation time but the order of magnitude is the same.

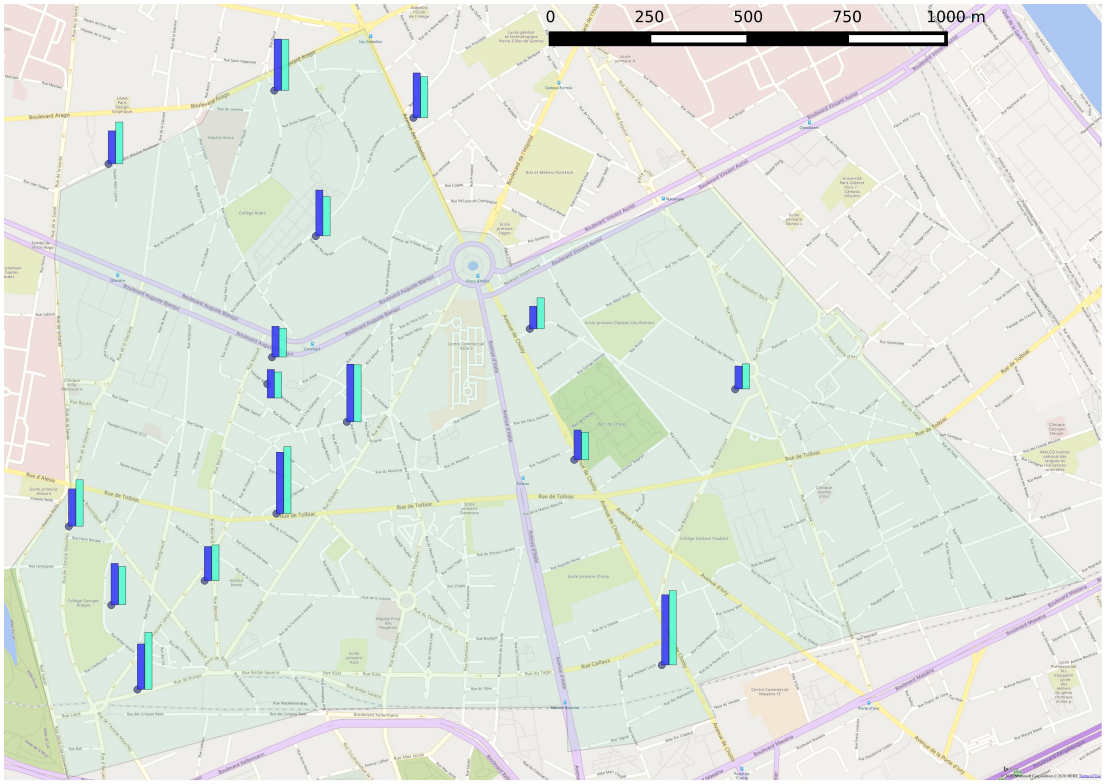


Figure IV.8. – RMSE of the leave-one-out cross-validation of the IM (cyan) and the JSPE (blue) without *a priori* parameters, the highest bar represents a RMSE of 4.7 dB

### IV.5.2. Frequency band

In this study, different noise sources with different noise spectra are mixed together to give only one noise indicator, the $LA_{eq,1h}$ in dB(A). A new approach might be to build several optimization algorithms for several frequency bands. With these tools it might be possible to build a larger quantity of indicators and to give a more precise description of the surrounding soundscape, for example by discriminating the type of vehicles.

### IV.5.3. Estimation of the parameters

The estimation of the parameters shown in figure IV.11 suggests that this method might help to estimate the input parameters of the meta-model such as traffic data for the study area. However, the estimated parameters are a proxy of the input parameter and are not as useful as the resulting noise map. This method is not adapted to reconstruct the input data with noise observation. Studies like (Eerden et al., 2014) have been specifically designed to construct emission sources such as traffic data with noise observations.

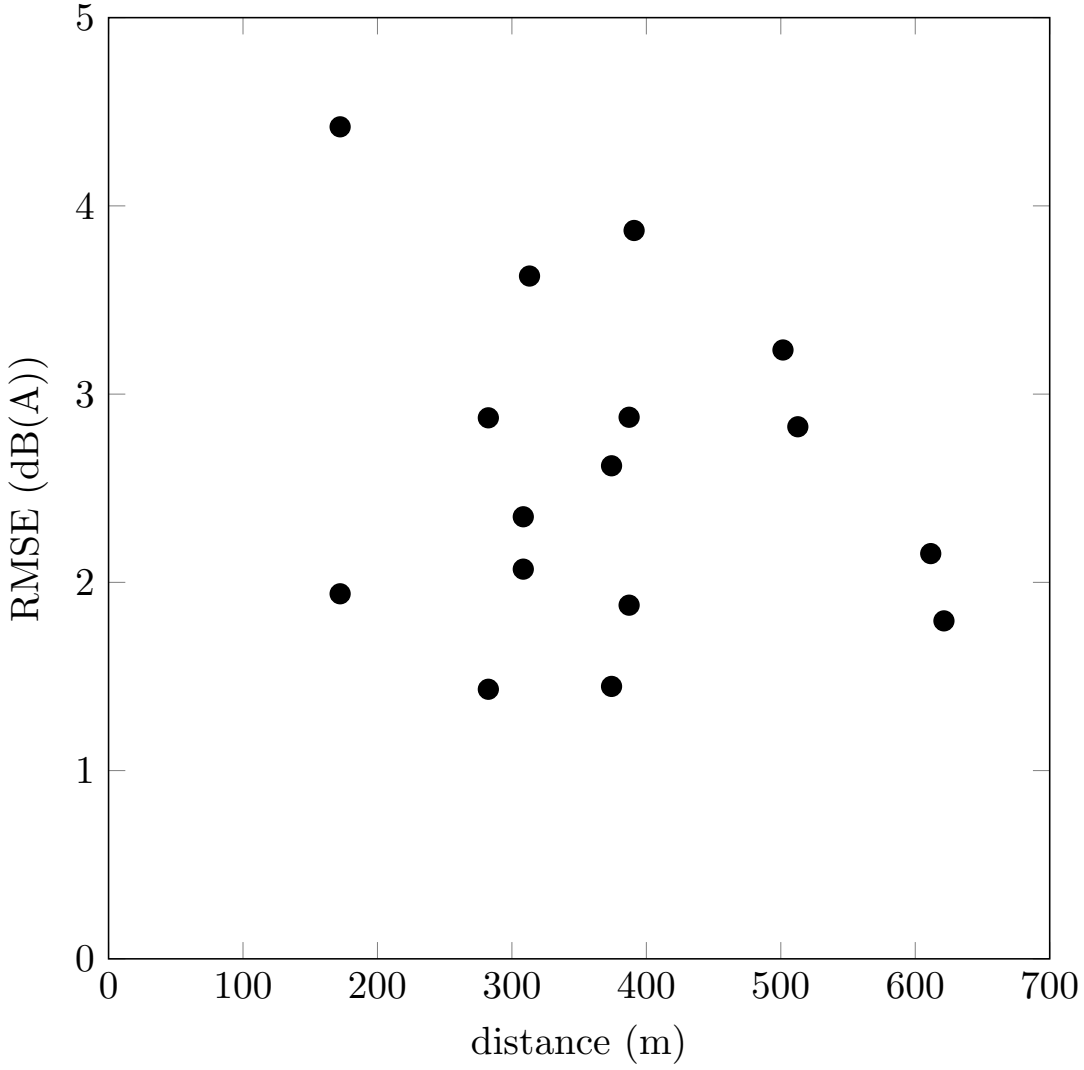


Figure IV.9. – RMSE performance of the JSPE without *a priori* parameters algorithm against the distance of the observation points to the network

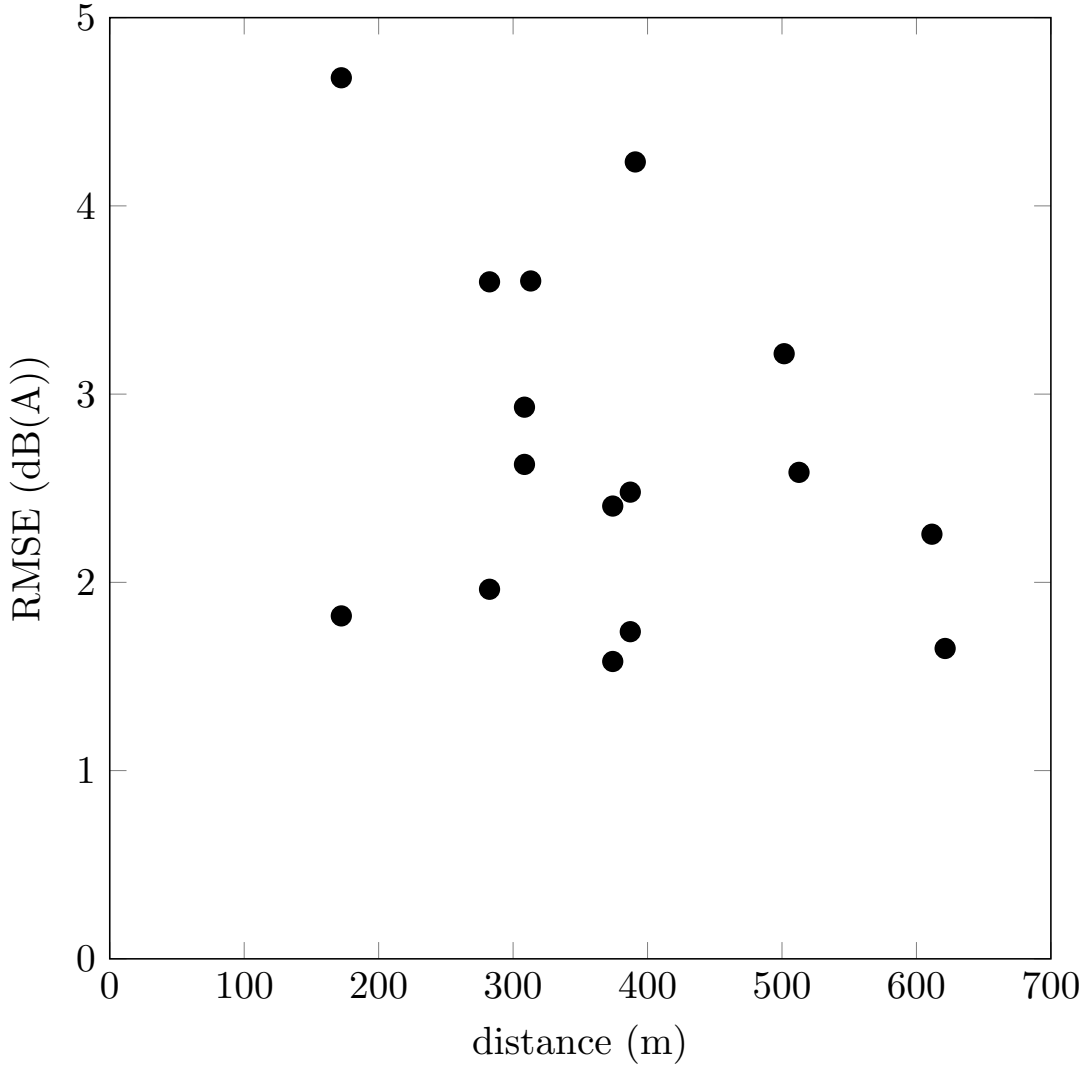


Figure IV.10. – RMSE performance of the IM without *a priori* parameters algorithm against the distance of the observation points to the network

Table IV.5. – Compared RMSE performance in dB(A)of the reception points when categorized according to the road section they are attached to (minor roads or major roads) for the JSPE and IM without *a priori* parameters algorithm.

| Algorithm | Minor Roads | Major Roads |
|---|---|---|
| IM | 2.72 dB(A) | 2.83 dB(A) |
| JSPE | 2.50 dB(A) | 2.63 dB(A) |

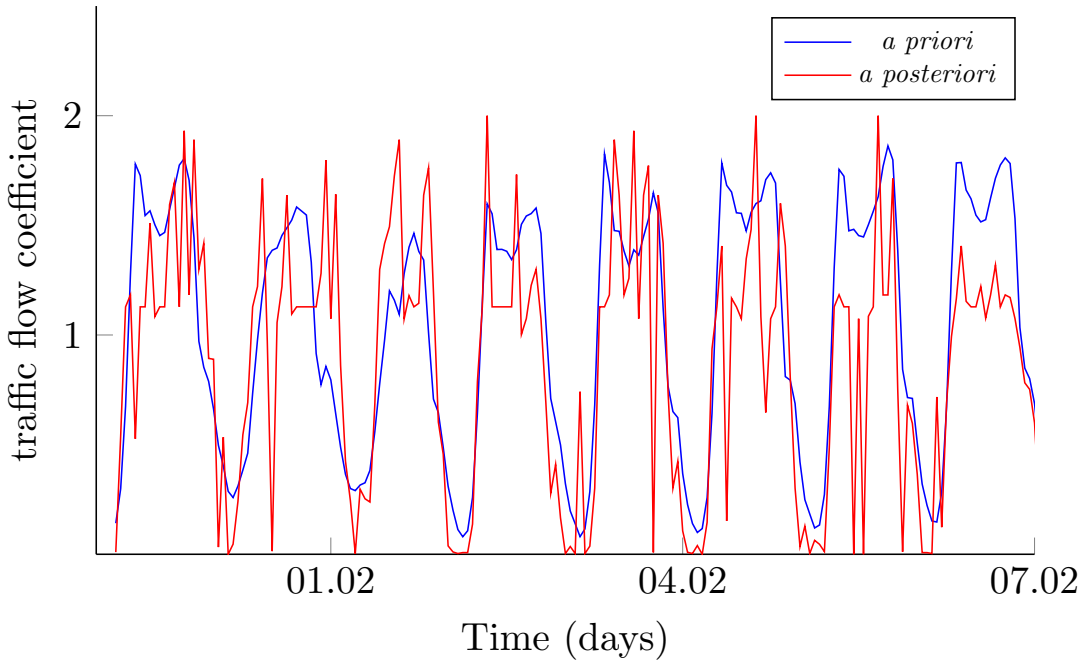


Figure IV.11. – *a priori* and *a posteriori* light vehicule flow of the meta-model.

### IV.5.4. Complexity of the meta-model

The meta-model used for this study has been built with a low number of parameters. It is possible to build a meta-model with a higher number of parameters, for example by discriminating the roads sections into categories by traffic densities with major, medium and minor road sections. Since it has been shown in (Barrigón Morillas et al., 2005) that road categories have different hourly traffic profiles. This might improve the accuracy of the process by giving different traffic parameters for each road sections.

### IV.5.5. difference between the IM and the JSPE

This study has shown a slightly better accuracy of the JSPE algorithm. However, it remains hard to tell what differs between the JSPE and the IM algorithms. The fact that the JSPE takes into account $\mathbf{B}$, the background error covariance matrix allows to think that the JSPE may shows sensibly better results when the observed noise differs a lot from the predicted noise. Since the IM algorithm is simpler to implement, it can be recommanded to quickly perform the data assimilation process. However, the JSPE

is more robust to high variation between noise predictions and observations.

### IV.5.6. Network optimisation

The observation points have originally been distributed over the study area for an other study, and then originally fulfilled an other purpose than the data assimilation technique discussed in this paper. An insightful development following this paper would be to discuss the optimal distribution of the network across the study area and see whether some strategic locations significatively enhance the performance of the data assimilation procedure.

## IV.6. Conclusion

The inverse modeling (IM) and joint state parameter estimation (JSPE) without *a priori* parameters improved the performance of the prediction of respectively 21 % and 26 % compared to the regular meta-model prediction. The inverse modeling shows similar results as the data assimilation method BLUE (Best Linear Unbiased Estimator) only, whereas the JSPE always has a lower RMSE value than the state estimation only. The performance of IM and JSPE without *a priori* parameters remains very similar to the performance with these parameters. This means that the operators who wish to obtain a dynamic noise mapping for a given area can get a satisfying level of accuracy without the need to get real-time traffic and weather data, and extend the availability of dynamic noise mapping to areas where no traffic measurement is available. The method explored in this paper only requires a meta-model of a noise mapping software for the study area and a set of observed noise levels at given locations. It is then possible to compute dynamic noise maps with only a noise model and noise observations with no *a priori* knowledge of the traffic and weather data.

## Acknowledgments

The authors would like to thank the CENSE project (ref. ANR-16-CE22-0012) team for initiating this project, the Agence Nationale de la Recherche (ANR) for their financial support, the city of Paris and Bruitparif [5] for granting access to their data, the ADEME project GRAFIC (contract number 1317C0028) for the measurement campaign whose noise data were used for the data assimilation and Nicolas Fortin of the UMRAE department of the Université Gustave Eiffel for his technical support on the NoiseModelling software.

5. See https://www.bruitparif.fr/bruitparif/

# V. Estimation and diagnosis of the parameters of a covariance matrix

*This chapter is dedicated to the presentation of a new diagnosis method whose goal is to evaluate the accuracy of the estimated parameters of a covariance matrix. The covariance matrix of the error between the predicted and the observed noise level is a fonction which depends on unknown parameters. In chapter III, these parameters are the global variance $\sigma_b$, the characteristic legth $L_b$ and the characteristic noise level $L_d$. It is an enhancement of the so-called $\chi^2$ diagnosis which has been introduced in chapter III. This new method, called canonical estimation of the parameters will be introduced, a framing of its convergence speed will be proven. Then a numerical test of the canonical estimation will be conducted and compared to the current $\chi^2$ diagnosis on a toy model in order to highlight the benefits of this new method compared to the limitations of the $\chi^2$ diagnosis.*

**Table of contents**

This chapter is a transcription of the following article: Antoine Lesieur, Vivien Mallet, and Julien Salomon. Estimation and diagnosis of the parameters of a covariance matrix (submitted). *Quarterly Journal of the Royal Meteorological Society*, b.

## Abstract

Many data assimilation techniques based on the so-called best linear unbiased estimator often use parametrized covariance matrices to describe the background errors and the observational errors. The accuracy of the covariance matrices is often evaluated by the so-called $\chi^2$ diagnosis which is computationally cheap but may give rise to false positive. This paper proposes a new kind of diagnosis based on the canonical estimation of the covariance matrix whose optimal value uniquely defines the associated parameters. In addition, we show that this approach allows to approximate the parameters with an error of the form $\mathcal{O}(\frac{1}{\sqrt{n}})$, when $n$ observations are used. Finally, a numerical study is carried out on a particular class of covariance matrices chosen to be very sensitive to the method used, both for the approximation of the parameters and for the diagnosis.

