Master thesis and internship[BR]- Master's Thesis : Optimal criterion to define convergence in unsteady CFD simulations for applications in Formula 1 car aerodynamics[BR]- Internship (linked to master's thesis)

Auteur : Cotteleer, Léo
Promoteur(s) : Terrapon, Vincent
Faculté : Faculté des Sciences appliquées
Diplôme : Master en ingénieur civil en aérospatiale, à finalité spécialisée en "aerospace engineering"
Année académique : 2019-2020
URI/URL : http://hdl.handle.net/2268.2/9082

UNIVERSITY OF LIÈGE

FACULTY OF APPLIED SCIENCES

Optimal criterion to define convergence in unsteady CFD simulations for applications in Formula 1 car aerodynamics

Master thesis submitted in partial fulfillment of the requirements for the degree of Master in Aerospace Engineering

By Léo Cotteleer

Advisor:	Prof. Dr. Vincent E. Terrapon
Jury:	Dr. Thomas Andrianne
	Marco Hahn
	Gennaro Serino

Academic year 2019 - 2020

Abstract

Computational Fluid Dynamics (CFD) analyses have become a tool of choice to deepen physical understanding and a precious help to solve practical problems. In the light of the continuous enhancement in computer power, the use of this new numerical tool is not planned to stop. This growth in computing capacity is leading to an increasing interest in scale-resolving turbulence such as detached-eddy simulation (DES). The unsteady pattern of the flow in Formula 1 racing therefore results to simulations that are very computational resources consuming. The resolution of numerous turbulent scales requires small time steps. In addition, time-dependent simulations take time to develop from arbitrary initial conditions to a statistically stationary state such that long samples are necessary for reliable statistical estimates. The initial transient duration and the required simulation time are highly case-specific, user burden and a priori unknown.

This work introduces a methodology that automates the initial transient estimation during a numerical CFD simulation and the prediction of the required simulation time to obtain reliable statistics. The techniques rely on statistical considerations for aerodynamic quantities of interest and have shown satisfactory results, in line with the a posteriori intuitions. In the two first parts, a rigorous mathematical background is provided and the methodology is described as well as its previous versions. In particular, a chapter is dedicated to the estimation of the statistical error by means of a second order model. The discussions and results provide a wide range of implementation issues and test cases. Especially, Delayed Detached Eddy Simulations (DDES) are performed for common bluff bodies (high angle-of-attack airfoil, triangular cylinder, rotating circular cylinder). Finally, the methodology has been implemented as a User-Defined Function (UDF) in ANSYS Fluent and in the last chapter are exposed the different interesting quantifications that can be extrapolated from the methodology. A concise summary of the main outcomes is provided in the conclusion. The latter is written in such a way that each principle finding is linked to the corresponding section.

Acknowledgements

First of all, I would like to express my deep gratitude to Professor Vincent E Terrapon. As supervisor, he has really helped me with his useful critiques, relevant remarks and his support throughout this work. Also, i would like to thank RedBull Formula One for the opportunity to work on such a challenging and motivating project. Especially i would like to thank Mr Hahn and Mr Serino for their advices and their time.

Life is often referred as a succession of choices, instants and encounters and it is often claimed that it can be drastically influenced by the people we run alongside with. As far as I am concerned, I can truly relate to this adage as my life would probably not be the same if I had not been lucky enough to meet two smart and inspiring people whom I hold in high esteem and who are now my two best friends : Sarah and Ismael.

Since the day we met on the school benches in high school, this great bond has only grown and evolved. With them, I experienced my first parties, my first drinks on terraces and all theses life experiences that shape you. But beyond everything, it was with them that I calculated my very first derivatives, my first integrals, that I have spent my first (among many, after these five years of university) intensive days of study and that I have passed my first exams at university. It is for these last points especially that I would like to thank them. They have represented the best what friendship is and should be. For many years, they constantly pulled me up and inspired me : Sarah, through her passion for space, which led me to get involved in the aerospacial field and led her to become an astrophysicist ; Ismael, through his dream to become a civil engineer, which widely influenced me with my own choice of studies. It is therefore certain that without these two people, my journey could have been completely different and I might not have been, currently, writing these thanks in my aerospace engineering masters degree's thesis. From the bottom of my heart, I wish them the future that they deserve and may their future life as an astrophysicist and a civil engineer bring them the satisfaction and they joy to which they aspire.

Finally, i would like to thank my family, in particular my parents and my sister Louise for their support during these five years. I hope this final work, that ends my university education, make you proud.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Context

Fluid flows analysis appears to be a fundamental topic in numerous fields as it became an everyday problem in sciences and engineering. In the particular case of the extreme competitive environment of Formula 1 racing, every team seek to optimize their cars. One way to increase the performances is to analyse the aerodynamics. In addition with the very first theoretical and experimental observations of the last years, the use of Computational Fluid Dynamics (CFD) analyses has become a tool of choice to deeper physical understanding and a precious help to solve practical problems. In the light of the continuously improvement in computer power, the use of this new numerical tool is not planned to stop. However, due to a lack of accuracy for limited computer performances, CFD analyses are still into development and can not be considered as the perfect tool.

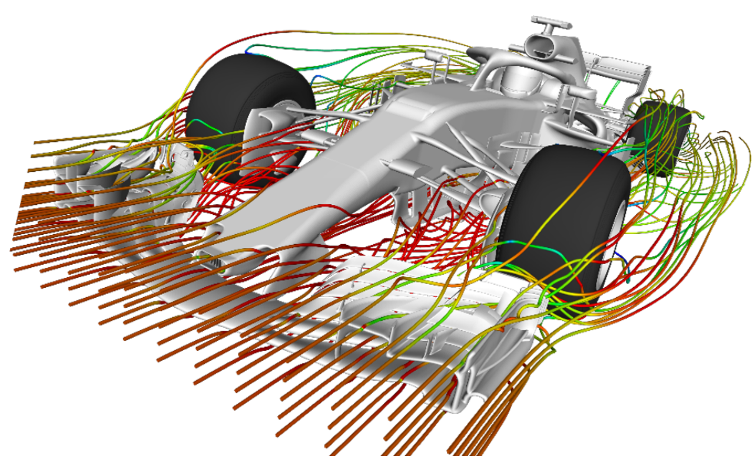

Figure 1.1: CFD simulation illustrating the flow around the RB14.

The inherent turbulent nature of the time-dependent flow around a Formula 1 car represents a barrier to the accuracy, efficiency and possibilities of CFD application (see Figure 1.1). Indeed, representing turbulence results in a trade-off between simulation quality and computational cost. Despite direct numerical simulations (DNS) allows the resolution of the whole range of turbulence scales in space and in time, the computational cost is still outside of the possible. Other computational approaches, such as the Unsteady Reynolds-averaged Navier-Stockes (URANS) equations, Large-eddy simulation (LES) or Detached-eddy simulation (DES) lead to a reduced cost whilst introducing uncertainties coming from modelling simplifications.

The unsteady pattern of the flow in Formula 1 racing therefore leads to simulations that are very computational resources consuming. The resolution of numerous turbulent scales needs small time steps. In addition, time-dependent simulations do not usually converge to a solution constant in time such that long samples are necessary for reliable statistical estimates. All this combined highlights the large amount of computing resources needed for CFD analyses and the challenges that CFD will have to face in the next decades.

1.2 Motivations

A special attention is usually required for unsteady simulations with scale resolution models. The resolved turbulence structures need a certain time to develop and to be transported through the fluid domain. After that, it is expected that the solution converges to a statistically steady solution. However no systematic methodology is able to correctly predict the time required to settle as it is usually very case specific. That is, the assessment of the needed time to settle is usually made by visual inspection of relevant quantities.

The usual guidelines to identify the initial transient is to monitor the simulation continuously during the running time of the simulation. Some monitoring points can be added at interesting locations and the time traces of their development in analysed. The amplitude and frequency of the local oscillations should be stabilized. However, the initial transient can correspond to several thousand of time steps and the self convergence checking can be time and user consuming.

In addition, engineers are rather interested in time-averaged quantities than in instantaneous details of the turbulence. Therefore, once the statistically steady state is reached, a time averaging operation is performed. Once again, there are no general guidelines to estimate accurately the number of time steps required to obtain relevant statistics. The common habit is to average a long time to be sure that the statistics are satisfactory.

It is therefore clear that the post-processing and averaging part of a numerical simulation is not robust and requires to pay attention continuously to the time evolution of the solution.

1.3 Objectives

In line with those issues, the objectives of the thesis are multiple and can be understood more schematically in Figure 4.1:

i. define and implement a convergence criterion to estimate the transient time t_t in a systematic and automatic way;

ii. define and implement a convergence criterion to estimate the necessary simulation time t_e to provide reliable statistics.

This methodology could not only simplify the post-processing and averaging process but also decrease strongly the computational cost of a simulation.

Indeed, a robust algorithm could allow to minimize both a priori unknown durations in the sense that no extra time steps would be considered due to a bad visual inspection. Moreover, by estimating the required simulation time t_e a quantification of the statistical error made on the averaging process can be made. This could allow to determine the averaged value with a certain confidence interval.

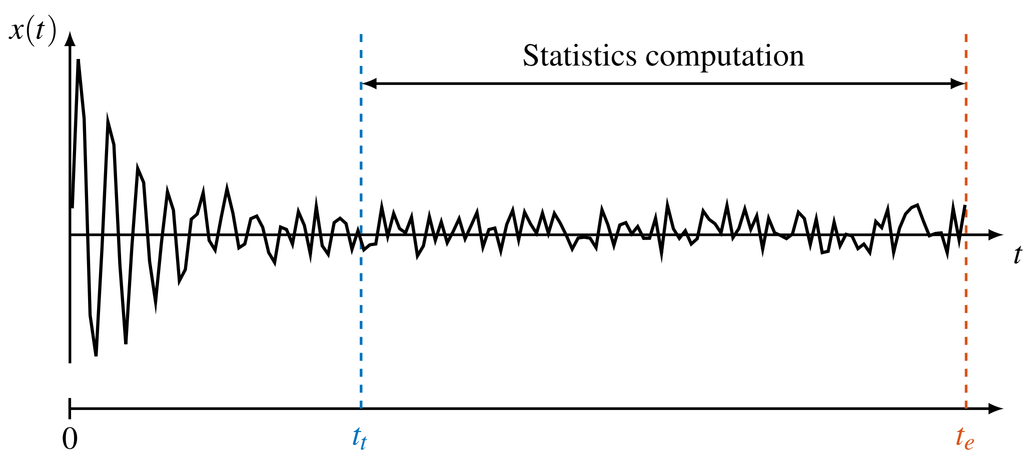

Figure 1.2: Schematic representation to highlight both the initial transient duration t_t and the required simulation time t_e.

To be really relevant, the algorithm should be able to estimate those durations during a simulation i.e on the fly. For that purpose, the flow monitor is implemented as a User-defined function (UDF) in ANSYS Fluent.

1.4 Outline of the thesis

The thesis is wanted to be presented in chronological order. Therefore, informations are provided step by step to reach the final form of the implementation. Hence, the latter is organized into three different parts divided into chapters. The first part is dedicated to the state of the art and a literature review. It aims at providing the relevant background to understand the latter discussions (Parts 2 and 3). The theoretical part is then divided into two chapters : chapter 2 provides a brief summary of the probability and statistic theories while chapter 3 extends those theories to the function of time, i.e the stochastic processes.

The second part explains the methodology used to define the convergence criteria. In chapter 4 is presented and reproduced a previous methodology of initial transient estimate [27]. Based on this are highlighted several drawbacks which give rise to chapter 5 and 6. The chapter 5 is dedicated to a theoretical support to expend the methodology of chapter 4 while in chapter 6 is presented an adapted version of the initial algorithm. Finally, chapter 7 is dedicated to the implementation of the algorithm as a User-Defined function in ANSYS Fluent.

The numerical results are presented and discussed in Part 3. The goal is to make a collection of investigations to perform a validation of key implementation features. In chapter 7 are collated the several test cases investigated in ANSYS fluent to check the validity of the implementation. Chapter 8 is devoted to the discussions and validations of the methodology.

Part I

State of the art

Outline of the first part

If you think that statistics has nothing to say about what you do or how you could do it better, then you are either wrong or in need of a more interesting job.

Stephen Senn

Turbulence is a state of fluid motion governed by the well-known Navier-Stokes equations. Although turbulence should simply be a solution of those equations, it results in a high complexity of description. This intricate flow represents a huge amount of information that cannot usually be fully understood. The engineers thus have to compress the information of this random, chaotic flow into a set of more accessible and relevant quantities. Therefore, the analysis of turbulence is directly linked with a statistical description.

The first chapters of the thesis are thus devoted to introduce some basic theoretical background of probability and statistical theories. The first section of the chapter about probabilities is wanted to give a rigorous mathematical formalism and to introduce the different notations that are used in the further chapters. The second section collates the different methods to extract relevant quantities from a large set of data. Particularly, it also quantify the possible errors that are made during this compression of information.

In fluid dynamics, the main quantities of interest are function of the time (e.g velocity) and thus the theories introduced in the first chapter have to be extended. This succession of random variables along time is referred to as a stochastic process. That is, the second chapter is dedicated to those processes and make the bridge with the first one to extend the theories to time-function. In particular, the concept of possible errors made during the compression of information is extended to be able to quantify these errors as a function of the time.

2 Probabilities and statistics background

The theory of probabilities is a branch of mathematics that aims to investigate random phenomenons. Like every mathematical theory, probability is based upon axioms and uses mathematical tools to prove theorems. In the other hand, statistics consist in collecting, treating and interpreting output data from a physical system (real as well as simulated). In particular, statistics help to build a model (probabilist or not) which represents accurately the measurable physical world. There is a fundamental interdependency between probabilities and statistics which uses probability framework as a tool. However, there is also a key point that differentiates both theories : probability is deductive while statistics are inductive.

In this chapter are presented the mathematical bases of probabilities and statistics. It allows to introduce the different notations and quantities of interest that will be used further in the thesis. For that purpose, the chapter is divided into two main sections, namely : Probabilist model and Statistics theory. As stated before, a bridge can clearly be established between both theories, that is, the chapter has to be read as two pieces from the same puzzle.

2.1 Probabilist model

An experiment is said to be random if there is no a priori prediction of its result, that is, repeated in the same conditions the latter can lead to a different result [44]. To investigate such an random experiment one has to define a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ where :

i. Ω is the sample state which contains all possible outcomes of the experiment denoted ω and called sample point;

ii. $\mathcal{F}$ is a σ-algebra and represents all subsets of Ω which can be realized;

iii. $\mathbb{P}$ is probability measure which respects the Kolmogorov axioms i.e a function $\mathbb{P} : \mathcal{F} \to [0,1]$ such that $\mathbb{P}(\Omega) = 1$ and $\forall F_1, F_2, ... \in \mathcal{F}$ there is $\mathbb{P}\left(\bigcup_i F_i\right) = \sum_i \mathbb{P}(F_i)$.

In the following, the theoretical concepts are developed for the function $\mathcal{X}(\cdot)$, also called random variable, defined on Ω, with real values and a σ-algebra $\mathcal{F}$ such that :

$$\forall F \in \mathcal{F} : \{\omega \in \Omega \,|\, \mathcal{X}(\omega) \in F\} \in \mathbb{R} \tag{2.1}$$

for sake of simplicity, the random variable is denoted $\mathcal{X}$ while $\{\omega \in \Omega \mid \mathcal{X}(\omega) \in F\}$ is denoted $\mathcal{X}^{-1}(F)$.

2.1.1 Basic description and properties

A real random variable $\mathcal{X}$ is said to be continuous if there is a function $f_{\mathcal{X}}(\cdot)$ defined on the real line such that $\forall a \leq b$ that is

$$\mathbb{P}(\mathcal{X} \in]a;b[) = \mathbb{P}(\mathcal{X} \in [a;b]) = \mathbb{P}(\mathcal{X} \in [a;b[) = \mathbb{P}(\mathcal{X} \in]a;b]) = \int_a^b f_{\mathcal{X}}(x)dx. \tag{2.2}$$

The function is called the probability density function which is positive and such that the integral on $\mathbb{R}$ is equal to 1.

Expected value

In a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, the expected value of a random variable $\mathcal{X}$, noted $E\{\mathcal{X}\}$ or equivalently $\mu_{\mathcal{X}}$ is defined as

$$E\{\mathcal{X}\} = \mu_{\mathcal{X}} = \int_{\Omega} \mathcal{X}(\omega) d\mathbb{P}(\omega), \tag{2.3}$$

which is totally equivalent to

$$E\{\mathcal{X}\} = \mu_{\mathcal{X}} = \int_{\mathbb{R}} x d\mathbb{P}(x) = \int_{\mathbb{R}} x f_{\mathcal{X}}(x) dx. \tag{2.4}$$

Variance and standard deviation

In the specific case of a finite[1] expected value $\mu_{\mathcal{X}} = E\{\mathcal{X}\}$, the variance of the random variable $\mathcal{X}$ is defined as

$$V\{\mathcal{X}\} = \sigma^2_{\mathcal{X}} = E\{(\mathcal{X} - \mu_{\mathcal{X}})^2\}, \tag{2.5}$$

when this quantity if finite i.e $V\{\mathcal{X}\} < \infty$. By definition, the standard deviation $\sigma_{\mathcal{X}}$ is the positive square root of the variance.

Expected value and variance represent respectively the central tendency and dispersion of the sample. A quantification of the combined tendency and dispersion can be obtained by defining the mean square value $\psi^2_{\mathcal{X}}$ as the variance plus the square of the mean. Moreover, they are

[1]NB : When the sample state Ω is finite itself, the expected value is always finite. On the contrary, for an infinite sample state, the expected value may not exist. Indeed, for a discrete random variable, the series might not converge and thus tends to infinity while for a continuous one the integral might not be defined.
Finally, it is worth to note that if $E\{\mathcal{X}^2\}$ is finite, then the expected value is also finite and the variance exists. [44]

linked by the Bienaymé-Tchebyshev inequality which states that $\forall a > 0$:

$$\mathbb{P}(|\mathcal{X} - \mu_{\mathcal{X}}| \geq a\sigma_{\mathcal{X}}) \leq \frac{1}{a^2}. \tag{2.6}$$

Covariance and correlation

Let's consider a couple of real random variables $\mathcal{X}$ and $\mathcal{Y}$ defined on the same probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. The covariance of $\mathcal{X}$ and $\mathcal{Y}$, denoted $\mathrm{cov}\{\mathcal{X};\mathcal{Y}\}$ or $C_{\mathcal{X}\mathcal{Y}}$, is defined as :

$$\mathrm{cov}\{\mathcal{X};\mathcal{Y}\} = C_{\mathcal{X}\mathcal{Y}} = E\{(\mathcal{X} - \mu_{\mathcal{X}})(\mathcal{Y} - \mu_{\mathcal{Y}})\} = E\{\mathcal{X}\mathcal{Y}\} - E\{\mathcal{X}\}E\{\mathcal{Y}\}, \tag{2.7}$$

when this quantity is finite. Further, the correlation function denoted $\mathrm{corr}\{\mathcal{X};\mathcal{Y}\}$ or $R_{\mathcal{X}\mathcal{Y}}$ is defined as $\mathrm{corr}\{\mathcal{X};\mathcal{Y}\} = E\{\mathcal{X}\mathcal{Y}\}$ such that both functions are linked through :

$$\mathrm{cov}\{\mathcal{X};\mathcal{Y}\} = \mathrm{corr}\{\mathcal{X};\mathcal{Y}\} - E\{\mathcal{X}\mathcal{Y}\}. \tag{2.8}$$

2.1.2 Convergence theorem

Central limit theorem

According to the central limit theorem the normal or Gaussian distribution often results from the sum of a large number of independent random variables [45]. More precisely, let's consider that $\{\mathcal{X}_n\}$ is a sequence of n mutually independent random variables of mean and variance $\mu_{\mathcal{X}}$ and $\sigma^2_{\mathcal{X}}$, respectively. The central theorem assets that :

$$\left(\frac{\sum_{i=1}^{n} \mathcal{X}_i - n\mu_{\mathcal{X}}}{\sigma_{\mathcal{X}}\sqrt{n}}\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \tag{2.9}$$

where $(\mathcal{X}_n) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ refers that the sequence $(\mathcal{X}_n)$ converges in law toward a normal distribution of 0 mean and unit variance.

In other words, the sampling distribution of the sequence $\{\mathcal{X}_n\}$ tends to a Gaussian distribution as the sample size becomes large and this, regardless of the distribution of the initial random variables $\mathcal{X}_i$, with $i = 1, ..., n$. In practice, the conclusion of the central limit theorem is mostly verified for $n > 4$ and becomes true for $n > 10$ [4].

2.2 Statistics theory

Any variable used to represent a physical phenomenon or any observed data can be qualified as being either deterministic or random *i.e* non-deterministic. A quantity is classified as deterministic if it is perfectly known, without any incertitudes or imperfections. Therefore these data can be described by a mathematical function. However, making the assumption of a fully

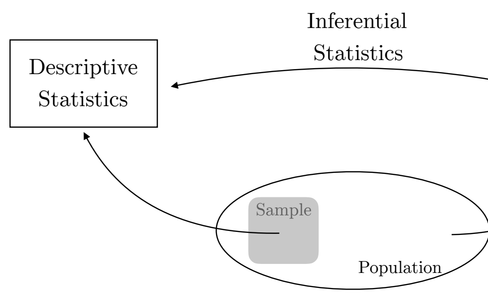

Figure 2.1: Schematic description of the connexions between inferential, descriptives statistics and probabilities.

predictive world is an illusion and can not match with reality. There is no way to predict exactly the world at a further time. Indeed, numerous external parameters can influence a phenomenon and it is therefore impossible to reproduce two times the same experiment with the same precision.

The result of an experiment can thus not be deterministic and should be treated as a random variable or random function. In other words, a given observation will represent a value among any other possible results that might have occurred. A quantity is then described in terms of its mean tendency and the dispersion around this mean value *i.e* in terms of probability statements and statistical averages. This variability of the results is quantified by means of two different approaches. In the one hand, descriptive statistics describe and quantify the variability. On the other hand, inferential statistics adjust a probability model for the measured data and thus is the bridge between probability theory and descriptives statistics (see Fig.2.1).

In the end, treating a phenomenon as deterministic is simply a specific case of probabilistic model. If several realisations of a same phenomenon always give the same result, the dispersion is null and the random variable is known with no incertitudes.

This section is thus divided in two main subsections : Sampling theory and Estimation theory. The first subsection is a branch of the descriptive statistics which aims to sum-up data collected during an experiment into quantities more easily understandable. In the other hand, the second subsection refers to inferential statistics which aims, by means of probabilist assumptions, to estimate and predict the behaviour of future experiments.

2.2.1 Sampling theory

Let consider the sample[2] $\mathbf{D}_n$ containing the n realisations $x_1, ..., x_n$ of the random variable $\mathcal{X}(\omega) : x_i = \mathcal{X}(\omega_i)$ defined on the probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. In the specific case of an independent and identically distributed sample, the subset of events $(\Omega_{\mathbf{D}_n}, \mathcal{F}_{\mathbf{D}_n}, \mathbb{P}_{\mathbf{D}_n})$ can be seen as the product, n times, of the subset $(\Omega_{\mathcal{X}}, \mathcal{F}_{\mathcal{X}}, \mathbb{P}_{\mathcal{X}})$.

A descriptive statistic is defined as every real random variable defined on $(\Omega_{\mathbf{D}_n}, \mathcal{F}_{\mathbf{D}_n}, \mathbb{P}_{\mathbf{D}_n})$ for $n \in \mathbb{N}_0$. In the following are described the main used sample statistics. Their naive definitions in this subsection will be more accurately explained in the next subsection : Estimation theory. Moreover, the descriptive statistics computed from a sample are denoted by a hat e.g $\hat{s}_{\mathcal{X}}(\mathbf{D}_n)$ is the descriptive statistical quantity s computed from a sample $\mathbf{D}_n$ from the parent random variable $\mathcal{X}$. This notation is part of a more general one also explicated in the next subsection.

Sample mean

The sample mean, noted $\hat{\mu}_{\mathcal{X}}$, is the random variable defined as :

$$\hat{\mu}_{\mathcal{X}}(\mathbf{D}_n) = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{2.10}$$

Under the assumption of a finite variance i.e $\sigma_{\mathcal{X}} < \infty$, the expected value and variance of the mean sample are respectively $E\{\hat{\mu}_{\mathcal{X}}\} = \mu_{\mathcal{X}}$ and $V\{\hat{\mu}_{\mathcal{X}}\} = \sigma^2_{\mathcal{X}}/n$. Therefore, $E\{\hat{\mu}_{\mathcal{X}}\}$ converges to $E\{\mathcal{X}\}$ as $n \to \infty$ and according to the law of large numbers $E\{\hat{\mu}_{\mathcal{X}}\}$ will converge almost surely (and thus in probability also) to the expected value, that is :

$$\hat{\mu}_{\mathcal{X}}(\mathbf{D}_n) \overset{a.s}{\to} \mu_{\mathcal{X}} \quad \equiv \quad \mathbb{P}\left(\lim_{n \to \infty} \hat{\mu}_{\mathcal{X}}(\mathbf{D}_n) = \mu_{\mathcal{X}}\right) = 1. \tag{2.11}$$

Finally, according to the central limit theorem, the expected value of the sample mean converges in law to a Gaussian distribution.

Sample variance

The sample variance, noted $\hat{\sigma}_{\mathcal{X}}$, is the random variable defined[3] as :

$$\hat{\sigma}^2_{\mathcal{X}}(\mathbf{D}_n) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu}_{\mathcal{X}})^2. \tag{2.12}$$

[2] To be very general, the sample is commonly of size $(n \times p)$: $\mathbf{D}_{n \times p}$ where n are the individuals (sample points) and p the different measured numerical parameters. In this chapter are only introduced the results for a uni-dimensional sample.

[3] Several definitions exist for the sample variance. This definition is chosen since this leads to a quantity unbiased and which does not require any a priori knowledge on the expected value $E\{\mathcal{X}\}$. The notion of unbiased is introduced in subsection 2.2.2.

Under the assumption of a finite variance i.e $\sigma_{\mathcal{X}} < \infty$, the expected value of the variance sample are respectively $E\{\hat{\sigma}^2_{\mathcal{X}}\} = \sigma^2_{\mathcal{X}}$. Therefore, $E\{\hat{\sigma}^2_{\mathcal{X}}\}$ converges to $V\{\mathcal{X}\}$ as $n \to \infty$ and according to both laws of large numbers $E\{\hat{\sigma}^2_{\mathcal{X}}\}$ will converge in probability and almost surely to the variance value, that is :

$$\hat{\sigma}^2_{\mathcal{X}}(\mathbf{D}_n) \overset{a.s}{\to} \sigma^2_{\mathcal{X}} \quad \equiv \quad \mathbb{P}\left(\lim_{n\to\infty} \hat{\sigma}^2_{\mathcal{X}}(\mathbf{D}_n) = \sigma^2_{\mathcal{X}}\right) = 1. \tag{2.13}$$

Although the sample variance converges to the variance, the descriptive statistic $\hat{\sigma}_{\mathcal{X}}$ underestimate the standard deviation according to the Jenssen inequality, namely $E\{\sqrt{\mathcal{X}}\} \leq \sqrt{E\{\mathcal{X}\}}$.

2.2.2 Estimation theory

In practice, the number of events is finite so that the computation of descriptive statistics can differ from the true values of the mean and the variance. Therefore, the goal is to estimate a numerical parameter that is as near as possible from its true value and to define an interval centred in this parameter that quantify the accuracy of the estimate.

Let consider that a population is analysed by means of a random variable $\mathcal{X}$ over the sample $\mathbf{D}_n$ and that its probability density function depends on a parameter θ to estimate, that is $f_{\mathcal{X}}(x;\theta)$. The numerical parameter is assumed to lie in the subset of possible $\Theta \subset \mathbb{R}$. If θ_0 is the true value of the parameter, the goal is to compute from $\mathbf{D}_n$:

i. a point estimation $\hat{\theta}$, near θ_0 in average;

ii. an interval estimation $[\theta_{\text{low}};\theta_{\text{up}}]$ which includes θ_0 with a certain uncertainty (resp. probability) γ (resp. $1-\gamma$).

Point estimation

Let $P(\cdot)$ a function, called estimator, which compute $\hat{\theta}$ with the knowledge of the sample $\mathbf{D}_n$. Hence, the estimator defines for a fixed sample size n an associated random variable $\mathcal{P}_n$.

The sample mean and the sample variance exposed previously are actually point estimators $\hat{\mu}_{\mathcal{X}}$, $\hat{\sigma}^2_{\mathcal{X}}$ of the true values $\mu_{\mathcal{X}}$ and $\sigma^2_{\mathcal{X}}$ respectively. This thus justifies the hat notation used in the previous section.

Error in estimation

The estimation error of the estimator $P(\cdot)$ is equal to the difference $P(\mathbf{D}_n) - \theta_0$ or equivalently $\hat{\theta} - \theta_0$. In terms of the associated random variable $\mathcal{P}_n$, the estimation error can be quantified by means of the mean square error. Indeed, the latter is the expected value of the dispersion of the estimate from its expected value :

$$\mathrm{MSE}\{\mathcal{P}_n\} = E\{(\mathcal{P}_n - \theta_0)^2\} \tag{2.14}$$

which can also be written as the sum of the variance of the estimate and the square of the bias of the estimate[4]:

$$E\{(\mathcal{P}_n - \theta_0)^2\} = V\{\mathcal{P}_n\} + b^2\{\mathcal{P}_n\}. \tag{2.15}$$

The bias of the estimate corresponds to the difference between the expected value of the estimate and the true value. The estimate is said to be unbiased if the expected value equals the true value. The measure is thus intrinsic to the estimator and is a systematic error that appears in the same magnitude for each experiment. That is :

$$b^2\{\mathcal{P}_n\} = E\{(E\{\mathcal{P}_n\} - \theta_0)^2\}. \tag{2.16}$$

In the other hand, the variance is referred as the expected value of the squared difference between the estimate and its mean value. Therefore, the variance is a random error which differs from a measurement to another (in magnitude as well as in direction). That is :

$$V\{\mathcal{P}_n\} = E\{(\mathcal{P}_n - E\{\mathcal{P}_n\})^2\}. \tag{2.17}$$

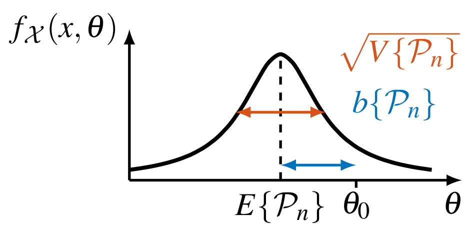

Figure 2.2: Schematic representation of the bias (blue) and the standard deviation (orange) of an estimator $\mathcal{P}_n$ of the statistical quantity θ generated with the random variable $\mathcal{X}$ characterised by its probability structure $f_{\mathcal{X}}(x,\theta)$.