## V.1. Introduction

In the data assimilation processes, a diagnosis of the covariance parameters is often used to check the accuracy of a covariance matrix, using samples supposed to follow a probability distribution with such covariance matrix. This numerical tool is crucial in the design of data assimilation processes such as BLUE (Best Linear Unbiased Estimator) (Liu et al., 2008). In this process, a key element, the Kalman matrix (Kalman, 1960), is often considered. Introduced in the 60's, this tool is widely used in automatic control. Since 1979 (Bryson et al., 1979), its application to data assimilation has also been proposed. The Kalman matrix is built from the prediction error covariance matrices. These matrices sometimes need to be appropriately parametrized in order to guarantee the performance of the assimilation process. Several applications make use of the covariance parameters diagnosis to tackle applications where data assimilation is required, as ,e.g., in meteorology, for assimilation of tracers (Ménard and Chang, 2000) or 4DVAR meteorological forecasting approaches (Ménard and Daley, 1996), air quality mapping (Tilloy et al., 2013), noise pollution mapping (Ventura et al., 2018),

The diagnosis used in all those works relies on the so-called $\chi^2$ diagnosis. This method has a low computational cost, hence, allows the data assimilation operator to quickly check the accuracy of the parameters she works with. However, this method suffers from flaws which might result in a wrong evaluation of the matrix parameters. As an example, the solution of the equation $\mathbb{E}[\chi^2(\boldsymbol{\theta})] = p$ is not uniquely defined, though at the basis of the approach. It follows that one might get a good score for a set of parameters which actually poorly describes the covariance matrix.

In this paper, we circumvent this limitation by proposing an alternative method which gives the best score only for the exact parameters of the covariance matrix and thus guarantees that a good diagnosis score actually corresponds to accurate estimated parameters. This method is inspired by the work presented in (Desroziers et al., 2005), and relies on the *canonical estimation* of the covariance between the terms of the covariance matrix itself. As a by-product, it can be adapted to find the best parameters by means of an optimization algorithm. We will see that the $\chi^2$ is unable to fulfill this task.

This paper is organized as follows. The case study is presented in section V.2. In

section V.3, we recall the basics on $\chi^2$ Diagnosis and detail its strengths and weaknesses. The canonical estimation method is presented in section V.4. We then illustrate the performance of this method by numerical experiments in section V.5, where we compare it to $\chi^2$ and test its robustness, study its convergence behavior for the parameters estimation, and its reliability for the diagnosis.

Thorough out this paper, $\mathrm{tr}(\mathbf{M})$ denotes the trace of a matrix $\mathbf{M}$, $\mathrm{Cov}(\phi)$ denotes the covariance matrix associated with a random sample $\phi$. Given two integers $k$ and $k'$, the space of continuously differentiable function form $\mathbb{R}^k$ to $\mathbb{R}^{k'}$ is denoted by $\mathcal{C}^1(\mathbb{R}^k,\ \mathbb{R}^{k'})$. The set of real symmetric matrices of size $p$ is denoted by $\mathcal{S}_p(\mathbb{R})$.

## V.2. Presentation of the case study

Given $n > 0$, we consider a sequence $(\boldsymbol{\nu}^{(k)})_{k\in[\![1,\ n]\!]}$ of independent identically distributed random vectors of dimension $p$. For the sake of simplicity, we often denote by $\boldsymbol{\nu}$ any arbitrary $\boldsymbol{\nu}^{(k)}$, in what follows. Its distribution is unknown with $\mathbb{E}[\boldsymbol{\nu}] = 0$. Let us denote by $\mathbf{S}^\star$ its covariance matrix, i.e., $\mathbf{S}^\star := \mathbb{E}[\boldsymbol{\nu}\boldsymbol{\nu}^T]$. Assume the latter belongs to a class of parametrized matrices $\mathbf{S}(\boldsymbol{\theta})$ with a smooth dependence with respect to $\boldsymbol{\theta} \in \mathbb{R}^\ell$, say $\mathbf{S} \in \mathcal{C}^1(\mathbb{R}^\ell,\ \mathbb{R}^{\frac{p(p+1)}{2}})$, since $\dim(\boldsymbol{\nu}) = p,\ \mathbf{S}^\star \in \mathbb{R}^{p\times p}$ with $\mathbf{S}$ symetrical, hence $\mathbf{S}^\star \in \mathcal{S}_p(\mathbb{R}) \hookrightarrow \mathbb{R}^{\frac{p(p+1)}{2}}$. We suppose moreover that there exists a unique $\boldsymbol{\theta}^\star$ such that $\mathbf{S}^\star = \mathbf{S}(\boldsymbol{\theta}^\star)$. In the case where the sequence $(\boldsymbol{\nu}^{(k)})_{k\in[\![1,\ n]\!]}$ is associated with $n$ observed data, we aim at building two processes, namely

— Estimation: find the best estimation $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}^\star$;

— Diagnosis: for a given $\tilde{\boldsymbol{\theta}}$, check if it approaches well $\boldsymbol{\theta}^\star$.

In section V.3, we explain how $\chi^2$ diagnosis applies to the later issue and show that it fails to tackle the former. We then present in section V.4 an alternative method which outperforms the $\chi^2$ approach in both of these tasks.

## V.3. State of the art: $\chi^2$ diagnosis

$\chi^2$ diagnosis has been introduced in (Dee, 1995) to test the so-called *perfect-model* assumption for 4DVAR applications (Courtier et al., 1994) in atmospheric data assimilation. For the sake of completeness, we recall how this approach proceeds.

### V.3.1. Presentation of $\chi^2$ diagnosis

$\chi^2$ diagnosis is a method used in data assimilation to check the accuracy of estimated covariance parameters. It proceeds as follows. For a sample of size $n$ and a parameters vector $\tilde{\boldsymbol{\theta}}$, define:

$$\chi^2_n(\tilde{\boldsymbol{\theta}}) := \frac{1}{n}\sum_{k=1}^{n} {\boldsymbol{\nu}^{(k)}}^T \mathbf{S}^{-1}(\tilde{\boldsymbol{\theta}})\boldsymbol{\nu}^{(k)}.$$

**Lemma 2.** *Keeping the notation of section V.2, we have:*

$$\forall k \in \mathbb{N},\ \mathbb{E}[{\boldsymbol{\nu}^{(k)}}^T \mathbf{S}^{-1}(\boldsymbol{\theta}^\star)\boldsymbol{\nu}^{(k)}] = p. \tag{V.1}$$

The proof of this result is given in C.1.1.1. All $\boldsymbol{\nu}^{(k)}$ are i.i.d with an expected value of 0, the use of Lemma 2 and the law of large numbers gives:

$$\chi^2_n(\boldsymbol{\theta}^\star) \xrightarrow{\mathcal{L}} p.$$

The value $\frac{\chi^2_n(\tilde{\boldsymbol{\theta}})}{p}$ consequently reads as a score which indicates how the estimate parameters are compliant with the observed data. When this quantity is close to one, the operator shall consider that the set of parameters $\tilde{\boldsymbol{\theta}}$ accurately reflects $\boldsymbol{\theta}^\star$. However, the set $E := \left\{\boldsymbol{\theta} | \frac{\mathbb{E}[\chi^2(\boldsymbol{\theta})]}{p} = 1\right\}$ is generally not a singleton, but a manifold of dimension $p-1$. Considering that $E_n := \left\{\boldsymbol{\theta} | \frac{\chi^2_n(\boldsymbol{\theta})}{p} = 1\right\}$ approaches $E$ a $n$ goes to infinity, it follows that finding an element $\boldsymbol{\theta}'$ of $E_n$ does not mean that the latter is close to $\boldsymbol{\theta}^\star$, i.e., $\lim_{n\to\infty} \|\boldsymbol{\theta}' - \boldsymbol{\theta}^\star\| = 0$, but that $\lim_{n\to\infty} \inf_{\mathbf{x}\in E} \|\boldsymbol{\theta}' - \mathbf{x}\| = 0$, with $\|\cdot\|$ the Euclidian norm. For example, if $\mathbf{S}(\boldsymbol{\theta}^\star) = \begin{pmatrix} \theta_1^\star & 0 \\ 0 & \theta_2^\star \end{pmatrix}$, then $\mathbb{E}[\boldsymbol{\nu}^T \mathbf{S}^{-1}(\boldsymbol{\theta}^\star)\boldsymbol{\nu}] = 2$. However, any parameter of the form $\boldsymbol{\theta}' = (\frac{\theta_1}{1+a\theta_1} \ \frac{\theta_2}{1+b\theta_2})^T$ where $a\theta_1 + b\theta_2 = 0$ also satisfies $\mathbb{E}[\boldsymbol{\nu}^T \mathbf{S}^{-1}(\boldsymbol{\theta}')\boldsymbol{\nu}] = 2$. Hence, a good score does not implies that $\tilde{\boldsymbol{\theta}}$ fairly approximates $\boldsymbol{\theta}^\star$.

### V.3.2. $\chi^2$ for parameter estimation

Besides diagnosis, one might try to use the $\chi^2$ diagnosis to estimate the parameter associated with a random vector. This approach proceeds as follows. Starting with the functional

$$\mathcal{J}_{n,\chi^2}(\boldsymbol{\theta}) := (\chi^2_n(\boldsymbol{\theta}) - p)^2,$$

one defines the approximation $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}^\star$ as a minimizer of $\mathcal{J}_{n,\chi^2}$, i.e., in such a way that $\chi^2_n(\hat{\boldsymbol{\theta}}_n)$ is optimally close to $p$. Since $\boldsymbol{\theta} \mapsto \mathbf{S}(\boldsymbol{\theta})$ is assumed to be differentiable, a Newton-like gradient algorithm such as BFGS ((Fletcher, 2013), (Bertsekas, 1997)) can be applied to compute a minimum value. Recall that in such a procedure, the gradient $\nabla\mathcal{J}_{n,\chi^2}$ of the functional is required. In our case, it can be computed by

$$\begin{aligned} \nabla\mathcal{J}_{n,\chi^2}(\boldsymbol{\theta}) &= 2(\chi^2_n(\boldsymbol{\theta}) - p)\nabla\bar{\chi}^2_n(\boldsymbol{\theta}) \\ &= -\frac{2}{n}(\chi^2_n(\boldsymbol{\theta}) - p)\sum_{i=1}^{n} \boldsymbol{\nu}^{(k)T} \mathbf{S}(\boldsymbol{\theta})^{-1} \nabla\mathbf{S}(\boldsymbol{\theta})\mathbf{S}(\boldsymbol{\theta})^{-1}\boldsymbol{\nu}^{(k)}. \end{aligned} \tag{V.2}$$

Because of the above mentioned possible local minima, the final value may depend on the initialization of the algorithm. This procedure is tested in section V.5.

## V.4. A new approach: canonical estimation parametrization

In this section, we propose a new approach, the *canonical estimation parametrization* whose purpose is to estimate accurately the parameters of the covariance matrix. This approach proceeds in three steps, namely, the canonical estimation of the covariance function, which is presented in section V.4.1, the computation of the covariance matrix of the vectorized estimated covariance matrix, which is detailed in section V.4.2 and the solving of a minimization problem in the framework of parameters estimation and diagnosis. This last step is presented in section V.4.3.

### V.4.1. Canonical estimation the covariance matrix

Define $\mathbf{S} := \mathrm{Cov}(\boldsymbol{\nu}, \boldsymbol{\nu})$ and $\hat{\mathbf{S}}_n$ its canonical estimation. Recall that $(\boldsymbol{\nu}^{(k)})_{k\in[\![1,\ n]\!]}$ denotes a sequence of independent identically distributed random vectors of dimension $p$, with an unknown distribution and a covariance matrix of the form $\mathbf{S}(\boldsymbol{\theta}^\star)$. We define the *canonical estimator* $\hat{\mathbf{S}}_n \in \mathcal{S}_p(\mathbb{R})$ of $\mathbf{S}(\boldsymbol{\theta}^\star)$, by:

$$[\hat{\mathbf{S}}_n]_{i,j} := \frac{1}{n-1}\sum_{k=1}^{n}(\boldsymbol{\nu}_i^{(k)} - \bar{\boldsymbol{\nu}}_i)(\boldsymbol{\nu}_j^{(k)} - \bar{\boldsymbol{\nu}}_j), \tag{V.3}$$

where $\boldsymbol{\nu}_i^{(k)}$ and $\bar{\boldsymbol{\nu}}_i$ denotes the $i$-th components of $\boldsymbol{\nu}^{(k)}$ and $\bar{\boldsymbol{\nu}}$ respectively, with $\bar{\boldsymbol{\nu}} := \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\nu}^{(k)}$.

### V.4.2. Expression and estimation of the covariance matrix of $\hat{\mathrm{s}}_n$

Define the projection $\phi$ of $\mathcal{S}_p(\mathbb{R})$ onto $\mathbb{R}^{\frac{p(p+1)}{2}}$ by:

$$\begin{array}{rccc} \phi & : & \mathcal{S}_p(\mathbb{R}) & \to & \mathbb{R}^{\frac{p(p+1)}{2}} \\ & & \mathbf{M} & \mapsto & \mathbf{m} = ([\mathbf{M}]_{ij})_{\substack{(i,j)\in[\![1,p]\!]^2 \\ i\leqslant j}}. \end{array}$$

Introduce then $\mathbf{s}(\boldsymbol{\theta}) := \phi(\mathbf{S}(\boldsymbol{\theta}))$ and $\hat{\mathbf{s}}_n := \phi(\hat{\mathbf{S}}_n)$. We now wish to find an expression of the covariance matrix of $\hat{\mathbf{s}}_n$ $\boldsymbol{\Gamma}_n := \mathrm{Cov}(\hat{\mathbf{s}}_n)$, in order to derive an approximation $\widehat{\boldsymbol{\Gamma}}_n$ based on the random sample $(\boldsymbol{\nu}^{(k)})_{k\in[\![1,\ n]\!]}$.

#### V.4.2.1. Exact expression of $\boldsymbol{\Gamma}_n$

Define the centralized random vector $\boldsymbol{\eta}^{(k)}$ by:

$$\forall (k,i) \in [\![1,\ n]\!] \times [\![1,\ p]\!],\ \boldsymbol{\eta}_i^{(k)} := \boldsymbol{\nu}_i^{(k)} - \mathbb{E}[\boldsymbol{\nu}_i].$$

Given $i,\ j,\ \ell, m \in [\![1,\ p]\!]^4$, and $a,b,c,d \in \{0,1\}^4$, we introduce the notation:

$$\begin{aligned} \sigma_{abcd} &:= \sum_{k=1}^{n} \boldsymbol{\eta}_i^{(k)a}\boldsymbol{\eta}_j^{(k)b}\boldsymbol{\eta}_\ell^{(k)c}\boldsymbol{\eta}_m^{(k)d}, \\ \mu_{abcd} &:= \mathbb{E}[\boldsymbol{\eta}_i^a\boldsymbol{\eta}_j^b\boldsymbol{\eta}_\ell^c\boldsymbol{\eta}_m^d]. \end{aligned}$$

Assuming further that $i \leqslant j$ and $m \leqslant \ell$, we define $i',\ j' \in [\![1,\ \frac{p(p+1)}{2}]\!]$ such that $\hat{\mathbf{s}}_{n,i'} = \hat{\mathbf{S}}_{n,i,j}$ and $\hat{\mathbf{s}}_{n,j'} = \hat{\mathbf{S}}_{n,m,\ell}$ and introduce:

$$\boldsymbol{\Gamma}_n := \mathrm{Cov}(\hat{\mathbf{s}}_n) \in \mathcal{S}_{\frac{p(p+1)}{2}}(\mathbb{R}),$$

so that:

$$[\boldsymbol{\Gamma}_n]_{i'j'} = \mathrm{Cov}([\hat{\mathbf{s}}_n]_{i'}, [\hat{\mathbf{s}}_n]_{j'}) = \mathrm{Cov}([\hat{\mathbf{S}}_n]_{i,j}, [\hat{\mathbf{S}}_n]_{m,\ell}).$$

**Lemma 3.** *The matrix $\boldsymbol{\Gamma}_n$ satisfies:*

$$[\boldsymbol{\Gamma}_n]_{i'j'} = -\frac{(n-1)^2}{n^3}\mu_{1100}\mu_{0011} + \frac{n-1}{n^3}(\mu_{1001}\mu_{0110} + \mu_{1010}\mu_{0101}) + \frac{(n-1)^2}{n^3}\mu_{1111}. \tag{V.4}$$

The proof is detailed in C.1.2. Thanks to (V.4), we see that $\mathbf{\Gamma}_n$ takes the form $\mathbf{\Gamma}_n = \frac{\mathbf{A}}{n} + \frac{\mathbf{B}}{n^2} + \frac{\mathbf{C}}{n^3}$, where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are some given matrices. Here, $\mathbf{A}$ is called the *asymptotical covariance matrix*. Because of (V.4), we get:

$$\mathbf{A} := \mu_{1111} - \mu_{1100}\mu_{0011}. \tag{V.5}$$

We deduce that the inverse of $\mathbf{\Gamma}_n$ satisfies:

$$\mathbf{\Gamma}_n^{-1} = n\mathbf{A}^{-1} - \mathbf{A}^{-2}\mathbf{B} + o(1).$$

This expansion will be used in section V.4.3.1.

#### V.4.2.2. Canonical estimation of $\mathbf{\Gamma}_n$

We now use (V.4) to define $\widehat{\mathbf{\Gamma}}_n$. Since the moments $\mu_{abcd}$ appearing in (V.4) are unknown, we approximate them by their empirical approximation and introduce:

$$[\widehat{\mathbf{\Gamma}}_n]_{i'j'} := -\frac{(n-1)^2}{n^3}\hat{\mu}_{1100}\hat{\mu}_{0011} + \frac{n-1}{n^3}(\hat{\mu}_{1001}\hat{\mu}_{0110} + \hat{\mu}_{1010}\hat{\mu}_{0101}) + \frac{(n-1)^2}{n^3}\hat{\mu}_{1111} \tag{V.6}$$

where:

$$\hat{\mu}_{abcd} := \frac{1}{n-1}\sum_{k=1}^{n}\left(\boldsymbol{\nu}_i^{(k)} - \bar{\boldsymbol{\nu}}_i\right)^a\left(\boldsymbol{\nu}_j^{(k)} - \bar{\boldsymbol{\nu}}_j\right)^b\left(\boldsymbol{\nu}_\ell^{(k)} - \bar{\boldsymbol{\nu}}_k\right)^c\left(\boldsymbol{\nu}_m^{(k)} - \bar{\boldsymbol{\nu}}_\ell\right)^d.$$

As opposed to $\mu_{abcd}$, $\hat{\mu}_{abcd}$ is computable from the random sample $(\boldsymbol{\nu}^{(k)})_{k\in[\![1,\ n]\!]}$.

Let us now recall the Landau notation $o_{\mathbb{P}}$ for the negligibility in probability. Given a sequence of random variables $(X_n)_n$ and $(u_n)_n$ a sequence of real numbers, we consider the formal following definition:

$$X_n = o_{\mathbb{P}}(u_n) \Leftrightarrow \forall\epsilon > 0,\ \forall\eta > 0,\ \exists n_0 \in \mathbb{N}, \forall n \geq n_0,\ \mathbb{P}\left(\frac{|X_n|}{u_n} \leqslant \epsilon\right) \geqslant 1 - \eta.$$

We then have $\hat{\mathbf{A}} = \mathbf{A} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$, $\hat{\mathbf{B}} = \mathbf{B} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$ and $\hat{\mathbf{C}} = \mathbf{C} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$, so that

$$\hat{\mathbf{\Gamma}}_n = \mathbf{\Gamma}_n + o_{\mathbb{P}}\left(\frac{1}{n\sqrt{n}}\right),\ \hat{\mathbf{\Gamma}}_n^{-1} = \mathbf{\Gamma}_n^{-1} + o_{\mathbb{P}}\left(\frac{1}{n\sqrt{n}}\right) = n\mathbf{A}^{-1} + o_{\mathbb{P}}(n). \tag{V.7}$$

### V.4.3. Applications to parameters estimation and diagnosis

In this section, we present two kinds of uses of our canonical estimation: the parameters estimation in section V.4.3.1 and the diagnosis in section V.4.3.2.

#### V.4.3.1. Parameters estimation

As done with the $\chi^2$ approach in section V.3.2, we formulate the parameters estimation problem in terms of an optimization problem. In this way, we introduce the functional

$$\mathcal{J}_{n,\mathbf{I}}(\boldsymbol{\theta}) := \|\mathbf{s}(\boldsymbol{\theta}) - \hat{\mathbf{s}}_n\|^2, \tag{V.8}$$

that has to be minimized. This method aims at computing the $\hat{\boldsymbol{\theta}}_n$ which satisfies the Least Mean Square criterion regarding the estimated covariance matrix $\hat{\mathbf{S}}_n$. The functional is differentiable, and the components of its gradient are:

$$\nabla \mathcal{J}_{n,\mathbf{I}}(\boldsymbol{\theta})_i = \left( 2 \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \theta_i}^T \left( \mathbf{s}(\boldsymbol{\theta}) - \hat{\mathbf{s}}_n \right) \right)_i . \tag{V.9}$$

It is then possible to compute $\hat{\boldsymbol{\theta}}$ with a Newton-like gradient algorithm. However, this method is not robust if the condition number is high, as observed in section V.5.1.2, where a numerical experiment shows how this cost function is sensitive to the condition number. Hence, a covariance matrix with a high condition number may lead to a poor estimation of $\hat{\boldsymbol{\theta}}_n$.

Let us now consider:

$$\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}(\boldsymbol{\theta}) := \left( \mathbf{s}(\boldsymbol{\theta}) - \hat{\mathbf{s}}_n \right)^T \widehat{\boldsymbol{\Gamma}}_n^{-1} \left( \mathbf{s}(\boldsymbol{\theta}) - \hat{\mathbf{s}}_n \right).$$

The functional $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$ is the usual log-likelihood of a Gaussian distribution $\mathcal{N}(\mathbf{s}(\boldsymbol{\theta}), \hat{\boldsymbol{\Gamma}}_n)$ (Fisher, 1922), hence this estimation method is called the *maximum likelihood estimation*. The associated minimization problem is a more sophisticated version of (V.8), in the sense that it does take into account the covariance of $\hat{\mathbf{s}}_n$, which is grasped in (V.8) by $\widehat{\boldsymbol{\Gamma}}_n$. Note that estimating $\boldsymbol{\theta}^\star$ using $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$ is not as straightforward as using $\mathcal{J}_{n,\mathbf{I}}$ because it requires the computation of $\hat{\boldsymbol{\Gamma}}_n$ which is a dense matrix. As was the case for the functionals $\mathcal{J}_{n,\chi^2}$ and $\mathcal{J}_{n,\mathbf{I}}$, its gradient can also be computed explicitly. In this case, its components are given by:

$$\nabla \mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}(\boldsymbol{\theta})_i = \left( 2 \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \theta_i}^T \hat{\boldsymbol{\Gamma}}_n^{-1} (\mathbf{s}(\boldsymbol{\theta}) - \hat{\mathbf{s}}_n) \right)_i . \tag{V.10}$$

As was the case for $\mathcal{J}_{n,\mathbf{I}}$, we can consequently make use of a steepest descent algorithm to optimize $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$.

We now wish to bound the error $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\star$. Consider the *confidence region*:

$$\Lambda_n^\alpha = \left\{ \mathbf{s} \in \mathbb{R}^{\frac{p(p+1)}{2}} | n(\mathbf{s} - \hat{\mathbf{s}}_n)^T \mathbf{A}^{-1} (\mathbf{s} - \hat{\mathbf{s}}_n) \leqslant z_\alpha \right\}, \tag{V.11}$$

where $\mathbf{A}$ is defined by (V.5) and $z_\alpha$ is the quantile of order $1-\alpha$ of the $\chi^2$ law with $\frac{p(p+1)}{2}$ degrees of freedom.

The following preliminary results will be used in the proof of Theorem 1.

Define $\mathcal{O}_{\mathbb{P}}$ the Landau notation for the domination in probability by

$$X_n = \mathcal{O}_{\mathbb{P}}(u_n) \Leftrightarrow \forall \eta \in [0,1],\ \exists M_\eta,\ \forall n \in \mathbb{N},\ \mathbb{P}\left( \frac{X_n}{u_n} \leqslant M_\eta \right) \geqslant 1 - \eta,$$

where $(X_n)_n$ is a sequence of random variables and $(u_n)_n$ a sequence of real numbers. We have the following result.

**Lemma 4.** *Let* $\mathbf{s}(\boldsymbol{\theta})$ *introduced in section V.4.2,* $\widehat{\boldsymbol{\Gamma}}_n$ *defined in* (V.6)*. If:*

$$\boldsymbol{v} := \frac{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta})}{\partial \theta_i} \widehat{\boldsymbol{\Gamma}}_n^{-1}}{\left\| \frac{\partial \mathbf{s}^T(\boldsymbol{\theta})}{\partial \theta_i} \widehat{\boldsymbol{\Gamma}}_n^{-1} \right\|_1}$$

*then:*

$$\boldsymbol{v} = \boldsymbol{v}_1 + (\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \mathbf{D}^T_{\partial\theta_i \mathbf{s}}(\boldsymbol{\theta}^\star)\mathcal{M}(\mathbf{p}) + \|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|_1 \boldsymbol{v}_2^T + o(\boldsymbol{\theta} - \boldsymbol{\theta}^\star) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right)$$

*with*

- $\boldsymbol{v}_1 := \dfrac{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1}$,
- $\mathcal{M}(\mathbf{p}) := \dfrac{\mathbf{A}^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1}$
- $\boldsymbol{v}_2^T = \dfrac{\left\|\left\|\mathbf{D}^T_{\partial\theta_i \mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}\right\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1^2}$.

The proof is given in C.1.3.1.

**Lemma 5.** *Let $(P_n)_n$, be a Boolean random variable. Define $\alpha \in [0,\ 1[$, such that $\forall n \in \mathbb{N},\ \mathbb{P}(P_n = True) \geq 1 - \alpha$. Let $(\sigma(n))_n \in \mathbb{N}^{\mathbb{N}}$ an increasing sequence of integers, $X_n$ be a real random variable and $(u_n)_n$ a real sequence such that*

$$\forall n \in \mathbb{N},\ (P_{\sigma(n)} = True \Rightarrow X_{\sigma(n)} = \mathcal{O}_{\mathbb{P}}(u_{\sigma(n)})). \tag{V.12}$$

*Then*

$$X_n = \mathcal{O}_{\mathbb{P}}(u_n).$$

The proof is given in C.1.3.2.

**Lemma 6.** *Let $\mathbf{s} : \mathbb{R}^\ell \to \mathbb{R}^p$ with $p \geq \ell$. Define $\boldsymbol{\theta}^\star \in \mathbb{R}^\ell$ such that $\mathbf{s}$ is differentiable at $\boldsymbol{\theta}^\star$ and whose Jacobian matrix $D_{\mathbf{s}}(\boldsymbol{\theta}^\star)$ satisfies $\dim(\Im(D_{\mathbf{s}}(\boldsymbol{\theta}^\star))) = \ell$. Then there exists $\lambda > 0$ such that for all $\boldsymbol{\theta} \in \mathcal{V}(\boldsymbol{\theta}^\star)$ a neighborhood of $\boldsymbol{\theta}^\star$:*

$$\frac{\|\mathbf{s}(\boldsymbol{\theta}) - \mathbf{s}(\boldsymbol{\theta}^\star)\|}{\|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|} \geqslant \lambda. \tag{V.13}$$

The proof is given in C.1.3.3. Let us now state our main result.

**Theorem 1.** *Let $(\boldsymbol{\nu}^{(k)})_{k \in [\![1,\ n]\!]}$ be a sequence of random vectors i.i.d. of dimension $p$. Let $\left\{\mathbf{S}(\boldsymbol{\theta}), \boldsymbol{\theta} \in \mathbb{R}^\ell\right\}$ be a class of parametrized covariance matrices, with $\boldsymbol{\theta} \mapsto \mathbf{S}(\boldsymbol{\theta}) \in \mathcal{C}^1(\mathbb{R}^\ell, \mathcal{S}_p(\mathbb{R}))$. Assume that there exists a unique $\boldsymbol{\theta}^\star \in \mathbb{R}^\ell$, such that $\mathbf{S}(\boldsymbol{\theta}^\star)$ is the covariance matrix of the $\boldsymbol{\nu}^{(k)}$, and that the Jacobian matrix $D_{\mathbf{s}}(\boldsymbol{\theta}^\star)$ of $\mathbf{s} : \boldsymbol{\theta} \to \mathbf{s}(\boldsymbol{\theta})$ is invertible. Suppose finally that for some $\epsilon > 0$, the sequence of approximations $(\hat{\boldsymbol{\theta}}_n)_{n \in \mathbb{N}}$ satisfies $\|\nabla \mathcal{J}_{n, \widehat{\boldsymbol{\Gamma}}_n}(\hat{\boldsymbol{\theta}}_n)\| \leqslant \epsilon$.*

*Then:*

$$\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\star\| = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).$$

*Proof.* In view of (V.11), a direct application of the central limit theorem gives:

$$\lim_{n \to \infty} \mathbb{P}(\mathbf{s}(\boldsymbol{\theta}^\star) \in \Lambda_n^\alpha) = 1 - \alpha. \tag{V.14}$$

As a consequence, there exists $(\sigma(n))_n \in \mathbb{N}^{\mathbb{N}}$ an increasing sequence of integers such that

$$\forall n \in \mathbb{N},\ \mathbf{s}(\boldsymbol{\theta}^\star) \in \Lambda^\alpha_{\sigma(n)},$$

and

$$\lim_{m\to\infty} \mathbb{P}(\exists n \in \mathbb{N}, \sigma(n) = m) = 1 - \alpha.$$

In view of Lemma 5, the purpose of this demonstration is to show that:

$$\forall n \in \mathbb{N}, \mathbf{s}(\boldsymbol{\theta}^\star) \in \Lambda^\alpha_{\sigma(n)} \Rightarrow \|\hat{\boldsymbol{\theta}}_{\sigma(n)} - \boldsymbol{\theta}^\star\| = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{\sigma(n)}}\right)$$

Let us then assume that $\forall n \in \mathbb{N}, m := \sigma(n),\ \mathbf{s}(\boldsymbol{\theta}^\star) \in \Lambda^\alpha_m$ and define $c := \sqrt{z_\alpha \lambda_{max}}$ and

$$I_m := \prod_{i=1}^{\frac{p(p+1)}{2}} \left[[\hat{\mathbf{s}}_m]_i - \frac{c}{\sqrt{m}},\ [\hat{\mathbf{s}}_m]_i + \frac{c}{\sqrt{m}}\right],$$

where $\lambda_{max}$ is the greatest eigenvalue of $\mathbf{A}$ (see (V.5)), hence $\Lambda^\alpha_m \subset I_m$. Define also for $i \in [\![1,\ \ell]\!]$ and $m \in \mathbb{N}$:

$$\boldsymbol{a}^T_{m,i} := 2\frac{\partial \mathbf{s}^T(\hat{\boldsymbol{\theta}}_m)}{\partial \theta_i}\widehat{\boldsymbol{\Gamma}}^{-1}_m. \tag{V.15}$$

In order to find a framing of the quantity $\|\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^\star\|$, we start by framing of the quantity $\boldsymbol{a}^T_{m,i}(\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \mathbf{s}(\boldsymbol{\theta}^\star))$. This one can be decomposed as follows:

$$\boldsymbol{a}^T_{m,i}(\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \mathbf{s}(\boldsymbol{\theta}^\star)) = \boldsymbol{a}^T_{m,i}(\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \hat{\mathbf{s}}_m) + \boldsymbol{a}^T_{m,i}(\hat{\mathbf{s}}_m - \mathbf{s}(\boldsymbol{\theta}^\star)).$$

Let us bound each of the two terms in the left-hand side. In view of (V.10) and $\|\nabla \mathcal{J}_{m,\widehat{\boldsymbol{\Gamma}}_m}(\hat{\boldsymbol{\theta}}_m)\| \leqslant \epsilon$, the former can be bounded by:

$$\forall i \in [\![1,\ \ell]\!],\ \left|\boldsymbol{a}^T_{m,i}(\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \hat{\mathbf{s}}_m)\right| \leqslant \epsilon \tag{V.16}$$

Given $i \in [\![1,\ \ell]\!]$, consider now the term $\boldsymbol{a}^T_{m,i}(\hat{\mathbf{s}}_m - \mathbf{s}(\boldsymbol{\theta}^\star))$. If the event $\mathbf{s}(\boldsymbol{\theta}^\star) \in \Lambda^\alpha_m$ happens, since $\Lambda^\alpha_m \subset I_m$,then the inner product of $\boldsymbol{a}_{m,i}$ with $\hat{\mathbf{s}}_m - \mathbf{s}(\boldsymbol{\theta}^\star)$ satisfies:

$$\left|\boldsymbol{a}^T_{m,i}(\hat{\mathbf{s}}_m - \mathbf{s}(\boldsymbol{\theta}^\star))\right| \leqslant \|\boldsymbol{a}_{m,i}\|_1 \frac{c}{\sqrt{m}}. \tag{V.17}$$

Combining equation (V.16) and (V.17), we get:

$$\left|\boldsymbol{a}^T_{m,i}(\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \mathbf{s}(\boldsymbol{\theta}^\star))\right| \leqslant \epsilon + \|\boldsymbol{a}_{m,i}\|_1 \frac{c}{\sqrt{m}}. \tag{V.18}$$

As an illustration, a 2D representation of the confidence region $\Lambda^\alpha_n$ is shown in Figure V.1. This region is an ellipse whose axes lengths are $\ell_i = \sqrt{\frac{z_\alpha \lambda_i}{m}}$, where $\lambda_i$ are the eigenvalues of $\mathbf{A}$ and directions are the eigenvectors of $\mathbf{A}$. Here, assume that the event $\mathbf{s}(\boldsymbol{\theta}^\star) \in \Lambda^\alpha_m$ occurs. with $(\mathbf{p}_i - \hat{\mathbf{s}}_m)_i$ the vertices of $I_m$, $d_0$ is the isolevel line of value $\boldsymbol{a}^T_{m,i}(\mathbf{s}(\boldsymbol{\theta}^\star) - \hat{\mathbf{s}}_m)$, $d_{-1}$ is the isolevel line of value $\boldsymbol{a}^T_{m,i}(\boldsymbol{p}_1 - \hat{\mathbf{s}}_m)$ and $d_1$ is the isolevel line of value $\boldsymbol{a}^T_{m,i}(\boldsymbol{p}_3 - \hat{\mathbf{s}}_m)$ for the function $\boldsymbol{x} \to \boldsymbol{a}^T_{m,i}\boldsymbol{x}$.

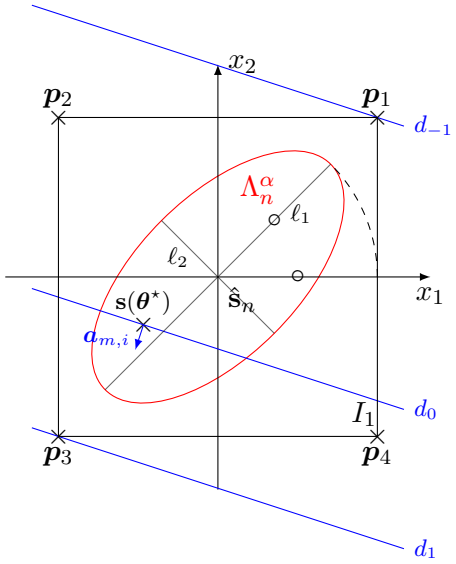


Figure V.1. – 2D representation of the confidence region $\Lambda_n^\alpha$.