A normalised root mean square error or coefficient of variation can be defined to express the

[4]Expanding Equation (2.14) :

$$E\{(\mathcal{P}_n - \theta_0)^2\} = E\{(\mathcal{P}_n - E\{\mathcal{P}_n\} + E\{\mathcal{P}_n\} - \theta_0)^2\} = E\{(\mathcal{P}_n - E\{\mathcal{P}_n\})^2\} + 2E\{(\mathcal{P}_n - E\{\mathcal{P}_n\})(E\{\mathcal{P}_n\} - \theta_0)\} + E\{(E\{\mathcal{P}_n\} - \theta_0)^2\}$$

By linearity of the expected value operator, the middle term is zero $E\{\mathcal{P}_n - E\{\mathcal{P}_n\}\} = E\{\mathcal{P}_n\} - E\{\mathcal{P}_n\}$, the mean square error reduces to

$$\mathrm{MSE}\{\mathcal{P}_n\} = E\{(\mathcal{P}_n - E\{\mathcal{P}_n\})^2\} + E\{(E\{\mathcal{P}_n\} - \theta_0)^2\} \equiv V\{\mathcal{P}_n\} + b^2\{\mathcal{P}_n\}$$

with the introduction of $V\{\mathcal{P}_n\} = E\{(\mathcal{P}_n - E\{\mathcal{P}_n\})^2\}$ and $b^2\{\mathcal{P}_n\} = E\{(E\{\mathcal{P}_n\} - \theta_0)^2\}$.

error of an estimate in terms of a percentage of the estimated :

$$\varepsilon\{\mathcal{P}_n\} = \frac{\sqrt{\text{MSE}\{\mathcal{P}_n\}}}{\theta_0} = \frac{\sqrt{V\{\mathcal{P}_n\} + b^2\{\mathcal{P}_n\}}}{\theta_0} \tag{2.18}$$

Interval estimation

In point estimation, there is no information about how does this estimate approach the parameter of interest. Another common way to proceed in estimation theory is to provide an interval $a < \theta < b$, in which the estimated parameter lies with a certain uncertainty.

Let $\mathcal{X}$ be a random variable of probability density function $f_{\mathcal{X}}(x;\theta)$, θ be the statistical quantity of $\mathcal{X}$ to be estimated and $P(\cdot)$ the estimator function. A confidence interval for θ with a given uncertainty γ, is an interval $[p_1; p_2]$ of probability $1-\gamma$ for the estimator $\mathcal{P}_n$, that is :

$$\mathbb{P}(\mathcal{P}_n \in [p_1; p_2]) = 1 - \gamma. \tag{2.19}$$

The functions that bound the interval $p_1(\theta)$ and $p_2(\theta)$ are function of γ as well as the estimator function $P(\cdot)$ used. In practice, the functions are chosen to build a symmetric interval centred in θ i.e $p_1 = \theta - \Delta\theta$ and $p_2 = \theta + \Delta\phi$.

As represented in Figure 2.3, the true value θ_0 lies thus in the interval $[\theta_{\text{low}}; \theta_{\text{up}}]$ where $\theta_{\text{low}} = p_2^{-1}(\theta)$ and $\theta_{\text{up}} = p_1^{-1}(\theta)$

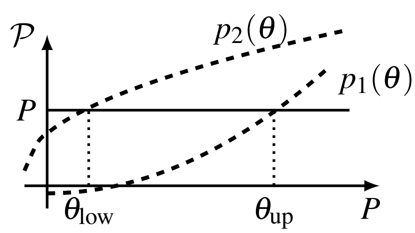

Figure 2.3: Schematic representation of the confidence interval for an estimator function $P(\cdot)$ of the statistical quantity θ built with the lower and upper functions $p_1(\theta)$, $p_2(\theta)$. Adapted from [33].

In the absence of a priori information about the probabilistic distribution of the sample, the Bienaymé-Chebyshev inequality (Equation 2.6) can be used to define confidence intervals. The named equation remains valid whatever the distribution of the parent variable $\mathcal{X}$ characterized by a finite expected value $\mu_{\mathcal{X}}$ and a finite non-zero variance $\sigma^2_{\mathcal{X}}$.

3 Random processes theory

A stochastic process or also called random process is a collection, a family, of random variables indexed by the time. From every experiment is produced a time history, or also called a sample function, of a time-dependent phenomenon [4]. By repeating this physical phenomenon, a collection of all possible sample functions is produced. This collection is called a random process or a stochastic process.

A random variable can either be represented by an infinite sample of realisations or probability density function. The same principle can be applied for a random process which can be totally described by an infinite series of sample or by a group of probabilistic quantities. However, in the real world there is no way to produce an infinity of the same physical phenomenon for an infinite duration. Therefore, when observed over a finite time interval, the sample function is rather called a sample record. By means of statistical procedures, the descriptive properties of the sample can be defined. These properties would be known without incertitudes if the sample time would be infinite. Only estimates, can be obtained with finite amount of observations as introduced in the previous chapter.

3.1 Basic description and properties

To really link both first chapters, the same mathematical framework is kept to introduce the different theoretical concepts. That is, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The collection of random variables indexed by the time set T represent a stochastic process which can be defined as [15] :

$$\{\mathcal{X}(t,\omega) : t \in T \quad \text{and} \quad \omega \in \Omega\} \tag{3.1}$$

A sample function is a single outcome of the stochastic process, it corresponds to the time series formed by taking a possible value of each random variable that constitutes the stochastic process, assumed to be real. For the random process $\{\mathcal{X}(t,\omega) : t \in T \quad \text{and} \quad \omega \in \Omega\}$, then $\forall \omega \in \Omega$ the mapping :

$$\mathcal{X}(\cdot,\omega) : T \to \mathbb{R} \tag{3.2}$$

is called a sample function. If the set T is the real line, the process is a continuous-time process. On the contrary, if T is a set of integers, then the process is a discrete-time process. In this case, the process can be seen as a succession of n realisations $x_1, ..., x_n$ of the random variable $\mathcal{X}(t,\omega) : x_i = \mathcal{X}(t_i,\omega)$.

3.1.1 Statistics of stochastic processes

As stated previously, a random process is a noncountable infinity or a countable number of random variables $\mathcal{X}(t,\omega) : x_i = \mathcal{X}(t_i,\omega)$. Therefore, the results exposed in the Chapter 2 can easily be extended to the specific case of a time-dependent problem. In particular, the probability density function $f_{\mathcal{X}}(x,t)$ becomes function of the time index $t \in T$.

The mean and variance definitions, Equations (2.3) and (2.5), are adapted to introduce the new time-dependent density which yields to time functions $\mu_{\mathcal{X}}(t)$ and $\sigma^2_{\mathcal{X}}(t)$. Regarding the covariance and correlation functions, Equations (2.7) and (2.8), they are now functions of the time t as well as a time-shift $\tau \in \mathbb{R} : C_{\mathcal{X}\mathcal{Y}}(t,t+\tau)$ and $R_{\mathcal{X}\mathcal{Y}}(t,t+\tau)$. Indeed, both functions are built on the joint probability density function of the random processes $\mathcal{X}(t,\omega) : x = \mathcal{X}(t,\omega)$ and $\mathcal{Y}(t,\omega) : y = \mathcal{Y}(t+\tau,\omega)$.

3.1.2 Classification of stochastic processes

A physical phenomenon has been classified as either deterministic or random. A further classification is made for the specific case of stochastic processes. They can be categorized in two main families : stationary or nonstationary (see Fig.3.1). Stationary random processes are themselves divided into ergodic or nonergodic processes.

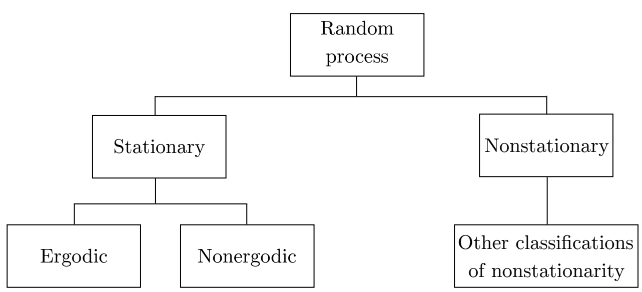

Figure 3.1: Classification of random data. Adapted from [4].

The observations in a experiment can potentially be spaced by an infinitesimal time-shift. The statistical properties of the process could then be estimated at any instant of time t. If the expected values $\mu_{\mathcal{X}}(t)$ and $\mu_{\mathcal{Y}}(t)$ of the random process $\mathcal{X}(t,\omega)$ and $\mathcal{Y}(t,\omega)$ are time-independent together with the covariance function $C_{\mathcal{X}\mathcal{Y}}(t,t+\tau)$ independent of any arbitrary time-shift then the processes are said to be weakly-sense stationary (abbreviated WSS). If further, the probability density functions related to the processes are also time-translation independent, then the processes are said to be strongly-sense stationary (abbreviated SSS). Because of the definitions of the mean and covariance, SSS is a subclass of WSS. In the particular case of a Gaussian process, where every possible probability density functions can be derived with the mean and covariance, both concepts coincide.

The statistical properties can be evaluated by computing ensemble averages at different moments of time. In other words, they can be computed by averaging over all the realisations of the process. However, it is also possible to compute statistical properties by averaging in a different plan, by computing time averages. The difference between ensemble averaging and time averaging is schematically represented in Fig.3.2

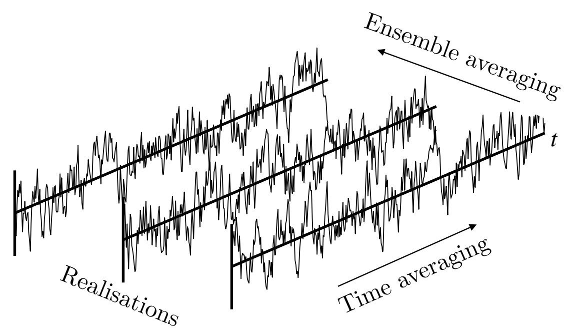

Figure 3.2: Schematic illustration of ergodicity. Adapted from [13].

According to the ergodicity theorem [4], a random process is said to be ergodic if the ensemble averaging provides the same result as the time averaging for the sample time tending to infinity. The main advantage is that all properties of an ergodic random process can be computed by performing time averages for a single sample function *i.e* with only one realisation of the physical phenomenon.

In practice, most of stationary random processes are assumed to be ergodic for their simplicity of description. In the following, every random process will be assumed to be ergodic.

3.2 Ergodic processes

In this section are derived the basic statistical procedures to estimate the descriptive properties of an assumed ergodic record $\mathcal{X}(t,\omega)$ where $t \in T \subset \mathbb{R}^+$ and $\omega \in \Omega$. In particular, a specific attention is given to the autocovariance and autocorrelation functions and the way they can be linked to the frequency content of the signal. Moreover, as mentioned below, the properties of an ergodic process can be assessed by performing time averages for a single experiment. Therefore, for sake of readability, the variable ω related to the sample space is omitted : $\mathcal{X}(t)$.

3.2.1 Mean and standard deviation

Because an ergodic process is a specific case of a stationary process, the mean and variance are constant and :

$$\mu_{\mathcal{X}} = \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathcal{X}(t)dt \quad \text{and} \quad \sigma_{\mathcal{X}}^2 = \lim_{T\to\infty} \frac{1}{T}\int_0^T (\mathcal{X}(t)-\mu_{\mathcal{X}})^2 dt. \tag{3.3}$$

3.2.2 Autocorrelation and autocovariance functions

The autocorrelation function $R_{\mathcal{X}\mathcal{X}}(\tau)$ of the ergodic process $\mathcal{X}$ measures the correlation between data spaced by a fixed time delay τ. Mathematically, the function corresponds to the average of the product of the time record and a delay copy of itself by a time τ. This function can be extended to take into account an arbitrary mean $\mu_{\mathcal{X}}$ so that :

$$R_{\mathcal{X}\mathcal{X}}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathcal{X}(t)\mathcal{X}(t+\tau)dt \quad \text{and} \quad C_{\mathcal{X}\mathcal{X}}(\tau) = R_{\mathcal{X}\mathcal{X}}(\tau) - \mu_{\mathcal{X}}^2. \tag{3.4}$$

Some properties can be observed from equation (3.4). In the specific case of a zero time-shift $\tau = 0$, the autocovariance function reduces to the variance $C_{\mathcal{X}\mathcal{X}}(0) = \sigma_{\mathcal{X}}^2$ while the autocorrelation reduces to the mean square error $R_{\mathcal{X}\mathcal{X}}(0) = \psi_{\mathcal{X}}^2$. Moreover, from the stationary hypothesis, the autocorrelation is a non-negative even function[1] of the time delay τ : $R_{\mathcal{X}\mathcal{X}}(-\tau) = R_{\mathcal{X}\mathcal{X}}(\tau)$.

Theses properties are represented schematically in Figure 3.3 for an arbitrary autocorrelation/autocovariance function.

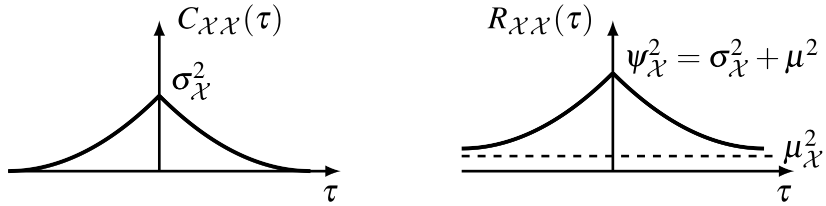

Figure 3.3: Schematic representation of autocovariance $C_{\mathcal{X}\mathcal{X}}(\tau)$ (left) and autocorrelation $R_{\mathcal{X}\mathcal{X}}(\tau)$(right) functions of a stationary random process $\mathcal{X}(t)$ with mean $\mu_{\mathcal{X}}$ and variance $\sigma_{\mathcal{X}}^2$. Adapted from [13].

[1] The propriety can easily be demonstrated by starting from equation (3.4) with $\tau = -\tau$. Since the process is assumed to be stationary, it is independent of time translation and the property is proved with the change of variable $t \to t + \tau$.

3.2.3 Spectral density functions

Under the assumption that the autocorrelation function is Lebesgue-measurable on the real line, namely that the integral of its absolute value is finite :

$$\int_{-\infty}^{\infty} |R_{\mathcal{X}\mathcal{X}}(\tau)|\, d\tau < \infty \tag{3.5}$$

the Fourier transform of $R_{\mathcal{X}\mathcal{X}}(\tau)$ exists and is defined by

$$S_{\mathcal{X}\mathcal{X}}(f) = \int_{\infty}^{\infty} R_{xx}(\tau) e^{-j2\pi f \tau} d\tau \tag{3.6}$$

with $j = \sqrt{-1}$ and f is the frequency in Hertz. In practice, the integrability condition (3.5) is always true for finite record lengths and the Fourier transform $S_{\mathcal{X}\mathcal{X}}$ exists. This function is the two-sided spectral density function, with $f \in]-\infty;+\infty[$. Given that the autocorrelation function is a continuous positive real function, the spectral density function is also positive and real according to Bochner's Theorem [30]. Moreover, by Fourier transform property, $S_{\mathcal{X}\mathcal{X}}(\tau)$ is an even function.

By Fourier inversion theorem :

$$R_{\mathcal{X}\mathcal{X}}(\tau) = \int_{-\infty}^{\infty} S_{\mathcal{X}\mathcal{X}}(f) e^{j2\pi f \tau} d\tau \tag{3.7}$$

so that functions $R_{\mathcal{X}\mathcal{X}}(\tau)$ and $S_{\mathcal{X}\mathcal{X}}(f)$ form a Fourier transform pair, Equations (3.6) and (3.7) are also called the Wiener-Khinchine relations [15]. However, the spectral density function does not require to be integrable. For instance, the Fourier transform of a step function is the *sinc* function. Therefore, the spectral density function of a constant function is bounded and continuous but not integrable in the Lebesgue sense.

In the case of a non-zero mean, namely $\mu_{\mathcal{X}} \neq 0$, the spectral density function is function of this mean value and contains an impulse at $f = 0$. To avoid this, it is convenient to define the power spectral density by means of the autocovariance function. Because of their mutual definition, see Equation (3.4), it follows :

$$S_{\mathcal{X}\mathcal{X}}(f) = S^{c}_{\mathcal{X}\mathcal{X}}(f) + \mu^2_{\mathcal{X}} \delta(f). \tag{3.8}$$

The function $S^{c}_{\mathcal{X}\mathcal{X}}(f)$ is called the autocovariance spectrum of $\mathcal{X}(t)$.

Finally, the one-sided autospectral density function $G_{xx}(f)$ can be defined for $f \in [0;+\infty[$ as

$$G_{xx}(f) = 2S_{xx}(f) \quad ,0 < f < +\infty \quad ,\text{otherwise } 0. \tag{3.9}$$

3.3 Important types of random process

In this section are enumerated several analytic random processes that can be useful for theoretical studies. A time, as well as frequency, domain analysis is performed. Two main processes are presented, namely : the white noise and the sine wave as they represent both ends of the frequency spectrum.

3.3.1 White noise

The white noise $\mathcal{W}(t)$ is the simplest stationary random process analytically. It is characterized by a constant autospectral density function $G_{\mathcal{W}\mathcal{W}}(f) = G_0$ for $f > 0$ only or $S_{\mathcal{W}\mathcal{W}}(f) = G_0/2$ over all frequencies. By Fourier transform, the associated autocovariance function is a centred Dirac impulsion $R_{\mathcal{W}\mathcal{W}}(\tau) = (G_0/2)\delta(\tau)$ (see Figure. 3.4). The Dirac impulsion indicates that two observations of the random process are uncorrelated, regardless the duration between both.

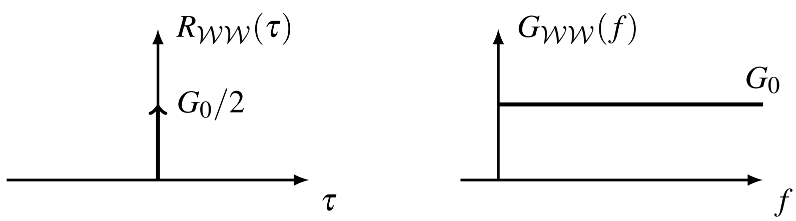

Figure 3.4: Autocorrelation function (left) and one-sided autospectral density function (right) of a white noise with constant $G_{\mathcal{W}\mathcal{W}}(f) = G_0$.

This kind of process is non-physical since it has an infinite mean square value (see Equations (3.4) and (3.7)). Therefore, such an analytic random process cannot be Gaussian because a Gaussian process is well defined for finite variance value. However, under certain conditions, real data can be approached by a white noise. This approximation can only be local i.e over a limited bandwidth, denoted B.

If f_0 is the center frequency of the filter of bandwidth B, the limited bandwidth or also called bandpass white noise is defined by the autospectral density function

$$G_{\mathcal{W}\mathcal{W}}(f) = \begin{cases} G_0 & 0 \leq f_0 - (B/2) \leq f \leq f_0 + (B/2) \\ 0 & \text{otherwise} \end{cases} . \tag{3.10}$$

The associated autocorrelation function is [24]

$$R_{\mathcal{W}\mathcal{W}}(\tau) = \int_0^{\infty} G_{\mathcal{W}\mathcal{W}}(f) e^{j2\pi f t} df = G_0 B \left(\frac{\sin(\pi B \tau)}{\pi B \tau} \right) \cos(2\pi f_0 \tau). \tag{3.11}$$

For the specific case of a low-pass white noise, where $f_0 = B/2$, the autospectral and autocorrelation functions become

$$G_{\mathcal{WW}}(f) = \begin{cases} G_0 & 0 \leq f \leq B \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \quad R_{\mathcal{WW}}(\tau) = G_0 B \left(\frac{\sin(\pi B \tau)}{\pi B \tau} \right) \tag{3.12}$$

and are represented in Figure 3.5. Finally, given that $R_{\mathcal{WW}}(0) = BG_0 = \psi_x^2$, the bandwidth-limited white noise signals have a finite mean square value.

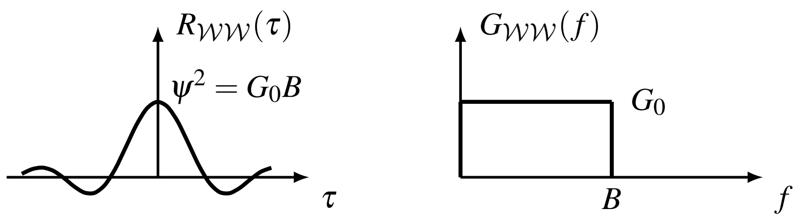

Figure 3.5: Autocorrelation function (left) and one-sided autospectral density function (right) of a low-pass white noise with constant $G_{\mathcal{WW}}(f) = G_0$ and bandwidth B.

3.3.2 Sine wave

At the other end of the frequency spectrum, the sine wave $\mathcal{Y}(t)$ of amplitude A is characterized by an autospectral density function modelled by a Dirac impulsion at the wave frequency f_0 : $G_{\mathcal{YY}}(f) = \frac{A^2}{2}\delta(f - f_0)$. By Fourier transform, the associated autocovariance function is a cosine $R_{\mathcal{YY}}(\tau) = \frac{A^2}{2}\cos(2\pi f_0 \tau)$ (see Figure. 3.6).

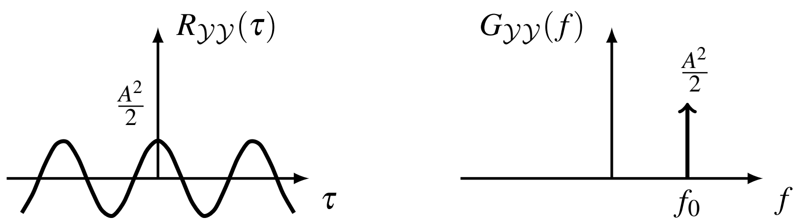

Figure 3.6: Autocorrelation function (left) and one-sided autospectral density function (right) of a sine wave with of frequency f_0 and amplitude A.

It is worth to note that :

$$R_{xx}(0) = \psi^2 = \frac{A^2}{2}. \tag{3.13}$$

3.4 Estimation theory for stochastic processes

In this section is introduced analytic developments to quantify the coefficient of variation of an unbiased estimate as a function of the sample length. Regarding Equation (2.18), since the bias is null, the coefficient corresponds to the statistical error i.e the variance or random error. The two estimators considered are the mean sample and the standard deviation of the sample. These analytic derivations are based on the autocovariance function of the process $\mathcal{X}$: $C_{\mathcal{X}\mathcal{X}}$. Those developments are directly inspired by [4].

3.4.1 Mean value estimate

Under the assumption of integrability of the autocovariance function, the variance of the mean estimator can be obtained by

$$V\{\hat{\mu}_{\mathcal{X}}\} = \frac{1}{T}\int_{-T}^{T}\left(1-\frac{|\tau|}{T}\right)C_{\mathcal{X}\mathcal{X}}(\tau)d\tau \tag{3.14}$$

where $\mathcal{X}(t)$ is an ergodic stochastic process with $t \in T$.
For large T, and thus $|\tau| << T$, Equation (3.14) becomes

$$V\{\hat{\mu}_{\mathcal{X}}\} \approx \frac{1}{T}\int_{-\infty}^{\infty}C_{\mathcal{X}\mathcal{X}}(\tau)d\tau \tag{3.15}$$

3.4.2 Standard deviation estimate

Following the same steps, the variance of the mean square error estimate can be obtained by :

$$\begin{aligned} V\{\hat{\psi}^2_{\mathcal{X}}\} &= \frac{2}{T}\int_{-T}^{T}\left(1-\frac{|\tau|}{T}\right)\left[C^2_{\mathcal{X}\mathcal{X}}(\tau)+2\mu^2_{\mathcal{X}}C_{\mathcal{X}\mathcal{X}}(\tau)\right]d\tau \\ &\approx \frac{2}{T}\int_{-\infty}^{\infty}\left[C^2_{\mathcal{X}\mathcal{X}}(\tau)+2\mu^2_{\mathcal{X}}C_{\mathcal{X}\mathcal{X}}(\tau)\right]d\tau \end{aligned} \tag{3.16}$$

And the variance of the variance estimate is thus given by :

$$V\{\hat{\sigma}^2_{\mathcal{X}}\} \approx \frac{2}{T}\int_{-\infty}^{\infty}C^2_{\mathcal{X}\mathcal{X}}(\tau)d\tau \tag{3.17}$$

3.4.3 Bandwidth-limited white noise

Considering the bandwidth-limited white noise introduced in subsection 3.3.1, it can be rewritten as a function of the descriptive statistics $\mu_{\mathcal{W}}$ and $\sigma^2_{\mathcal{W}}$:

$$G_{\mathcal{W}\mathcal{W}}(f) = \begin{cases} \frac{\sigma^2_{\mathcal{W}}}{B}+\mu^2_{\mathcal{W}}2\pi\delta(f) & 0 \le f \le B \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad R_{\mathcal{W}\mathcal{W}}(\tau) = \sigma^2_{\mathcal{W}}\left(\frac{\sin(2\pi B\tau)}{2\pi B\tau}\right)+\mu^2_{\mathcal{W}} \tag{3.18}$$

Often, it is more convenient to work with the autocovariance function $C_{\mathcal{W}\mathcal{W}}(t) = R_{\mathcal{W}\mathcal{W}}(t) - \mu^2_{\mathcal{W}}$. The latter cancel for $\tau = (k/2B)$, where k is an integer. It means that points spaced by

$(1/2B)$ are uncorrelated. If the random process $\mathcal{W}(t)$ is Gaussian, they will be also statistically independent. Under this assumption, by use of Equations (3.15) and (3.17) it yields :

$$\varepsilon_{\text{ana}}\{\hat{\mu}_{\mathcal{W}}\} \approx \frac{1}{\sqrt{2BT}}\left(\frac{\sigma_{\mathcal{W}}}{\mu_{\mathcal{W}}}\right) \quad \text{and} \quad \varepsilon_{\text{ana}}\{\hat{\sigma}_{\mathcal{W}}\} \approx \frac{1}{\sqrt{4BT}}. \tag{3.19}$$

Equations (3.19) allow to assess the value of the statistical errors on the mean and the standard deviation estimates as a function of the record length T for the specific case of a low-pass white noise characterized by a bandwidth B, a mean $\mu_{\mathcal{W}}$ and a standard deviation $\sigma_{\mathcal{W}}$. The errors both tend to zero as T approaches infinity, therefore $\hat{\mu}_{\mathcal{W}}$ and $\hat{\sigma}_{\mathcal{W}}$ are consistent estimates of the mean and the standard deviation, respectively.

Part II

Methodology : previous work and contributions

Outline of the second part

The pictures and vocabulary are the key to physical intuition, and the intuition makes the sudden leaps of insight... mathematics, of course is the ultimate arbiter; only it can say whether the leaps of insight are correct.

Richard Price and Kip Thorn

The second part of the thesis is dedicated to the description of the methodology employed to estimate the initial transient duration as well as the required simulation time. This part is wanted to be written in a chronology order to understand the path followed during the work. For that purpose, the part is divided into three chapters : previous work and reproduction, a theoretic analysis to expand the previous work methodology and lastly the final version of the methodology and its numerical implementation in `C` programming language and as a Fluent User-Defined Function. Finally, several post-processing operations have been performed with functions implemented in `Matlab`.

The chapter 4 focuses on the previous work on the transient duration estimation. Particularly, the methodology introduced by Mocket et. al in 2009 [27] is explained. The latter is reproduced for several test cases. From it are highlighted some drawbacks which open tracks of further ameliorations.

In Chapter 5 is introduced the first part of the contributions to the existing methodology. In his paper, Mockett estimates the transient duration by means of a white-noise scaling of the lift coefficient (this notion will takes more sense in Chapter 4). The Chapter 5 is therefore dedicated to the introduction of a more general model based on physical inspections and physical laws.

In Chapter 6 are presented all the contributions to the methodology. A complete summary of the algorithm is provided and the numerical implementation is briefly described. Finally, the translation of the methodology to a Fluent UDF is explained in Chapter 7.

4 Transient duration estimation - Previous work

4.1 Introduction

In unsteady simulations, it is often necessary to detect when the solution has reached a statistically stationary state. Indeed, the flow exhibits an initial transient due to arbitrary initial conditions. To obtain valid sampling, statistics have to be performed on a sampling window (see Figure 4.1) free of transient fluctuations.

Usually, the approaches to highlight the latter are based on visual inspections of relevant variables time records. Such methods lead to a lack of robustness and accuracy. On the other hand, in the specific case of an LES simulation, looking at previous experiences can be useful and particularly the number of large eddies versus time. Indeed, a certain number of periods is needed to progress the flow and then perform averaging. For instance, for jets and swirling jets, Shur, Spalart and al. [36] and Tucker [40] required time periods of around $1000L/U$ and a total simulation time needed of $2000L/U$, respectively, where L is the jet diameter and U the jet velocity. Therefore, duration considerations are strongly case specific and may become of poor relevance in an other kind of flow studied.

When the simulations are periodic, i.e with no stochastic behaviour, the periodicity of the flow can be assessed by fast Fourier transformation (FFT) according to Ahmed & Barber [2]. The methodology is based on the amplitudes of the dominant frequencies which are used to evaluate the convergence to a periodic state. However, this FFT based methodology can require a huge amount of data and eventually case specific. Furthermore, LES method do not allow a clear periodic solution which makes this procedure impractical.