Then we have in figure V.1:

$$\boldsymbol{a}_{m,i}^T(\boldsymbol{p}_1 - \hat{\mathbf{s}}_n) \leq \boldsymbol{a}_{m,i}^T(\mathbf{s}(\boldsymbol{\theta}^\star) - \hat{\mathbf{s}}_m) \leq \boldsymbol{a}_{m,i}^T(\boldsymbol{p}_3 - \hat{\mathbf{s}}_m) \tag{V.19}$$

then for the equation (V.19):

$$\boldsymbol{a}_{m,i}^T(\boldsymbol{p}_1 - \hat{\mathbf{s}}_n) = -\|\boldsymbol{a}_{m,i}\|_1 \frac{c}{\sqrt{m}},\ \boldsymbol{a}_{m,i}^T(\boldsymbol{p}_3 - \hat{\mathbf{s}}_n) = \|\boldsymbol{a}_{m,i}\| \frac{c}{\sqrt{m}}$$

Define now

$$d_m := \|\boldsymbol{a}_{m,i}\|_1 \tag{V.20}$$

$$\boldsymbol{v}_m{}^T := \frac{\boldsymbol{a}_{m,i}^T}{d_m c} = \frac{2}{d_m c} \frac{\partial \mathbf{s}^T(\hat{\boldsymbol{\theta}}_m)}{\partial \theta_i} \widehat{\boldsymbol{\Gamma}}_m^{-1}. \tag{V.21}$$

Equation (V.18) leads to the following inequalities concerning the inner product of $\boldsymbol{v}_m$ with $\hat{\mathbf{s}}_m - \mathbf{s}(\boldsymbol{\theta}^\star)$:

$$\left| \boldsymbol{v}_m{}^T(\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \mathbf{s}(\boldsymbol{\theta}^\star)) \right| \leqslant \frac{\epsilon}{d_m} + \frac{1}{\sqrt{m}} \tag{V.22}$$

On the other hand, we get from Lemma 4 that

$$c\boldsymbol{v}_m^T = \boldsymbol{v}_1^T + (\boldsymbol{\theta} - \boldsymbol{\theta}^\star)^T \mathbf{D}_{\partial\theta_i \mathbf{s}}^T(\boldsymbol{\theta}^\star)\mathcal{M}(\mathbf{p}) + \|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|_1 \boldsymbol{v}_2^T + o(\boldsymbol{\theta} - \boldsymbol{\theta}^\star) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{m}\right). \tag{V.23}$$

We then consider the scalar product of each sides of (V.23) with $(\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \mathbf{s}(\boldsymbol{\theta}^\star))$. In this way, let $\zeta_m \in [0,\ 2\pi[$ be such that:

$$\boldsymbol{v}_1{}^T(\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \mathbf{s}(\boldsymbol{\theta}^\star)) = \|\boldsymbol{v}_1{}^T\| \|\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \mathbf{s}(\boldsymbol{\theta}^\star)\| \cos(\zeta_m).$$

Using the Cauchy-Schwarz inequality we have:

$$\begin{aligned}&\left((\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{M}+\|\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star\|_1\boldsymbol{v}_2^T\right)(\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star))\\ \leq\ &\|(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{M}+\|\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star\|_1\boldsymbol{v}_2^T\|\|\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star)\|.\end{aligned}$$

The definition of $\boldsymbol{v}_m$ given by (V.21) gives:

$$\begin{aligned}c{\boldsymbol{v}_m}^T(\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star))=({\boldsymbol{v}_1}^T+(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{M}+\|\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star\|_1\boldsymbol{v}_2^T)(\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star))+\\ o(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)+\mathcal{O}_{\mathbb{P}}(\frac{1}{m})\\ \leq\|\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star)\|(\|{\boldsymbol{v}_1}^T\|\cos(\zeta_m)+\\ \|(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{M}+\|\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star\|_1\boldsymbol{v}_2^T\|)\\ +o(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)+\mathcal{O}_{\mathbb{P}}(\frac{1}{m})\end{aligned}\tag{V.24}$$

Combining (V.15) and (V.20), we get $d_m=\left\|2\frac{\partial\mathbf{s}^T(\hat{\boldsymbol{\theta}}_m)}{\partial\theta_i}\hat{\boldsymbol{\Gamma}}_m^{-1}\right\|$. According to (V.7), $\hat{\boldsymbol{\Gamma}}_m^{-1}=m\mathbf{A}^{-1}+o_{\mathbb{P}}(m)$, with $\lim_{m\to\infty}\frac{\partial\mathbf{s}^T(\hat{\boldsymbol{\theta}}_m)}{\partial\theta_i}=\frac{\partial\mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial\theta_i}$. It follows that we have $d_m=\left\|2\frac{\partial\mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial\theta_i}\mathbf{A}^{-1}m+o_{\mathbb{P}}(m)\right\|=m+o_{\mathbb{P}}(m)$. Hence

$$\frac{\epsilon}{d_m}=o_{\mathbb{P}}\left(\frac{1}{m}\right)=\mathcal{O}_{\mathbb{P}}\left(\frac{1}{m}\right)$$

Combining this last equality with (V.22) and (V.24), we find that:

$$\begin{aligned}&\frac{1}{c}\|\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star)\|(\|\boldsymbol{v}_1^T\|\,|\cos(\zeta_m)|+\|(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{M}+\|\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star\|_1\boldsymbol{v}_2^T\|\,||)\\ &\leqslant\frac{1}{\sqrt{m}}+o(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)+\mathcal{O}_{\mathbb{P}}\left(\frac{1}{m}\right).\end{aligned}\tag{V.25}$$

Since $\|\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star\|\to 0$ by construction of the output of the optimization algorithm, we have $\lim_{m\to\infty}\|(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{M}+\|\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star\|_1\boldsymbol{v}_2^T\|=0$.Consider now the quantity $\|\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star)\|\frac{\|\boldsymbol{v}_{m,1}^T\|\cos(\zeta_m)}{2c}$ (with 2 chosen arbitrarily). We have

$$\begin{aligned}&\|\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star)\|\frac{\|\boldsymbol{v}_{m,1}^T\|\,|\cos(\zeta_m)|}{2c}\\ &\leqslant\|\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star)\|(\|\boldsymbol{v}_{m,1}^T\|\,|\cos(\zeta_m)|+\|(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{M}+\|\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star\|_1\boldsymbol{v}_2^T\|\,||).\end{aligned}\tag{V.26}$$

Combining it with (V.25), we obtain

$$\|\mathbf{s}(\hat{\boldsymbol{\theta}}_m)-\mathbf{s}(\boldsymbol{\theta}^\star)\|\frac{\|\boldsymbol{v}_{m,1}^T\|\,|\cos(\zeta_m)|}{2c}\leqslant\frac{1}{\sqrt{m}}+o(\hat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}^\star)+\mathcal{O}_{\mathbb{P}}\left(\frac{1}{m}\right).$$

Define $A^{\phi}_{\boldsymbol{v}_1} = \{\mathbf{x} \in \mathbb{R}, |(\mathbf{x}, \boldsymbol{v}_1) - \frac{\pi}{2}| \leq \phi\} \subset \mathbb{R}^{\frac{p(p+1)}{2}}$. Thanks to the properties of the ellipsoid $\Lambda^{\alpha}_n$, we have

$$\mathbb{P}(\mathbf{s}(\boldsymbol{\theta}^\star) \in \Lambda^{\alpha}_n \backslash A^{\phi}_{\boldsymbol{v}_1}) \geq \mathbb{P}(\mathbf{s}(\boldsymbol{\theta}^\star) \in \Lambda^{\alpha}_n \backslash A^{\phi}_{\boldsymbol{v}_{\min}}) = 1 - \alpha',$$

with $\boldsymbol{v}_{\min}$ the eigenvector associated with $\lambda_{\min}$, the smallest eigenvalue of $\mathbf{A}$. Choose $\sigma(n) = m$ such that $\mathbf{s}(\boldsymbol{\theta}^\star) \in \Lambda^{\alpha}_m \backslash A^{\phi}_{\boldsymbol{v}_1}$ , see (V.14), the previous calculation remains valid, as well as the assumptions of Lemma 5. Using Lemma 6, combining (V.13) with (V.26) and $|\cos(\zeta_m)| \geq \cos(\phi)$, with $\cos(\phi) > 0$, we obtain that for $m$ large enough:

$$\begin{aligned}\|\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^\star\| \leqslant \frac{\|\mathbf{s}(\hat{\boldsymbol{\theta}}_m) - \mathbf{s}(\boldsymbol{\theta}^\star)\|}{\lambda} &\leqslant \frac{2}{\lambda\|\boldsymbol{v}_1\|\cos(\zeta_m)\sqrt{m}} + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right) + o(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^\star) \\ &\leqslant \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right) + o(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^\star),\end{aligned}$$

meaning that

$$\|\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^\star\| = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right).$$

The result follows from Lemma 5. □

#### V.4.3.2. Diagnosis

The canonical estimation diagnosis consists in quantifying the quality of $\tilde{\boldsymbol{\theta}}$, an estimation of $\boldsymbol{\theta}^\star$, by computing the value $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}(\tilde{\boldsymbol{\theta}})$. This method has pros and cons compared to $\chi^2$ diagnosis. On the one hand, the minimum is expected to be unique so that a small score shall imply a vicinity with the global optimum. On the other hand, the value 0 is an asymptotic expectation, by construction of $\mathcal{J}_{\widehat{\boldsymbol{\Gamma}}_n}$, and the optimal value with a finite sample size is unknown. An experimental comparison of the performance of the two strategies is proposed in section V.5.

## V.5. Performance

We first consider a case where $\boldsymbol{\nu} \in \mathbb{R}^5$, i.e., $p = 5$. The parameters are estimated from a random sample $(\boldsymbol{\nu}_i)_i$ iid where $\forall i,\ \boldsymbol{\nu}_i \sim \mathcal{N}(0, \mathbf{S}(\boldsymbol{\theta}^\star))$ where:

$$\mathbf{S}(\boldsymbol{\theta}) := \begin{pmatrix} \theta_1 & 0 & 0 & 0 & 0 \\ 0 & \theta_1 & 0 & 0 & 0 \\ 0 & 0 & \theta_1 & 0 & 0 \\ 0 & 0 & 0 & \theta_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda\theta_1\theta_2 \end{pmatrix},$$

whose condition number is $\kappa(\mathbf{S}(\boldsymbol{\theta})) = \lambda\theta_2$. We then fix $\boldsymbol{\theta}^\star = (1, 1)^T$, $\lambda = 10^5$ and set $\boldsymbol{\theta}_{init} = (20, 30)^T$ as initialization parameters for the optimization procedure. The random sample $(\boldsymbol{\nu}^k)_{k\in[\![1,\ n]\!]}$ is obtained by a pseudo random number generatorwhich simulates a Gaussian sample whose mean is set to 0 and covariance matrix is set to $\mathbf{S}(\boldsymbol{\theta}^\star)$. We consider a sample of size $n = 10^4$. The choice of $\mathbf{S}(\boldsymbol{\theta})$ is motivated by the fact that $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$ is sensitive to $\kappa(\mathbf{S}(\boldsymbol{\theta}))$ whereas $\mathcal{J}_{n,\mathbf{I}}$ does not depend on it, hence the possibility to asset the influence $\kappa(\mathbf{S}(\boldsymbol{\theta}))$ on the performance of the two approaches.

### V.5.1. Experimental Setup

The estimated value of $\boldsymbol{\theta}$ is computed using BFGS algorithm, meaning that both $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$ (resp. $\mathcal{J}_{n,\mathbf{I}}$) and its gradient $\nabla\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$ (resp. $\nabla\mathcal{J}_{n,\mathbf{I}}$) are required. Recall that the latter can be computed using the results of sections V.3.2 and V.4.3.1, see equations (V.2), (V.10), (V.9). The implementation is carried out by means of the Python **Scipy** library ((Virtanen et al., 2020)).

In this section, we first evaluate the robustness of the optimization algorithm with respect to the initialization value $\boldsymbol{\theta}_{init}$ (see section V.5.1.1). We then compare the efficiency of $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$ and $\mathcal{J}_{n,\mathbf{I}}$ regarding the performance of the respective optimization algorithms, in term of

— condition number of the matrix $\kappa(\mathbf{S}(\boldsymbol{\theta}))$, see section V.5.1.2,

— required size $n$ of the random sample, see section V.5.1.3.

Finally, as a parameter diagnosis process, the accuracy of our new diagnosis is compared to $\chi^2$ diagnosis in section V.5.2.

#### V.5.1.1. Robustness of the optimization algorithms

As an iterative process, the performance of the parameters identification algorithms shall depend on its initialization $\boldsymbol{\theta}_{init}$. The goal of our first test is to see how the resulting approximation $\hat{\boldsymbol{\theta}}_n$ is sensitive to $\boldsymbol{\theta}_{init}$. In our tests, a random sample of 100 initialization parameters $\boldsymbol{\theta}_{init}$ has been picked with $\boldsymbol{\theta}_{init}$ following a uniform distribution on $\prod_{i=1}^{k}[0, 2\theta_i^\star]$. The optimization algorithms have been applied to the same random sample of $(\boldsymbol{\nu}^{(k)})_{k\in[\![1,n]\!]}$ with $n = 10^3$. The results for the first component $|\hat{\theta}_1 - \theta_1^\star|$ are displayed in Figure V.2. We see that the canonical estimation method outperforms the $\chi^2$ approach since the distribution of the error in the optimal value $\hat{\boldsymbol{\theta}}_n$ is much more narrow with the former. Similar results have been obtained with $\theta_2$.

#### V.5.1.2. Performance with respect to $\kappa(\mathbf{S}(\boldsymbol{\theta}))$

In this section, we demonstrate the efficiency of our approach when dealing with large values of $\kappa(\mathbf{S}(\boldsymbol{\theta}))$, i.e., with poorly conditioned problems. Recall that in the considered case, $\kappa(\mathbf{S}(\boldsymbol{\theta})) = \lambda\theta_2$. Hence, if the order of magnitude of $\theta_2$ is 1, the order of magnitude of the condition number $\kappa(\mathbf{S}(\boldsymbol{\theta}))$ is equal to the order of magnitude of $\lambda$.

The functionals $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$ and $\mathcal{J}_{n,\mathbf{I}}$ have been designed using a fixed sample size $n = 10^4$ . The error obtained by both procedures is shown in Figure V.3, where we represent the median and the interdecile interval for each value of $\lambda$ and each optimization algorithm. Recall that the initial parameter $\boldsymbol{\theta}_{init}$ is fixed in all this test to $\boldsymbol{\theta}_{init} = (20,\ 30)^T$. For each evaluated condition number $\kappa(\mathbf{S}(\boldsymbol{\theta}))$, 20 trials have been carried out with different random samples with the same random seed (but with a different covariance matrix, hence the samples are different), and the optimization algorithm has been run with the same initilization value. We see that the performance (here measured by the error) of the optimization algorithm based on the functional $\mathcal{J}_{\widehat{\boldsymbol{\Gamma}}_n}$ computed with $\hat{\boldsymbol{\Gamma}}_n$ remains stable with respect to $\kappa(\mathbf{S}(\boldsymbol{\theta}))$ whereas when using $\mathcal{J}_{\mathbf{I}}$, the error increases when $\kappa(\mathbf{S}(\boldsymbol{\theta}))$ increases.

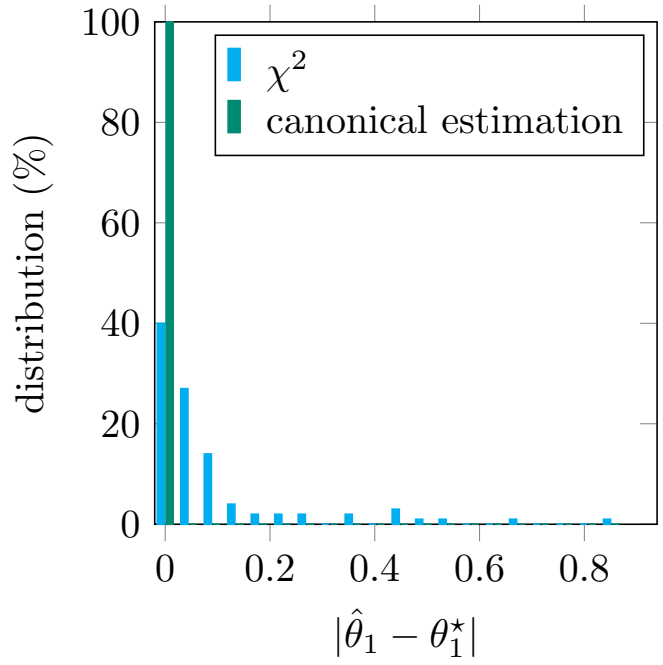


Figure V.2. – Distribution of the error $|\hat{\theta}_1 - \theta_1^\star|$ in the optimal parameter found with the $\chi^2$ procedure and the canonical estimation method.

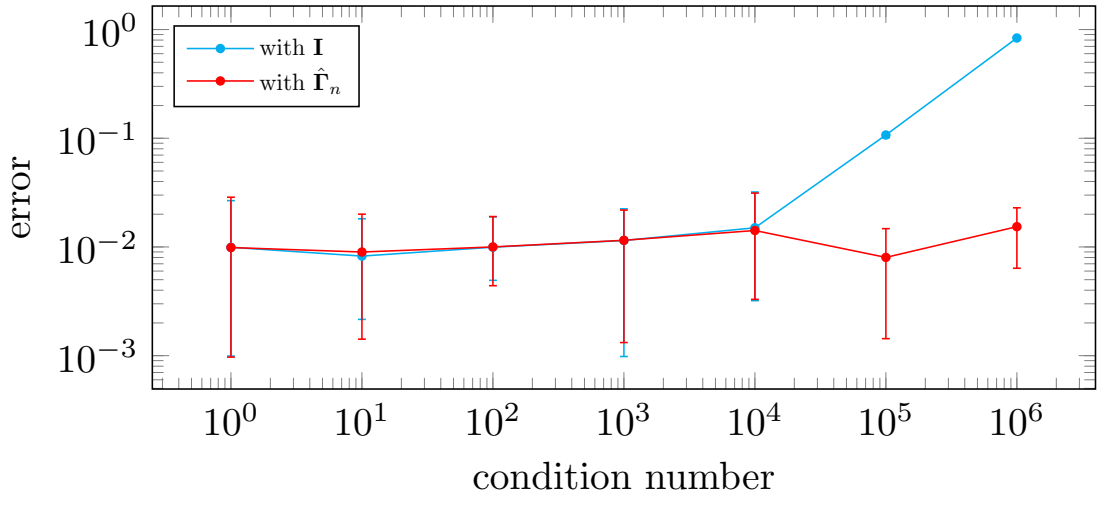


Figure V.3. – Estimation error with respect to the condition number $\kappa(\mathbf{S}(\boldsymbol{\theta}))$ of $\mathbf{S}$, the intervals correspond to the interdecile interval and the highlighted dots represents the median values.

### V.5.1.3. Rate of convergence

As shown is section V.4.3.1, the optimal value $\hat{\boldsymbol{\theta}}_n$ found with the canonical estimation approach converges toward the real parameters $\boldsymbol{\theta}^\star$ at a rate dominated by $\mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$. The purpose of this test is to check numerically this bound and to compare it with

the $\chi^2$ method. In addition, we compare the rate of convergence of the procedures associated to each cost functional. For each fixed sample size, 20 random samples have been tested for each optimization algorithm. The condition number $\kappa(\mathbf{S}(\boldsymbol{\theta}))$ has been fixed to $10^5$. The convergence rate for every procedure is shown in Figure V.4, where the median and the interdecile interval for each sample size and each optimization algorithm are displayed. As opposed to section V.5.1.1 and as in the previous section, the initialization parameter is the same for all the computations, with $\boldsymbol{\theta}_{init} = (20, 30)^T$.

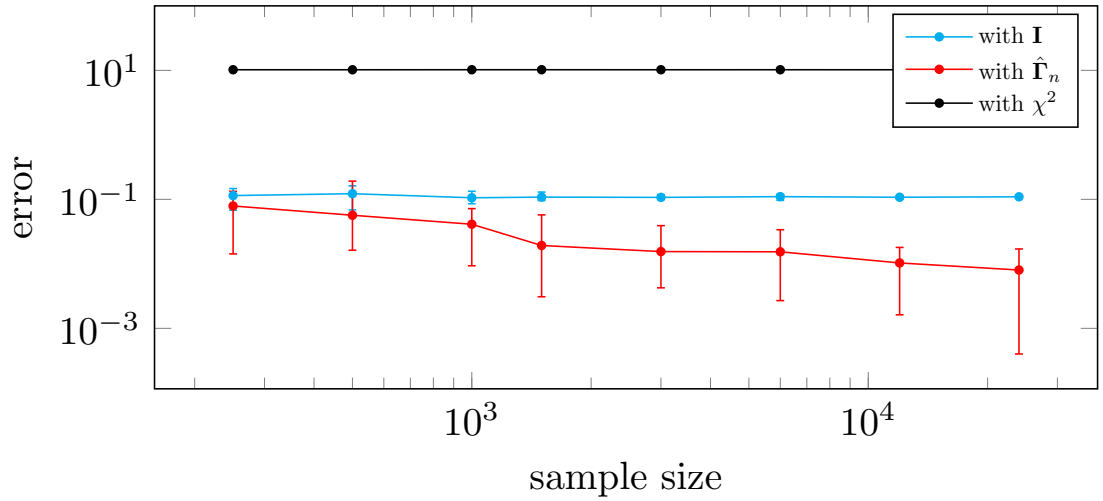


Figure V.4. – Estimation error against the sample size of $\boldsymbol{\nu}$ for several optimization algorithms

The $\chi^2$ method and the method based on $\mathcal{J}_{n,\mathbf{I}}$ are not influenced by the size of the sample regarding the accuracy of the results, whereas the accuracy of the method based on $\widehat{\boldsymbol{\Gamma}_n}$ improves as the sample size increases, which is inline with the convergence rate obtained in Theorem 1.