Finally, a more reliable and robust technique based on statistical considerations has been proposed by Mockett and al. [27] to estimate, *a posteriori*, the initial transient duration. It is assumed that the transient causes a distortion in combination or in isolation of the statistical errors for the mean and the standard deviation estimates, $\varepsilon\{\hat{\mu}_{C_L}\}$ and $\varepsilon\{\hat{\sigma}_{C_L}\}$ respectively (see Equation (2.18)) of the lift coefficient.

The methodology that has been developed (presented in the next chapters) is mainly based on Mockett's paper. The latter has been adapted to be used on the fly in the aim of monitoring an unsteady simulation. Therefore, this chapter aims to introduce the named methodology. Especially, the drawbacks of the latter are highlighted to be used as a foundation to the contributions added to the methodology.

4.2 Formulation of the method

Let consider[1] a generic time record with $t \in T \subset \mathbb{R}^+$ such that $\mathcal{X}(t) : T \to \mathbb{R}$. It does not represent any physical phenomenon (see Figure 4.1) and is only used for sake of visibility. The signal is assumed to be ergodic which allows the analysis of a single sample function and to omit the sample space index. Finally, $\mathbf{D}_T$ denotes the sample i.e the succession of the realisations $x_t = \mathcal{X}(t)$.

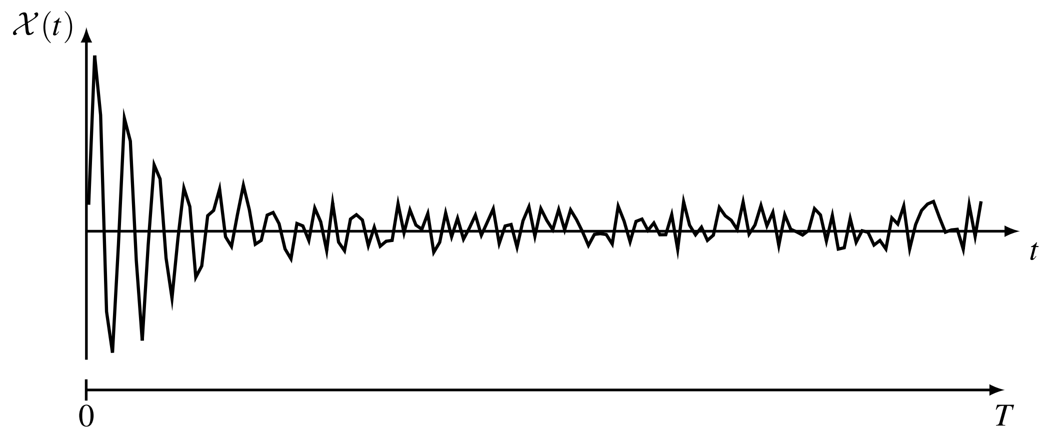

Figure 4.1: Generic time record $\mathcal{X}(t)$, of total length T, which does not represent any physical phenomenon.

Two different point estimators are used to sum-up the information contained in the signal : the sample mean $\hat{\mu}_{\mathcal{X}}(\mathbf{D}_T)$ and the sample standard deviation $\hat{\sigma}_{\mathcal{X}}(\mathbf{D}_T)$ as defined in Equations (2.10) and (2.12). The method is based on the normalised estimation error of both estimators $\varepsilon\{\hat{\mu}_{\mathcal{X}}\}$ and $\varepsilon\{\hat{\sigma}_{\mathcal{X}}\}$ defined in Equation (2.18). Because both point estimators are unbiased, the estimation error is equivalent to the statistical error and :

$$\varepsilon\{\hat{\mu}_{\mathcal{X}}\} = \frac{\sqrt{V\{\hat{\mu}_{\mathcal{X}}\}}}{\mu_{\mathcal{X}}} \quad \text{and} \quad \varepsilon\{\hat{\sigma}_{\mathcal{X}}\} = \frac{\sqrt{V\{\hat{\sigma}_{\mathcal{X}}\}}}{\sigma_{\mathcal{X}}} \tag{4.1}$$

where $\mu_{\mathcal{X}}$ and $\sigma_{\mathcal{X}}$ are respectively the true mean and the true standard deviation.

The methodology can be divided into two main parts : the statistical error estimation and the initial transient detection. The estimation is performed numerically by means of the available

[1]For homogeneity, the notations adopted in the Mockett's paper have been adapted to respect the ones introduced in the state of the art chapter.

data and a prognosis for longer sample lengths is achieved through the use of analytically-derived relationships for theoretical low-pass white-noise. This part of the methodology is explained in detail in subsection 4.2.1. Further, the transient stage is assumed to cause a distortion in combination or isolation of the random errors. A criterion is therefore used to highlight this duration.This part of the methodology is explained in detail in subsection 4.2.2.

4.2.1 Quantification of random error

Let $P(\cdot)$ be a point estimator of a numerical parameter; in this case the mean or the standard deviation, and $\mathcal{P}_t$ the associated random variable such that $\mathcal{P}_T = P(\mathbf{D}_T)$. The statistical error time evolution can be assessed by dividing the sample into shorter windows of length $T_w \in T$. For $T_w << T$, the coefficient of variation can be estimated by

$$\varepsilon_{T_w}\{\mathcal{P}_{T_w}\}(T_w, T) = \frac{\sqrt{\left\langle (\mathcal{P}_{T_w} - \mathcal{P}_T)^2 \right\rangle}}{\mathcal{P}_T}, \tag{4.2}$$

where $< >$ represents averaging over the available windows. Equation (4.2) therefore provides an error which varies with T and T_w. The error estimation becomes less accurate as the window size tends to the sample length T since there are only few windows available to average.

To further clarify, T_w varies to obtain a measure of the error for different T_w. For each fixed window size, the descriptive statistics are estimated inside each windows $(\mathcal{P}_{T_w})$. Then, the sample statistics of the whole available data $(\mathcal{P}_T)$ is subtracted. Finally, each difference is squared and an average is performed. A schematic example is presented in Figure 4.2 for the generic signal.

In addition, analytical expressions derived in the first part of this master thesis allow to prognosticate the error trend for longer sample time. Indeed, Equations (3.19) represent the coefficient of variation as a function of sample length for the specific case of a low-pass Gaussian white noise :

$$\varepsilon_{\text{ana}}\{\hat{\mu}_{\mathcal{X}}(\mathbf{D}_T)\}(T) = \frac{1}{\sqrt{2BT}}\left(\frac{\sigma_{\mathcal{X}}}{\mu_{\mathcal{X}}}\right) \quad \text{and} \quad \varepsilon_{\text{ana}}\{\hat{\sigma}_{\mathcal{X}}(\mathbf{D}_T)\}(T) = \frac{1}{\sqrt{4BT}} \tag{4.3}$$

Since the true descriptive statistics μ_x and σ_x are unobservable, the coefficient of variation is also unobservable. However, an estimation of the latter can be made by replacing these by the best available estimate corresponding to the descriptive statistics using the entire sample : $\hat{\mu}_{\mathcal{X}}(\mathbf{D}_T)$ and $\hat{\sigma}_{\mathcal{X}}(\mathbf{D}_T)$. Therefore, since the statistical error is estimated by windows averaging, the only remaining unknown is the frequency spectrum parameter B.

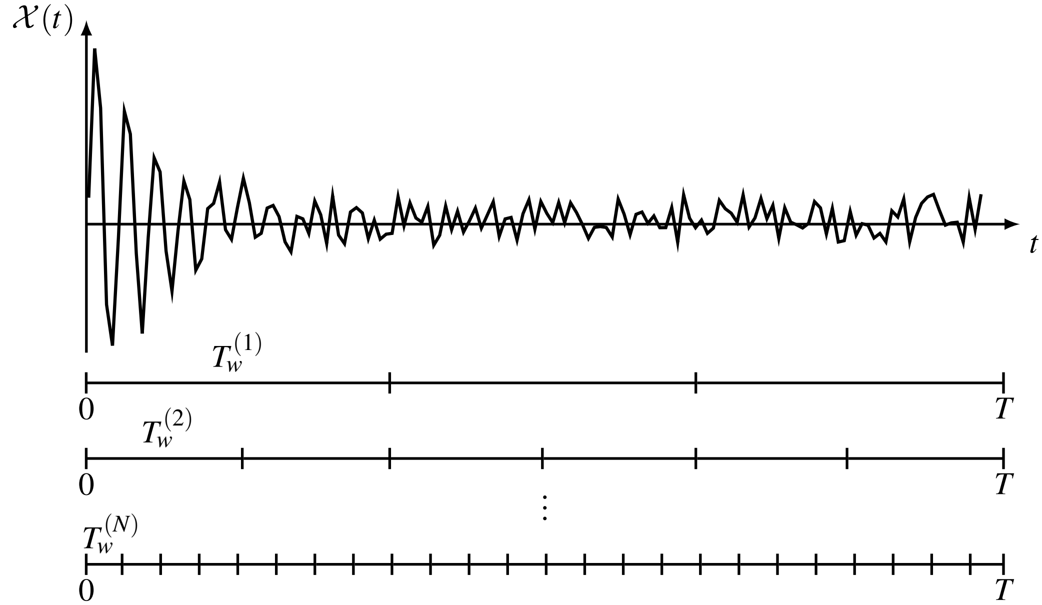

Figure 4.2: Schematic representation of the division process into series of windows to estimate the statistical error.

To obtain the value for this parameter, both error expressions defined by equations (4.2) and (4.3), are equalised for each $T_w = T$. It results a value of B for each equality. For the method to be conservative, the lowest value is chosen. For instance, considering the statistical error on the mean estimate for a Gaussian white noise :

$$\varepsilon_{T_w}\{\hat{\mu}_{\mathcal{X}}\}(T_w, T_w) = \varepsilon_{\text{ana}}\{\hat{\mu}_{\mathcal{X}}\}(T_w) \Leftrightarrow B(T_w) = \frac{1}{2T_w[\varepsilon\{\hat{\mu}_{\mathcal{X}}(\mathbf{D}_{T_w})\}(T_w, T_w)]^2}\left(\frac{\hat{\sigma}_{\mathcal{X}}}{\hat{\mu}_{\mathcal{X}}}\right)^2 \quad (4.4)$$

and the final value of B is chosen as $\min[B(T_w)]$. At the end, an estimation of the statistical error is obtained for longer sample length that the initial one i.e larger than T.

4.2.2 Estimation of initial transient

The first part of the methodology presented a technique to estimate the statistical error of a time record as a function of the sample length T and prognosis the trend of the error by a white noise scaling. In this subsection, a technique to estimate the transient duration is proposed.

Let denote the transient duration t_t such that all the transient content lies in the interval $0 < t < t_t$ with $t \in [0, T]$. The transient content is assumed to cause a distortion in the statistical error of the descriptive statistics estimates. To highlight this duration, the start sample is defined as t_0. Moving this initial time forward through the time record is equivalent to shorten the sample by t_0 as represented in Figure 4.3. For each new sample of length $(T - t_0)$, generated from the initial one, an estimation of the statistical error is made following the methodology presented in the previous subsection as t_0 is varied.

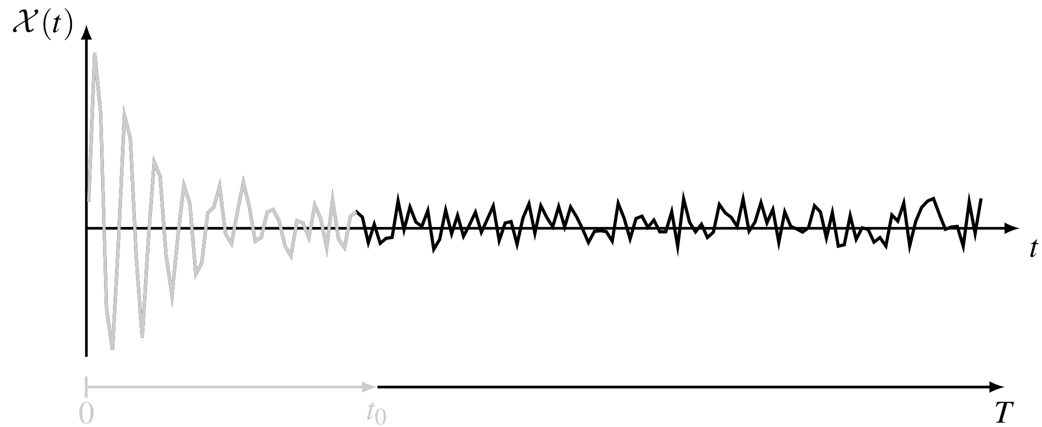

Figure 4.3: Generic time record $\mathcal{X}(t)$, of total length T shortened with a shifting start time t_0. The suppressed part of the sample is plotted in light gray while the conserved part is plotted in black.

The initial content in the sample present for $t_0 < t_t$ is assumed to lead to an increasing in $\varepsilon\{\mathcal{P}_T\}$ compared to the same sample free of any transient i.e for $t_0 \geq t_t$. Once the initial transient has been suppressed, the dispersion and thus the statistical error is assumed to rise given the reduction in the sample length $(t - t_0)$. Therefore, the transient duration should correspond to a minimum in $\varepsilon\{\mathcal{P}_T\}(t_0)$. To take into account both variations in mean and in standard deviation, the transient duration corresponds to the minimum of the product $\Pi(t_0)$ of $\varepsilon\{\hat{\mu}_{\mathcal{X}}(\mathbf{D}_{T-t_0})\}(t_0)$ and $\varepsilon\{\hat{\sigma}_{\mathcal{X}}(\mathbf{D}_{T-t_0})\}(t_0)$ as represented in Figure 4.4.

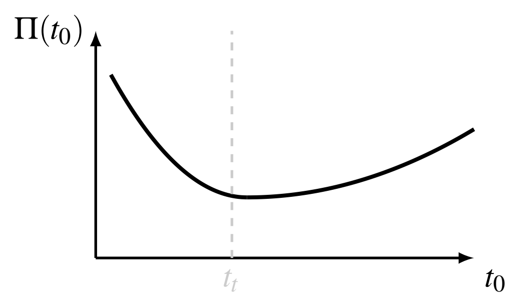

Figure 4.4: Schematic representation of the product of statistical errors on the mean and variance estimates to highlight the transient duration.

As mentioned in the previous section, the error estimation becomes decreasingly accurate when the number of windows needed for the average is low. Therefore, when the sample is strongly shortened, i.e when t_0 tends to T, the method is limited. Hence, t_0 is limited such that $0 < t_0 < T/2$.

4.2.3 Summary

In the end, Mockett introduced a methodology to estimate the transient duration by means of statistical error considerations, with the time series as the sole input. A flowchart summarizing the methodology is provided in Figure 4.5.

4.3 Method implementation : Reproduction, validation and drawbacks

This section is dedicated to the numerical results obtained with an in-house `C` implementation of the previously introduced methodology. From these results are highlighted drawbacks of the method and are introduced different tracks of resolution for an adapted methodology introduced in Chapter 6. The analysis is not wanted to go in deep details but only to highlight thread followed by the thesis in the further chapters.

The methodology relies on the Gaussian white-noise scaling assumption, hence several synthetic relevant cases that are and are not in agreement with the assumptions are investigated. Moreover, the algorithm is also applied to results obtained by experiment.

4.3.1 White noise

Let consider a low-pass Gaussian white noise characterised by $\mu_{\mathcal{W}} = 1$ and $\sigma_{\mathcal{W}} = 0.1$ generated during 1 second with a time-step $\Delta t : 10^{-5}$ [s]. Therefore, the data set is composed of $n = 10^5 + 1$ samples.

In theory, two distinct realisations of a white noise are uncorrelated. However, in practice a numerical discretization is used, corresponding to the non-zero time step Δt. Thus, there is no information on what happens between two successive data. According to the Nyquist-Shannon theorem, for a sampling frequency $f_s = 1/\Delta t$, the frequency content outside of the interval $[-f_s/2\,; f_s/2]$ cannot be represented [28]. Hence, the analytic bandwidth, called the Nyquist Bandwidth, of a white noise corresponds to $B_{\text{nyq}} = 1/2\Delta t = 5 \times 10^4$ [Hz] in the present case.
Therefore, the analytic statistical errors ε_{ana} are:

$$\begin{cases} \varepsilon_{\text{ana}}\{\hat{\mu}_{\mathcal{W}}\}(t) = \left(\frac{\sigma_{\mathcal{W}}}{\mu_{\mathcal{W}}}\right) \frac{1}{\sqrt{2B_{\text{nyq}}t}} \approx \frac{3.2\times10^{-4}}{\sqrt{t}} \\ \\ \varepsilon_{\text{ana}}\{\hat{\sigma}_{\mathcal{W}}\}(t) = \frac{1}{\sqrt{4B_{\text{nyq}}t}} \approx \frac{2.24\times10^{-3}}{\sqrt{t}} \end{cases} \tag{4.5}$$

Figure 4.6 shows a really good agreement between analytic derivation, computation of the error by windows division ε_{T_w} and prediction with a curve-fitting parameter ε_{fit}. However,

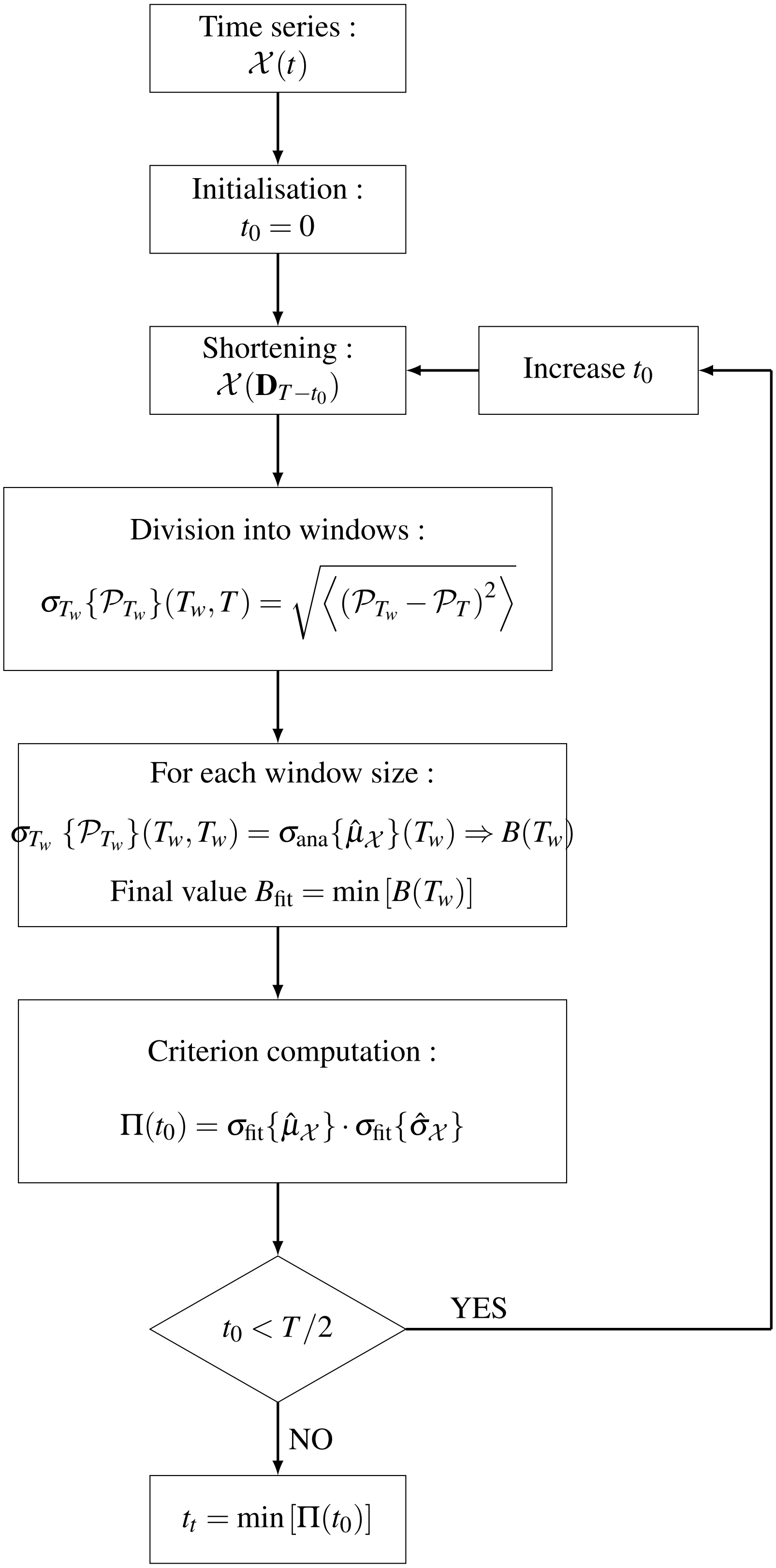

Figure 4.5: Flowchart of the methodology introduced by Mockett.

small divergences can be seen for large T_w, T. Indeed, as the window size tends to the sample time, the number of windows to average and compute the error becomes small and the method is less accurate. Despite these spurious differences, the methodology remains efficient for the theoretical signal of a Gaussian white noise.

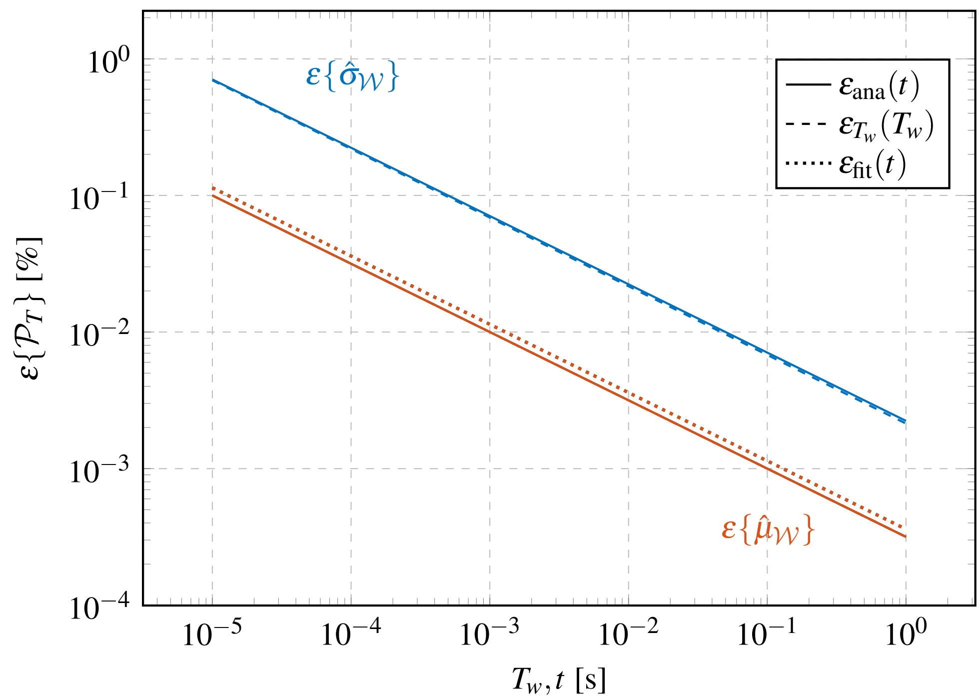

Figure 4.6: Comparison of analytic, estimated and fitted statistical errors in the special case of a Gaussian white noise with $\sigma_\mathcal{W} = 0.1$, $\mu_\mathcal{W} = 1$ and $\Delta t = 10^{-5}$. The analytic error is in dotted line, the estimated in dashed line and the fitted in continuous line. The errors on the mean estimate are plotted in orange while the standard deviations ones are plotted in blue.

It is important to highlight the influence of the time-discretization on the error curves. Indeed, the bandwidth is directly function of the time-step through the Nyquist-Shannon theorem. In addition, the first value (i.e the first point the x-axis) computed corresponds to the value for the smallest window which is directly limited by the time-discretization.

Finally, the methodology has also been tested for a white noise whose pseudo-random numbers are extracted from a uniform density of probability. It provides a similar level of satisfaction.

4.3.2 Sine wave

Let consider a sine wave $\mathcal{Y}(t)$ characterised by $\mu_\mathcal{Y} = 1$, $\sigma_\mathcal{Y} = 0.1$ and $f_0 = 10$ Hz generated during 1 second with a time-step $\Delta t : 10^{-5}$. Therefore, the data set is composed of $n = 10^5 + 1$ samples.

The periodic oscillations of the statistical error ε_{T_w} (Figure 4.7) are expected due to the periodic nature of the signal. Indeed, the average of a periodic signal over an entire period is zero which explains the sharp decrease of the mean error every period $1/f_0$, while the standard deviation is zero each half period $2/f_0$. To highlight this behaviour, the statistical error is plotted with respect to a normalized time T_w^* or t^* obtained by multiplying the time with the wave frequency f_0. It can thus easily be seen in Figure 4.7 that the sharp drops occur when the normalized time is an integer for the mean and half of this integer for the standard deviation.

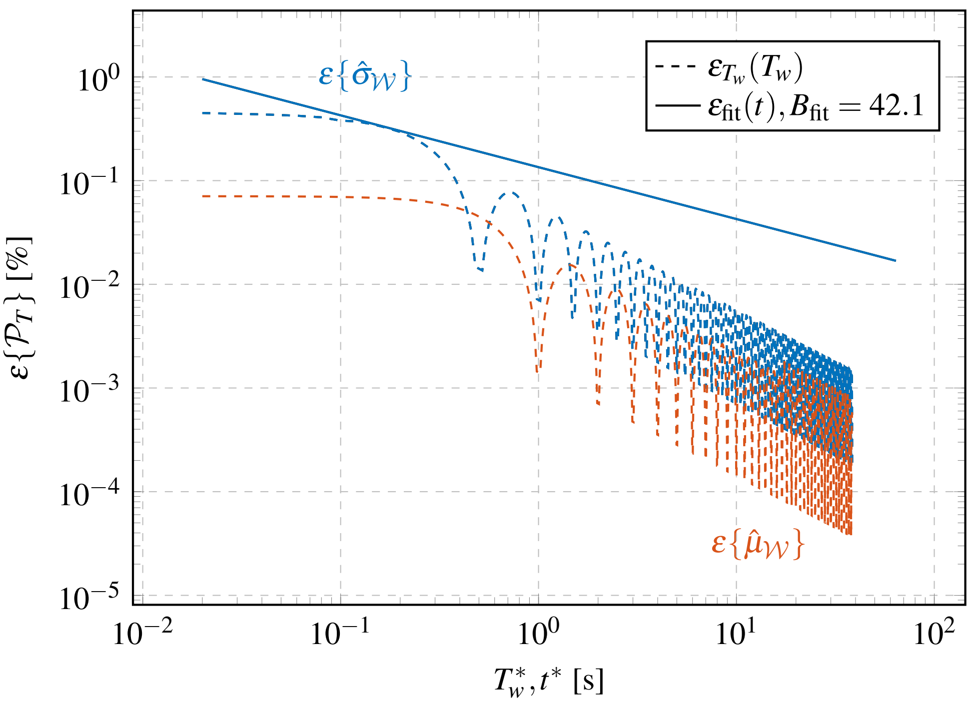

Figure 4.7: Comparison of estimated and fitted statistical errors in the special case of a sine wave with $\sigma_{\mathcal{Y}} = 0.1$, $\mu_{\mathcal{Y}} = 1$, $f_0 = 10$ [Hz] and $\Delta t = 2 \times 10^{-2}$. The estimated error is in dashed line and the fitted in continuous line. The errors on the mean estimate are plotted in orange while the standard deviations ones are plotted in blue. The time (x-axis) is multiplied by the wave frequency : $T_w^* = T_w f_0$ and $t^* = t f_0$.

Once again, the time-discretization has a real influence on the error curves. It can be seen a first horizontal plateau (in a log-log plot, which is actually an illusion due to logarithm scale) whose first value (i.e the first point the x-axis) computed corresponds to the value for the smallest window, which is directly limited by the time-discretization. Therefore, the smaller the time step, the longer the plate is (for a same frequency f_0). Moreover, two sine waves which have multiple frequencies can produce the same sample in function of the sampling frequency, which is called the aliasing. The discretization should thus be chosen to be able to catch the actual frequency of the sine wave. This is also a direct application of the Nyquist-Shannon theorem. In a more general form, the latter states that if the signal contains no frequencies higher than B_{Nyq}, it is completely covered by use of $\Delta t = 1/(2B_{\text{Nyq}})$. In the specific case of a deterministic signal with frequency f_0, the bandlimit should therefore be $B_{\text{Nyq}} = f_0$.

It can clearly be noticed that the latter cannot be used to estimate the curve fitting parameter ($B_{\mathrm{Nyq}}/B_{\mathrm{fit}} \approx 0.25$). Finally, different curve fitted parameters have been computed for different time step and it appears that the parameter B_{fit} is independent of the time discretization.

A further and more important observation is that the white noise scaling introduced by Mockett tends to strongly overestimate the error for a periodic signal which presents a scaling as $\varepsilon \sim t^{-1}$ instead of $\varepsilon \sim t^{-1/2}$ for the white noise (for the sake of visibility, only the fitted error for the standard deviation is represented). This particular drawback should be strongly kept in mind since it has driven the major part of the next chapter of the thesis.

4.3.3 Experimental data

After having investigated synthetic signals, the methodology is applied for a real turbulent signal. As emphasised in subsection 4.3.2, the white-noise assumption is not verified for periodic signal, a relevant investigation is therefore bluff-body flow, with a strong quasi-periodic behaviour from the large scale vortex shedding added to a broadband turbulent fluctuation. The measurements[2] of the unsteady aerodynamic coefficients for a NACA0021 airfoil in deep stall at angle-of-attack $\alpha = 60^\circ$ and Reynolds number 2.7×10^5 provided by Swalwell et al. appears to be an ideal test. Time traces of the aerodynamic coefficients are computed from the pressured measured by a Pitot upstream of and above the model[3].

Time histories of the sectional coefficients correspond to 32.000 points over the time interval $T^* \approx 9000(c/|U_\infty|)$ with c the chord and $|U_\infty|$ the free stream velocity. This is equivalent to around 1760 vortex shedding periods. The sectional lift coefficient $C_l(t^*)$ is represented in Figure 4.8 with respect to the non-dimensionalized time. For clarity, the signal is truncated such that $0 \leq t^* \leq 500$.

[2] The experiment data are available in open source access at http://qnet-ercoftac.cfms.org.uk/w/index.php/UFR_2-11_Test_Case, Last view May 22th 2020.

[3] A complete description of the test facility and measurement techniques used in the experiments are available in Annex.

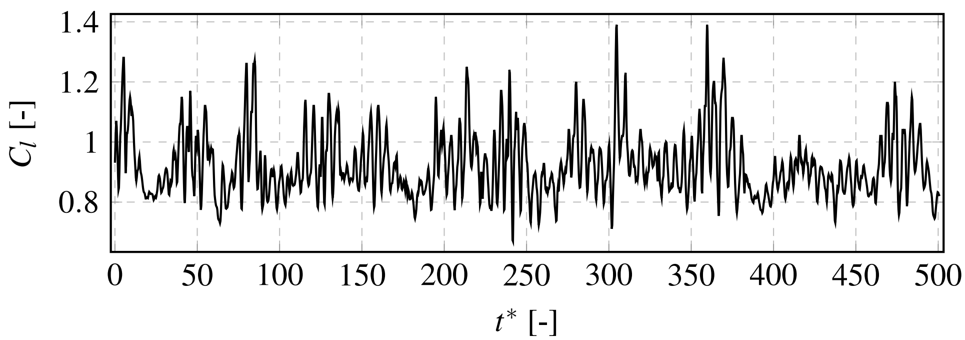

Figure 4.8: Time trace of the measured sectional lift coefficient (C_l) from [39]. The time axis is non-dimensionalised as $t^* = t(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity. The data is truncated as $0 \leq t^* \leq 500$.

In Figure 4.9 are represented the estimated errors ε_{T_w} and the corresponding fitted curves ε_{fit} for the mean and standard deviation. The same curve-fitting parameter is considered for both estimators (the one for the mean). The agreement with the white noise assumption is satisfactory for large t^*; typically for $t^* > 18$ which barely corresponds to 4 periods of vortex shedding. The theoretical limitations of the analytic formula for ε_{ana} introduced in section 3.4 states that the formula is valid for $BT \geq 5$. This lower limit, represented in dotted green line in Figure 4.9, allows therefore to quantify the deviation from the white noise scaling $\varepsilon^{-1/2}$.

Intuitively, those divergences from the white-noise scaling could be explained by the strong correlation of the data at the begin of the sampling. The analytic error expressions being derived by means of the autocovarriance function[4] $R_{\mathcal{X}\mathcal{X}}(\tau^*)$ of a white noise, it is expected that the estimated error ε_{T_w} divergences from the fitted ε_{fit} for time scales where the correlation is the more important i.e at usually at the beginning of the sampling. A usual measure of correlation corresponds to the autocorrelation coefficient $\rho_{\mathcal{X}\mathcal{X}}(\tau^*)$ which is defined as a the ratio of the $R_{\mathcal{X}\mathcal{X}}(\tau^*)$ and its value at the origin $R_{\mathcal{X}\mathcal{X}}(0) = \sigma^2_{\mathcal{X}}$.

In practice, two data shifted by τ^* are said to be weak-correlated if $\rho_{\mathcal{X}\mathcal{X}}(\tau^*) \in [-0.2; 0.2]$ (see [4] for more details). A representation of the autocorrelation coefficient[5] is represented in Figure 4.10 as well as the sigma-interval. It is interesting to highlight that the time needed for the data to be said weakly-correlated corresponds to $\tau^* \approx 20$ which is the value from which the white-noise scaling becomes accurate. This observation thus reinforce the fact that the methodology produces accurate results for uncorrelated signals (which is the definition of a white noise) or at least for signals with a highly decreasing autocorrelation function.

[4] The notation τ^* is employed in this context to make reference to the non-dimensionalised time t^* such that τ^* corresponds to the time shift between two different non units time .

[5] The autocorrelation coefficient has been computed using the `Matlab` function `autocorr` with a pre-defined number of lags $T-1$.

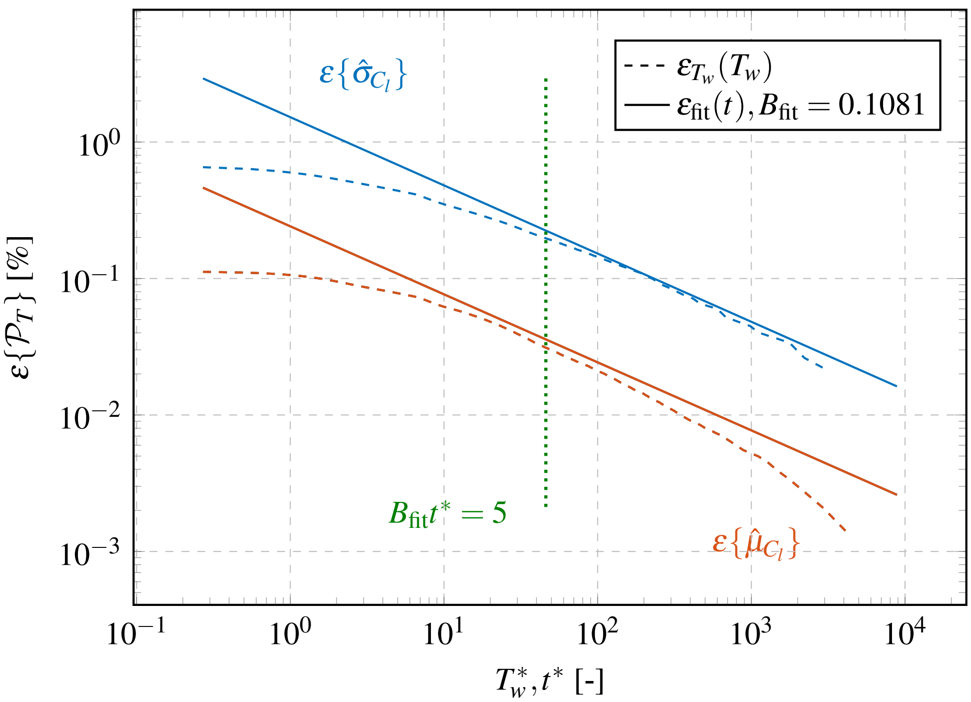

Figure 4.9: Comparison of estimated and fitted statistical errors for the measured sectional lift coefficient (C_l) from [39]. The estimated error is in dashed line and the fitted in continuous line. The errors on the mean estimate are plotted in orange while the standard deviations ones are plotted in blue. The time axis is non-dimensionalised as $t^* = t(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity.

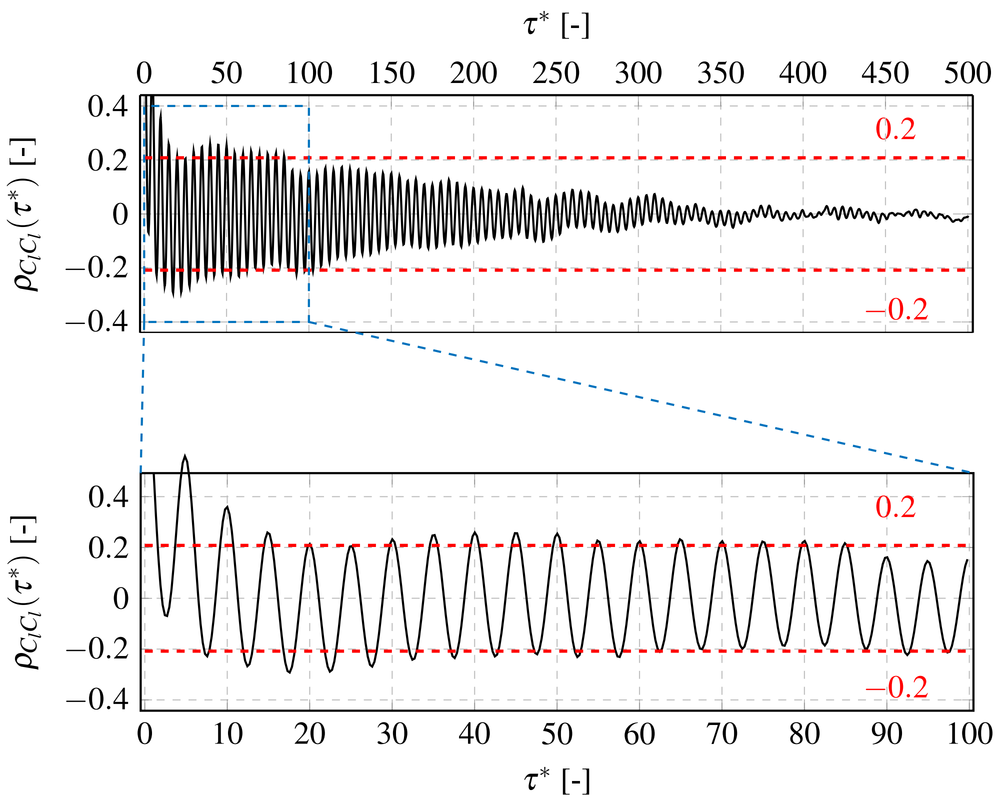

Figure 4.10: Autocorrelation coefficient as a function of the time shift τ^* for the sectional lift coefficient C_l measured by Swalwell et al. for a NACA0021 at angle-of-attack $\alpha = 60°$ [39]. The data is truncated as $0 \leq \tau^* \leq 500$ (up) and a zoom is performed between 0 and 100 non-dimensional units (low).

Regarding probabilistic considerations, the probability density function of the sectional lift coefficient is clearly non-Gaussian (Figure 4.11). This is mainly due to the turbulent nature of the flow. However, as expected according to the central limit theorem, the distribution of the window averages tends to fit a normal distribution. If $T_w^{(n)}$ refers to the window size with n sample points inside, it can be seen that for low n the sampling distribution is similar to the parent random variable C_l while for larger n the distribution becomes normal. This is in line with expectations since a certain number of windows is needed to overcome correlation while a certain number of points inside each window is also required. Therefore for $n <<$ and $n >>$ the correlation between sample points leads to a non-Gaussian probability density function.

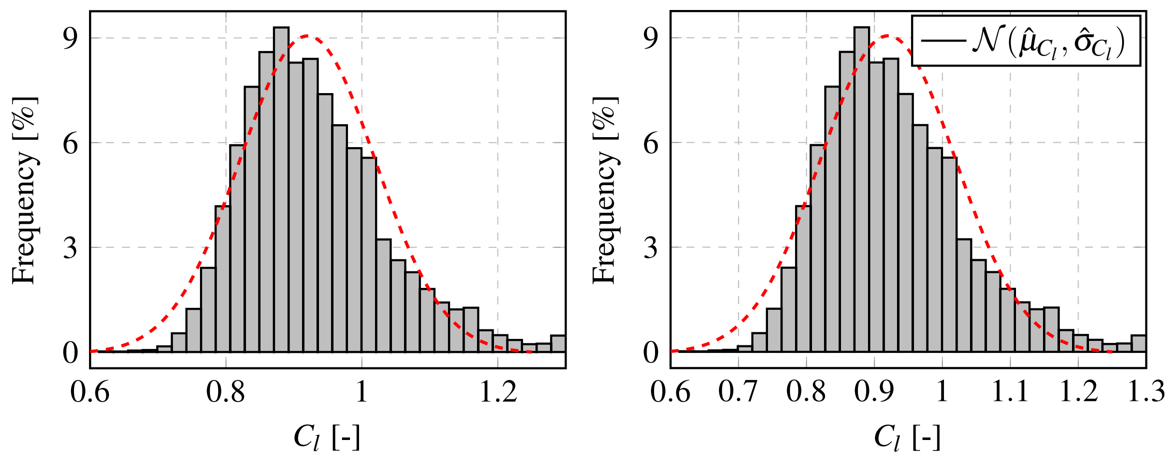

Figure 4.11: Frequency histograms of the sectional lift coefficient C_l (left) and of the sample mean in each window for different sizes of windows (right). The probability distributions are compared to the Gaussian one (dashed red line) characterised by $\hat{\mu}_{C_l}$ and $\hat{\sigma}_{C_l}$.

The operation of windows division introduced in section 4.2.1 has its importance regarding probabilities and statistics. The statistical error $\varepsilon_{T_w}\{\mathcal{P}_T\}$ of a point estimator $\mathcal{P}_T$ represents a random variable corresponding to a succession of averaged values for a fixed window size T_w. To clarify, when the sample of length T is divided into windows of size T_w it consists in sampling the original random variable into random variables $\mathcal{P}_{T_w}$ such that $\bigcup_{i=1}^{N_w} T_w^{(i)} = T$. Each random variable $\mathcal{P}_{T_w}$ is therefore characterised by its own probability structure f_{T_w}.

According to the central limit theorem, the distribution of the window averages converges in law toward a Gaussian random variable. Therefore, the fitted error ε_{fit} can be used to build a confidence interval for the true value θ_0 as introduced in Equation (2.19). The latter is rewritten for convenience :

$$\mathbb{P}\left(\mathcal{P}_T \in [p_1; p_2]\right) = 1 - \gamma, \tag{4.6}$$

which yields in the specific case of a Gaussian random variable and small statistical error magnitude[6]

$$\frac{\mathcal{P}_T}{1+\varepsilon_{\text{fit}}} \leq \theta_0 \leq \frac{\mathcal{P}_T}{1-\varepsilon_{\text{fit}}} \quad \text{with 68\% confidence} \tag{4.7}$$

or

$$\frac{\mathcal{P}_T}{1+2\varepsilon_{\text{fit}}} \leq \theta_0 \leq \frac{\mathcal{P}_T}{1-2\varepsilon_{\text{fit}}} \quad \text{with 95\% confidence} \tag{4.8}$$

The confidence intervals for the points estimator $\hat{\mu}_{C_l}$ and $\hat{\sigma}_{C_l}$ are represented in Figure 4.12. For both estimators, the running statistics is almost completely enveloped for the 68% confidence interval and totally enveloped for the 95%. It confirms that the methodology could be suitable to predict the total run time in terms of a targeted statistical accuracy.

[6] b

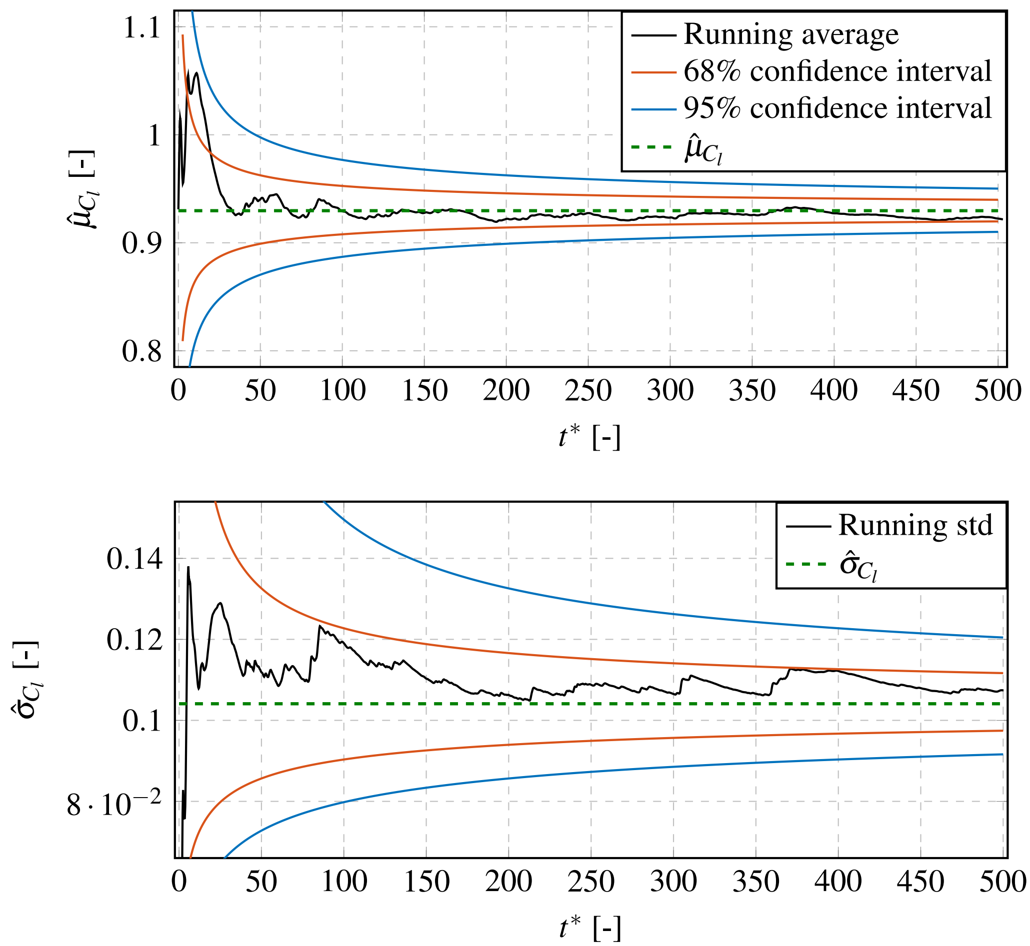

Figure 4.12: Running average (above) and standard deviation (below) of the measured sectional lift coefficient (C_l) from and confidence intervals (orange 68% and blue 95%) computed by means of the fitted error ε_{fit}. The best estimate, $\hat{\mu}_{C_l}$ and $\hat{\sigma}_{C_l}$, is represented in green dashed line. The time axis is non-dimensionalised as $t^* = t(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity. The data is truncated as $0 \leq t^* \leq 500$.

4.3.4 Transient estimation

In the present subsection, the transient detection algorithm is tested for simple synthetic transient signals. In particular are investigated oscillatory negative exponentials with and without additional white-noise. The previous list of transient signals is obviously far from exhaustive but allows to have a first insight of the methodology capabilities. Moreover, the negative exponential are characterised by a time-constant, denoted τ_c that quantify the duration before the signal stabilizes and becomes fully developed. Therefore, those kind of signals allow to compare the results of the methodology to a well-known quantity. Indeed, for a negative exponential of time-constant τ_c and amplitude A :

$$\mathcal{X}(t) = Ae^{-t/\tau_c}, \tag{4.9}$$

the expected value after which the signal converges to an asymptotic value is approximately $A\tau_c$.

Let consider an oscillatory negative exponential of time-constant $\tau_c = 10$, frequency 5 Hz, an amplitude $A = 4$ to which is added a pseudo-random white noise of unit amplitude (see above graph in Figure 4.13) simulated during 300 seconds. Therefore, the expected transient duration is around 40 seconds. After that, the signal stabilizes to the noisy contribution.

The middle graph in Figure 4.13 represents the product of both statistical errors for the mean and the standard deviation for the total sample length i.e $T = 300$ s. As expected, the product decreases as the sample is shortened or similarly as t_0 increases. Then a minimum is reached for $t_0 = t_t = 35.5$ s, which corresponds more or less to the expected transient duration from analytic definition of the time-constant. After that, the product increases with t_0 because of the shortening of the sample. It can also be seen that there are some spurious drop for $t_0 \approx T$ which is a drawback intrinsic to the methodology itself and its numerical implementation. As the sample is shortened drastically, less windows are available to average and estimate the statistical error. It justifies therefore the fact that the transient duration estimate is computed as the minimum of the product in the interval $[0;T/2]$. However, the truncation can be a bit too conservative as the unwanted drop occurs for very large t_0.

The third graph in Figure 4.13 represents the estimated transient duration t_t as a function of the sample length T i.e as the simulation progresses. It can be seen that the methodology requires a certain time to estimate accurately the transient duration ($\approx 2.2t_t$). Before converging to the asymptotic value, the transient duration varies linearly as $t_t = T/2$. It is due to the fact that the shortening variable t_0 is limited to $T/2$. Moreover, the methodology is derived in the specific assumption that the signal is free of any initial transient. Therefore, it requires a certain sample of stationary signal to become accurate and efficient. However, once the actual transient duration is estimated, only small oscillations occur but the value of t_t does not vary any more.

4.3.5 Summary

In the end, the methodology introduced by Mockett to estimate *a posteriori* the initial transient has demonstrated satisfactory outcomes as well as major drawbacks. This drawbacks will definitely influence the further chapters and a subsection is thus dedicated to summarize the latter.

First of all, the white noise scaling introduced by Mockett tends to strongly overestimate the error for a periodic signal which presents a scaling as $\varepsilon \sim t^{-1}$ instead of $\varepsilon \sim t^{-1/2}$ for the white noise. This drawback is the purpose of the next Chapter. Parallel to this point, the agreement with the scaling is satisfactory for large simulation time, when the correlation between data has significantly decreased.

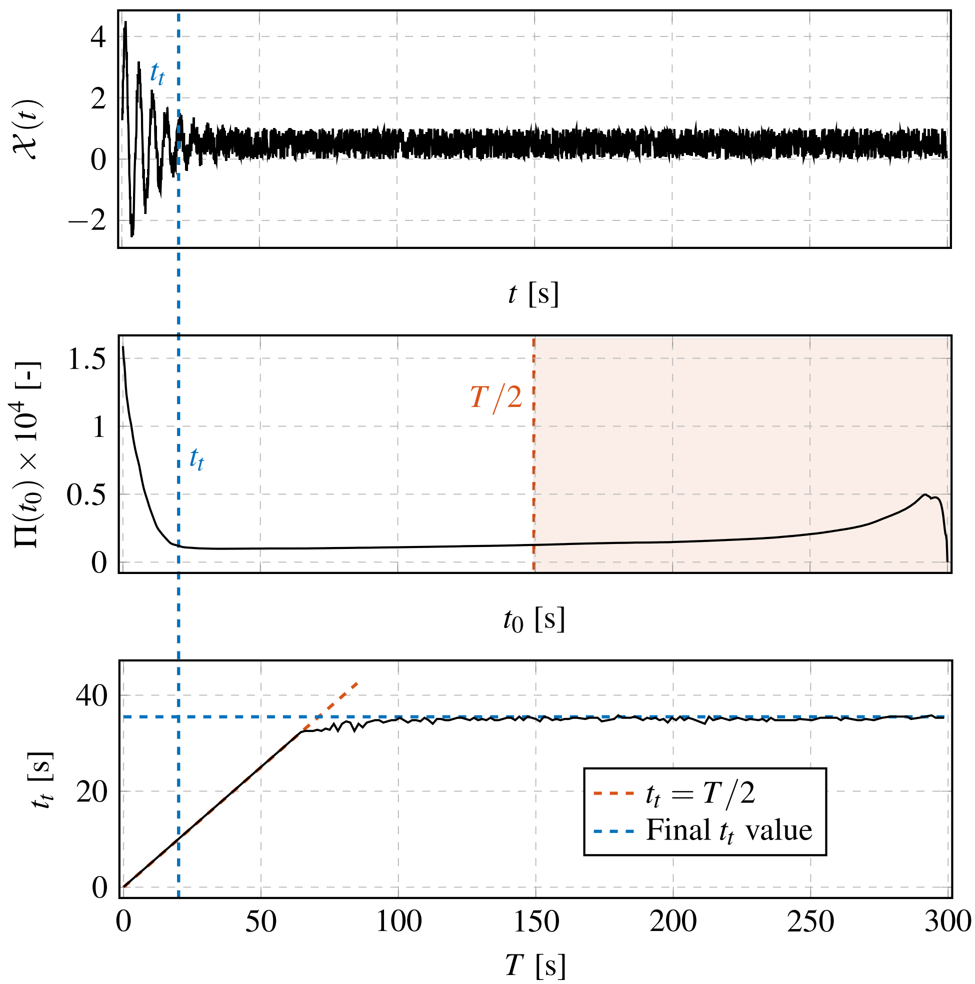

Figure 4.13: Time trace of the synthetic signal $\mathcal{X}(t)$ (above). Variation of the product $\Pi(t_0)$ highlighting a minimum at t_t (mid). Prediction of the transient estimate t_t as a function of the sample length T (below).

Finally, due to the unwanted variations in the product of $\sigma\{\hat{\mu}_{C_L}\}$ and $\sigma\{\hat{\mu}_{C_L}\}$ for large t_0 (compared to the simulation time T) the interval on which the minimum corresponding to t_t is computed is truncated to $T/2$. This increases the required time to estimate the converged value of t_t. In addition, the estimation of t_t can converge toward different values. Highlighting the more realistic one on the fly therefore requires engineering judgement or additional informations.

5 Influence of the frequency content on the statistical error

A major outcome of the Mockett's methodology introduced in the previous chapter is that it produces highly accurate results for a white noise while it cannot handle the periodic signals, or more precisely, it highly overestimates the statistical error. From this drawback comes the intuition that the frequency content has a direct influence on the time-dependence of ε. It actually makes sense since the time variation of $V\{\hat{\mu}_{\mathcal{X}}\}$ (resp. $V\{\hat{\sigma}_{\mathcal{X}}\}$) is computed by means of the time integral of the autocovariance function $C_{\mathcal{X}\mathcal{X}}(\tau)$ in Equation (3.14) (resp. Equation (3.17)) which forms a Fourier pair with the power spectral density $S_{\mathcal{X}\mathcal{X}}(f)$. This short reminder allows to introduce the main mathematical tools that will be used in this analysis and the chronological order in which they are employed.

The path followed through this chapter is then :

$$S_{\mathcal{X}\mathcal{X}}(f) \longrightarrow C_{\mathcal{X}\mathcal{X}}(\tau) \longrightarrow \varepsilon\{\mathcal{P}_T\} = g(t), \tag{5.1}$$

where $g(t)$ denoted a function of time.

In other words, a pre-defined frequency content of a signal $\mathcal{X}(t)$ is imposed through its power spectral density. From the latter can be derived the autocovariance function and finally the statistical error.

However, the analytic derivation of the autocovariance function by means of Fourier transform can be a hard task and sometimes even not possible (at least with the author knowledge of mathematical tools). Therefore, the analysis is performed firstly for well-known synthetic signals which forms the both ends of the PSD range : the white noise and the sine wave. Based on those results is introduced an intermediate case that allows to approach, to the limit, the white noise or the sine wave.

Finally, the idea that the bandwidth of a signal can influence the mean square error made through estimation is not new. Steen Krenk from the Denmark Technical University has introduced pieces of answer in the context of the dynamic response to pedestrian loads on a bridge modelled by a 1 degree-of-freedom oscillator [22]. He has highlighted that the displacement response amplitude for the pedestrian load modelled as a deterministic periodic signal varied like ξ^{-1} whereas the amplitude for a stochastic load varied as $\xi^{-1/2}$ with ξ the structural damping ratio of the structure. The parallelism with the variation of ε as t^{-1} for a sine wave and $t^{-1/2}$ for a white noise is therefore immediate. The further analysis is hence performed with this similarity in mind and the intermediate case is directly influenced by the structural engineering field.

5.1 Context and motivations

Turbulent flows are defined as three-dimensional and unsteady in nature. The fluctuating velocity takes the form of overlapping eddies due to vortex stretching over a wide range of scales, called the spectrum of turbulence. The largest scales in the spectrum are the most energetic. They generally have a length scale similar to macroscopic quantities such as the geometry or flow features. These large eddies are produced by the velocity gradients and thus strongly depends on the mean flow. Further, the energy is continuously transferred to smaller scales by vortex stretching. Finally, the energy is dissipated to heat at smaller scales by molecular viscosity. The origin of the cascade of energy is usually attributed to Richardson while twenty years later, Kolmogoroff provided relevant dimensional analysis of the energy spectrum. Particularly, Kolmogoroff's theory sates than the turbulent spectrum can be scaled as a «$-5/3$ law »

$$E(\kappa) \propto k^{-5/3}. \tag{5.2}$$

For instance, let consider the experimental sectional lift coefficient introduced in subsection 4.3.3. The $-5/3$ variation can clearly be seen in Figure 5.1 representing the power spectral density of C_l with respect to the Strouhal number.

The key physics of this type of flow is predominantly characterised by the unsteady, massively-separated wake region. The flow can be seen as the superposition of a nominally periodic shedding of large scale and of a finer random turbulent fluctuation. There is thus an energetic contribution for lowest frequencies as well as an energetic contribution in the neighbourhood of the shedding frequency.

Therefore, it can easily be understood that the assumption of a constant spectrum for all frequencies (white noise scaling) can sometimes be strong and lead to a misunderstanding of the whole flow physics. In particular it has been demonstrated in previous chapter that the white noise scaling was not efficient for sine wave. Hence the methodology is assumed to provide inaccurate results for flows with strong periodic characteristic e.g flows around a cylinder with an URANS simulation.