## V.5.2. Diagnosis

In this section, we aim at evaluating the benefit brought by $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$ compared to $\chi^2$ diagnosis only. In this way, we study the behavior of $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}(\boldsymbol{\theta})$ and $\chi^2(\boldsymbol{\theta})$ in a neighborhood of $\boldsymbol{\theta} = \boldsymbol{\theta}^\star$. Here, the sample size of $(\boldsymbol{\nu}^{(k)})_{k\in[\![1,n]\!]}$ and the condition number $\kappa(\mathbf{S}(\boldsymbol{\theta}))$ have been fixed to $10^3$ and $10^5$, respectively. The functionals $\mathcal{J}_{n,\chi^2}$ and $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}$ are represented as functions of $\theta_1$ and $\theta_2$ in Figures V.5 and V.6.

On the one hand, the canonical estimation method shows much better results than $\chi^2$ diagnosis. Contrary to the results obtained with $\chi^2$ diagnosis, the obtained values are centered around the real value $\boldsymbol{\theta}^\star$ and remain in a narrower domain of validity, as shown in the figures by the blue areas. On the other hand, with this diagnosis, the optimal value is not known *a priori*, whereas this value is $dim(\boldsymbol{\nu})$ with $\chi^2$ diagnosis. As a consequence, we recommend to combine the two approaches as follows:

— First, test the value $\tilde{\boldsymbol{\theta}}_{\chi^2}$, which shall provide an *a priori* value based on the knowledge of the problem we are studying. In this way, $\chi^2$ diagnosis enables to

check if the diagnosed $\tilde{\boldsymbol{\theta}}_{\chi^2}$ lies in an small enough interval, i.e., if $\chi^2$ diagnosis value is close enough to $dim(\boldsymbol{\nu})$.

— Second, check with the canonical estimation method new values within a neighborhood of $\tilde{\boldsymbol{\theta}}_{\chi^2}$ with a sieve or an optimization algorithm. The smaller $\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}(\tilde{\boldsymbol{\theta}}_{\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}})$, the diagnosis result is, the closer the diagnosed parameter $\tilde{\boldsymbol{\theta}}_{\mathcal{J}_{n,\widehat{\boldsymbol{\Gamma}}_n}}$ is to $\boldsymbol{\theta}^\star$.

The canonical estimation method clearly appears as an improvement over the $\chi^2$ diagnosis. The shape of the functional observed in Figure V.6 shows that usual steepest descent algorithms will converge faster to the real value $\boldsymbol{\theta}^\star$. However, because the optimal value of the functional $\mathcal{J}_{\widehat{\boldsymbol{\Gamma}}_n}$ is unknown, it cannot be a substitute of $\chi^2$ diagnosis.

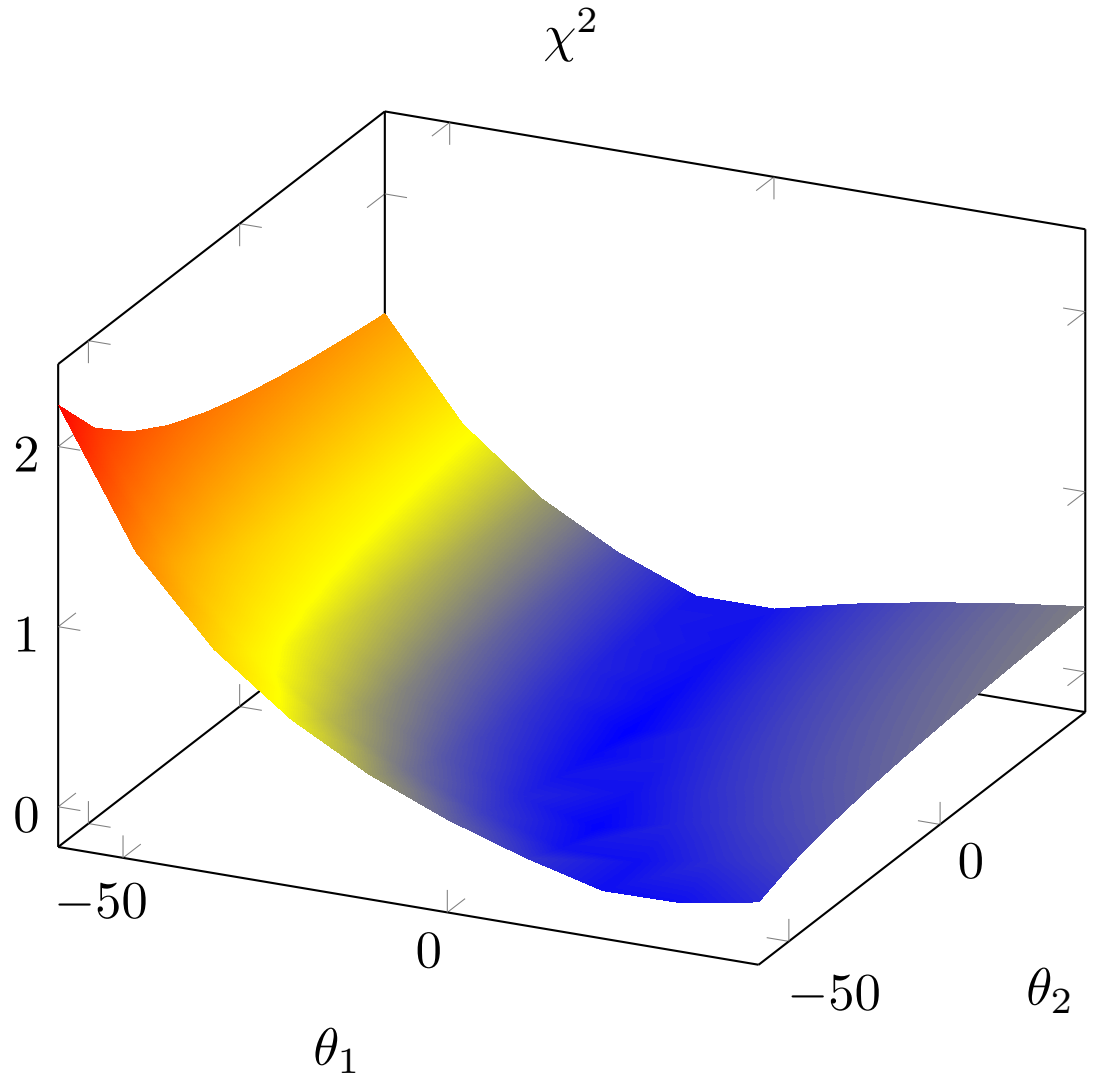


Figure V.5. – diagnosis value against the $\boldsymbol{\theta}$ to be diagnosed with $\chi^2$ diagnosis. The values are expressed as a percentage of their relative errors.

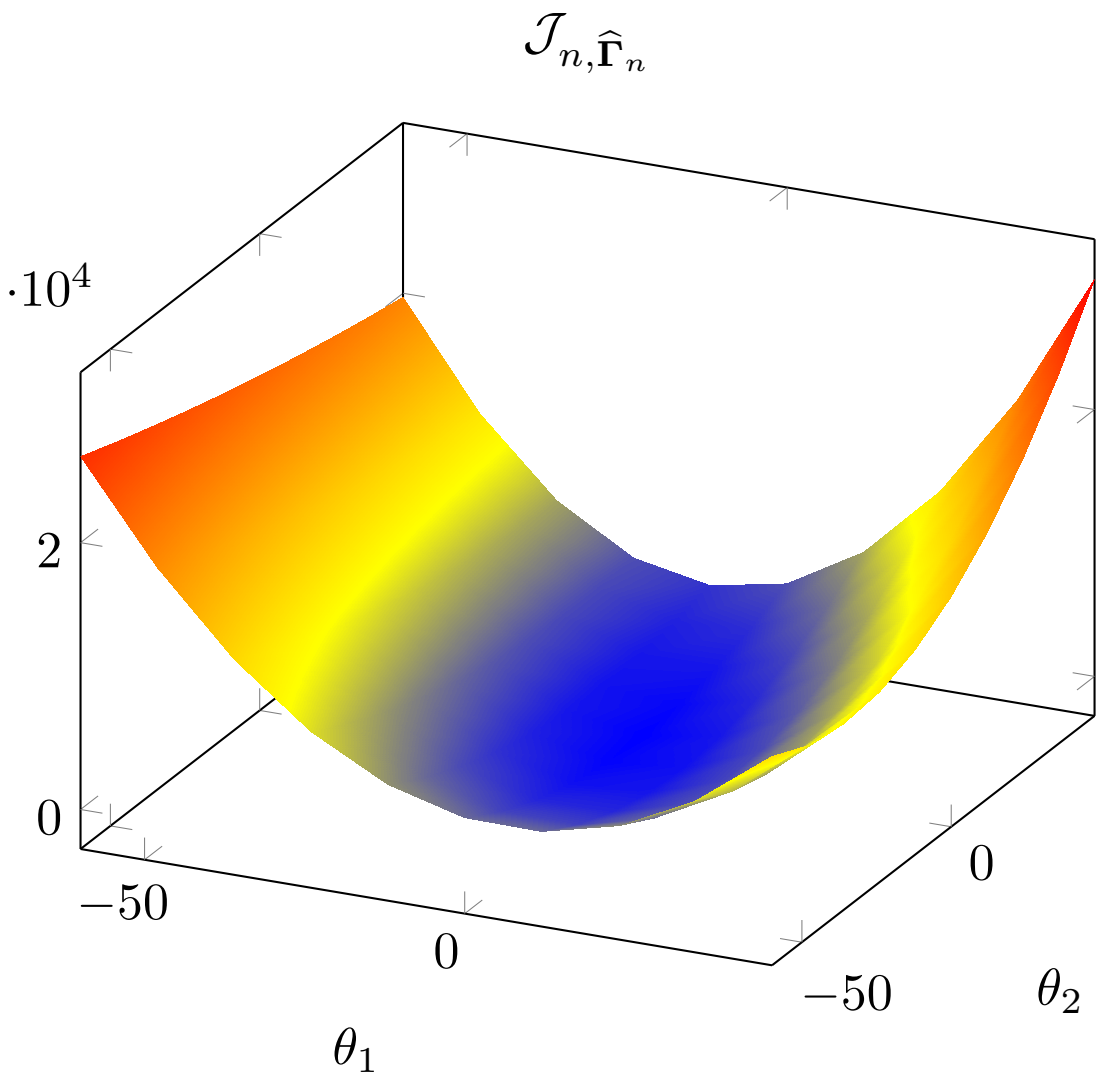


Figure V.6. – diagnosis value against the $\boldsymbol{\theta}$ to be diagnosed with the canonical estimation method. The values are expressed as a percentage of their relative errors.

## V.6. Discussion

### V.6.1. Use of Gaussian synthetic samples

Further studies are required to tackle actual data, as considered, e.g., in weather forecasting. In the same way, the canonical estimation method should be tested with other distributions since Gaussian distributions are thoroughly determined by their two first moments. Indeed, the convergence rate of the estimation of the covariance matrix $\widehat{\mathbf{\Gamma}}_n$, when $\boldsymbol{\nu}$ is not a Gaussian vector, might be lower since the canonical estimator $\hat{\mu}_{abcd}$ might have lower convergence rate when approximating the coefficients. However, it remains dominated by $\frac{1}{\sqrt{n}}$.

### V.6.2. Use of the estimated covariance matrix of s

The use of $\widehat{\mathbf{\Gamma}}_n$ rather than $\mathbf{\Gamma}_n$ is motivated by the fact that the probability distribution of $\boldsymbol{\nu}$ is never fully known by the operator, thus a functional using $\mathbf{\Gamma}_n$ would not have any practical application. However, although $\widehat{\mathbf{\Gamma}}_n$ converges toward $\mathbf{\Gamma}_n$, the accuracy of $\widehat{\mathbf{\Gamma}}_n$ for a given sample size $n$ is unknown. It is therefore impossible to fully evaluate the accuracy of the estimated parameter $\widehat{\boldsymbol{\theta}}$ especially in the case where the probability distribution of $\boldsymbol{\nu}$ is unknown.

## V.7. Conclusion

In this paper, we have shown how the limitations of $\chi^2$ diagnosis can be circumvented by a new method that the authors called the canonical estimation method. In addition, this method can be implemented in an algorithm that finds the optimal parameters for covariance matrices. The study has shown that this parameters finding process has a convergence speed dominated by $\frac{1}{\sqrt{n}}$ which allows to evaluate the accuracy of the estimated parameters regarding the sample size it is based on. However, $\chi^2$ diagnosis is not outdated since the diagnosis phase which uses the canonical estimation is a complement to $\chi^2$ diagnosis as shown in section V.5.2.

This study has been carried out on synthetic data with a set of i.i.d. random vectors and a fixed covariance matrix. This simplification is not realistic regarding real life observation, as, e.g. meteorological data. Some further study shall verify the validity of this method for actual data. However, $\chi^2$ diagnosis is based on the same assumptions and is still practically used, so this limitation does not undermine the validity of the study.

## Acknowledgments

The authors would like to thank the CENSE project (ref. ANR-16-CE22-0012) team for initiating this project, the Agence Nationale de la Recherche (ANR) for their financial support.

# VI. Conclusions

## VI.1. Thesis review

In this study, several ways of generating noise maps have been evaluated, in order to take into account the influence of an observation network in the noise mapping generation process. The first step, developed in this thesis in chapter II consists in designing a meta-model built in order to reproduce the output of a noise mapping software of a given area where the inputs are variations of the noise sources parameters, such as the speed and the flow rate of the vehicles, or the noise propagation, such as the reflection coefficient of the buildings or the temperature. It is generated by a learning sample of input parameters and their resulting noise maps computed by the noise mapping software. The resulting meta-model is a linear combination of a reduced set of statistically significant normalized noise maps. Their weights are outputs of interpolation functions generated with the projection of the output noise maps onto the maps of the reduced basis. This method allows to fairly reproduce the outputs of the noise mapping software with a much lower computation time, and then, among other applications, to compute dynamic noise maps in real time.

The second step is to include the noise level observed through an array of noise level meters scattered across the study area with different methods, the Best Linear Unbiased Estimator (BLUE) in chapter III, and the inverse modeling and Joint State Parameter Estimation in chapter IV.

The BLUE method shown in chapter III adds a correction layer to a noise map generated by the meta-model, called background, whose input parameters are based on ground traffic and weather observations. The correction layer is built with an estimation of the background matrix $\mathbf{B}$ and the observation matrix $\mathbf{R}$. The RMSE is reduced by 30 % when estimating the performance of this process through a leave-one-out cross-validation method.

The inverse modeling is studied in chapter IV. This method uses an optimization algorithm to compute the input parameters of the metamodel which best fit with the observations of the sound level meters. This method generally gives better results than the BLUE method at the cost of a longer computation time, due to the exploration part of the optimization algorithm. In addition, this method is suitable for areas where no traffic nor weather ground measurements are taken. It is then available for a wider range of study areas.

Finally, chapter V proposes a new technique for diagnosing the parameters of a covariance matrix, such as the matrix $\mathbf{B}$ used in chapters III and IV. This method is a complement to the currently used $\chi^2$ diagnosis. In addition, it is possible to adapt the process in order to estimate the parameters instead of performing a diagnosis, the estimation being out of reach for the $\chi^2$ diagnosis. This method is more robust than the $\chi^2$ diagnosis and offers new perspectives.

## VI.2. Perspectives

The study has been conducted for one specific type of noise indicator, the $LA_{eq,1h}$ which has been described in (I.12). This indicator was chosen because it is recommended for noise mapping by the END (European Commission, 2015). However, recent works in sound perception have highlighted the limits of this indicator to characterize the perceived quality of sound environments. Data assimilation could be conducted for other type of indicators such as some frequency bands or the $L_{10,1h}$ and the $L_{90,1h}$ which respectively correspond to the noise level of the first and last decile of the noise distribution on a 1-hour period. These indicators cannot be estimated through regular noise mapping methods, but some studies have shown how to rely on regular noise simulators to estimate these indicators (Aumond and Can, 2018). The covariance structure however may vary since a different phenomenon is observed. The highest centiles of noise distributions are due indeed to close noise sources while background noise is the accumulation of numerous but remote sound sources. Mixed noise indices such as the harmonica index (Prascevic et al., 2015) could also be produced. In addition, the sources used in this study were limited to the road traffic noise sources, acknowledged as the most annoying noise sources in urban areas. However, the END also recommends the production of noise maps for railway or aircraft noise. In addition, several sound perception models emerged, in which sound sources such as birds or voices intervene (Lavandier and Defréville, 2006). Hybridizing the data assimilation processes for different noise sources would be of great interest. This would be made possible because recent works on noise observations have allowed to discriminate noise sources with deep learning methods (Gontier et al., 2019). This will require some adaptation of the procedure, as the typology of sources considered most likely impacts the structure of the covariance matrices. This is probably the case, for example, for air traffic because of its elevation.

Finally, several new applications may emerge following the results of this study.

First, the meta-model created in chapter II which has been used for data assimilation purposes can be used for several other applications which have not been explored yet in this work. Some applications can be done in sensitivity analysis, and uncertainty propagation in relation with other works such as Aumond et al. (2019, 2021). Indeed, the extremely short computation time of the generation of a noise map allows to compute the output of a very large number of parameters and to study the probability distribution of some outputs based on the probability distribution of the inputs. An other domain of sensitivity analysis requires a very large amount of calls to the model, the variance based sensitivity analysis and the computation of the so-called Sobol indices (Sobol, 1990). The Sobol indices of the input parameters of a noise mapping reference software can only be computed with the help of a meta-model.

Second, the meta-model used in this study for data-assimilation only required a limited amount of input parameters ($\dim \mathbf{p} = 7$). However, it is possible to build a richer meta-model with a larger amount of input parameters. For example, in chapter II different types of roads whose traffic behaviors are uncorrelated, according to Barrigón Morillas et al. (2005) have been ttreated separately, which led to an input vector of higher dimension ($\dim \mathbf{p} = 16$). In addition, it is possible to enrich the weather data. In this study, only the temperature has been considered as an input, but it is

possible to add wind direction and intensity, hygrometry, probability of occurrence of favorable atmospheric condition, etc.) which would lead to a much richer and more complex meta-model and might better approach the real dynamic noise distribution. The number of input should not affect the implementation of the meta-model, under the condition that the interpolation functions still correctly approach the projection over every basis vector of the reduced model. By treating separately different types of roads, it should be possible to get better results since there is a strong correlation between the behavior of the different road sections of the same type and that the behavior of the different types of roads are uncorrelated.