Several studies aimed at developing mathematical tools to model the turbulent spectrum namely those of Poppe, Von Karman [42] or Davenport [9]. However, this studies often lead to ex-

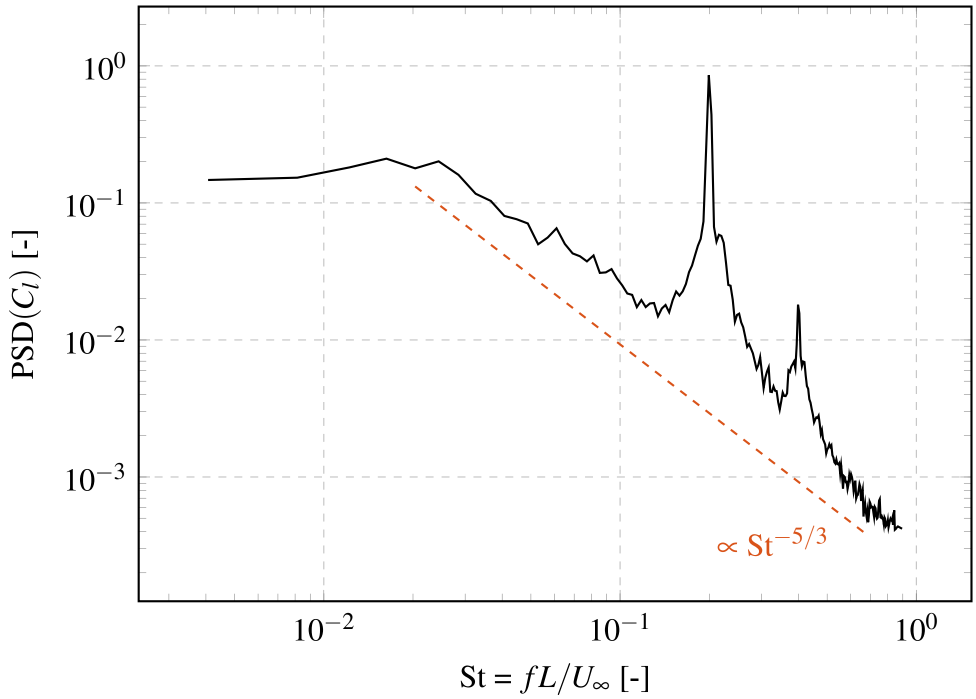

Figure 5.1: Power spectral density of the measured sectional lift coefficient C_l for a high angle-of-attack (60°) NACA0021 as a function of the Strouhal number. The measurements are provided by Swalwell et al [39].

pressions for the for the power spectral density that are hard to manipulate. In this chapter, a rather simple expression is used to derive analytical expressions that allow to quantify both effects : periodicity and broadband random fluctuations. For that purpose, the two main limit cases, namely the white noise and the sine wave are reminded and/or further derived. Based on these cases is introduced a mathematical model that is able to model both ends of the frequency spectrum. Finally, this mathematical model is introduced in the previous methodology to make the latter more general.

5.2 Time signal generation

In this section is briefly explained how the signal in the time domain is generated based on its frequency content. Under the assumption of an ergodic process, the frequency content $G_{\mathcal{X}\mathcal{X}}(f)$ and its corresponding random process $\mathcal{X}(t)$ are linked by [15]

$$G_{\mathcal{X}\mathcal{X}}(f) = \lim_{T\to\infty} \frac{1}{T} |X(f,T)|^2, \tag{5.3}$$

where T is the total length of the signal and

$$X(f,T) = \int_{-T/2}^{T/2} \mathcal{X}(t) e^{-j2\pi f t} dt \tag{5.4}$$

is the truncated Fourier transform of the process.

The limit expression in (5.3) expresses the fact that a finite sample length cannot contain enough information to model the power spectrum at low frequencies. However, if the frequency content is assumed to lie in a finite frequency interval, say $[f_{\min}; f_{\max}]$, the lowest frequencies can be caught for a sample length of at least $1/f_{\min}$ [34]. Given that, the limit can be omitted provided a long enough length record. From (5.3), it yields :

$$|X(f,T)| = \sqrt{T G_{\mathcal{X}\mathcal{X}}(f)}. \tag{5.5}$$

A solution of this equation is thus

$$X(f,T) = \sqrt{T G_{\mathcal{X}\mathcal{X}}(f)} e^{j\phi(f)}, \tag{5.6}$$

where $\phi(f)$ is a random variable with uniform probability density function between 0 and 2π. Finally, the time signal is obtained by inverse Fourier transform.

5.3 Limit cases

As introduced in section 3.3, a white noise is characterized by a constant autospectral density function and an associated centred Dirac autocovariance function. Given a infinite variance, the process is unreal but can nevertheless be approximated locally in a limited bandwidth, denoted B (see Figure 3.5). At the other end of the frequency spectrum, the sine wave is characterized by an autospectral density function modelled by a Dirac impulsion at the wave frequency while the associated autocovariance function is a cosine (see Figure 3.6). To really emphasize the difference, both power spectral densities are depicted in Figure 5.2 with $\mathcal{W}(t)$ the band limited white noise and $\mathcal{Y}(t)$ a sine of amplitude A and frequency f_0.

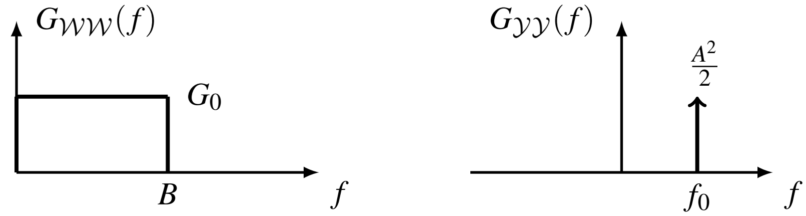

Figure 5.2: One-sided autospectral density functions of a low-pass white noise $\mathcal{W}(t)$ (left) with constant $G_{\mathcal{W}\mathcal{W}}(f) = G_0$ and bandwidth B and of a sine wave $\mathcal{Y}(t)$ (right) of amplitude A and frequency f_0.

Their corresponding autocorrelation functions (Fourier pair) are represented in Figure 5.3. That is, the correlation between sample points decreases rapidly for the band limited white-noise and the rate of decrease varies with the bandwidth B. Regarding the sine wave, the magnitude of the autocorrelation remains constant with time; in the sense that the amplitude of the peaks does not decrease.

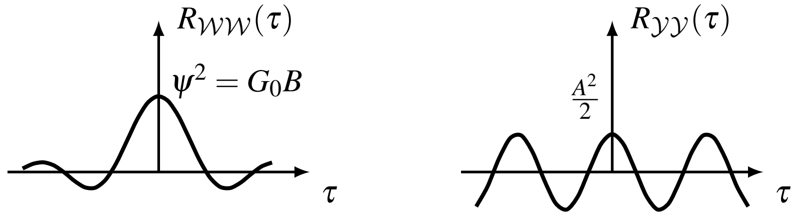

Figure 5.3: Autocorrelation functions of a low-pass white noise $\mathcal{W}(t)$ (left) with bandwidth B and of a sine wave $\mathcal{Y}(t)$ (right) of amplitude A and frequency f_0.

5.3.1 Bandwidth limited Gaussian white noise

A complete description of the bandwidth limited Gaussian white noise has already been provided in Part 1 and Chapter 4. Therefore, only the main analytical relations are re-written :

$$\varepsilon_{\text{ana}}\{\hat{\mu}_{\mathcal{W}}\} = \left(\frac{\sigma_{\mathcal{W}}}{\mu_{\mathcal{W}}}\right)\frac{1}{\sqrt{2\pi Bt}} \quad \text{and} \quad \varepsilon_{\text{ana}}\{\hat{\sigma}_{\mathcal{W}}\} = \frac{1}{\sqrt{4Bt}}. \tag{5.7}$$

5.3.2 Sine wave

An analytic formula for the time-evolution of the statistical error can be derived for the sine wave following almost the same steps as for the white noise in section 3.4.2. The variance of the mean estimate and the variance of the variance estimate are given respectively by :

$$V\{\hat{\mu}_{\mathcal{Y}}\} = \frac{1}{T}\int_{-T}^{T}\left(1-\frac{|\tau|}{T}\right)C_{\mathcal{Y}\mathcal{Y}}(\tau)d\tau \tag{5.8}$$

and

$$V\{\hat{\sigma}^2_{\mathcal{Y}}\} = \frac{2}{T}\int_{-T}^{T}\left(1-\frac{|\tau|}{T}\right)\left[C^2_{\mathcal{Y}\mathcal{Y}}(\tau)+2\mu^2_{\mathcal{Y}}C_{\mathcal{Y}\mathcal{Y}}(\tau)\right]d\tau. \tag{5.9}$$

In the specific case of white noise, the derivation has been simplified by assuming $\tau << T$. The assumption directly makes sense regarding the autocorrelation function. Indeed, assuming τ greatly lower than T corresponds to assuming an autocorrelation that decreases rapidly. Hence, this assumption is not acceptable anymore for a sine wave due to its non decreasing autocorrelation function (see Figure 5.3). Therefore, whole terms of (5.8) and (5.9) should be treated. After some derivations, it yields for a sine wave $\mathcal{Y}(t)$ of descriptive statistics $\mu_{\mathcal{Y}}$,$\sigma_{\mathcal{Y}}$

and frequency f_0 :

$$\varepsilon_{\text{ana}}\{\hat{\mu}_{\mathcal{Y}}\} = \left(\frac{\sigma_{\mathcal{Y}}}{\mu_{\mathcal{Y}}}\right)\frac{\sqrt{1-\cos(2\pi f_0 t)}}{2\pi f_0 t} \quad \text{and} \quad \varepsilon_{\text{ana}}\{\hat{\sigma}_{\mathcal{Y}}\} = \frac{1}{2}\frac{|\sin(2\pi f_0 t)|}{2\pi f_0 t}. \tag{5.10}$$

Equations (5.10) allow to assess the value of the statistical errors on the mean and the standard deviation estimates as a function of the time t for the specific case of a sine wave characterized by a frequency f_0, a mean $\mu_{\mathcal{Y}}$ and a standard deviation $\sigma_{\mathcal{Y}}$. The periodic oscillations of the statistical error are expected regarding the periodic nature of the signal. Indeed, the average of a periodic signal over an entire period is zero which explains the sharp decreasing of the mean error every period $1/f_0$ while the standard deviation is zero each half period $2/f_0$.

Hopefully, the analytic derivations correspond to the error computed using windows averaging in the previous chapter (see Figure 4.7). To be convinced, both analytic ε_{ana} and window averaging ε_{T_w} have been represented on the same graph in Figure 5.4. It can be seen that there is a perfect correspondence as both errors are almost superimposed (the only divergences come from the numerical appreciation of the sharp drop to $\varepsilon = 0$) . Therefore, the error computed by windows averaging can be assumed to be a «reference »in the further analysis.

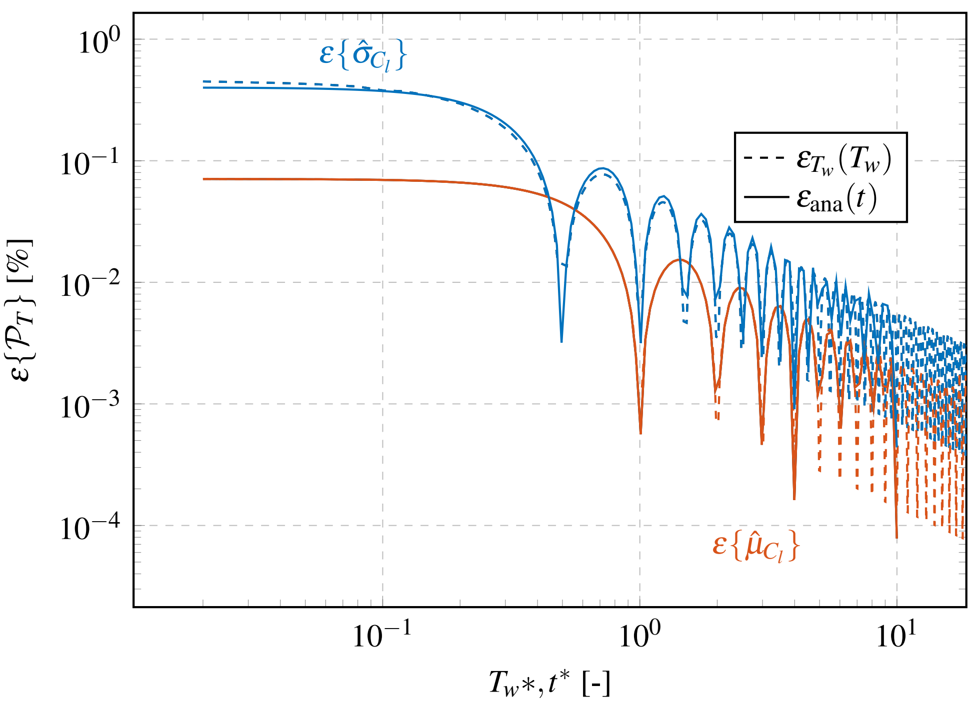

Figure 5.4: Comparison of estimated ε_{T_w} and analytic ε_{ana} statistical errors in the special case of a sine wave with $\sigma_{\mathcal{Y}} = 0.1$, $\mu_{\mathcal{Y}} = 1$, $f_0 = 10$ [Hz] and $\Delta t = 2 \times 10^{-2}$ [s]. The estimated error is in dashed line and the analytic in continuous line. The errors on the mean estimate are plotted in orange while the standard deviations ones are plotted in blue. The time (x-axis) is multiplied by the wave frequency : $T_w^* = T_w f_0$ and $t^* = t f_0$.

The analytic formulas allow to highlight some different features more accurately in the behaviour of the error for a sine wave. First, the plateau (in a log-log plot) can be quantified by the limit :

$$\lim_{t \to 0} \varepsilon_{\text{ana}}\{\hat{\mu}_{\mathcal{Y}}\} = \lim_{t \to 0} \left(\frac{\sigma_{\mathcal{Y}}}{\mu_{\mathcal{Y}}}\right) \frac{|\sin(2\pi f_0 t)|}{2\pi f_0 t} = \left(\frac{\sigma_{\mathcal{Y}}}{\mu_{\mathcal{Y}}}\right), \tag{5.11}$$

where the well know result for the limit of the cardinal sine[1] as $t \to 0$ is used. The first value of the statistical error for the mean estimate is thus expected to correspond to this ratio. It is interesting to notice that a similar plateau has been highlighted in Figure 4.9 corresponding to the statistical error for the measured lift coefficient. Furthermore, the value corresponding to this plateau is around 0.11 which exactly corresponds to the ratio $(\hat{\sigma}_{C_l}/\hat{\mu}_{C_l})$. That is, it confirms that the energetic contribution of the vortex shedding in Figure 5.1 can definitely play a role.

This is with this intuition and final motivation that comes the next section. A mathematical model will thus be built to take into account both contributions : periodic and broadband.

5.4 Intermediate case

The frequency content and its Fourier transform : the autocorrelation function, definitely influence how the statistical error varies with the sample length. To highlight this relationship a bit more, the statistical error can be rewritten as the product of a function of time and a spectral parameter :

$$\varepsilon_{\text{ana}}\{\hat{\mu}_{\mathcal{X}}\} = \left(\frac{\sigma_{\mathcal{X}}}{\mu_{\mathcal{X}}}\right) K_{\mu_{\mathcal{X}}} g_{\mathcal{X}}(t) \quad \text{and} \quad \varepsilon_{\text{ana}}\{\hat{\sigma}_{\mathcal{X}}\} = K_{\sigma_{\mathcal{X}}} g_{\mathcal{X}}(t) \tag{5.12}$$

where $K_{\mu_{\mathcal{X}}}$ and $K_{\sigma_{\mathcal{X}}}$ are the spectral parameters respectively corresponding to the mean and the standard deviation and $g(t)$ a function of time which is directly defined by the frequency spectrum $G_{\mathcal{X}\mathcal{X}}$. The summary of these parameters are available in Table 5.1.

[1]By definition of the cardinal sine $\text{sinc}(t)$:

$$\text{sinc}(t) = \frac{\sin(t)}{t} \quad \text{and} \quad \lim_{t \to 0} \text{sinc}(t) = 1$$

	$\varepsilon_{\text{ana}}\{\mathcal{P}_{\mathcal{X}}\}$	$K_{\mathcal{P}_{\mathcal{X}}}$	$g_{\mathcal{X}}(t)$
White Noise	$\hat{\mu}_{\mathcal{W}}$	$\frac{1}{\sqrt{2B}}$	$\frac{1}{\sqrt{t}}$
$\mathcal{W}(t)$	$\hat{\sigma}_{\mathcal{W}}$	$\frac{1}{\sqrt{4B}}$	
Sine wave	$\hat{\mu}_{\mathcal{Y}}$	$\frac{1}{2\pi f_0}$	$\frac{\lvert\sin(2\pi f_0 t)\rvert}{t}$
$\mathcal{Y}(t)$	$\hat{\sigma}_{\mathcal{Y}}$	$\frac{1}{4\pi f_0}$	

Table 5.1: Summary of the spectral parameters and time dependencies for the statistical error on the mean and standard deviation estimates for a bandpass white noise and a sine wave.

The choice of the mathematical model for the intermediate case is motivated by many factors. First of all, it should be able to model both limit cases when a quantity of interest of the spectrum is pushed to the limits. This parameter is naturally chosen as a pseudo-bandwidth of the signal such that when the parameter tends to zero, the signal is periodic and when the parameter tends to infinity, the signal is purely random. Secondly, the expression of the spectrum should be such that the analytic derivations are feasible. Indeed, the process of the statistical error derivation can be tricky and the expression should be easily manipulable. Finally, the spectrum should be line with de physical expectations of the turbulence.

5.4.1 Physical considerations about turbulence and correlation

[2]Correlation between random variables has a central position in turbulence modelling. In the more general case, the fluid random variables are fields e.g velocity, pressure etc. Despite they are statistics, the correlation functions are also fields and thus deterministic. Usually, these are functions of position (or relative position) and time. In this short insight to feel the physics of turbulence, only time correlation is considered.

To approach a first insight of the autocorrelation function in the context of turbulence, let consider an instantaneous turbulent velocity $\tilde{u}$ and a mean velocity U, linked by $U \equiv E\{\tilde{u}\}$. Therefore, the fluctuating velocity u is given by :

$$u = \tilde{u} - U. \tag{5.13}$$

This simply corresponds to a decomposition of the total velocity into a mean and a fluctuation $\tilde{u} = U + u$ with $E\{u\} = 0$.

[2]This discussion is mainly inspired by [14].

Lagrangian approach for correlation

This section focuses on the time-autocorrelation i.e the correlation that can exist between a velocity component and itself delayed by a time shift τ :

$$R_{uu}(\tau) = \text{corr}\{u(t+\tau)u(t)\} = E\{u(t+\tau)u(t)\}. \tag{5.14}$$

Let consider a simple discrete random process that predicts the fluid velocity at time shifts τ :

$$u(t+\tau) = \rho u(t) + sw(t) \tag{5.15}$$

with ρ and s real coefficients and $w(t)$ a white noise of unit magnitude and zero mean. Under the assumption of a statistically stationary process, the variance is a constant and $E\{u(t+\tau)^2\} = E\{u(t)^2\} = E\{u^2\}$. Therefore, squaring (5.15) and applying the expected value operator on both sides :

$$E\{u(t+\tau)^2\} = \rho^2 E\{u(t)^2\} + s^2 E\{w^2\} + 2rsE\{u(t)w(t)\}. \tag{5.16}$$

Given that w is a white noise, the latter is uncorrelated with the velocity field u and the last term is zero. Furthermore, by the intrinsic definition of w, $E\{w^2\} = 1$ and thus

$$s = \sqrt{(1-r^2)E\{u(t)^2\}}. \tag{5.17}$$

If now (5.15) is multiplied by $u(t)$ and the expected value operator is applied :

$$E\{u(t+\tau)u(t)\} = \rho E\{u(t)^2\} \tag{5.18}$$

and it is directly understood that a is the correlation coefficient[3]. For a stationary process, the correlation is only a function of the time shift τ. By expanding ρ in a McLauren series :

$$\rho = 1 + \frac{d\rho}{dt}(0)\tau + o(\tau^2) \tag{5.19}$$

and by dimensional analysis, $d\rho/dt(0)$ has unit t^{-1} and thus a correlation length-scale T_c is introduced such that $d\rho/dt(0) \equiv 1/T_c$. Finally, by pre-multiplying (5.15) by $u(t-\tau)$, a differential equation for $R_{uu}(\tau)$ can be obtained :

$$\frac{dR_{uu}(\tau)}{d\tau} = -\frac{R_{uu}(\tau)}{T_c}. \tag{5.20}$$

[3]It justifies also the nomenclature ρ used, to stay coherent with the previous introduction to the correlation coefficient in section 4.3.3.

It appears thus that the autocorrelation of the simple random process introduced in (5.15) has a negative exponential behaviour and that the correlation length-scale is :

$$R_{uu}(\tau) = e^{-|\tau|/T_c} \quad \text{and} \quad T_c = \int_0^{\infty} R(\tau')d\tau'. \tag{5.21}$$

This short investigation for a simple random process allows to highlight a time dependency for the autocorrelation function as well as a quantity of interest : the correlation length scale. Both conclusions have mainly driven the choice of further introduced mathematical model.

Frequency spectrum of the correlation

The frequency spectrum $G_{uu}(f)$ can be obtained by inverse Fourier transform of the autocorrelation function. It yields :

$$G(f) = 2\int_{-\infty}^{+\infty} R_{uu}(\tau)e^{-j2\pi f\tau}d\tau = \frac{E\{u(t)^2\}T_c}{\pi(1+4\pi^2 f^2 T_c^2)}. \tag{5.22}$$

5.4.2 Mathematical model for the frequency spectrum

Equations (5.21) and (5.22) indicate that the time autocorrelation function of a turbulent velocity field could be in the form of a negative exponential and thus a rational power spectrum density with $G \propto f^{-2}$. Based on these two insights, a general one-sided frequency spectrum and its associated autocorrelation function can be defined for a random process $\mathcal{X}(t)$:

$$G_{\mathcal{X}\mathcal{X}}(f) = \frac{\sigma_{\mathcal{X}}^2}{\pi}\frac{\beta}{(f-f_0)^2+\beta^2} \quad \text{and} \quad R_{\mathcal{X}\mathcal{X}}(\tau) = \frac{\sigma_{\mathcal{X}}^2}{\sqrt{2\pi}}e^{-\beta|\tau|}\cos(2\pi f_0\tau) \tag{5.23}$$

where the parameters β, f_0 and $\sigma_{\mathcal{X}}$ are positive and real (see Figure 5.5).

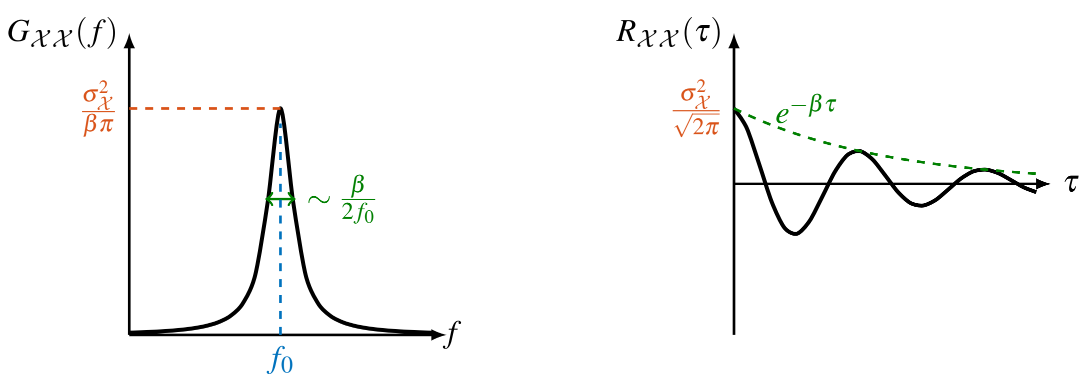

Figure 5.5: Schematic representation of the one-sided power spectral density $G_{\mathcal{X}\mathcal{X}}(f)$ (left) and the autocorrelation function $R_{\mathcal{X}\mathcal{X}}(\tau)$ (right) of a random process $\mathcal{X}(t)$ with variance $\sigma_{\mathcal{X}}^2$, main frequency f_0 and bandwidth parameter β.

Each parameter of (5.23) has its own importance and signification :

i. $\sigma_{\mathcal{X}}^2$ is the variance of the parent signal $\mathcal{X}(t)$. It corresponds to the area under $G_{\mathcal{X}\mathcal{X}}(f)$ i.e the integral of G for all frequencies. It directly influences the magnitude of the power spectrum i.e the energy of the time signal. This explains therefore the multiplying factor $(1/\pi)$ that ensures the validity of the property $\int_{\mathbb{R}} G_{\mathcal{X}\mathcal{X}}(f)df = \sigma_{\mathcal{X}}^2$.

ii. f_0 is the frequency on which the peak is centred, it corresponds to the periodic energetic contribution.

iii. β is a bandwidth parameter which defines the acuity of the peak. The half-power width is proportional to β/f_0. In addition, it influences the time autocorrelation as it appears in the exponential exponent. It can thus be understood as the inverse of a time constant. That is, a greater β would tend to decrease rapidly the correlation while a small β would do the opposite. This is one of the justifications for this parameter to be called «bandwidth »parameter. When the bandwidth is large, the autocorrelation is presumed to be weak (white noise) in contrast to a sine wave with high correlation and Dirac bandwidth.

For convenience, the frequency spectrum and the autocorrelation function are non-dimensionalized by introducing the parameters :

$$\beta^* = \beta/f_0 \quad , \quad f^* = f/f_0 \quad \text{and} \quad \tau^* = \tau f_0 \tag{5.24}$$

so that (5.23) can re-written as :

$$G(f^*) = \frac{\sigma_{\mathcal{X}}^2}{\pi f_0} \frac{\beta^*}{(f^*-1)^2 + \beta^{*2}} \quad \text{and} \quad R(\tau^*) = \frac{\sigma_{\mathcal{X}}^2}{\sqrt{2\pi}} e^{-\beta^*\tau^*} \cos(2\pi\tau^*). \tag{5.25}$$

The parameter of interest is therefore β^* corresponding to the ratio between the bandwidth parameter and the centred frequency. Even more, it defines the acuity of the peak and is hence useful to model the limit cases. A signal with a small β^* is qualified as a narrow-band signal while a large β^* corresponds to a wide-band signal.

Influence of β^*

Different regimes can thus be highlighted with respect to β^*. For large $\beta^* \gg 1$ the system is highly uncorrelated and the peak width is important so that the system behaves like a pseudo white-noise. On the contrary, for small $\beta^* \ll 1$ the correlation between sample points decreases slowly and the peak of the frequency spectrum is very sharp, thus the system acts like a pseudo periodic signal. The intermediate range $\beta^* \approx 1$ is the real interest of the analysis and a dedicated section will be presented further.

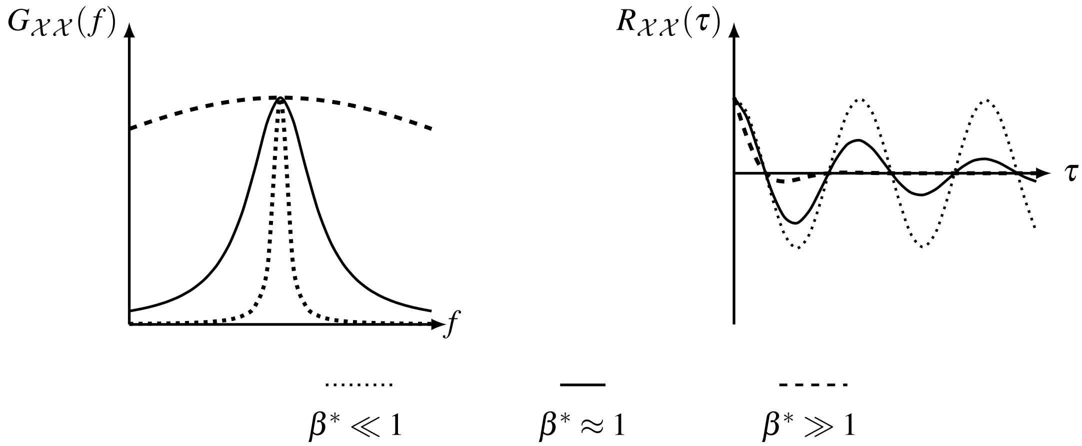

Figure 5.6: Schematic representation of the one-sided power spectral density $G_{\mathcal{XX}}(f)$ (left) and the autocorrelation function $R_{\mathcal{XX}}(\tau)$ (right) for the three different regimes $\beta^* \ll 1$ (dotted), $\beta^* \approx 1$ (continuous) and $\beta^* \gg 1$ (dashed). The power spectral densities have been scaled for the sake of visibility.

A schematic representation of the three regimes is represented in Figure 5.6. The power spectral densities have been scaled for the sake of visibility. It can be seen that both frequency spectrum and autocorrelation functions are really similar to those of limit cases (see Figures 5.2 and 5.3).

Analogy to the structural engineering

An interesting comparison can be made between the present mathematical model and frequency response function of a standard second order mechanical system [18]. Indeed, by symmetrically extending the one-sided power spectral density to the full frequency domain :

$$S_{\mathcal{XX}}(f) = \frac{1}{2}\left[G_{\mathcal{XX}}(f) + G_{\mathcal{XX}}(-f)\right] = \frac{\sigma_{\mathcal{X}}^2 \beta}{\pi} \frac{f_0^2 + \beta^2 + f^2}{\left(f_0^2 + \beta^2 - f^2\right)^2 + 4\beta^2 f^2} \tag{5.26}$$

Introducing the parameters $f_n^2 = f_0^2 + \beta^2$ and $\xi f_n = \beta$, it yields :

$$S_{\mathcal{XX}}(f) = \frac{\sigma_{\mathcal{X}}^2 f_n}{\pi} \frac{f_n^2 + f^2}{\left(f_n^2 - f^2\right)^2 + 4\xi^2 f_n^2 f^2} \tag{5.27}$$

The relationship with the response of a simple spring-mass system with natural frequency f_n and damping ratio ξ is therefore immediate. Moreover, the damping ratio takes the well-known form :

$$\xi = \frac{\beta/f_0}{\sqrt{1+(\beta/f_0)2}} = \frac{\beta^*}{\sqrt{1+\beta^{*2}}} \tag{5.28}$$

5.5 Analyses and discussions

Section 5.3 introduced the both ends of the frequency spectrum which allow by their simplicity to derive useful equations while 5.4 introduced a mathematical expression for the frequency content that allows to model different configurations as a function of the quantity of interest β^*. In this section is investigated how the statistical error varies with this parameter and how the latter can be quantify to be used in the algorithm presented in the previous chapter.

For that purpose, both limit cases $\beta^* \gg 1$ and $\beta^* \ll 1$ are investigated. In particular, it is verified that the model can represent the both ends of the frequency spectrum. After that, the intermediate case $\beta^* \approx 1$ is studied and the link between the statistical error and the bandwidth parameter is performed.

5.5.1 Case $\beta^* \ll 1$

In the case of a small normalized bandwidth parameter, the power spectrum is a very sharp peak and the autocorrelation function a pseudo cosine wave (see Figure 5.6). Further, in the limit when $\beta^* \to 0$:

$$G_{\mathcal{X}\mathcal{X}}(f^*) \sim \frac{\sigma_{\mathcal{X}}^2}{\pi f_0}\delta(1) \quad \text{and} \quad R_{\mathcal{X}\mathcal{X}}(\tau^*) \sim \frac{\sigma_{\mathcal{X}}^2}{\sqrt{2\pi}}\cos(2\pi\tau^*) \tag{5.29}$$

which corresponds to the analytic characteristic of a sine wave of frequency f_0 and amplitude $\sigma_{\mathcal{X}}/\pi$ i.e the multiplicative factor in the definition of $G_{\mathcal{X}\mathcal{X}}(f)$.

It can be seen in Figure 5.7 that the mathematical model can represent a sine wave in the limit $\beta^* \to 0$. In this case, the statistical error is similar to the statistical error of a sine wave at frequency $f^* = 1$ with an amplitude $\sigma_{\mathcal{X}}^2/(\pi)$. Moreover, the statistical error decreases as $1/t$ as expected.

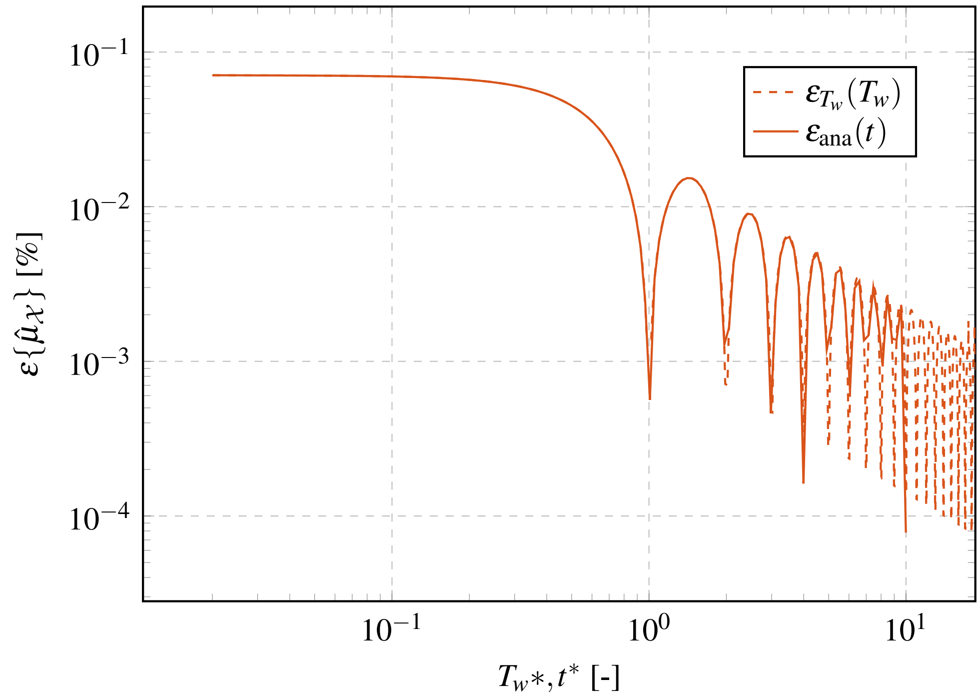

Figure 5.7: Time evolution of the statistical error on the mean estimate for the signal $\mathcal{X}(t)$ characterised by the couple of parameters $f_0 = 1$, standard deviation $\sigma_{\mathcal{X}} = 0.1$ and $\beta^* = 10^{-4}$.

5.5.2 Case $\beta^* \gg 1$

In the case of a large normalized bandwidth parameter, the power spectrum is a flat and the autocorrelation function decreases rapidly to zero (see Figure 5.6).

It can be seen in Figure 5.8 that the mathematical model can represent a white noise in the limit $\beta^* \to \infty$. In particular, the estimated statistical error is almost superimposed with the analytic formula taking into account the Nyquist bandwidth.

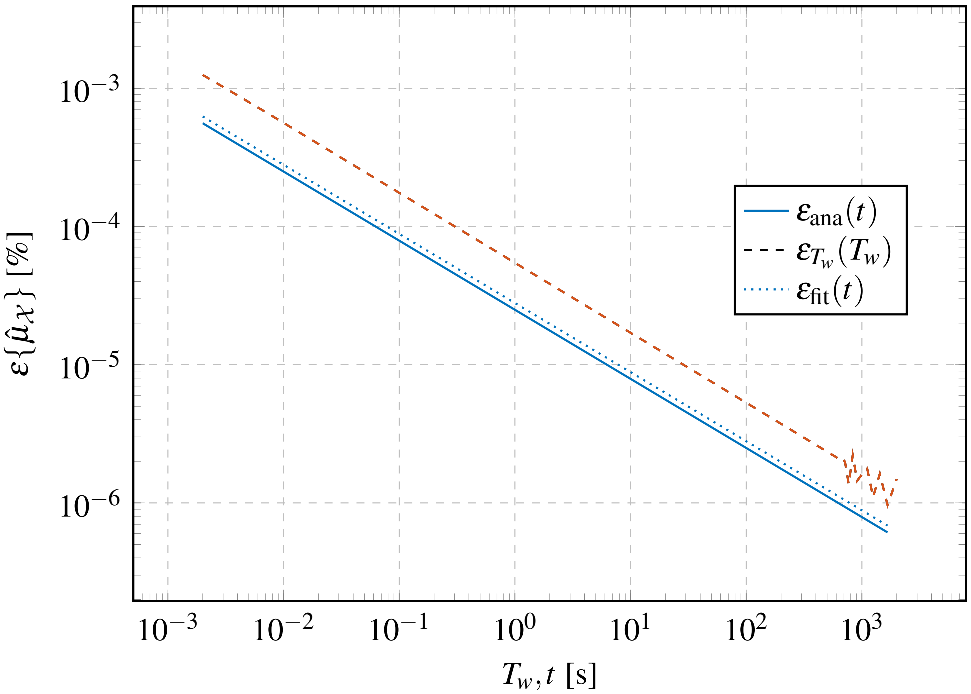

Figure 5.8: Time evolution of the statistical error on the mean estimate for the signal $\mathcal{X}(t)$ characterised by the couple of parameters $f_0 = 1$, standard deviation $\sigma_{\mathcal{X}} = 0.1$ and $\beta^* = 5$.

5.5.3 Case $\beta^* \approx 1$

The case $\beta^* \approx 1$ is the one of main interest since it corresponds to the range usually met in practice. It also corresponds to the range with a lack of understanding. Therefore, this subsection aims to describe qualitatively and quantitatively the behaviour of the statistical error for an intermediate bandwidth parameter.

General overview and problem statement

As a first introduction, the shapes of the autocorrelation functions, the power spectra and the statistical errors for two different values of β^* (1 and 0.05) have been represented in Figure 5.9.

The influence of the normalized bandwidth is therefore not negligible in the statistical error computation. In the first row, it can be seen that the correlation is much more important for a lower β^*. For the same x-axis scale, the correlation decreases to almost zero at $\tau^* = 4$ for $\beta^* = 1$ while it remains considerable, around 0.5 for $\beta^* = 0.05$.

Regarding the frequency content, that acuity of the peak at the resonant frequency $f^* = 1$ is smaller for a smaller β^*. In addition, the magnitude of the peak i.e the value at its maximum is larger for a smaller β^*. A key value of the bandwidth parameter to analyse would thus be the one corresponding to a "flat peak". Since the peak is assumed to correspond to the periodic contribution to the time signal, the key value of β^* would hence be the value at which there is no periodicity at all and the signal would be like a white noise.

Finally, the shapes of both statistical error are different but a common behaviour can be highlighted. For a short time, the statistical error tends to behave like a sine wave and some oscillations similar to those in Figure 5.7 can be seen. As the time increases, the oscillations disappears to become a straight line. More interestingly, different slopes in ε_{T_w} can be noticed as the time is increasing. Firstly, when oscillations are noticeable, the error varies like $1/t^*$ and thus when the error is a straight line it varies like $1/\sqrt{t^*}$. It is therefore presumed that for low t^* the error is mainly driven by the periodic content while for large t^*, the random broadband part is the most predominant.

Behaviour for a small t

For a small t, the behaviour of the statistical error is totally driven by the periodic content and the time signal corresponds to :

$$\mathcal{X}(t^*) \sim \frac{\sigma_{\mathcal{X}}^2}{\pi} \sin(2\pi t^*). \tag{5.30}$$

Therefore, the statistical error can directly be deduced by the analytic formula for a sine wave.

Behaviour for a large t

For a large t, additional derivations have to be done since the first part of the statistical error is influenced by the periodic content. In particular, the curve-fitting parameter is not relevant anymore.

For large t, the white noise assumption can be applied i.e it can be assumed that $|\tau^*| << T$ in (5.8) :

$$V\{\hat{\mu}_{\mathcal{X}}\} \approx \frac{1}{T}\int_{-\infty}^{\infty} C_{\mathcal{X}\mathcal{X}}(\tau^*)d\tau^* = \frac{2}{T}\int_0^{\infty} C_{\mathcal{X}\mathcal{X}}(\tau^*)d\tau^*, \tag{5.31}$$

where the last equality is due to the parity of the autocorrelation function. The computation of the integral can be direct if one remembers that $C_{\mathcal{X}\mathcal{X}}$ forms a Fourier pair[4] with $G_{\mathcal{X}\mathcal{X}}$. Indeed, it can be reduced to :

$$V\{\hat{\mu}_{\mathcal{X}}\} \approx \frac{1}{T}\frac{G_{\mathcal{X}\mathcal{X}}(0)}{2} = \frac{1}{T}\frac{\sigma_{\mathcal{X}}^2}{\pi f_0^2}\frac{\beta^*}{1+\beta^{*2}}. \tag{5.32}$$

[4] By definitions of a Fourier pair and of the one-sided power spectrum density:

$$G_{\mathcal{X}\mathcal{X}}(f) = 4\int_0^{\infty} R_{\mathcal{X}\mathcal{X}}(\tau)\cos(2\pi f\tau)d\tau \quad \rightarrow \quad G_{\mathcal{X}\mathcal{X}}(0)/4 = \int_0^{\infty} R_{\mathcal{X}\mathcal{X}}(\tau)d\tau$$

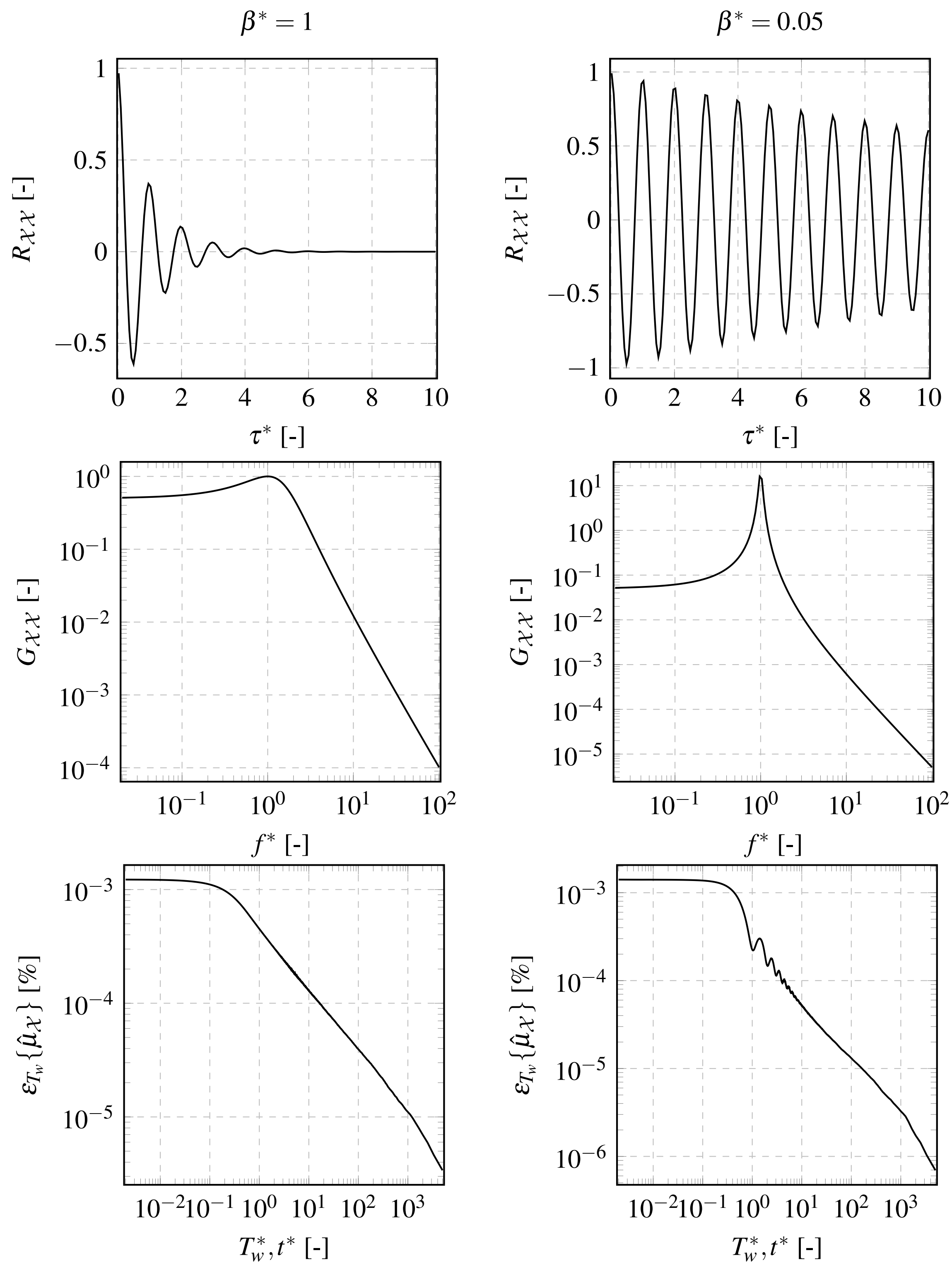

Figure 5.9: Autocorrelation functions (first row), one-sided frequency spectrum (middle row) and statistical error for the mean estimate (last row) in the specific cases of $\beta^* = 1$ (left column) and $\beta^* = 0.05$ (right column). Other parameters $\sigma^2_{\mathcal{X}}$ and f_0 have been set to unity.

Therefore, the statistical error for a large t, denoted $\varepsilon_{t>}$ is given by :

$$\varepsilon_{t>}\{\hat{\mu}_{\mathcal{X}}\}(t^*) = \frac{\sqrt{V\{\hat{\mu}_{\mathcal{X}}\}}}{\mu_{\mathcal{X}}} = \left(\frac{\sigma_X}{\mu_X}\right)\frac{1}{\sqrt{2B_{t>}t^*}} \quad \text{with} \quad B_{t>} = \frac{\pi f_0(1+\beta^{*2})}{\beta^*}. \tag{5.33}$$

As far as the standard deviation is concerned, the statical error defined in (5.9) can be reduced as :

$$V\{\hat{\sigma}_{\mathcal{X}}\} \approx \frac{2}{T}\int_{-\infty}^{\infty} C^2_{\mathcal{X}\mathcal{X}}(\tau)d\tau. \tag{5.34}$$

According to Perseval's theorem, the integral of the square of a function is equal to the integral of the square of its Fourier transform (i.e the energy is conserved from time-domain to frequency domain). Therefore, the previous equation can be re-written as :

$$V\{\hat{\sigma}_{\mathcal{X}}\} \approx \frac{1}{T}\int_{0}^{\infty} G^2_{\mathcal{X}\mathcal{X}}(f)df = \frac{8}{T}\frac{\sigma^2_{\mathcal{X}}}{2\pi}\left[\frac{\beta^{*2}+2\pi^2}{2\beta^{*3}+8\pi^2\beta^*}\right]. \tag{5.35}$$

Finally, the statistical error is thus given by :

$$\varepsilon_{t>}\{\hat{\sigma}_{\mathcal{X}}\}(t^*) = \frac{\sqrt{V\{\hat{\sigma}_{\mathcal{X}}\}}}{\sigma_{\mathcal{X}}} = \frac{1}{\sqrt{4B_{t>}t^*}} \quad \text{with} \quad B_{t>} = \frac{\pi}{16}\left[\frac{2\beta^{*3}+8\pi^2\beta^*}{\beta^{*2}+2\pi^2}\right]/ \tag{5.36}$$

Scaling transition

If ε_{sin} denotes the analytic expression of the statistical error for a sine wave and ε_{wn} for a white noise, it has been highlighted that the statistical error of the process $\mathcal{X}(t^*)$ is similar to $\varepsilon_{\text{sin}}(t^*)$ for a small t^* and similar to $\varepsilon_{\text{wn}}(t^*)$ for a large t^*. Let consider a time $t^* = t^*_{\text{tr}}$, called transition time, which corresponds to the intersection of both curves ε_{sin} and ε_{wn}.
It can therefore be written :

$$\varepsilon(t^*) \sim \begin{cases} \varepsilon_{\text{sin}}(t^*) & ,t^* \in [0;t^*_{\text{tr}}] \\ \varepsilon_{\text{wn}}(t^*) & ,t^* \in]t^*_{\text{tr}};+\infty[\end{cases}. \tag{5.37}$$

A more visual definition of the transition time is available in Figure **??**. Actually, the transition between the $1/t^*$ and the $1/\sqrt{t^*}$ scaling is not punctual. The latter is realised on a interval of time centred in t^*_{tr}. However, assessing the punctual parameter t^*_{tr} is easier as a first step.

Mathematically, this trend is well understood if the statistical error of $\mathcal{X}(t^*)$ can be assumed to be the sum of two contributions : the periodic and the broadband one. In that way, it can be written :

$$\varepsilon(t^*) \sim \varepsilon_{\text{sin}} + \varepsilon_{\text{wn}} = \varepsilon_{\text{sin}}\left(1+\frac{\varepsilon_{\text{wn}}}{\varepsilon_{\text{sin}}}\right) = \varepsilon_{\text{wn}}\left(1+\frac{\varepsilon_{\text{sin}}}{\varepsilon_{\text{wn}}}\right) \tag{5.38}$$

Therefore, for a small t^* :

$$\lim_{t^*\to 0} \varepsilon(t^*) \sim \lim_{t^*\to 0} \varepsilon_{\text{sin}} \left(1+\frac{\varepsilon_{\text{wn}}}{\varepsilon_{\text{sin}}}\right) \sim \lim_{t^*\to 0} \varepsilon_{\text{sin}} \left(1+\frac{t^{*-1/2}}{t^{*-1}}\right) = \varepsilon_{\text{sin}} \tag{5.39}$$

so that $\varepsilon_{\text{wn}} = o\left(\varepsilon_{\text{sin}}\right)$ for $(t^* \to 0)$ and thus $\varepsilon(t^*) \sim \varepsilon_{\text{sin}}(t^*)$.
On the contrary, for a large t^* :

$$\lim_{t^*\to \infty} \varepsilon(t^*) \sim \lim_{t^*\to \infty} \varepsilon_{\text{wn}} \left(1+\frac{\varepsilon_{\text{sin}}}{\varepsilon_{\text{win}}}\right) \sim \lim_{t^*\to \infty} \varepsilon_{\text{wn}} \left(1+\frac{t^{*-1}}{t^{-1/2}}\right) = \varepsilon_{\text{wn}} \tag{5.40}$$

so that $\varepsilon_{\text{sin}} = o\left(\varepsilon_{\text{wn}}\right)$ for $(t^* \to \infty)$ and thus $\varepsilon(t^*) \sim \varepsilon_{\text{wn}}(t^*)$.

Therefore, a point estimate of the scaling transition can be obtained by expressing the latter as the intersection of $\varepsilon_{t<}$ and $\varepsilon_{t>}$. For convenience, the sine dependence of $\varepsilon_{t<}$ is omitted to only consider the $1/t^*$ variation. It is thus equivalent to the straight line passing through every peaks. In the case of the mean estimate, it yields :

$$t^*_{tr} = \frac{\varepsilon_{t<}}{\sqrt{1-\cos(2\pi t^*)}} \cap \varepsilon_{t>} \quad \Leftrightarrow \quad t^*_{tr} = \frac{2B_{t>}}{\pi^2}. \tag{5.41}$$

Introducing the expression of $B_{t>}$ (5.33) :

$$t^*_{tr} = \frac{(1+\beta^{*2})}{2\pi f_0 \beta^*}. \tag{5.42}$$

Summary

In the current subsection has been derived a complete description of the statistical error in the case $\beta^* \approx 1$. It has been shown that the error could be seen as a succession of both contributions. A periodic behaviour for a small t^* with a $1/t^*$ dependence and a broadband contribution for a larger t^* with a $1/\sqrt{t^*}$ dependence. The transition of scaling t_{tr} has been estimated punctually. In the end, the statistical error for $\beta^* \approx$ can be defined as :

$$\varepsilon\{\hat{\mu}_X\}(t^*) = \begin{cases} \left(\frac{\sigma_\mathcal{X}}{\mu_\mathcal{X}}\right) \frac{\sqrt{1-\cos(2\pi t^*)}}{2\pi t^*} & , t^* \in [-\infty \,;\, t^*_{tr}] \\ \\ \left(\frac{\sigma_\mathcal{X}}{\mu_\mathcal{X}}\right) \frac{1}{\sqrt{2B_{t>}t^*}} & , t^* \in]t^*_{tr} \,;\, +\infty[\end{cases}, \tag{5.43}$$

where the bandwidth parameter and the transition point are given by :

$$B_{t>} = \frac{\pi f_0 (1+\beta^{*2})}{\beta^*} \quad \text{and} \quad t^*_{tr} = \frac{2B_{t>}}{\pi^2}. \tag{5.44}$$

A graphic representation is provided in Figure 5.10 for the parameter $\beta^* = 0.05$. It can be seen that the point estimation of the transition is a simplification and a further improvement would be to estimate a transition band centred in t^*_{tr} whose width would probably depends on β^*. However, there is a really good agreement between the estimated statistical error and the mathematical model derived which gives confidence in the scaling introduced. In particular, the bandwidth $B_{t>}$ provides a good scaling for a large t^* and the transition point is in line with visual inspection. Numerically, the values are, in the case $\beta^* = 0.05$ and $f_0 = 1$:

$$B_{t>} = \frac{\pi 1(1+0.05^2)}{0.05} \approx 62.99 \text{ Hz} \quad \text{and} \quad t^*_{tr} = \frac{2B_{t>}}{\pi^2} \approx 12.76. \tag{5.45}$$

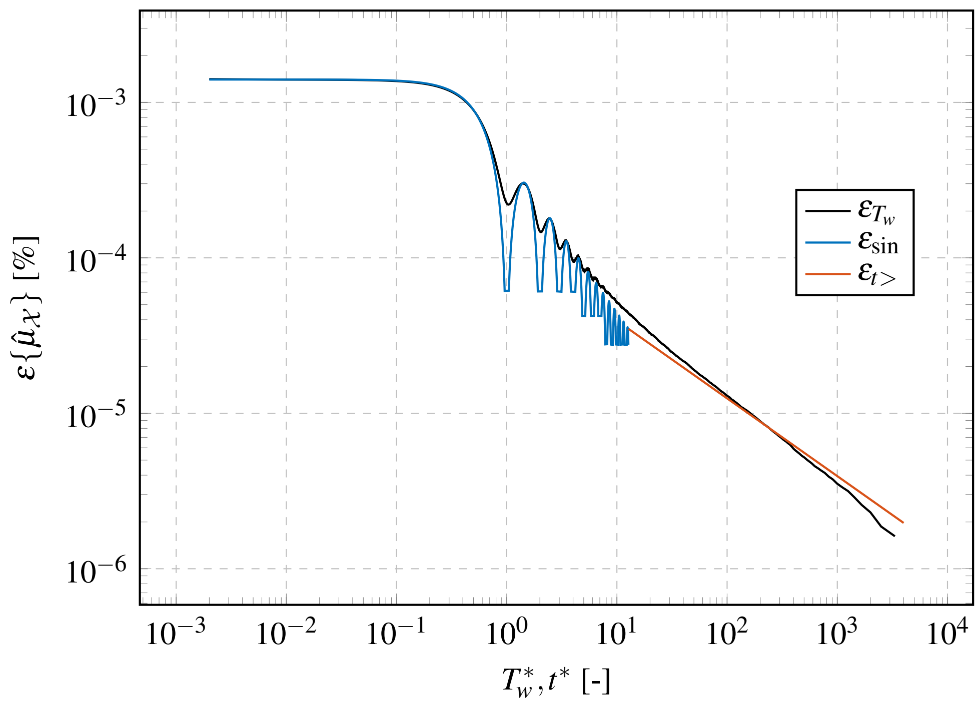

Figure 5.10: Time evolution of the statistical error on the mean estimate for the signal $\mathcal{X}(t)$ characterised by the couple of parameters $f_0 = 1$, standard deviation $\sigma_{\mathcal{X}} = 0.1$ and $\beta^* = 5$.

Eventually, the oscillations of the sine dependency can be omitted and the statistical error would be scaled by linear parts so that :

$$\varepsilon\{\hat{\mu}_X\}(t^*) = \begin{cases} \left(\frac{\sigma_{\mathcal{X}}}{\mu_{\mathcal{X}}}\right) \frac{1}{\pi t^*} & , t^* \in [-\infty\, ;\, t^*_{tr}] \\ \\ \left(\frac{\sigma_{\mathcal{X}}}{\mu_{\mathcal{X}}}\right) \frac{1}{\sqrt{2B_{t>} t^*}} & , t^* \in]t^*_{tr}\, ;\, +\infty[\end{cases} . \tag{5.46}$$

5.6 Prediction of the error

This section aims to make the bridge between the statistical error estimate and the new mathematical model introduced above. Instead of the simple curve fitting parameter B_{fit} considered in the Mockett's paper, there are now two parameters to estimate, namely f_0 and β. The first parameter directly influences the sine wave contribution while the second influences the broadband part. Moreover, a knowledge of both frequency spectrum parameters are required to estimate the transition point.

The first step is to check if the statistical error is composed of a periodic component. For that purpose, the time step is assumed small enough compared to the frequency f_0 to lead to a plateau (in a log-log plot) for a small t. In this specific case, the methodology introduced by Mockett to compute the curve fitting parameter as $\min[B(T_w)]$ can be used to estimate the frequency. Indeed, the first drop in the statistical error corresponds to a time $1/f_0$. Therefore, the time corresponding to the curve fitting parameter B_{fit} is thus $1/2f_0$ (see Figure 4.7). On the contrary, if the statistical error does not present any plateau, the simple white-noise scaling is applied and there is no need to estimate any parameter f_0.

The second step is to estimate the bandwidth parameter β. This can be made by estimating the slope for large T_w. If the first guess of the dependency validates the white noise scaling $1/\sqrt{t}$ then the transition point t_{tr} can be roughly estimated. After that, the estimation of the curve fitting parameter is made in the interval $[t_{tr}\,; T]$ and this provides an accurate value for β.

6 Adapted methodology and numerical implementation

In Chapter 4 has been introduced a methodology to estimate a posteriori the initial transient duration. The latter is used as a basis for the final, adapted methodology considered in the thesis. In particular, some drawbacks have been emphasized, namely : the incapability to handle periodic flows and some spurious oscillations when the window size increases. Chapter 5 therefore aimed to introduce a new mathematical model to overcome the first issue and provide a more general description of the fluid physics.

In this Chapter, the final methodology is summarized. In the present methodology, the statistical error is estimated by windows averaging coupled with analytical formulas to predict the further variation. This procedure requires optimal choices, the first section is therefore dedicated to this discussion. Finally, the initial methodology is extended to allow the estimation of the required simulation time t_e.

6.1 Statistical error computation

The window averaging methodology to estimate the statistical error has been described in section 4.2.1 and some limitations have been highlighted. Indeed, the estimate becomes poorly accurate as the window size tends to the signal length. Furthermore, this statistical error estimation directly influences the prediction by means of analytic formulas through the curve fitting parameter. Then, the product of both variances of the estimates which allows to highlight the initial transient is affected and can lead to a bad estimation of t_t due to spurious and unwanted oscillations.

6.1.1 Windows generation

Window generation is of a paramount importance since it influences directly the computational time and cost of the algorithm. Indeed, the more windows of different sizes are considered, the more the algorithm is time and resources consuming. It is therefore important to generate those in a way that minimizes the number of windows while keeping an acceptable overall accuracy. Let denote T_w the number of time steps per window, N_{T_w} the number of windows of

size T_w and N the number of data in the sample to be discretized.
The three quantities are obviously linked by

$$N_{T_w} = \frac{N}{T_w}. \tag{6.1}$$

However, due to the discretized nature of the sample, to one N_{T_w} can be associated several T_w since it is not necessary that N_{T_w} is a multiple of N. Therefore, a choice that can be made to limit the computational cost is to only consider a unique value of T_w and thus only consider the N_{T_w} associated. To illustrate, let consider the simple example where $N = 10$ and N_{T_w} is initially chose as $[1, \cdots, 10]$, then :

$$T_w = \frac{N}{N_{T_w}} = [10,\ 5,\ 3,\ 2,\ 2,\ 1,\ 1,\ 1,\ 1,\ 1]$$

Therefore, it has been chosen to only keep the unique value of T_w which decreases strongly the computational cost. Obviously, the inverse path could also have been considered i.e imposing the T_w and then compute the N_{T_w}. However, a particular attention should have been given to the T_w non-multiple of N.

In the top of Figure 6.1 is represented the evolution of the curve fitting parameter as a function of the simulation time for the experimental sectional lift coefficient introduced in Chapter 4. This parameter, denoted B_{all} is computed with the naive and consuming first guess for the windows generation. The bottom graph represents the evolution of the ratio between the curve fitting parameter computed with the uniqueness methodology of window generation and the naive one.

It can be seen that the robustness of the curve fitting estimators is almost the same. Indeed, both methodologies take the same time to converge to the same final value. Divergences only occur when even the parameter B_{all} has not converged yet. As soon as the latter has reached its asymptotic value, the ratio between both parameters is almost always equal to one. It drastically reduces the time of computation. For a naive window generation, the time of computation varies like N^2 while it varies like N with a negligible gradient for the uniqueness methodology. In particular, less than half a second is required[1] to estimate the statistical error for a sample of 3.5×10^4 points.