Third, in this study, the distribution of the observation points across the study area was given as an input. One may wonder if, for a given quantity of noise level meters, some distribution gives better results than others when conducting the data assimilation process. For instance, the CENSE project which is currently conducting an experimental campaign in the city of Lorient, France, in the area studied in chapter II, has displayed a very large quantity of noise level meters in the area. For a given number $n$ of sensors, it might be possible to use a combinatorial approach to find the best subset of sensors which gives the best results and to infer some qualitative characteristics over the best distributions. This research could highlight the most relevant features to take into account when an array of sound level meters is displayed across a study area, for instance, the street typology, the building distribution or the density of sensors. In addition it could be possible to find thresholds for the size of the subset, for instance a number of sensors from which the performance drastically improves or on the other side, where it stagnates and reaches a plateau.

# A. Appendix:Data assimilation for urban noise mapping with a meta-model

## A.1. $\chi^2$ diagnosis

### A.1.1. $\chi^2$ diagnosis

With $d_{ij}$ and $|x_i - x_j|$ known for all $(i,j) \in [\![1,n]\!]^2$, the characteristic values $\sigma_b$, $L_d$ and $L_b$ must be tuned to optimize the efficiency of the analytical covariance matrix.

To do so, a $\chi^2$ diagnosis is carried out. It consists in checking the consistency between the available innovations $\boldsymbol{\nu}_t = \mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^b$ and their variances $\mathbf{S}_t = \mathbf{H}_t\mathbf{B}\mathbf{H}_t^T + \mathbf{R}_t$ for each time step $t$. This method has proved to be useful in the meteorology (Ménard and Chang, 2000) and in other fields. Let $\chi_t^2$ be the scalar:

$$\chi_t^2 = \boldsymbol{\nu}_t^T \mathbf{S}_t^{-1} \boldsymbol{\nu}_t \tag{A.1}$$

Its expected value is

$$\begin{aligned} \mathbb{E}[\chi_t^2] &= \mathbb{E}[\boldsymbol{\nu}_t^T \mathbf{S}_t^{-1} \boldsymbol{\nu}_t] \\ &= \mathrm{tr}(\mathbf{S}_t^{-1}\mathbb{E}[\boldsymbol{\nu}\boldsymbol{\nu}^T]) \\ &= \mathrm{tr}(\mathbf{S}_t^{-1}\mathbf{S}_t) = \mathrm{tr}(\mathbf{I}_{p_t}) = p_t \end{aligned} \tag{A.2}$$

$p_t$ being the dimension of $\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^b$ at time $t$. The dimension of $\mathbf{y}_t$ may vary in time since the microphones may not be all simultaneously active, and so does $\mathbf{R}_t$, $\mathbf{H}_t$ and $\mathbf{S}_t$. Let $A$ be the value:

$$A = \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \frac{\chi_t^2}{p_t} \tag{A.3}$$

The parameters $\sigma_b$, $L_d$ and $L_b$ are chosen so that $A$ is close to 1.

### A.1.2. Note on the $\chi^2$ diagnosis

According to the definition of section III.2.5.2, the covariance matrix $\mathbf{B}_{ana}$ is no longer the simulation error covariance but the covariance of the vector $\boldsymbol{\Psi}^T\boldsymbol{\Psi}\mathbf{x}^t - \mathbf{x}^t$, with $\boldsymbol{\Psi}$ the matrix which concatenates the basis vector of the reduced subspace. Hence the characteristic parameters obtained with the $\chi^2$ diagnosis are ill defined. However, since $\mathbf{x}^b$ remains a fairly good approximation of $\boldsymbol{\Psi}^T\boldsymbol{\Psi}\mathbf{x}^t$, the orders of magnitude of the characteristic parameters obtained with this method are still satisfying.

## A.2. Sum of the analytical and statistical covariance matrices

By construction of the meta-model, $\mathbf{x}^b$ lies in the subspace generated by the maps set of the metamodel $\{\boldsymbol{map}_i\}_{i \in [\![1,k]\!]}$ with $k \ll n$. This suggests to decompose the

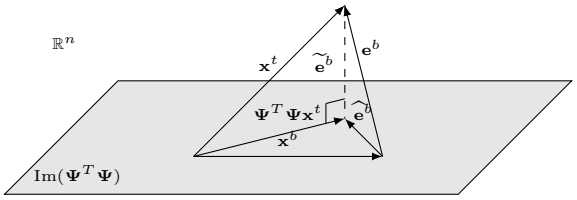


Figure A.1. – Geometrical interpretation of the relation between the statistical and the analytical covariance matrices

observation in two orthogonal components $\widehat{\mathbf{e}}^b \in \mathrm{Im}(\mathbf{\Psi}^T\mathbf{\Psi})$ and $\widetilde{\mathbf{e}}^b \in \mathrm{Im}(\mathbf{\Psi}^T\mathbf{\Psi})^\perp$ as illustrated in figure A.1. The current section shows that the construction of $\mathbf{x}^b$ allows that a new covariance matrix $\mathbf{B}$ can be generated by adding the analytical covariance matrix defined in section B.1.1 $\mathbf{B}_{ana}$ to a statistically defined covariance matrix $\mathbf{B}_{stat}$.

The simulation error can be decomposed as follows: $\mathbf{e}^b = \widehat{\mathbf{e}}^b + \widetilde{\mathbf{e}}^b$. $\widehat{\mathbf{e}}^b$ is the modeling error that grasps the error due to model approximations, and the discrepancy between the estimated input parameters and the true parameters. $\widetilde{\mathbf{e}}^b$ is the fluctuation error due to the noise level components which are not related to the traffic noise. Its covariance matrix is expressed with the analytical covariance matrix described in section B.1.1. Under the assumption that these two components are independent ($\mathbb{E}[\widehat{\mathbf{e}}^b\widetilde{\mathbf{e}}^{bT}] = \mathbb{E}[\widetilde{\mathbf{e}}^b\widehat{\mathbf{e}}^{bT}] = 0$), $\mathbf{B}$ can be expressed as

$$\begin{aligned}
\mathbf{B} = \mathbb{E}[\mathbf{e}^b\mathbf{e}^{bT}] &= \mathbb{E}[(\widehat{\mathbf{e}}^b + \widetilde{\mathbf{e}}^b)(\widehat{\mathbf{e}}^b + \widetilde{\mathbf{e}}^b)^T] \\
&= \mathbb{E}[\widehat{\mathbf{e}}^b\widehat{\mathbf{e}}^{bT} + \widetilde{\mathbf{e}}^b\widetilde{\mathbf{e}}^{bT} + \widetilde{\mathbf{e}}^b\widehat{\mathbf{e}}^{bT} + \widehat{\mathbf{e}}^b\widetilde{\mathbf{e}}^{bT}] \\
&= \mathbf{B}_{stat} + \mathbf{B}_{ana} + \underbrace{\mathbb{E}[\widetilde{\mathbf{e}}^b\widehat{\mathbf{e}}^{bT} + \widehat{\mathbf{e}}^b\widetilde{\mathbf{e}}^{bT}]}_{=0} \\
&= \mathbf{B}_{stat} + \mathbf{B}_{ana}
\end{aligned} \tag{A.4}$$

A way to check the validity of this assumption is to evaluate the norm of the projection $\Pi_{\mathrm{Im}(\mathbf{B}_{stat})}$ of the eigenvectors of $\mathbf{B}_{ana}$ onto $\mathrm{Im}(\mathbf{B}_{stat})$. In this case study, for each eigenvector $\boldsymbol{v}_i$ of $\mathbf{B}_{ana}$. The values remain low

$$\frac{\|\Pi_{\mathrm{Im}(\mathbf{B}_{stat})}(\boldsymbol{v}_i)\|}{\|\boldsymbol{v}_i\|} < 5\,\% \tag{A.5}$$

# B. Appendix:Meta-model aided inverse modeling and Joint State Parameter Estimation for noise data assimilation

## B.1. Matrices descriptions

The cost functions defined in section IV.3.2.1 make use of the output of the meta-model and apply a correction which depends on $\mathbf{B}$, the covariance matrix of the simulation error, $\mathbf{R}$, the covariance matrix of the observational error, and the discrepancy between observations and simulations at the observation location. The matrices $\mathbf{B}$ and $\mathbf{R}$ seize the spatial behavior of the model error and the observation error respectively.

### B.1.1. Description of the covariance matrix B

A parametrized simulation error covariance function has been successfully used in a previous work (Ventura et al., 2018). The simulation error cross correlations between two receivers $i$ and $j$ are expressed in terms of:

— $d_{ij}$, the length of the shortest path along the road network between the projection of receivers $i$ and $j$ on the closest road computed with a Johnson algorithm (Johnson, 1977)

— the nominal noise field itself, through a correlation function $\rho(x_i, x_j)$, where $x_i$ and $x_j$ are the values of $\mathbf{x}^b$ at points $i$ and $j$ as proposed by (Riishøjgaard, 2002). The corresponding simulation error covariance between points $i$ and $j$ reads:

$$[\mathbf{B}_{ana}]_{ij} = \sigma_b^2 \exp\left(-\frac{d_{ij}}{L_d}\right) \exp\left(-\frac{|x_i - x_j|}{L_b}\right) \tag{B.1}$$

where $\sigma_b^2$ is the characteristic variance of each point, $L_d$ is a characteristic distance in m and $L_b$ is a characteristic noise level value in dB.

### B.1.2. Description of the covariance matrix R

The observation errors at two different receivers are not correlated. The observation error essentially depends on the accuracy of the microphone which is described by its standard deviation $\sigma_r$, therefore:

$$\mathbf{R} = \sigma_r^2 \mathbf{I}_p \tag{B.2}$$

with $\mathbf{I}_p$ the identity matrix of dimension $p$. In this study, $\sigma_r$ =1 $\mathrm{dB}^2$, which corresponds to the indutrial accuracy of the sensor.

### B.1.3. Equivalence between $\mathcal{J}(\mathbf{p},\epsilon)$ and $\mathcal{J}'(\mathbf{p})$

The following approach aims to find a new cost function which only involves $\mathbf{p}$ to lower the computation time ($\dim(\mathbf{p}) \ll \dim(\boldsymbol{\epsilon})$). The first step is to decompose the gradient in 2 vectors:

$$\nabla\mathcal{J}(\mathbf{p}^\star,\boldsymbol{\epsilon}^\star)=0 \Rightarrow \left\{ \begin{array}{rcl} \nabla_{\mathbf{p}}\mathcal{J}(\mathbf{p}^\star,\boldsymbol{\epsilon}^\star) & = & 0 \\ \wedge & & \\ \nabla_{\boldsymbol{\epsilon}}\mathcal{J}(\mathbf{p}^\star,\boldsymbol{\epsilon}^\star) & = & 0 \end{array} \right. \tag{B.3}$$

$$\begin{aligned} \nabla_{\boldsymbol{\epsilon}}\mathcal{J}(\mathbf{p}^\star,\boldsymbol{\epsilon}^\star)=0 &\Rightarrow 2\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}^\star)+\mathbf{H}\boldsymbol{\epsilon}^\star)+2\mathbf{B}^{-1}\boldsymbol{\epsilon}^\star=0 \\ &\Rightarrow \boldsymbol{\epsilon}^\star=\mathbf{K}(\mathbf{y}-\mathbf{H}\widehat{\mathcal{M}}(\mathbf{p}^\star)) \end{aligned} \tag{B.4}$$

With $\mathbf{K}=(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}+\mathbf{B}^{-1})^{-1}\mathbf{H}^T\mathbf{R}^{-1}=\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})^{-1}$ using the Sherman-Morrisson-Woodbury (SMW) identity (Sherman and Morrison, 1950)

With the substitution of $\boldsymbol{\epsilon}$ by the expression of $\boldsymbol{\epsilon}^\star$

$$\nabla_{\mathbf{p}}\mathcal{J}(\mathbf{p},\boldsymbol{\epsilon}^\star)=\left(2\left((\mathbf{I}-\mathbf{H}\mathbf{K})(\mathbf{y}-\mathbf{H}\widehat{\mathcal{M}}(\mathbf{p})\right)\mathbf{R}^{-1}\mathbf{H}\frac{\partial\widehat{\mathcal{M}}(\mathbf{p})}{\partial p_j}\right)_{j\in[\![1,k]\!]} \tag{B.5}$$

The optimisation problem is reformulated as

$$\nabla\mathcal{J}'(\mathbf{p})=\left(2\left(\mathbf{y}-\mathbf{H}\widehat{\mathcal{M}}(\mathbf{p})\right)^T(\mathbf{I}-\mathbf{H}\mathbf{K})^T\mathbf{R}^{-1}\mathbf{H}\frac{\partial\widehat{\mathcal{M}}(\mathbf{p})}{\partial p_j}\right)_{j\in[\![1,k]\!]} \tag{B.6}$$

The integration of this formula gives

$$\mathcal{J}'(\mathbf{p})=\left\|\mathbf{y}-\mathbf{H}\widehat{\mathcal{M}}(\mathbf{p})\right\|^2_{(\mathbf{I}-\mathbf{H}\mathbf{K})^T\mathbf{R}^{-1}} \tag{B.7}$$

## B.2. Differentiation of the metamodel

### B.2.1. kriging

Formula of the interpolation function

$$\forall i\in[\![1,k]\!],\ \hat{\alpha}_i(\mathbf{p})=\boldsymbol{a}_i\cdot\boldsymbol{\gamma}^i+b_i,\ \boldsymbol{a}_i=\boldsymbol{\lambda}_i^T\boldsymbol{\Gamma}_i^{-1}\left[\boldsymbol{I}_r-\mathbf{1}_r\frac{\mathbf{1}_r^T\boldsymbol{\Gamma}_i^{-1}}{\mathbf{1}_r^T\boldsymbol{\Gamma}_i^{-1}\mathbf{1}_r}\right]\in\mathbb{R}^r,\ b_i=\frac{\boldsymbol{\lambda}_i^T\boldsymbol{\Gamma}_i^{-1}\mathbf{1}_r}{\mathbf{1}_r^T\boldsymbol{\Gamma}_i^{-1}\mathbf{1}_r}\in\mathbb{R} \tag{B.8}$$

$k$ is the dimension of the reduced basis

differentiation under $p_n$

$$\frac{\partial\hat{\alpha}_i}{\partial p_n}(\mathbf{p})=\boldsymbol{a}_i\cdot\frac{\partial\boldsymbol{\gamma}^i}{\partial p_n}(\mathbf{p}) \tag{B.9}$$

$\boldsymbol{\gamma}^i(\mathbf{p})=\left(\gamma_j^i(\mathbf{p})\right)_{j\in[\![1,r]\!]}$, $r$ is the trainins sample size and $\gamma_j^i(\mathbf{p})=\prod_{m=1}^p\phi\left(|p_m-p_m^{(j)}|,\theta_m^i\right)$, $p$ is the number of parameters of the meta-model.

Hence:

$$\frac{\partial \gamma_j^i}{\partial p_n} = \frac{\partial \phi}{\partial p_n}(|p_n - p_n^{(j)}|, \theta_n^i) \prod_{\substack{m=1 \\ m \neq n}}^{p} \phi(|p_m - p_m^{(j)}|, \theta_m^i) \tag{B.10}$$

$\phi(r, \theta) = \sigma^2 \exp(-\frac{r}{\ell})$, Hence $\frac{\partial \phi}{\partial r}(r, \theta) = -\frac{\sigma^2}{\ell} \exp(-\frac{r}{\ell}) = -\frac{1}{\ell}\phi(r, \theta)$ and

$$\frac{\partial \gamma_j^i}{\partial p_n}(\mathbf{p}) = -\frac{\operatorname{sgn}(p_n - p_n^{(j)})}{\ell_j^i} \gamma_j^i(\mathbf{p}) \tag{B.11}$$

$$\frac{\partial \hat{\boldsymbol{\alpha}}}{\partial p_n}(\mathbf{p}) = \left( \frac{\partial \hat{\alpha}_1}{\partial p_n}(\mathbf{p}) \ \cdots \ \frac{\partial \hat{\alpha}_k}{\partial p_n}(\mathbf{p}) \right)^T \tag{B.12}$$

$$\boxed{\frac{\partial \widehat{\mathcal{M}}}{\partial p_n}(\mathbf{p}) = \boldsymbol{\Psi} \frac{\partial \hat{\boldsymbol{\alpha}}}{\partial p_n}(\mathbf{p})} \tag{B.13}$$

### B.2.2. RBF

$$\hat{\alpha}_i(\mathbf{p}) = \sum_{j=1}^{r} \gamma_{i,j} \phi(\Lambda(\mathbf{p}, \mathbf{p}^{(j)})) \tag{B.14}$$

$$\hat{\boldsymbol{\alpha}}(\mathbf{p}) = \boldsymbol{\Gamma} \boldsymbol{\Phi}(\mathbf{p}) \tag{B.15}$$

with $\boldsymbol{\Gamma} = (\gamma_{i,j})_{i,j} \in \mathbb{R}^{k \times r}$ and $\boldsymbol{\Phi}(\mathbf{p}) = \left( \phi \left( \Lambda \left( \mathbf{p}, \mathbf{p}^{(j)} \right) \right) \right)_j \in \mathbb{R}^r$

$$\begin{aligned} \frac{\partial \boldsymbol{\Phi}_i}{\partial p_j}(\mathbf{p}) &= \frac{\partial \Lambda}{\partial p_j}(\mathbf{p}, \mathbf{p}^{(i)}) \phi'(\Lambda(\mathbf{p}, \mathbf{p}^{(i)})) \\ &= \frac{p_j}{\Lambda(\mathbf{p}, \mathbf{p}^{(i)})} \phi'(\Lambda(\mathbf{p}, \mathbf{p}^{(i)})) \end{aligned} \tag{B.16}$$

if $\phi(r) = r^3$, $\phi'(r) = 3r^2$ then

$$\frac{\partial \boldsymbol{\Phi}_i}{\partial p_j}(\mathbf{p}) = 3 p_j \Lambda(\mathbf{p}, \mathbf{p}^{(i)}) \tag{B.17}$$

$$\boxed{\frac{\partial \widehat{\mathcal{M}}}{\partial p_j}(\mathbf{p}) = \boldsymbol{\Psi} \boldsymbol{\Gamma} \frac{\partial \boldsymbol{\Phi}}{\partial p_j}(\mathbf{p}) = 3 p_j \boldsymbol{\Psi} \boldsymbol{\Gamma} \boldsymbol{\Lambda}(\mathbf{p})} \tag{B.18}$$

with $\boldsymbol{\Lambda}(\mathbf{p}) = (\Lambda(\mathbf{p}, \mathbf{p}^{(i)}))_i \in \mathbb{R}^r$