6.1.2 Statistics computation

The estimated error ε_{T_w} computation requires a huge number of windows and for every window, the descriptive statistics have to be calculated. Therefore, the operation can become very consuming in terms of computer resources and time, if an adequate strategy is not employed.

[1] Computation time estimated with a laptop MacBook Pro 2.7 GHz Intel Core i5.

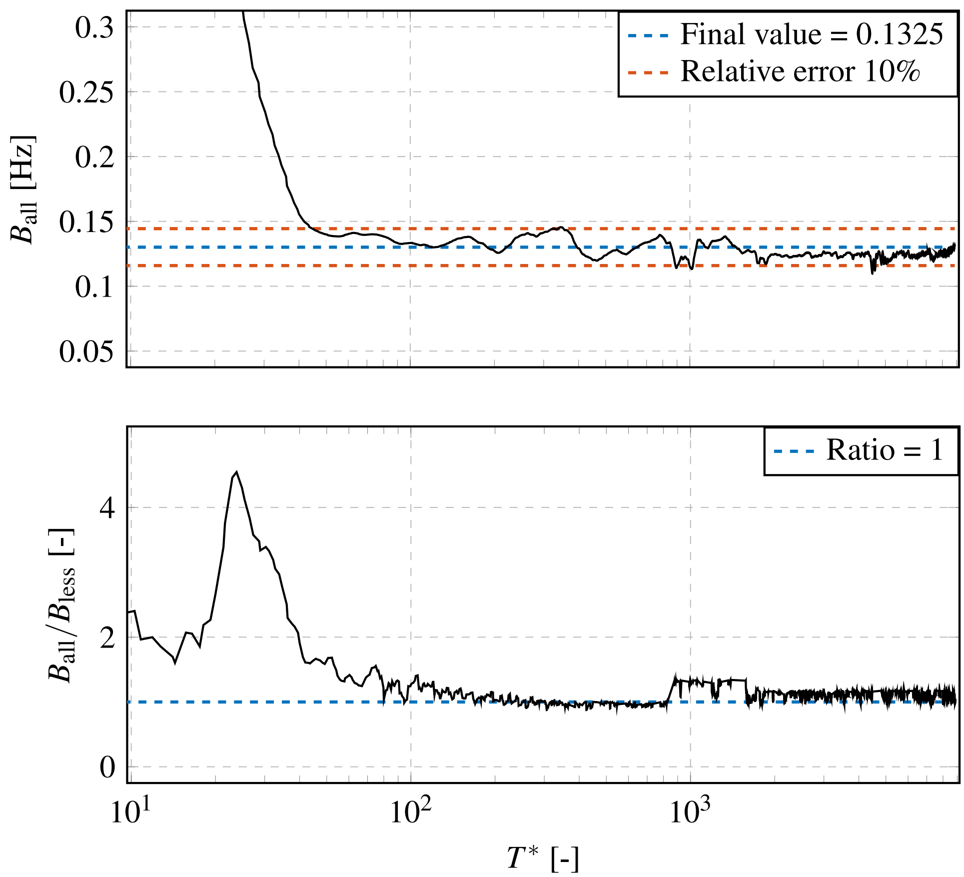

Figure 6.1: Running curve fitting parameter B_{all} computed with a naive windows generation (above) and running ratio between B_{all} and B_{less}, the curve fitting parameter computed with the uniqueness windows generation methodology. The time axis is non-dimensionalised as $R^* = T(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity. The data is truncated as $0 \leq T^* \leq 500$.

A foolish and dirty methodology would have been to loop in each window to compute the mean and the standard deviation. In the end, the sample has to be divided into N_w windows and N different window sizes T_w are required, this leads to an amount of windows where statistics have to be computed which is not acceptable.

The strategy used in the numerical implementation of the methodology is based on a cumulative sum of the input signal.

Mean computation

Let consider a cumulative sum function $cumsum_\mu(t) = \sum_{i=1}^{i=t} x(i)$. Assuming that a window T_w extends from a time $t = t_1$ to a time $t = t_2$, the sample mean value over a window $\hat{\mu}_{C_l}(\mathbf{D}_{T_w})$ thus corresponds to[2] :

$$\hat{\mu}_{C_l}(\mathbf{D}_{T_w}) = \frac{1}{t_2 - t_1 + 1} \sum_{i=t_1}^{i=t_2} x(i). \tag{6.2}$$

According to Chasles' rule :

$$\begin{aligned}
\sum_{i=1}^{i=t_2} x(i) &= \sum_{i=1}^{i=t_1-1} x(i) + \sum_{i=t_1}^{i=t_2} x(i) \\
\Leftrightarrow \hat{\mu}_{C_l}(\mathbf{D}_{T_w}) &= \frac{1}{t_2 - t_1 + 1} \left[\sum_{i=1}^{i=t_2} x(i) - \sum_{i=t_1}^{i=t_1-1} x(i) \right] \\
&= \frac{1}{t_2 - t_1 + 1} \left[cumsum_\mu(t_2) - cumsum_\mu(t_1 - 1) \right].
\end{aligned}$$

Thus, this strategy allows to convert the N_w loop on each window to a succession of arithmetic sums. In the particular case of $t_1 = 1$, the last term is zero and the mean of the whole sample is computed.

Standard deviation computation

Let consider a cumulative sum function $cumsum_\sigma(t) = \sum_{i=1}^{i=t} x^2(i)$. Assuming again that a window T_w extends from a time $t = t_1$ to a time $t = t_2$, the sample variance value over a window $\hat{\sigma}_{C_l}(\mathbf{D}_{T_w})$ thu corresponds to :

$$\hat{\sigma}^2_{C_l}(\mathbf{D}_{T_w}) = \frac{1}{t_2 - t_1} \sum_{i=t_1}^{i=t_2} \left[x(i) - \hat{\mu}_{C_l}(\mathbf{D}_{T_w}) \right]^2. \tag{6.3}$$

[2]The denominator $(t_2 - t_1 + 1)$ is due to the convention adopted that the very first sample point corresponds to $t = 1$ and thus an interval $t_2 - t_1 = N$ contains $N + 1$ sample points.

By expanding the square :

$$\begin{aligned}\hat{\sigma}^2_{C_l}(\mathbf{D}_{T_w}) &= \frac{1}{t_2-t_1}\left[\sum_{i=t_1}^{i=t_2} x^2(i) - 2\sum_{i=t_1}^{i=t_2} x(i)\hat{\mu}_{C_l}(\mathbf{D}_{T_w}) + \sum_{i=t_1}^{i=t_2}\hat{\mu}_{C_l}(\mathbf{D}_{T_w})\right] \\ &= \frac{1}{t_2-t_1}\left[\sum_{i=t_1}^{i=t_2} x^2(i) - 2(t_2-t_1+1)\hat{\mu}^2_{C_l}(\mathbf{D}_{T_w}) + (t_2-t_1+1)\hat{\mu}^2_{C_l}(\mathbf{D}_{T_w})\right].\end{aligned}$$

Finally, using the Chasles' rule to expand the first sum yields to the final expression :

$$\hat{\sigma}^2_{C_l}(\mathbf{D}_{T_w}) = \frac{1}{t_2-t_1}\left[cumsum_\sigma(t_2) - cumsum_\sigma(t_1-1) - (t_2-t_1+1)\hat{\mu}^2_{C_l}(\mathbf{D}_{T_w})\right]. \tag{6.4}$$

In the particular case of $t_1 = 1$, the variance of the whole sample is computed.

6.1.3 Moving average filter

In the windows averaging process, the resulting estimated statistical error is noisy and is therefore characterised by spurious oscillations. This numerical issue can be of a paramount importance since the shape of the statistical error directly influences the curve fitting parameter. Hence, the transient estimation, as well as the required simulation time, can become less accurate. Indeed, if a spurious maximum occurs in the statistical error, the curve fitting parameter is computed to be conservative and therefore adapts itself to this maximum. As a consequence, the predictive error i.e the fitting error, can be strongly overestimated or not, depending on the magnitude of this local maximum, the statistical error.

The strategy implemented to overcome this issue is to apply a moving average filter to the statistical error. In this way, the resulting estimation ε_{T_w} is free of any spurious oscillation and the curve-fitting parameter is adapted to the smoothed signal. For instance, the Figure 6.2 highlights the difference between the statistical error of the experimental lift coefficient smoothed and with the spurious oscillations.

6.2 Required simulation time

The methodology introduced in Chapter 4 can be extended to estimate the required simulation time to reach a predefined accuracy for the statistical error, say ε_{tol}.

Firstly, the initial transient is removed from the sample to only consider the assumed steady state (see Figure 6.3). Then, the statistical error is estimated as a function of the running time and to each t is computed a $B_{\text{fit}}(t)$. The procedure is continued until the series $\{B_{\text{fit}}(t)\}$ has converged to an asymptotic value, denoted B. Then, the required simulation times are

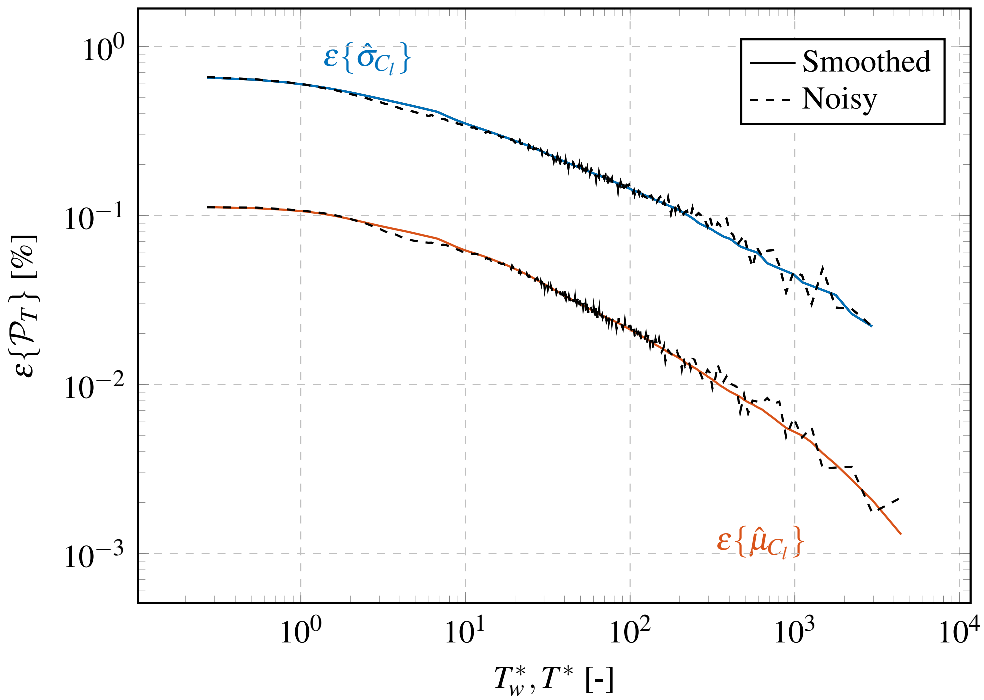

Figure 6.2: Comparison of estimated and smoothed statistical errors for the measured sectional lift coefficient (C_l) from [39]. The smoothed errors are in continuous lines and the noisy in dashed lines. The smoothed error is in orange for the mean estimator and in blue for the standard deviation. The time axis is non-dimensionalised as $t^* = t(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity.

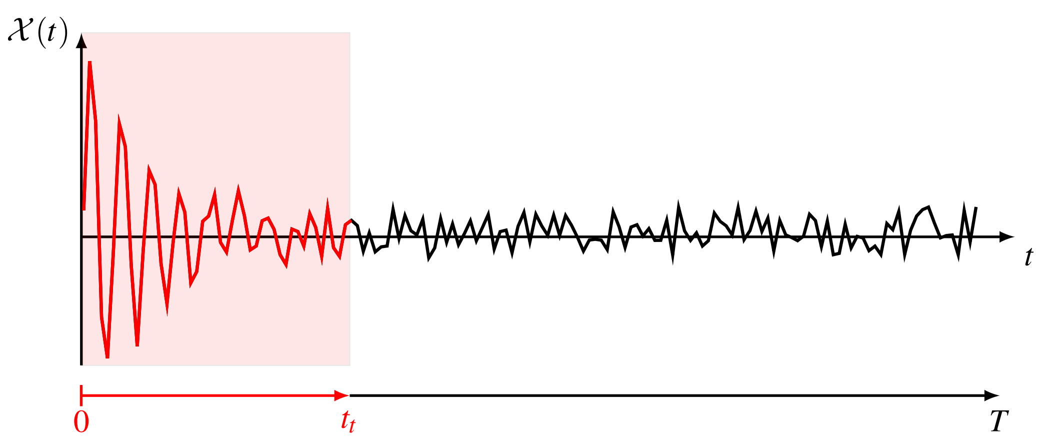

Figure 6.3: Generic time record $\mathcal{X}(t)$, of total length T with an initial transient duration t_t. The transient part of the sample is plotted in light red while the conserved part is plotted in black.

estimated following Equations (4.3) :

$$t_{e,\mu} - t_t = \frac{1}{2B\varepsilon_{\text{tol}}^2}\left(\frac{\hat{\sigma}_{\mathcal{X}}}{\hat{\mu}_{\mathcal{X}}}\right)^2 \qquad t_{e,\sigma} - t_t = \frac{1}{4B\varepsilon_{\text{tol}}^2}. \tag{6.5}$$

The final value considered for t_e is chosen as $\max(t_{e,\mu}, t_{e,\sigma})$. Eventually, a new curve fitting parameter can be estimated further during the simulation $(t < t_e)$ to gain some accuracy.

From this pre-defined accuracy, confidence intervals can thus be build as introduced in equations (4.7) and (4.8). In that way, the time-average quantity of interest can be expressed as the final averaged value (in t_e) with a certain residual statistical error.

6.3 Summary

A flowchart of the complete methodology is provided in Figure 6.4. In the end, the latter is able to estimate the initial transient duration as well the required simulation time with the time signal as a sole input. Moreover, the methodology is adapted to be used on the fly in order to monitor an unsteady simulation.

Chronologically, a succession of windows averaging and sample shortening is realized in order to estimate the initial transient. Once an asymptotic value is reached for the transient duration, the data sampling and time statistics is triggered to only take into account the assumed steady-state solution. On this steady solution a succession of windows averaging is applied again to enable the prediction of the statistical error for a longer sample length than the current time step. Hence, an estimation of the simulation time required to reach a pre-defined statistical accuracy can be made and the simulation is stopped as soon as this accuracy is reached.

Due to the high repetition of windows averaging and shortening of the sample, the algorithm is a bit more costly in terms of time than the simple statistical error estimation. In particular, for a sample of 2×10^4 points, approximately 500 seconds are required to estimate the transient duration during the whole simulation i.e for a sample length T. Considering that the methodology is stopped as soon as the transient estimation has converged, it should obviously take less time. Hence, this computation time, compared to the duration of a CFD numerical simulation (which lasts several hours) is judged to be acceptable.

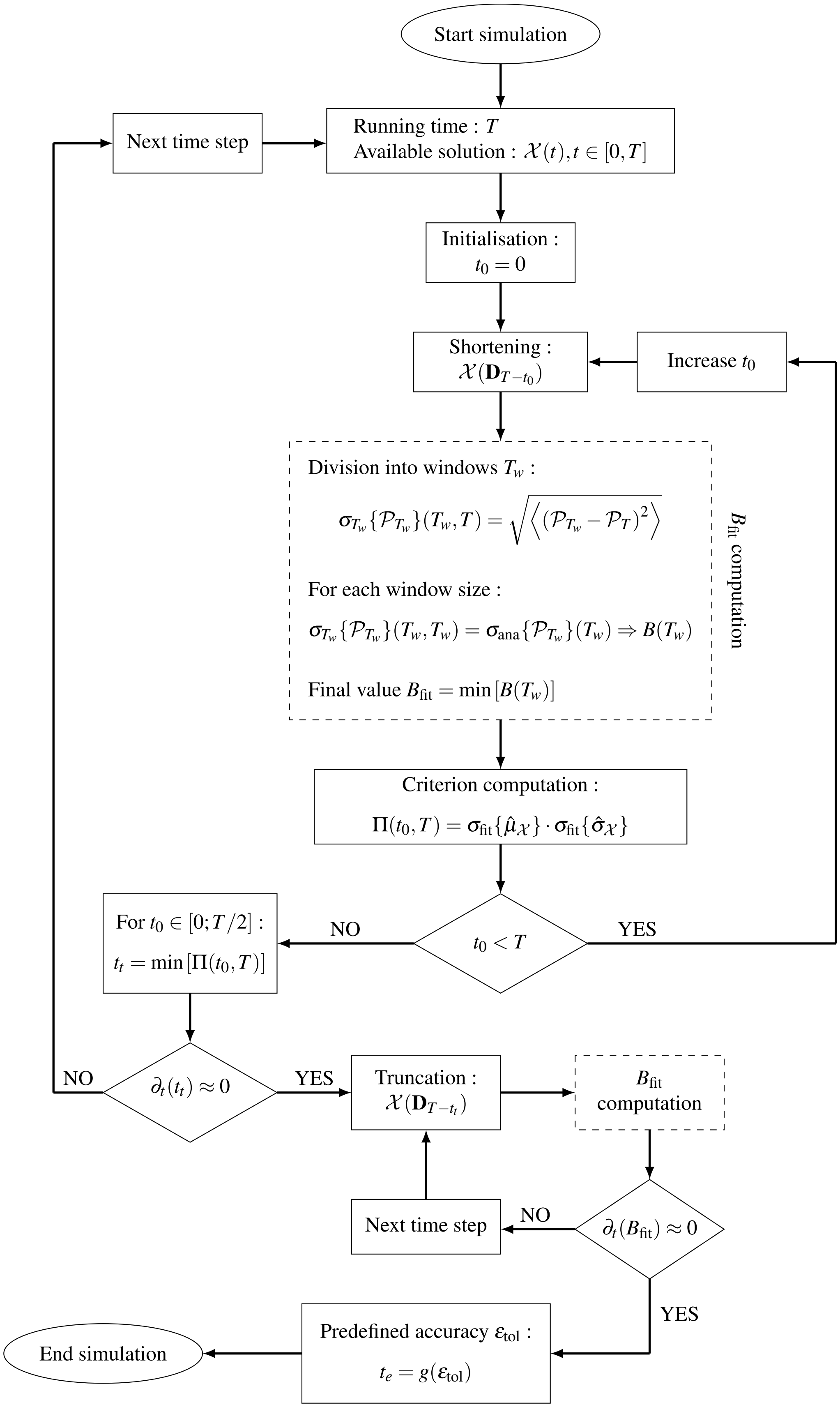

Figure 6.4: Flowchart of the adapted methodology.

7 ANSYS Fluent User-Defined function implementation

In Chapter 6 has been introduced the complete methodology to estimate on the fly the initial transient duration and the required simulation time to reach a pre-defined statistical accuracy. The final purpose of the thesis is to implement this algorithm as a User-Defined Function (UDF) in the commercial software ANSYS Fluent. The main goal is therefore to monitor an unsteady simulation.

A UDF is a function that the user program and that can be dynamically loaded with the Fluent solver. The access to data solver is made using predefined macro that are supplied by Fluent [16] and contained in the derive `udf.h`. In the present Chapter are therefore presented the main choices for the translation of the algorithm in a UDF. In particular, the main steps in the flowchart 6.4 are described and justified in terms of the provided macro.

Finally, this Chapter is not wanted to be a description line per line of the implementation but only to describe in the broad lines the architecture and the fundamental choices.

7.1 General considerations

Two main types of UDF exist in Fluent, namely the interpreted or the compiled UDF. In the present case, a compiled UDF has been preferred since it executes faster than interpreted UDFs and it is not restricted in the use of the `C` programming language.

Furthermore, each UDF has to be defined by means of the predefined macros provided by Fluent. These macros, denoted by `DEFINE_NAME` are defined in the derive `udf.h`. The algorithm is based on statistical considerations for the lift coefficient in unsteady simulations. Therefore, the general solver macro `DEFINE_EXCECUTE_AT_END` has been chosen. The latter is general purpose macro that is executed at the end of an iteration or at the end of a time step depending on if the simulation is a steady state run of a transient run. The algorithm is thus run at each end of time step.

That is, the implementation of the UDF can be schematically divided into three fundamentals steps :

i. Access the solver data;

ii. The transient detection algorithm ;

iii. The UDF output transmission to Fluent Inc.

Indeed, the algorithm requires to have access to the time history of the lift coefficient. After that, the statistical errors are computed to estimate the initial transient. Once the transient duration has been clearly identified, the data sampling and the time averaging are triggered. Finally, the simulation is stopped when a pre-defined accuracy is reached. These three steps are therefore investigated in the three following sections.

7.2 Access the solver data

Access the time history of the lift coefficient in a User-Defined Function in ANSYS can be a hard task. Indeed, the `C` variables in a UDF are volatile i.e they are automatically freed after the UDF has been called. In other words, they cannot be stored in memory during the whole simulation. Therefore, artifices have to be used to store the time trace of the lift coefficient.

The lift coefficient computation is made in fluent through a *report definition* whose output is a text file with the current simulation time, time step and value of the lift. A predifined macro is implemented in fluent to access the data from a report : `Get_Report_Definition_Value`. The latter accesses the last calculated value i.e the value of the current time step. Therefore, one way to have access to the time history of the lift is to write and read the value from an external file. However, this technique is limited by the rate of I/O and can be really time consuming.

Another technique is to use User-Defined Memory (UDM) Storage which allows to store values computed by the UDF in memory to be used later by another UDF or for post-processing. However, normally, UDM are not meant for that purpose, they usually store field variables. Hence, storing the lift coefficient in a UDM is not the most efficient way to proceed since it requires a loop on all threads and faces of the fluid domain.

Despite those drawbacks, user-defined memories have been considered in the implementation and further improvements should then be carried out to obtain a more efficient UDF. In particular, two UDM have been considered to store the cumulative sum of the lift coefficient and the squared cumulative sum of the lift coefficient as justified in section 6.1.2.

7.3 Transient detection algorithm

The transient detection algorithm aims to trigger the data sampling and the time statistics. However, to assess its convergence, the transient estimation has to be stored in memory during the simulation. Therefore, a UDM is dedicated to the transient estimation. The algorithm is hence performed until convergence of t_t i.e a typical while condition. After that, the data sampling and statistics are automatically triggered by means of the `execute command` panel of the GUI or by text commands.

7.4 UDF output transmission

The last major step in the methodology is to estimate the required simulation time. Hence, once the sampling is triggered, the statistical error is estimated on the fly until the curve fitting parameter B_{fit} has converged. To make the simulation stop when a certain accuracy is reached, the convergence option of Fluent[1] has been used. For that purpose, a boolean variable is defined and the convergence is imposed to the solver when the boolean variable becomes true. Hence, this variable is set to be true as soon as the simulation time becomes equal or greater than the estimated required time t_e.

[1] Available for ANSYS R18 and more recent.

Part III

Discussions and Validations

Outline of the third part

Progress is made by trial and failure; the failures are generally a hundred times more numerous than the successes; yet they are usually left unchronicled.

William Ramsay

The final part of the manuscript is dedicated to the results provided by the implementation of the methodology introduced in the second part. The main outcomes of the methodology are respectively the detection and the duration estimation of an initial transient as well as an estimation of the required simulation time to obtain relevant statistics. Therefore, the analysis and discussions of the results are mainly focused on these two quantities.

The Chapter 8 gathers the different test cases investigated during the thesis. A particular attention has been devoted to bluff body flows. The computational domains used for numerical simulations are described as well as main numerical parameters.

The final Chapter aims to expose the different results for the transient duration and the simulation time estimations. These results are discussed and analysed to possibly highlight further improvements of the methodology. Moreover, the different choices made in the methodology are validated and justified. Finally, the results obtained by the User-Defined Function are sensitively the same as those obtained with the `C` and `Matlab` implementations. Hence, only the results obtained by an a posteriori computation are presented for the sake of visibility.

8 Description of test cases

The first chapter of the discussions and validations aims to introduce the different test cases employed for the further investigations. The choice of a dedicated chapter to gather the cases is considered advantageous compared to a mix of descriptions and results since the validations are investigated by themes.Hence, results for different cases can be cited in multiple sections.

For each test case, an overview of the flow type as well as a description of the numerical parameters of interest are given. In particular, the turbulence modelling, the numerical settings and the grids or setup issues are provided and justified.

Finally, it has been chosen to investigate particularly the bluff body flows. They are characterised by a strong separated flow and an unsteady wake. Even if highly unsteady, the wake downstream of the body is characterised by a periodic and asymmetric vortex shedding known as the von Karman vortex street. It is therefore expected that the lift force has a fluctuating behaviour. The vortex street frequency f is commonly described by the non-dimensional Strouhal number

$$\mathrm{St} = \frac{fL}{|U_\infty|}, \tag{8.1}$$

where L is a characteristic length of the geometry e.g the diameter for a cylinder of rather $c \sin \alpha$ for an airfoil at angle-of-attack α.

8.1 General considerations

In the formula one context, modelling turbulent flows as accurately as possible is of a paramount importance. The unsteady pattern of the flow in Formula 1 racing therefore leads to simulations that are very computational resources consuming. The resolution of numerous turbulent scales needs small time steps. A wide range of turbulence models exist and it is increasingly clear that the model should be chosen in line with the classes of flows that are investigated. Although today's CFD simulations are mainly performed with (Usteady) Reynolds-Averaged Navier-Stokes (U)RANS turbulence model, it is not suitable for the present application. Indeed, it is shown by experience that these models do no provide enough information of the frequency content of the flow, even with a very small grid resolution. It is actually a direct consequence of the RANS averaging procedure that acts as a "filter of turbulence" in the velocity field. In the specific case of bluff bodies flow, the latter is globally unstable and is characterised by the apparition of a «new »turbulence downstream of the body. RANS models usually provide poor results for that typical flows as they only highlight the main frequency contribution to the flow i.e the pseudo-periodic vortex shedding downstream of the body. However, experiences (see Figure 5.1) rather show a broadband turbulence spectrum with an additional periodic contribution (represented by a peak in a frequency spectrum). Therefore, it can be concluded that RANS formulation is not an optimal choice to solve turbulence in the Formula 1 context.

The need for additional knowledge on the flow frequency content that RANS model cannot provides motivates the use of Scale-Resolving Simulation (SRS). In those models, only certain scales of turbulence are resolved e.g the Large Eddy Simulation (LES) model resolves the largest turbulence scales. This model is particularly suitable for free shear flows as the resolved scales are of the order of the shear layer thickness. However, for wall boundary layers flows the turbulence length scale decreases strongly near the wall and becomes small compared to the boundary layer thickness. This effect is increasingly noticeable as the Reynolds number increases. Therefore, LES model can become outside of the possible in an academic context (and even for most of the computer power available in industry). Hence, the choice of turbulence model that is considered in the following is an hybrid model which combines large eddies resolution far from the wall and RANS in the wall boundary layer.

ANSYS fluent provides different hybrid models as for instance Detached Eddy Simulation (DES) or Scale-Adaptive Simulation (SAS). From those two other more specific models have been developed such as the Delayed Detached Eddy Simulation (DDES), the Shielded Detached Eddy Simulation (SDES). In the following, DDES have been considered.

8.1.1 Meshing requirements

The choice of an hybrid model of turbulence directly influences the grid refinement. Indeed, the attached flow regions are solved in RANS while detached regions away from walls are resolved in LES. Therefore, the grid should be of RANS quality inside the boundary layer and of LES quality far from walls.

Different guidelines for the RANS grid have been considered. First, a value of $y^+ \approx 1$ at the wall is required which directly influences the first cell thickness. Then, the whole boundary layer is covered by a high grid resolution : typically 15 structured cells in the boundary.

Regarding the LES grid zones, a first approximation of the mesh resolution can be obtained under some assumptions. For a bluff body of characteristic length D, the largest relevant scales are assumed to be of the same order as the this instability zone. Experiences provide the following guideline for the maximum cell size [25] :

$$\Delta_{\max} \leq 0.05D \tag{8.2}$$

which roughly corresponds to more than 20 cells per characteristic length D. Moreover, cells with high aspect ratio are avoided as possible.

8.2 Case 1 : NACA0012 airfoil in deep stall

The first test case is the three-dimensional flow around a symmetric NACA0012 airfoil at angle of attack $\alpha = 60°$ and a Reynolds number $\mathrm{Re}_c = U_\infty c/\nu = 2.7 \times 10^5$ with a chord of 1m and air standard condition. The NACA airfoil is symmetric and characterised by a 12% thickness-to-chord ratio (see Figure 8.1). Several studies have already investigated this test case which makes it easy to compare and verify (see Spalart et al. [35] or Mockett [26]).

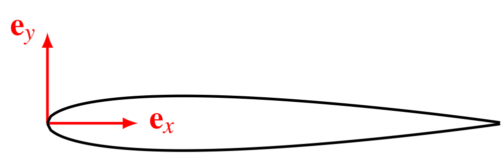

Figure 8.1: Schematic representation the geometry and coordinate system for NACA0012 test case. The vectors $\mathbf{e}_x$ and $\mathbf{e}_y$ are orthonormal vectors.

The computational grid is a C-topology with the origin coordinate system at the profile nose. The far field domain boundary is located at a distance $10c$ upstream and $20c$ downstream the airfoil and expanted of $1c$ in the spanwise direction. The NACA is treated as a no-slip wall condition. The grid is fine enough in the boundary layer to ensure the condition $y^+ < 1$ during the whole simulation (see Figure 8.2). The time step is fixed to $\Delta t = 0.0025c/U_\infty$ as preconised by [26]. It gives a Courant number of CFL ~ 0.7 inside de boundary layer and CFL ~ 1 in the major part of the unsteady wake. Finally, the simulation has been performed during $T^* = TU_\infty/c \approx 600$ i.e around $N = 27000$ time steps and hence represents a large statistic sample.