# C. Appendix: Estimation and diagnosis of the parameters of a covariance matrix

## C.1. Detailed calculus

### C.1.1. Detailed calculus of section V.3.1

#### C.1.1.1. details of equation V.1

$$\begin{aligned}
\forall k \in \mathbb{N},\ \mathbb{E}[{\boldsymbol{\nu}^{(k)}}^T \mathbf{S}^{-1}(\boldsymbol{\theta}^\star)\boldsymbol{\nu}^{(k)}] &= \operatorname{tr}\left(\mathbb{E}\left[{\boldsymbol{\nu}^{(k)}}^T \mathbf{S}^{-1}(\boldsymbol{\theta}^\star)\boldsymbol{\nu}^{(k)}\right]\right) \\
&= \mathbb{E}\left[\operatorname{tr}\left({\boldsymbol{\nu}^{(k)}}^T \mathbf{S}^{-1}(\boldsymbol{\theta}^\star)\boldsymbol{\nu}^{(k)})\right)\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left(\mathbf{S}(\boldsymbol{\theta}^\star)^{-1}\boldsymbol{\nu}^{(k)}{\boldsymbol{\nu}^{(k)}}^T\right)\right] \\
&= \operatorname{tr}\left(\mathbb{E}[\mathbf{S}(\boldsymbol{\theta}^\star)^{-1}\boldsymbol{\nu}^{(k)}{\boldsymbol{\nu}^{(k)}}^T]\right) \\
&= \operatorname{tr}\left(\mathbf{S}(\boldsymbol{\theta}^\star)^{-1}\mathbb{E}[\boldsymbol{\nu}^{(k)}{\boldsymbol{\nu}^{(k)}}^T]\right) \\
&= \operatorname{tr}\left(\mathbf{S}(\boldsymbol{\theta}^\star)^{-1}\mathbf{S}(\boldsymbol{\theta}^\star)\right) = \operatorname{tr}(\mathbf{I}_p) = p.
\end{aligned}$$

### C.1.2. Proof of Lemma 3

$$
\begin{aligned}
\boldsymbol{\Gamma}_{n,ijk\ell} =& \mathbb{E}[(\frac{1}{n}\sum_{m=1}^{n}\boldsymbol{\eta}_{i,m}\boldsymbol{\eta}_{j,m} - \frac{1}{n^2}\sum_{m=1}^{n}\boldsymbol{\eta}_{i,m}\sum_{m=1}^{n}\boldsymbol{\eta}_{j,m}) \\
& (\frac{1}{n}\sum_{m=1}^{n}\boldsymbol{\eta}_{k,m}\boldsymbol{\eta}_{\ell,m} - \frac{1}{n^2}\sum_{m=1}^{n}\boldsymbol{\eta}_{k,m}\sum_{m=1}^{n}\boldsymbol{\eta}_{\ell,m})] \\
& - \mathbb{E}[\frac{1}{n}\sum_{m=1}^{n}\boldsymbol{\eta}_{i,m}\boldsymbol{\eta}_{j,m} - \frac{1}{n^2}\sum_{m=1}^{n}\boldsymbol{\eta}_{i,m}\sum_{m=1}^{n}\boldsymbol{\eta}_{j,m}] \\
& \mathbb{E}[\frac{1}{n}\sum_{m=1}^{n}\boldsymbol{\eta}_{k,m}\boldsymbol{\eta}_{\ell,m} - \frac{1}{n^2}\sum_{m=1}^{n}\boldsymbol{\eta}_{k,m}\sum_{m=1}^{n}\boldsymbol{\eta}_{\ell,m}] \\
=& \mathbb{E}[\frac{1}{n^2}(\sigma_{1100}\sigma_{0011} - \frac{1}{n}(\sigma_{1100}\sigma_{0010}\sigma_{0001} + \sigma_{0011}\sigma_{1000}\sigma_{0100}) \\
& + \frac{1}{n^2}\sigma_{1000}\sigma_{0100}\sigma_{0010}\sigma_{0001})] \\
& - \frac{1}{n^2}(\mathbb{E}[\sigma_{1100}] - \frac{1}{n}\mathbb{E}[\sigma_{1000}\sigma_{0100}]) \\
& (\mathbb{E}[\sigma_{0011}] - \frac{1}{n}\mathbb{E}[\sigma_{0010}\sigma_{0001}]) \\
=& \frac{1}{n^2}\mathbb{E}[\sigma_{1100}\sigma_{0011}] - \frac{1}{n^3}(\mathbb{E}[\sigma_{1100}\sigma_{0010}\sigma_{0001}] + \mathbb{E}[\sigma_{0011}\sigma_{1000}\sigma_{0100}]) \\
& + \frac{1}{n^4}\mathbb{E}[\sigma_{1000}\sigma_{0100}\sigma_{0010}\sigma_{0001}] \\
& - \frac{1}{n^2}(\mathbb{E}[\sigma_{1100}] - \frac{1}{n}\mathbb{E}[\sigma_{1000}\sigma_{0100}]) \\
& (\mathbb{E}[\sigma_{0011}] - \frac{1}{n}\mathbb{E}[\sigma_{0010}\sigma_{0001}])
\end{aligned}
$$

According to (V.3), we have:

$$
\begin{aligned}
[\boldsymbol{\Gamma}]_{n,i'j'} =& \frac{1}{n^2}\mathbb{E}[\sigma_{1100}\sigma_{0011}] - \frac{1}{n^3}(\mathbb{E}[\sigma_{1100}\sigma_{0010}\sigma_{0001}] + \mathbb{E}[\sigma_{0011}\sigma_{1000}\sigma_{0100}]) \\
& + \frac{1}{n^4}\mathbb{E}[\sigma_{1000}\sigma_{0100}\sigma_{0010}\sigma_{0001}] \\
& - \frac{1}{n^2}(\mathbb{E}[\sigma_{1100}] - \frac{1}{n}\mathbb{E}[\sigma_{1000}\sigma_{0100}])(\mathbb{E}[\sigma_{0011}] - \frac{1}{n}\mathbb{E}[\sigma_{0010}\sigma_{0001}]).
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\Gamma}_{n,ijk\ell} =& \frac{1}{n^2}(n\mu_{1111} + n(n-1)\mu_{1100}\mu_{0011}) - \frac{2}{n^3}(n\mu_{1111} + n(n-1)\mu_{1100}\mu_{0011}) \\
& + \frac{1}{n^4}(n\mu_{1111} + n(n-1)(\mu_{1100}\mu_{0011} + \mu_{1010}\mu_{0101} + \mu_{1001}\mu_{0110})) \\
& - \left(\frac{n-1}{n}\right)^2\mu_{1100}\mu_{0011} \\
=& -\frac{(n-1)^2}{n^3}\mu_{1100}\mu_{0011} + \frac{n-1}{n^3}(\mu_{1001}\mu_{0110} + \mu_{1010}\mu_{0101}) + \frac{(n-1)^2}{n^3}\mu_{1111}
\end{aligned}
$$

The result then follows from:

$$\begin{aligned}
\mathbb{E}[\sigma_{1100}\sigma_{0011}] =& \mathbb{E}[\sigma_{1100}\sigma_{0010}\sigma_{0001}] = \mathbb{E}[\sigma_{0011}\sigma_{1000}\sigma_{0100}] \\
=& n\mu_{1111} + n(n-1)\mu_{1100}\mu_{0011} \\
\mathbb{E}[\sigma_{1000}\sigma_{0100}\sigma_{0010}\sigma_{0001}] =& n\mu_{1111} + n(n-1)(\mu_{1100}\mu_{0011} + \mu_{1010}\mu_{0101} + \mu_{1001}\mu_{0110}) \\
\mathbb{E}[\sigma_{1100}] =& \mathbb{E}[\sigma_{1000}\sigma_{0100}] = n\mu_{1100} \\
\mathbb{E}[\sigma_{0011}] =& \mathbb{E}[\sigma_{0010}\sigma_{0001}] = n\mu_{0011}.
\end{aligned}$$

## C.1.3. Detailed calculus of section V.4.3.1

### C.1.3.1. proof of Lemma 4

By construction, since $\mathbf{\Gamma} = \frac{\mathbf{A}}{n} + \frac{\mathbf{B}}{n^2} + \frac{\mathbf{C}}{n^3}$ and $\widehat{\mathbf{A}} = \mathbf{A} + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$, $\widehat{\mathbf{B}} = \mathbf{B} + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$ and $\widehat{\mathbf{C}} = \mathbf{C} + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$ hence $\widehat{\mathbf{\Gamma}}_n = \mathbf{\Gamma}_n + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n\sqrt{n}}\right)$ and $\widehat{\mathbf{\Gamma}}_n^{-1} = \mathbf{\Gamma}_n^{-1} + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n\sqrt{n}}\right)$

$$\begin{aligned}
\boldsymbol{v}^T &= \frac{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta})}{\partial \theta_i}\widehat{\mathbf{\Gamma}}_n^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta})}{\partial \theta_i}\widehat{\mathbf{\Gamma}}_n^{-1}\right\|_1} \\
&= \frac{\left(\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i} + (\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star) + o(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)\right)(n\mathbf{A}^{-1} + \mathbf{A}^{-2}\mathbf{B} + o(1) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n\sqrt{n}}\right))}{\left\|\left(\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i} + (\boldsymbol{\theta}-\boldsymbol{\theta}^\star)\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star) + o(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)\right)(n\mathbf{A}^{-1} + \mathbf{A}^{-2}\mathbf{B} + o(1) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n\sqrt{n}}\right))\right\|_1} \\
&= \frac{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1} + (\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1} + (\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}\right\|_1} + o(\boldsymbol{\theta}-\boldsymbol{\theta}^\star) + \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n^2\sqrt{n}}\right)
\end{aligned}$$

Since $\mathcal{O}\left(\frac{1}{n}\right) = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right)$ and $\mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n^2\sqrt{n}}\right) = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right)$

$$\begin{aligned}
&= \frac{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1} + (\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1 c + \left\|(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}\right\|_1} + o(\boldsymbol{\theta}-\boldsymbol{\theta}^\star) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right) \\
&= \frac{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1} + (\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1 c\left(1 + \frac{\left\|(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}\right\|_1}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1}\right)} + o(\boldsymbol{\theta}-\boldsymbol{\theta}^\star) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right) \\
&= \frac{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1} + (\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1}\left(1 - \frac{\left\|(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}\right\|_1}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1}\right) + o(\boldsymbol{\theta}-\boldsymbol{\theta}^\star) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right) \\
&= \frac{\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1} + (\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\frac{\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1} - \|\boldsymbol{\theta}-\boldsymbol{\theta}^\star\|_1\frac{\left\|\mathbf{D}^T_{\partial_{\theta_i}\mathbf{s}}(\boldsymbol{\theta}^\star)\mathbf{A}^{-1}\right\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}}{\left\|\frac{\partial \mathbf{s}^T(\boldsymbol{\theta}^\star)}{\partial \theta_i}\mathbf{A}^{-1}\right\|_1^2} \\
&+ o(\boldsymbol{\theta}-\boldsymbol{\theta}^\star) + \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right)
\end{aligned}$$

### C.1.3.2. proof of Lemma 5

Define $\eta \in [0,1]$. According to (V.12), if for every $n \in \mathbb{N}$, $P_{\sigma(n)}$ is true, then there exists $M_\eta$ such that for every $n \in \mathbb{N}$

$$\mathbb{P}(\frac{X_{\sigma(n)}}{u_{\sigma(n)}} \leq M_\eta) \geq 1 - \eta$$

Now for all $n \in \mathbb{N}$, consider the quantity $\mathbb{P}(\frac{X_n}{u_n} \leq M_\eta)$:

$$\begin{aligned}
\mathbb{P}(\frac{X_n}{u_n} \leq M_\eta) &\geq \mathbb{P}(\exists m \in \mathbb{N}, \sigma(m) = n \wedge \frac{X_{\sigma(m)}}{u_{\sigma(m)}} \leq M_\eta) \\
&\geq \mathbb{P}(\exists m \in \mathbb{N}, \sigma(m) = n)\mathbb{P}(\frac{X_{\sigma(m)}}{u_{\sigma(m)}} \leq M_\eta) \\
&\geq (1-\alpha)(1-\eta)
\end{aligned}$$

With $\eta' := 1 - (1-\alpha)(1-\eta)$, we have

$$\forall \eta', \forall n \in \mathbb{N},\ \mathbb{P}(\frac{X_n}{u_n} \leq M_\eta) \geq 1 - \eta'$$

### C.1.3.3. proof of Lemma 6

Under the assumption of Lemma 6, it is possible to perform a Taylor expansion of $\mathbf{s}$ around $\boldsymbol{\theta}^\star$:

$$\begin{aligned}
\mathbf{s}(\boldsymbol{\theta}^\star + \mathbf{x}) &= \mathbf{s}(\boldsymbol{\theta}^\star) + \mathbf{D}_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{x} + o(\mathbf{x}) \\
\|\mathbf{s}(\boldsymbol{\theta}^\star + \mathbf{x}) - \mathbf{s}(\boldsymbol{\theta}^\star)\| &= \|\mathbf{D}_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{x}\| + o(\|\mathbf{x}\|)
\end{aligned}$$

Let $\mathbf{D}'_\mathbf{s}(\boldsymbol{\theta}^\star) = \Pi_{\Im(\mathbf{D}_\mathbf{s}(\boldsymbol{\theta}^\star))}\mathbf{D}_\mathbf{s}(\boldsymbol{\theta}^\star)$, $\Pi_E$ being the canonical projection of $\mathbb{R}^p$ onto one of its subspace $E$. For all $\mathbf{x} \in \mathbb{R}^\ell$, $\|\mathbf{D}'_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{x}\| = \|\mathbf{D}_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{x}\|$. According to the assumption of the Lemma, $\mathbf{D}'_\mathbf{s}(\boldsymbol{\theta}^\star)$ is invertible. Since $\mathbf{D}'_\mathbf{s}(\boldsymbol{\theta}^\star)$ in invertible, for all $\mathbf{x} \in \mathbb{R}^\ell$, $\|\mathbf{D}'_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{x}\| \neq 0$, in addition, the application $\mathbf{x} \mapsto \|\mathbf{D}'_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{x}\|$ is continuous as a composition of a linear application and a norm, which are both continuous in finite dimension. Since the spheres are compact in finite dimension, $m = \min_{\|x\|=1} \|\mathbf{D}'_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{x}\|$ exists and is strictly positive. It can be shown that $m = \sqrt{\min \sigma(\mathbf{D}'^T_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{D}'_\mathbf{s}(\boldsymbol{\theta}^\star))}$ with $\sigma(\mathcal{M}(\mathbf{p}))$ being the spectrum of $\mathcal{M}(\mathbf{p})$. And so

$$\forall \mathbf{x} \in \mathbb{R}^\ell \backslash \{0\},\ \|\mathbf{D}_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{x}\| \geq m\|\mathbf{x}\|$$

Hence for any $\epsilon > 0$, there exists $\mathcal{V}_\epsilon(\boldsymbol{\theta}^\star)$ a neighborhood of $\boldsymbol{\theta}^\star$ such that $o(\|\mathbf{x}\|) \geq -\epsilon\|\mathbf{x}\|$. Hence if we choose $\epsilon$ such that $|m| - \epsilon > 0$ we have

$$\|\mathbf{s}(\boldsymbol{\theta}^\star + \mathbf{x}) - s(\boldsymbol{\theta}^\star)\| = \|\mathbf{D}_\mathbf{s}(\boldsymbol{\theta}^\star)\mathbf{x}\| + o(\|\mathbf{x}\|) \geq (|m| - \epsilon)\|\mathbf{x}\|$$

Finally, with $\lambda := |m| - \epsilon$ and $\boldsymbol{\theta} = \boldsymbol{\theta}^\star + \mathbf{x}$

$$\frac{\|\mathbf{s}(\boldsymbol{\theta}) - \mathbf{s}(\boldsymbol{\theta}^\star)\|}{\|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|} \geqslant \lambda.$$

# Bibliography


Chapter 4 stochastic limits and calculus in mean square. In T.T. Soong, editor, *Random Differential Equations in Science and Engineering*, volume 103 of *Mathematics in Science and Engineering*, pages 72 – 114. Elsevier, 1973. doi: https://doi.org/10.1016/S0076-5392(08)60159-9. URL http://www.sciencedirect.com/science/article/pii/S0076539208601599.

François Abbaléa, Savine Andry, Marine Baulac, Michel C Bérengier, Bernard Bonhomme, Jérôme Defrance, Jean Pierre Deparis, Guillaume Dutilleux, David Ecotière, Benoît Gauvreau, et al. *Road Noise Prediction, 2—Noise Propagation Computation Method Including Meteorological Effects (NMPB 2008)*. Sétra, 2009.

Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition, 1964.

Pierre Aumond and Arnaud Can. Probabilistic modeling framework to predict traffic sound distribution. In *Euronoise 2018*, pages –, Hersonissos, France, May 2018. URL https://hal.archives-ouvertes.fr/hal-02011918. Euronoise 2018 , Hersonissos, GRECE, 27-/05/2018 - 31/05/2018.

Pierre Aumond, Arnaud Can, Bert De Coensel, Dick Botteldooren, Carlos Ribeiro, and Catherine Lavandier. Modeling soundscape pleasantness using perceptual assessments and acoustic measurements along paths in urban context. *Acta Acustica united with Acustica*, 103:430–443, 05 2017. doi: 10.3813/AAA.919073.

Pierre Aumond, Arnaud Can, Vivien Mallet, Bert De Coensel, Carlos Ribeiro, Dick Botteldooren, and Catherine Lavandier. Kriging-based spatial interpolation from measurements for sound level mapping in urban areas. *The Journal of the Acoustical Society of America*, 143(5):2847–2857, 2018. doi: 10.1121/1.5034799. URL https://doi.org/10.1121/1.5034799.

Pierre Aumond, Arnauda Can, Vivienb Mallet, Benoita Gauvreau, and Gwenaël Guillaume. Global sensitivity analysis of a noise mapping model based on open-source software. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, volume 259, pages 4031–4035. Institute of Noise Control Engineering, 2019.

Pierre Aumond, Nicolas Fortin, and Arnaud Can. Overview of the noisemodelling open-source software version 3 and its applications. In *Proceedings of the 49th International Congress and Exhibition on Noise Control Engineering e-congress*, 2020a.

Pierre Aumond, Vivien Mallet, Benoît Gauvreau, and Gwenaël Guillaume. Sensitivity analysis for urban noise modeling. *Applied acoustics*, 13, 1 2020b. doi: Submittedforpublication.

Pierre Aumond, Arnaud Can, Viven Mallet, Benoit Gauvreau, and Gwenaël Guillaume. Global sensitivity analysis for road traffic noise modelling. *Applied Acoustics*, 176:107899, 2021. ISSN 0003-682X. doi: https://doi.org/10.1016/j.apacoust.2020.107899. URL https://www.sciencedirect.com/science/article/pii/S0003682X20310021.

Juan Miguel Barrigón Morillas, Valentín Escobar, Juan Antonio Sierra, Rosendo Vílchez-Gómez, José Vaquero, and José Trujillo Carmona. A categorization method applied to the study of urban road traffic noise. *The Journal of the Acoustical Society of America*, 117:2844–52, 06 2005. doi: 10.1121/1.1889437.

Mathias Basner, Wolfgang Babisch, Adrian Davis, Mark Brink, Charlotte Clark, Sabine Janssen, and Stephen Stansfeld. Auditory and non-auditory effects of noise on health. *Lancet (London, England)*, 383(9925):1325–1332, Apr 2014. ISSN 1474-547X. doi: 10.1016/S0140-6736(13)61613-X. URL https://www.ncbi.nlm.nih.gov/pubmed/24183105.

Olivier Baume, Benoit Gauvreau, Michel Berengier, Fabrice Junker, Hans Wackernagel, and Jean-Paul Chiles. Geostatistical modeling of sound propagation: Principles and a field application experiment. *The Journal of the Acoustical Society of America*, 126:2894–904, 12 2009. doi: 10.1121/1.3243301.

Páll Bergthörsson and Bo R Döös. Numerical weather map analysis. *Tellus*, 7(3): 329–340, 1955.

Dimitri P Bertsekas. Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334, 1997.

Françis Besnard, Jean-François Hamet, Joël Lelong, Emmanuel Le Duc, Vincent Guizard, Nathalie Fürst, Sonia Doisy, and Guillaume Dutilleux. *Road Noise Prediction, 1—Calculating sound emissions from road traffic (NMPB 2008)*. Sétra, 2009.

Erwan Bocher, Gwenaël Guillaume, Judicaël Picaut, Gwendall Petit, and Nicolas Fortin. Noisemodelling: An open source gis based tool to produce environmental noise maps. *ISPRS International Journal of Geo-Information*, 8(3), 2019. ISSN 2220-9964. doi: 10.3390/ijgi8030130. URL https://www.mdpi.com/2220-9964/8/3/130.

F. Bouttier and P. Courtier. *Meteorological Training Course Lecture Series*, chapter Data Assimilation Concepts and Methods, page 0. ECMWF, 2002. URL https://www.ecmwf.int/node/16928. Learning.

David S Broomhead and David Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. Technical report, Royal Signals and Radar Establishment Malvern (United Kingdom), 1988.

Arthur Bryson, Y.-C Ho, and George Siouris. Applied optimal control: Optimization, estimation, and control. *Systems, Man and Cybernetics, IEEE Transactions on*, 9: 366 – 367, 07 1979. doi: 10.1109/TSMC.1979.4310229.

Arnaud Can, Luc Dekoninck, and Dick Botteldooren. Measurement network for urban noise assessment: Comparison of mobile measurements and spatial interpolation approaches. *Applied Acoustics*, 83:32–39, 09 2014. doi: 10.1016/j.apacoust.2014.03.012.

CEAC. Doc 29: Report on standard method of computing noise contours around civil airports. Technical report, CEAC, ECAC, 1997.

Stephen E Cohn. An introduction to estimation theory. *Journal of the Meteorological Society of Japan. Ser. II ("Special Issue" Data Assimilation in Meteology and Oceanography: Theory and Practice)*, 75(1B):257–288, 1997.

P Courtier, Jn Thepaut, and A Hollingsworth. A Strategy For Operational Implementation Of 4d-Var, Using An Incremental Approach. *Quarterly Journal Of The Royal Meteorological Society*, 120(519, B):1367–1387, JUL 1994. ISSN 0035-9009. doi: {10.1002/qj.49712051912}.

Bert De Coensel, Kang Sun, Weigang Wei, Timothy Van Renterghem, Matthieu Sineau, Carlos Ribeiro, Arnaud Can, Pierre Aumond, Catherine Lavandier, and Dick Botteldooren. Dynamic noise mapping based on fixed and mobile sound measurements. In *Proceedings of the 10th European Congress and Exposition on Noise Control Engineering*, pages 2339–2344, 2015.

DP Dee. Testing the perfect-model assumption in variational data assimilation. In *Proc. Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography*, pages 225–228. Citeseer, 1995.

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3385–3396, 2005. doi: 10.1256/qj.05.108. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.108.

Frits Eerden, Freek Graafland, Peter Wessels, A.J. Segers, and Erik Salomons. Internoise 2014 - 43rd international congress on noise control engineering: Improving the world through noise control. In *Model based monitoring of traffic noise in an urban district*, page 0, 11 2014.

European Commission. Directive 2002/49/ec of the european parliament and of the council of 25 june 2002 relating to the assessment and management of environmental noise, 2002. URL http://data.europa.eu/eli/dir/2002/49/oj.

European Commission. Commission directive (eu) 2015/996 of 19 may 2015 establishing common noise assessment methods according to directive 2002/49/ec of the european parliament and of the council (text with eea relevance), 2015. URL http://data.europa.eu/eli/dir/2015/996/oj.

European Environment Agency. *Noise in Europe 2014*, 2014.

R. A. Fisher. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London Series A*, 222:309–368, January 1922. doi: 10.1098/rsta.1922.0009.

George S. Fishman. *Monte Carlo: Concepts, Algorithms and Applications*. Springer Verlag, New York, NY, USA, 1996.

Gregg G. Fleming, Amanda S. Rapoza, and Cynthia S. Y. Lee. Development of national reference energy mean emission levels for the fhwa traffic noise model (fhwa tnm), version 1.0. Nov 1995. URL https://rosap.ntl.bts.gov/view/dot/6290. DOT-VNTSC-FHWA-96-2;FHWA-PD-96-008.

Harvey Fletcher and WA Munson. Loudness, its definition, measurement and calculation. *The Journal of the Acoustical Society of America*, 5(2):82–108, 1933.

R. Fletcher. *Newton-Like Methods*, chapter 3, pages 44–79. John Wiley & Sons, Ltd, 2013. ISBN 9781118723203. doi: 10.1002/9781118723203.ch3. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118723203.ch3.

GKE Geddes. The assessment of noise in audio-frequency circuits. *BBC research division Report. EL-17 UDC*, 534, 1968.

Arthur Gelb. *Applied optimal estimation*. MIT press, 1974.

Natalia Genaro, Antonio Torija, A Ramos-Ridao, Ignacio Requena, Diego Ruiz, and Montserrat Zamorano. A neural network based model for urban noise prediction. *The Journal of the Acoustical Society of America*, 128:1738–46, 10 2010. doi: 10.1121/1.3473692.

Félix Gontier, Catherine Lavandier, Pierre Aumond, Mathieu Lagrange, and Jean-François Petiot. Estimation of the perceived time of presence of sources in urban acoustic environments using deep learning techniques. *Acta Acustica united with Acustica*, 105, 11 2019. doi: 10.3813/AAA.919384.

Caspar A. Hallmann, Martin Sorg, Eelke Jongejans, Henk Siepel, Nick Hofland, Heinz Schwan, Werner Stenmans, Andreas Müller, Hubert Sumser, Thomas Hörren, Dave Goulson, and Hans de Kroon. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLOS ONE*, 12(10):1–21, 10 2017. doi: 10.1371/journal.pone.0185809. URL https://doi.org/10.1371/journal.pone.0185809.

Carl Hart, Nathan Reznicek, David Wilson, Chris Pettit, and Edward Nykaza. Comparisons between physics-based, engineering, and statistical learning models for outdoor sound propagation. *The Journal of the Acoustical Society of America*, 139:2640–2655, 05 2016. doi: 10.1121/1.4948757.

Sabrina Havard, Brian J Reich, Kathy Bean, and Basile Chaix. Social inequalities in residential exposure to road traffic noise: An environmental justice analysis based on the record cohort study. *Occupational and Environmental Medicine*, 68(5):366–374, 2011. ISSN 1351-0711. doi: 10.1136/oem.2010.060640.

David Miles Huber and Robert E Runstein. *Modern recording techniques*. CRC Press, 2013.

Charlotte Hurtley, editor. *Night noise guidelines for Europe*. World Health Organization Europe, Copenhagen, Denmark, 2009. ISBN 978-92-890-4173-7. OCLC: ocn475454508.

Bertrand Iooss and Amandine Marrel. Advanced methodology for uncertainty propagation in computer experiments with large number of inputs. *Nuclear Technology*, 205:1–19, 03 2019. doi: 10.1080/00295450.2019.1573617.

Bertrand Iooss, Loïc Boussouf, Vincent Feuillard, and Amandine Marrel. Numerical studies of the metamodel fitting and validation processes. *International Journal of Advances in Systems and Measurements*, 3, 01 2010.

Alexandre Janon, Maelle Nodet, and Clementine Prieur. Uncertainties assessment in global sensitivity indices estimation from metamodels. *International Journal for Uncertainty Quantification*, 4:21–36, 01 2014. doi: 10.1615/Int.J.UncertaintyQuantification.2012004291.

Donald B. Johnson. Efficient algorithms for shortest paths in sparse networks. *J. ACM*, 24(1):1–13, January 1977. ISSN 0004-5411. doi: 10.1145/321992.321993. URL http://doi.acm.org/10.1145/321992.321993.

I.T. Jolliffe. *Principal Component Analysis*. Springer Verlag, 1986.

Gunnar Jónsson and Finn Jacobsen. A comparison of two engineering models for outdoor sound propagation: Harmonoise and nord2000. *Acta Acustica united with Acustica*, 94, 03 2008. doi: 10.3813/AAA.918031.

R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 03 1960. ISSN 0021-9223. doi: 10.1115/1.3662552. URL https://doi.org/10.1115/1.3662552.

Stylianos Kephalopoulos, Marco PAVIOTTI, and Fabienne Anfosso-Lédée. *Common noise assessment methods in Europe (CNOSSOS-EU)*. PUBLICATIONS OFFICE OF THE EUROPEAN UNION, January 2012. doi: 10.2788/31776. URL https://hal.archives-ouvertes.fr/hal-00985998.

Catherine Lavandier and Boris Defréville. The contribution of sound source characteristics in the assessment of urban soundscapes. *Acta Acustica united with Acustica*, 92(6):912–921, 2006.

Catherine Lavandier, Pierre Aumond, Saul Gomez, and Catherine Dominguès. Urban soundscape maps modelled with geo-referenced data. *Noise Mapping*, 3, 01 2016. doi: 10.1515/noise-2016-0020.

François-Xavier Le Dimet and Olivier Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, 38A(2):97–110, 1986. doi: 10.1111/j.1600-0870.1986.tb00459.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.1986.tb00459.x.

Loïc Le Gratiet, Stefano Marelli, and Bruno Sudret. *Metamodel-Based Sensitivity Analysis: Polynomial Chaos Expansions and Gaussian Processes*, chapter 0, pages 1289–1325. Springer International Publishing, Cham, 2017. ISBN 978-3-319-12385-1. doi: 10.1007/978-3-319-12385-138". URL https://doi.org/10.1007/978-3-319-12385-138.

Antoine Lesieur, Pierre Aumond, Arnaud Can, and Vivien Mallet. Meta-model aided inverse modelling and joint state parameter estimation for noise data assimilation (submitted). *The Journal of the Acoustical Society of America*, a.

Antoine Lesieur, Vivien Mallet, and Julien Salomon. Estimation and diagnosis of the parameters of a covariance matrix (submitted). *Quarterly Journal of the Royal Meteorological Society*, b.

Antoine Lesieur, Pierre Aumond, Vivien Mallet, and Arnaud Can. Meta-modeling for urban noise mapping. *The Journal of the Acoustical Society of America*, 148(6):3671–3681, 2020. doi: 10.1121/10.0002866. URL https://doi.org/10.1121/10.0002866.

Antoine Lesieur, Vivien Mallet, Pierre Aumond, and Arnaud Can. Data assimilation for urban noise mapping with a meta-model. *Applied Acoustics*, 178:107938, 2021. ISSN 0003-682X. doi: https://doi.org/10.1016/j.apacoust.2021.107938. URL https://www.sciencedirect.com/science/article/pii/S0003682X21000311.

Xu-Qing Liu, Jian-Ying Rong, and Xiu-Ying Liu. Best linear unbiased prediction for linear combinations in general mixed linear models. *Journal of Multivariate Analysis*, 99(8):1503 – 1517, 2008. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2008.01.004. URL http://www.sciencedirect.com/science/article/pii/S0047259X0800016X.

Hai Dang Mai. *The operational application of dynamic traffic assignment models and the convergence of equilibrium algorithms*. Theses, Ecole des Ponts ParisTech, December 2006. URL https://pastel.archives-ouvertes.fr/pastel-00002183.

Vivien Mallet, Anne Tilloy, David Poulet, Sylvain Girard, and Fabien Brocheton. Meta-modeling of adms-urban by dimension reduction and emulation. *Atmospheric Environment*, 184:37 – 46, 2018. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2018.04.009. URL http://www.sciencedirect.com/science/article/pii/S1352231018302346.

G. Matheron. *Traité de géostatistique appliquée*. Number vol. 1 in Memoires. Éditions Technip, 1962. URL https://books.google.fr/books?id=88YKAQAAMAAJ.

M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979. ISSN 00401706. URL http://www.jstor.org/stable/1268522.

Michael D. McKay. Latin hypercube sampling as a tool in uncertainty analysis of computer models. In *Proceedings of the 24th Conference on Winter Simulation*, WSC '92, page 557–564, New York, NY, USA, 1992. Association for Computing Machinery.

ISBN 0780307984. doi: 10.1145/167293.167637. URL https://doi.org/10.1145/167293.167637.

Richard Ménard and Lang-Ping Chang. Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part II: chi2-Validated Results and Analysis of Variance and Correlation Dynamics. *Monthly Weather Review*, 128(8):2672–2686, 08 2000. ISSN 0027-0644. doi: 10.1175/1520-0493(2000)128<2672:AOSCTO>2.0.CO;2. URL https://doi.org/10.1175/1520-0493(2000)128<2672:AOSCTO>2.0.CO;2.

Richard Ménard and Roger Daley. The application of kalman smoother theory to the estimation of 4dvar error statistics. *Tellus A*, 48(2):221–237, 1996. doi: 10.1034/j.1600-0870.1996.t01-1-00003.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1600-0870.1996.t01-1-00003.x.

Henk Miedema, S. Janssen, K. Rokho, and Lex Brown. Burden of disease from environmental noise - quantification of healthy life years lost in europe. *Book Burden of disease from environmental noise - Quantification of healthy life years lost in Europe*, pages 91–98, 01 2011.

Momir Prascevic, Darko Mihajlov, Dragan Cvetkovic, and Aleksandar Gajicki. Assessment of environmental noise by harmonica index – case study: The city of niš. In *Acoustics & Vibration of Mechanical Structures II*, volume 801 of *Applied Mechanics and Materials*, pages 51–59. Trans Tech Publications Ltd, 11 2015. doi: 10.4028/www.scientific.net/AMM.801.51.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

Guillermo Rey Gozalo and Juan Morillas. Analysis of sampling methodologies for noise pollution assessment and the impact on the population. *International Journal of Environmental Research and Public Health*, 13:18, 11 2016. doi: 10.3390/ijerph13050490.

Lars Riishøjgaard. A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus A*, 50:42–57, Sep 2002. doi: 10.1034/j.1600-0870.1998.00004.x.

Olivier Roustant, David Ginsbourger, and Yves Deville. Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software, Articles*, 51(1):1–55, 2012. ISSN 1548-7660. doi: 10.18637/jss.v051.i01. URL https://www.jstatsoft.org/v051/i01.

Erik Salomons, Dirk van maercke, Jérôme Defrance, and Foort Roo. The harmonoise sound propagation model. *Acta Acustica united with Acustica*, 97:62–74, 01 2011. doi: 10.3813/AAA.918387.

Jaume Segura Garcia, J.J. Perez-Solano, Máximo Serrano, Enrique Navarro, Santiago Felici-Castell, Antonio Soriano, and Francisco Montes. Spatial statistical analysis of

urban noise data from a wasn gathered by an iot system: Application to a small city. *Applied Sciences*, 6:380, 11 2016. doi: 10.3390/app6120380.

Xavier Sevillano, Joan Claudi Carrié, Francesc Alías, Patrizia Bellucci, Laura Peruzzi, Simone Radaelli, Paola Coppi, Luca Nencini, Andrea Cerniglia, Alessandro Bisceglie, R Benocci, and Giovanni Zambon. Dynamap – development of low cost sensors networks for real time noise mapping. *Noise Mapping*, 3, 01 2016. doi: 10.1515/noise-2016-0013.

Jack Sherman and Winifred J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Ann. Math. Statist.*, 21(1):124–127, 03 1950. doi: 10.1214/aoms/1177729893. URL https://doi.org/10.1214/aoms/1177729893.

Ligia Silva, Marta Oliveira, and José Ferreira Silva. Urban form indicators as proxy on the noise exposure of buildings. *Applied Acoustics*, 76, 01 2013. doi: 10.1016/j.apacoust.2013.07.027.

I Sobol. On sensitivity estimation for nonlinear mathematical models. *Keldysh Applied Mathematics Institute*, 1:112–118, 01 1990.

Albert Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, USA, 2004. ISBN 0898715725.

Quentin M Tenailleau, Nadine Bernard, Sophie Pujol, Hélene Houot, Daniel Joly, and Frédéric Mauny. Assessing residential exposure to urban noise using environmental models: does the size of the local living neighborhood matter? *Journal of Exposure Science and Environmental Epidemiology*, 25(1):89–96, 2015.

Anne Tilloy, Vivien Mallet, David Poulet, Céline Pesin, and F. Brocheton. Blue-based no 2 data assimilation at urban scale. *Journal of Geophysical Research: Atmospheres*, 118, 02 2013. doi: 10.1002/jgrd.50233.

H.J.A. van Leeuwen. Railway noise prediction models: a comparison. *Journal of Sound and Vibration*, 231(3):975 – 987, 2000. ISSN 0022-460X. doi: https://doi.org/10.1006/jsvi.1999.2570.

Raphael Ventura, Vivien Mallet, and Valerie Issarny. Assimilation of mobile phone measurements for noise mapping of a neighborhood. *The Journal of the Acoustical Society of America*, 144(3):1279–1292, 2018. doi: 10.1121/1.5052173. URL https://doi.org/10.1121/1.5052173.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Weigang Wei, Timothy Renterghem, Bert De Coensel, and Dick Botteldooren. Dynamic noise mapping: A map-based interpolation between noise measurements with high temporal resolution. *Applied Acoustics*, 101:127–140, 01 2016. doi: 10.1016/j.apacoust.2015.08.005.

Kohei Yamamoto. Road traffic noise prediction model "asj rtn-model 2008": Report of the research committee on road traffic noise. *Acoustical Science and Technology*, 31: 2–55, 01 2010. doi: 10.1250/ast.31.2.

Giovanni Zambon, R. Benocci, Alessandro Bisceglie, H. Roman, and Patrizia Bellucci. The life dynamap project: Towards a procedure for dynamic noise mapping in urban areas. *Applied Acoustics*, 11 2016. doi: 10.1016/j.apacoust.2016.10.022.

Chadia Zayane, Christian Lajaunie, Laurent Praly, Alexandre Girard, and Jean-Marc Jicquel. Joint state and parameter estimation: a bayesian approach. *IEEE ICMSC*, 01 2010.