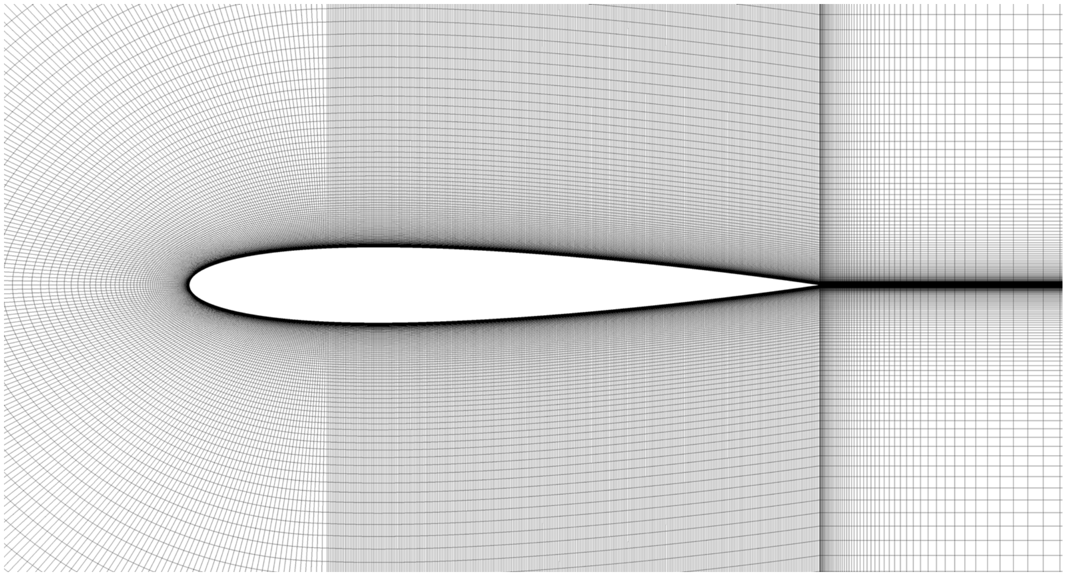

Figure 8.2: Computation grid for the NACA 0012 test case. A zoom is realised near the airfoil.

8.3 Case 2 : Flow past a triangular cylinder

The second test case is the two-dimensional flow around a tryangleand a Reynolds number $\mathrm{Re}_c = U_\infty c/\nu = 2 \times 10^6$ with a edge of 1m and air standard condition. The triangle is equilateral.

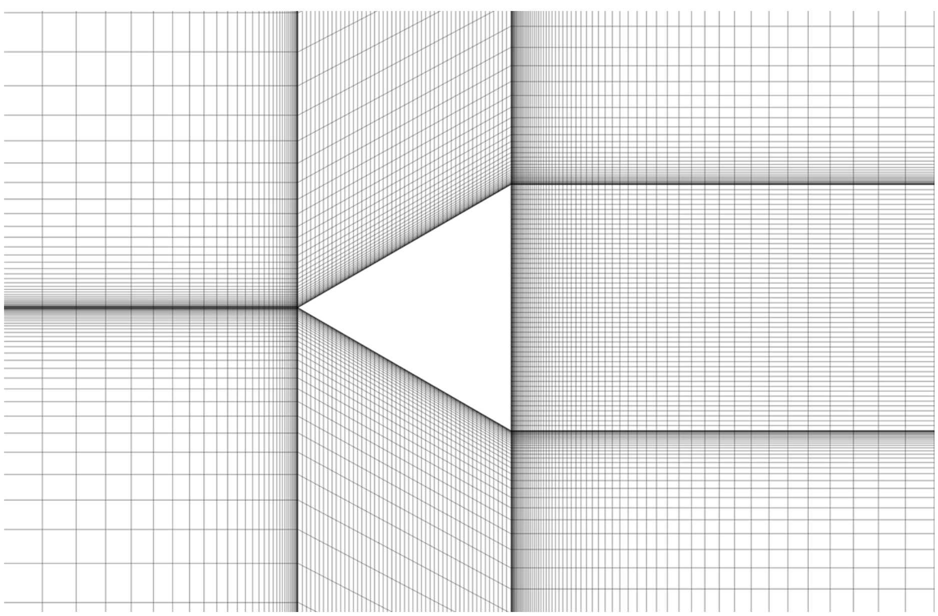

Figure 8.3: Computation grid for the triangular cylinder test case. A zoom is realised near the cylinder.

9 Demonstrations and analysis

In Chapter 6, the methodology to estimate the initial transient and the required simulation time has been completely described. The different choices of implementation have been summarized. Finally, in Chapter 8 have been collated the different test cases considered to analyse the robustness and the acuity of the algorithm.

In the present chapter, each individual section is devoted to a particular topic and is based on multiple test cases. This is why a dedicated chapter for the test cases description has been favoured. The results are wanted to be presented in chronological order i.e in the same order as in a numerical simulation. Hence, the first section is dedicated to the transient detection while the second one is dedicated to the required simulation time. Each section is then divided into subsections that investigate a different topic and introduce discussions.

9.1 Initial transient estimation

9.1.1 General results

The different test cases allow to investigate different fluid flows. The NACA0012 test case corresponds to highly separated flow in the downstream with an unsteady wake. Therefore, a superposition of a nominally periodic shedding of large scale and of a finer random turbulent fluctuations is expected. There is an energetic contribution for lowest frequencies as well as an energetic contribution in the neighbourhood of the shedding frequency. As far as the triangular cylinder is concerned, even if it is not the most relevant test case in the F1 framework, it enables to investigate the capabilities of the methodology for highly periodic flows which require a certain time to develop and reach a periodic oscillation almost constant in amplitude.

The results provided by the methodology are in line with the visual expectations for every test cases (above graphs in Figures 9.1, 9.2). Regarding the deep stall airfoil, an initial transient around $t_t^* \approx 78.5$ is computed. It approximately corresponds to the duration required for the fluid flow to become independent of the initial condition (uniform free stream velocity). Therefore, it gives confidence in the validity of the implementation. For the triangular test case, an initial transient of $t_t^* \approx 42$ is found. This duration corresponds to the time needed for the flow to settle down and for the periodic vortex shedding to reach an almost constant amplitude. Therefore, it is satisfactory to see that the capabilities of methodology can be extended to periodic flows even with a white noise scaling.

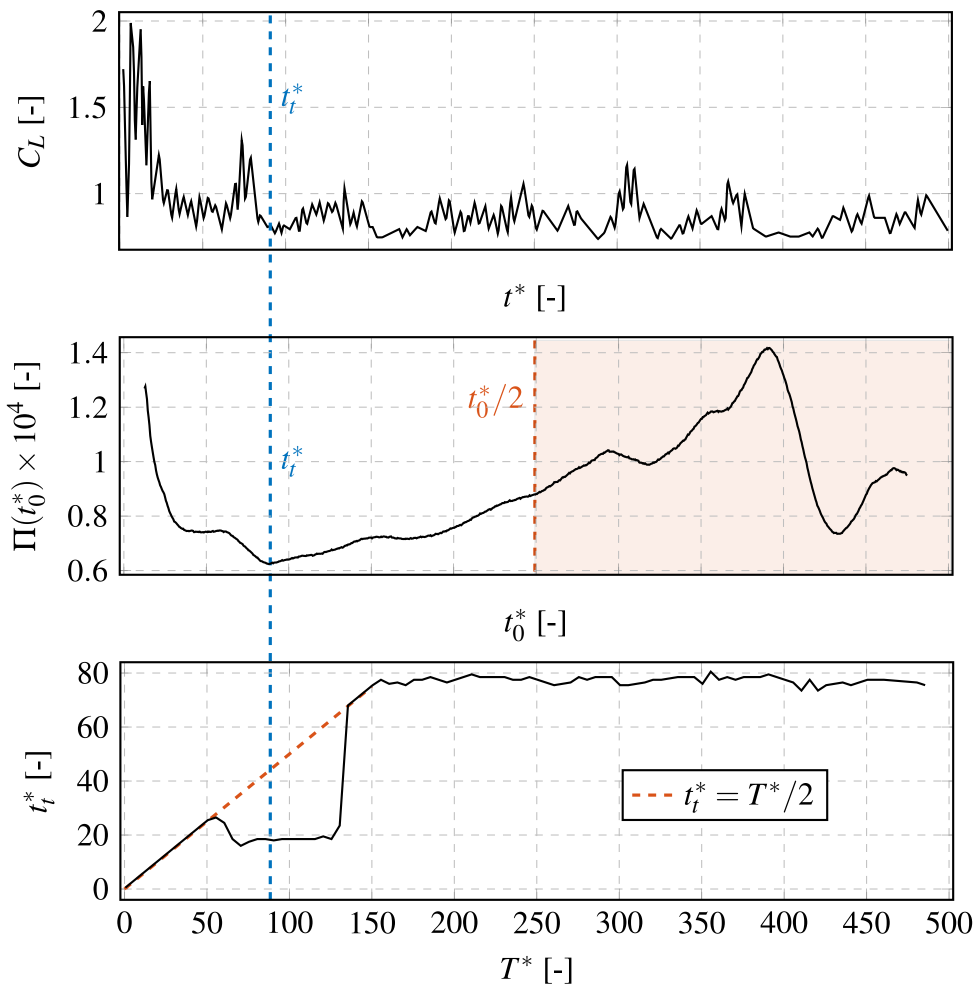

Figure 9.1: Time trace of the integrated sectional lift coefficient C_L (above) as a function of the non-dimensionalised time $t^* = t(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity. Variation of the product $\Pi(t_0^*)$ highlighting the minimum at t_t^* (mid). Prediction of the transient estimate as a function of the sample length (below). The data set is truncated as $0 \le t^* \le 500$.

The middle graph in Figure 9.1 shows some spurious discrepancies in the product $\Pi(t_0^*)$ for t_0^* increasing i.e when the signal is highly shortened. Indeed, the shorter the sample, the less windows are available for the estimation of the statistical error. Hence, the results become less accurate and can lead to totally wrong values for a t_0^* tending to T^*. It justifies therefore the truncation of the interval in which t_t^* is computed i.e $t_0^* \in [0; T^*/2]$. However, a smooth and continuous increasing shape after the transient duration can be observed for the triangular cylinder (see mid Figure 9.2). This is probably due to the «smoothness»of the signal itself. Indeed, for a deep stall airfoil, the lift coefficient has a strong noisy pattern due to the broad-band contribution in the unsteady wake. As far as the triangular cylinder is concerned, the lift coefficient is strongly periodic and weakly noisy. This double contribution in the NACA lift coefficient can therefore influence the curve fitting parameter and therefore the product. Hence, in the framework of the formula 1 with strongly noisy signals, the truncation with an upper limit $T^*/2$ is justified and even more : it is recommended.

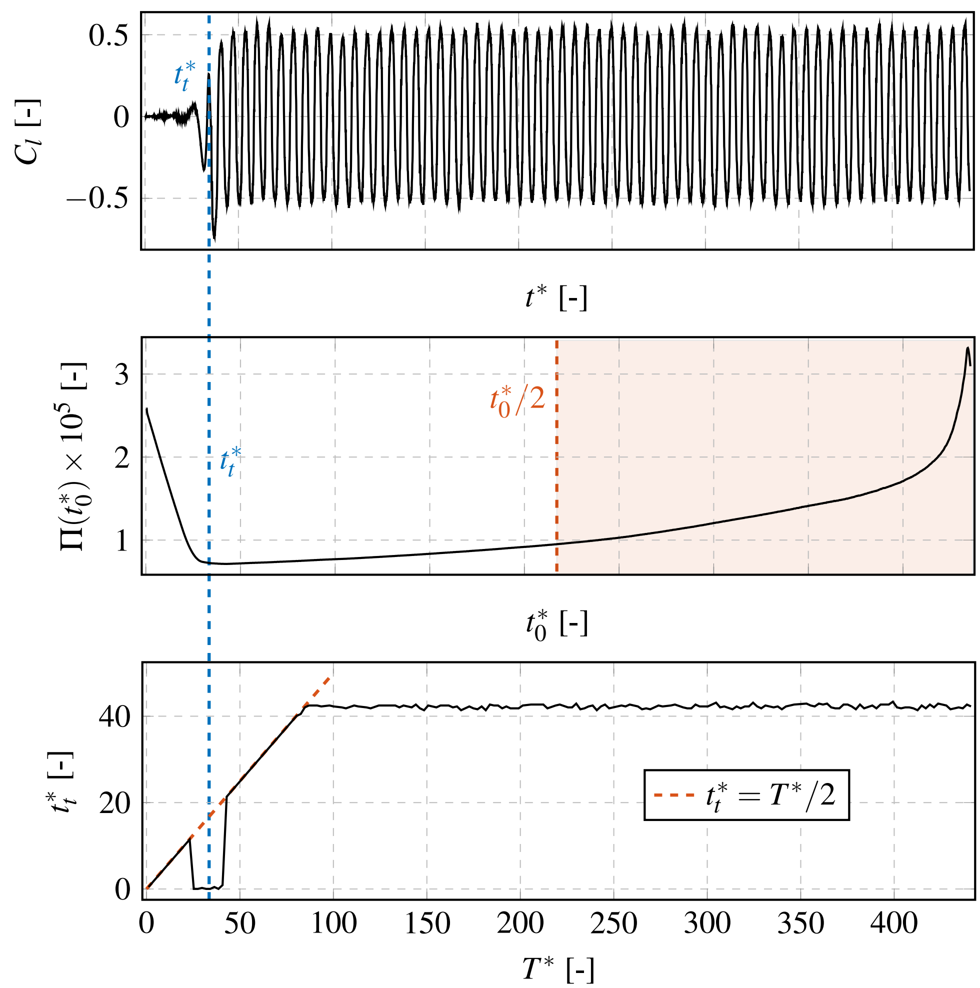

Figure 9.2: Time trace of the sectional lift coefficient C_l (above) as a function of the non-dimensionalised time $t^* = t(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity. Variation of the product $\Pi(t_0^*)$ highlighting the minimum at t_t^* (mid). Prediction of the transient estimate as a function of the sample length (below).

An insight of the robustness of the methodology can be obtained by the time evolution of the estimated transient duration (below graphs in Figure 9.1 and 9.2). The shape of the curve $t_t^*(T^*)$ is the same for both test cases. The transient duration initially varies linearly with the simulation time with a slope corresponding to the upper limit of the truncation interval, namely $t_t^* = T^*/2$. It was expected since the initial transient corresponds to the minimum of $\Pi(t_0^*)$. After that, the estimation drops to a lower value and then remains almost constant during several time units. This first plateau extends from $T^* \approx 60$ to $T^* \approx 125$ for the airfoil and from $T^* \approx 27$ to $T^* \approx 45$ for the cylinder. It is interesting to notice that the plateau always contains the final estimation for the initial transient. For the sample length values in the neighbourhood of the first plateau, the signal is almost totally composed of the initial transient. Hence, the methodology cannot distinguish the transient pattern to the following steady state signal. This also means that the product Π is a decreasing function before the lower limit of the first plateau. Then, the product is an increasing function and reaches a local maximum since its minimum corresponds to a lower value than $T^*/2$.

9.1.2 Additional constraint on the initial transient

The first plateau that occurs in the transient duration estimation as a function of the simulation time can be a dramatic issue since it yields to a highly underestimated value of the initial transient. Therefore, an additional constraint or information is required to assess the numerical convergence to the actual asymptotic value of t_t. The first point of the following has been implemented in the UDF while the two last are let as a further improvement.

Lower limit for the initial transient

In unsteady simulations, the flow requires some time to settle into a statistically stationary state. Indeed, the resolved turbulence must first develop itself and then be transported through the fluid domain. Therefore, the strategy adopted to avoid a convergence toward an underestimated value as in Figure 9.1 and 9.2 is based on physical and dimensional considerations. However, the initial transient is highly dependent of the flowfield and this strategy must be considered as a rule of thumb and not as a general guideline.

A first guess of the initial transient t_t can be obtained by estimation of the throughflow time, denoted t_{TF}. It corresponds to the required time for the mean flow to pass through the fluid domain. Therefore, by dimensional analysis, if L_{TF} is the length of the domain and U_∞ the upstream velocity it yields :

$$t_{LF} = \frac{L_{TF}}{U_\infty}. \tag{9.1}$$

In practice, the transient duration (even if highly case specific) corresponds to a rough estimate of 3-5 throughflow times [25]. Hence, the initial transient should be at least greater than t_{LF}.

For instance, regarding the high angle-of-attack airfoil simulation introduced in the previous chapter, a fluid domain of length $L_{TF} = 30c$ is used with c the chord and a free stream velocity U_∞. In terms of the non-dimensionalized time it yields :

$$\frac{ct^*_{TF}}{U_\infty} = \frac{30c}{U_\infty} \quad \rightarrow \quad t^*_{TF} = 30. \tag{9.2}$$

Therefore, by means of this first guess the first plateau at $t_t^* = 20$ in the curve $t_t^*(T^*)$ is judged to be unreliable and the only remaining acceptable converged value of the transient duration

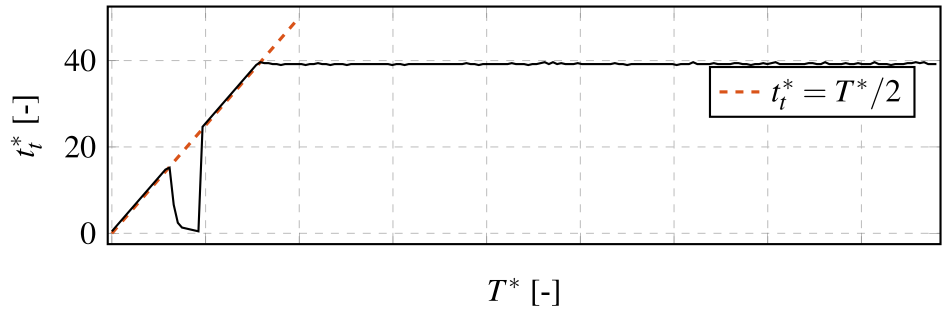

Figure 9.3: Time trace of the initial transient estimation for the triangular cylinder. Every time variable is non-dimensionalised as $T^* = T(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity.

is the actual one $t_t^* = 78.5$.

The same rule can be applied for the triangular cylinder which gives $t_{TF}^* = 20$. Hence, for every test cases, this rule of thumb provides satisfactory results. An additional advantage that can be obtained with this strategy is to reduce the number of numerical operations needed to estimate t_t. Indeed, if the rule of thumb is assumed to be verified, the operations to find $t_t(t)$ for $t < t_{LF}$ becomes not relevant and not necessary.

However, there is no certitude of this real dependence between the transient duration and the fluid domain. In particular, a very short domain does not insure a very short transient duration. Additional tests are then required as well as additional information about the fluid flow.

Additional aerodynamic quantity

An additional constraint has to be imposed for the transient duration to be able to catch numerically the convergence toward the actual asymptotic value. In the present methodology, the transient duration is estimated by means of statistical considerations for the lift coefficient. In particular, the transient duration corresponds to the minimum in the product of the statistical errors for both mean and standard deviation estimates. However, there is no certitude for this quantity to be the most relevant. After investigations, it appeared that the drag coefficient does not provide any additional information (see Figure 9.3).

Interesting quantities could also be those that would be zero for RANS simulation e.g spanwise forces since they are influenced by a scale resolution. Moreover taking into account the symmetry of the flow, they should oscillate around zero.

Continuously transient estimation

As soon as the transient estimate is assumed to have converged, the data sampling and the time statistics are triggered. Therefore, another option can be to continuously estimate the transient duration and allow a reset of the statistics if the transient estimate appears not to be the final one. Practically, the data sampling would be triggered as soon as the simulation reaches the first plateau. Once the latter reaches the second, and definitive one, the statistics are reset and then resume starting from 0. Even if this solution can be more costly in terms of time a priori, it allows to estimate the actual transient duration without any rule of thumb or additional aerodynamic quantity.

9.1.3 Time needed to estimate t_t

The methodology requires a certain time to estimate the transient duration, typically around $2t_t$. This is a direct consequence of the truncation of the shortening variable t_0 in the interval $[0; 0.5T]$. Therefore, a legitimate question would be to know if this truncation is highly conservative or on the contrary, highly necessary.

In Figure 9.4 are represented the initial transient estimation of a triangular cylinder for different upper limits of the truncated interval, namely $0.5T$, $0.7T$ and $0.9T$. The first linear variation of t_t is clearly seen for every truncations and the slopes correspond to every upper limit of the interval considered. While $0.7T$ and $0.5T$ curves are maximum at the transient duration, $0.9T$ reaches a higher value before converging to the asymptotic t_t. In the end, both curves $0.7T$ and $0.9T$ reach the asymptotic value almost at the same time ($\approx 1.3t_t^*$).

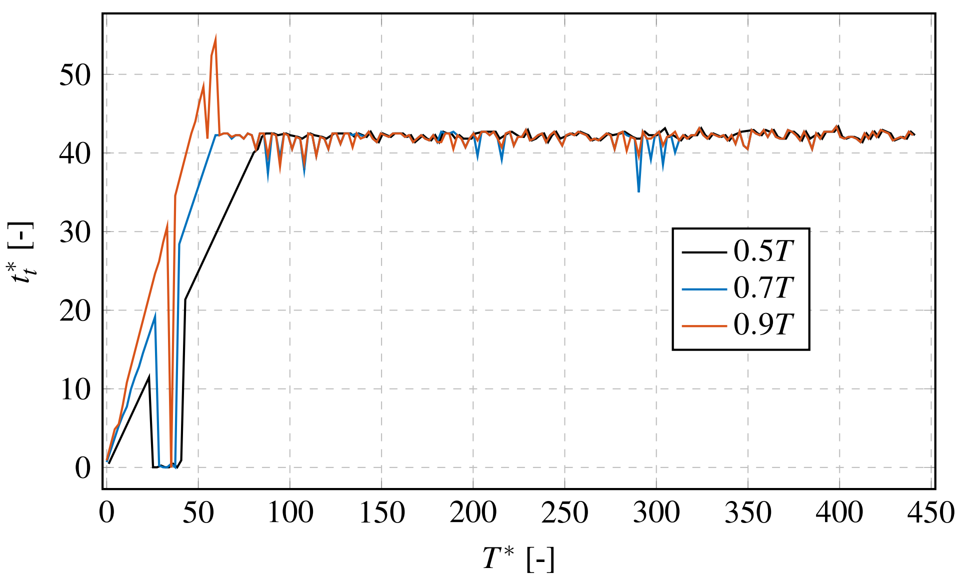

Figure 9.4: Time traces of the transient duration estimates of the triangular cylinder test case for different upper limits of the truncated interval for t_0^*.

Regarding the smoothness of the time traces, as soon as the latter have converged, additional oscillations are seen for larger upper limits as expected. Indeed, the larger the upper limit the lower the number of available windows to estimate the transient duration for large t_0.

More interestingly, the width of the first plateau decreases with the upper limit increasing. Hence, an evaluation of t_t for different upper limits at a same T^* could be a solution to the first plateau issue. Indeed, it can be assumed that if a plateau is noticed numerically for a low upper limit, then it should be noticed by a upper limit to confirm it. Regarding Figure 9.4, both curves $0.5T$ and $0.7T$ experience a plateau for $T^* \approx 30$ but not $0.9T$. Therefore, the first plateau could be treated as a numerical issue and not as an actual convergence.

9.1.4 Importance of the moving average filter

As introduced in Chapter 6, a moving average filter is applied to the statistical error to overcome some numerical oscillations due to the windows averaging. This choice appeared to be fundamental in terms of robustness and accuracy. The smoothness of the statistical error had led to a better curve fitting parameter estimation and therefore a more accurate and robust estimation of the initial transient.

To highlight these improvements, the evolution of the initial transient and the product Π are represented with and without the filter in Figure 9.5. It can be seen that even if the asymptotic value for t_t is almost the same between the filtered and the non-filtered transient estimation, some important oscillations occur when the statistical error is not filtered. This is a direct consequence of the fact that the product Π is greatly noisy (above graph Figure 9.5). That is, the minimum corresponding to the initial transient can be affected if a spurious oscillation is interpreted as this minimum. The same phenomenon can be highlighted for the other test cases.

These oscillations in the product may be due to the uniqueness windows generation introduced in Chapter 6 which reduces the number of windows. However, it has been judged advantageous in terms of computation time to favour this method and then apply a moving average filter. Indeed, the number of windows influences a lot more the computation time than the single loop required to filter the data. In other words, it is privileged to consider less information (less windows) which yields less accuracy and then filter the signal. To quantify and reinforce this choice, after numerous tests[1], it appeared that the windows average process represents 70% of the the total computation time while the filtering only 4%.

[1] The different tests have been performed with the tool `Run and Time` implemented in `Matlab`. It allows to quantify the time spent in each line of code and therefore highlight the most consuming part of the algorithm.

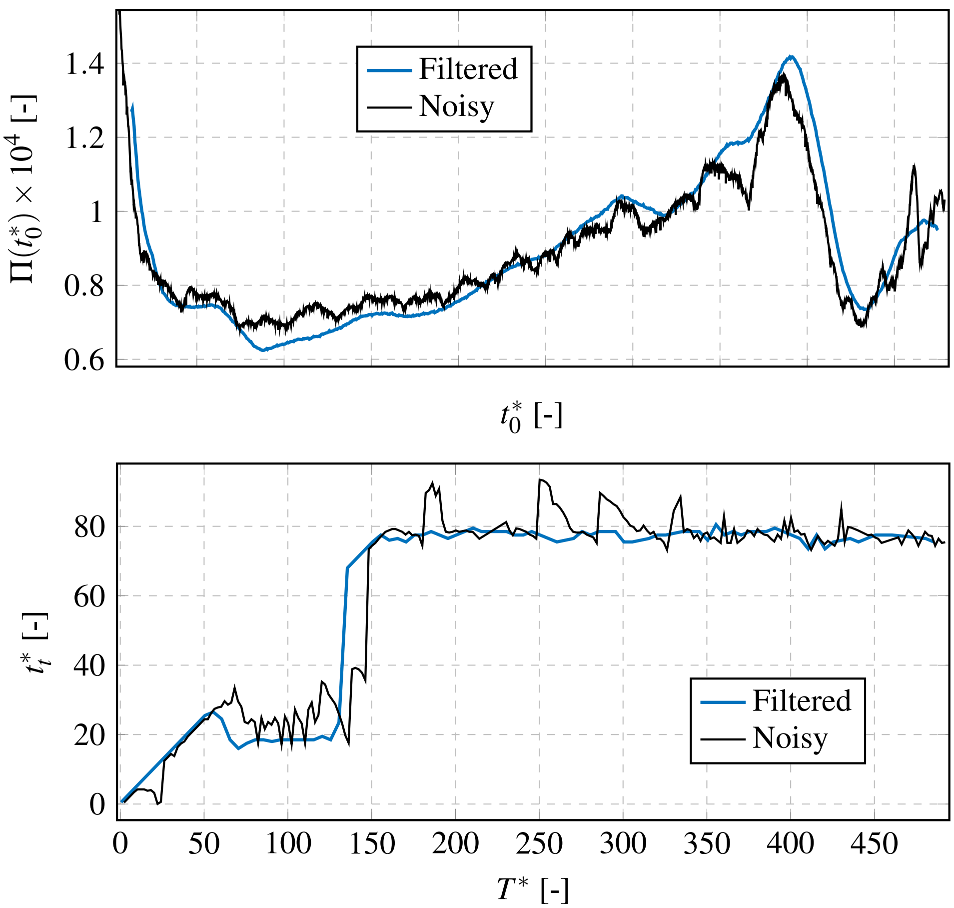

Figure 9.5: Comparison of the product $\Pi(t_0^*)$ of the statistical errors on the mean and the standard deviations estimators (above) and of the transient estimation (below) for the filtered statistical errors (blue) and non-filtered (black). The time axis is non-dimensionalised as $T^* = T(|U_\infty|/c)$ and $t_0^* = t_0(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity. The data is truncated as $0 \leq t^* \leq 500$.

9.2 Required simulation time estimation

The required simulation time t_e is computed as the time necessary to reach a pre-defined statistical accuracy ε_{tol}. However, obtaining accurate estimates can require a large statistical sample. In the framework of a highly computationally expensive turbulence-scale resolving, it is therefore also relevant to quantify the error bar on a statistical quantity if only a limited number of time-steps are executed. Indeed, as it has already been described in the beginning of the chapter, the lift coefficient of a deep airfoil is noisy and a long time samples can be required to obtain an accurate mean.

In Figure 9.6 are represented the truncated sectional lift coefficient (above) for the NACA airfoil, the time evolution of the curve fitting parameter (middle) and finally the evolution of the sampling window as a function of the predefined tolerance. It can be seen that even with the noisy nature of the signal, the curve-fitting parameter converges rapidly : ≈ 170 convective units that corresponds to roughly 90 additional units after the initial transient. From this parameter the required simulation time can be estimated by :

$$t_e^* = \left(\frac{\hat{\sigma}_{C_L}}{\hat{\mu}_{C_L}}\right)^2 \frac{4.55}{\varepsilon_{\text{tol}}^2} \times \frac{1}{100^2}, \tag{9.3}$$

where the factor $1/100^2$ is due to the fact that the coefficient of variation ε is normalized and thus expressed in percent.

Furthermore, it can be seen in the graph below in Figure 9.6 that a window sampling of 300 convective units is required to reach a statistical accuracy of the order of 10^{-4}. However, the solution seems to be already known to the second decimal after the transient duration since the sampling window is really small. As far as the triangular cylinder is concerned, the truncated signal (free of any initial transient) is almost purely periodic with small variations of amplitude and additional noise due to the numerical resolution (above graph in Figure 9.7). The curve fitting parameter requires more time to converge to its asymptotic value. Therefore, more time is required to estimate the simulation time t_e. In addition, it can be seen in the graph below in Figure 9.7 that a larger sampling window is required to reach a same accuracy compared to the NACA test case. This can be explained by the incapability of the methodology to handle purely periodic signals (section 4.3.2). In particular, the overestimation of the statistical error results in an overestimation of the required simulation time.

9.2.1 Importance of the statistical error scaling

As shown in Figure 9.7, the truncated signal for the triangular cylinder is strongly periodic. Given that the required simulation time is estimated by means of the statistical error, the white-noise scaling highly overestimates t_e (see Figure 9.8). It highlights the important im-

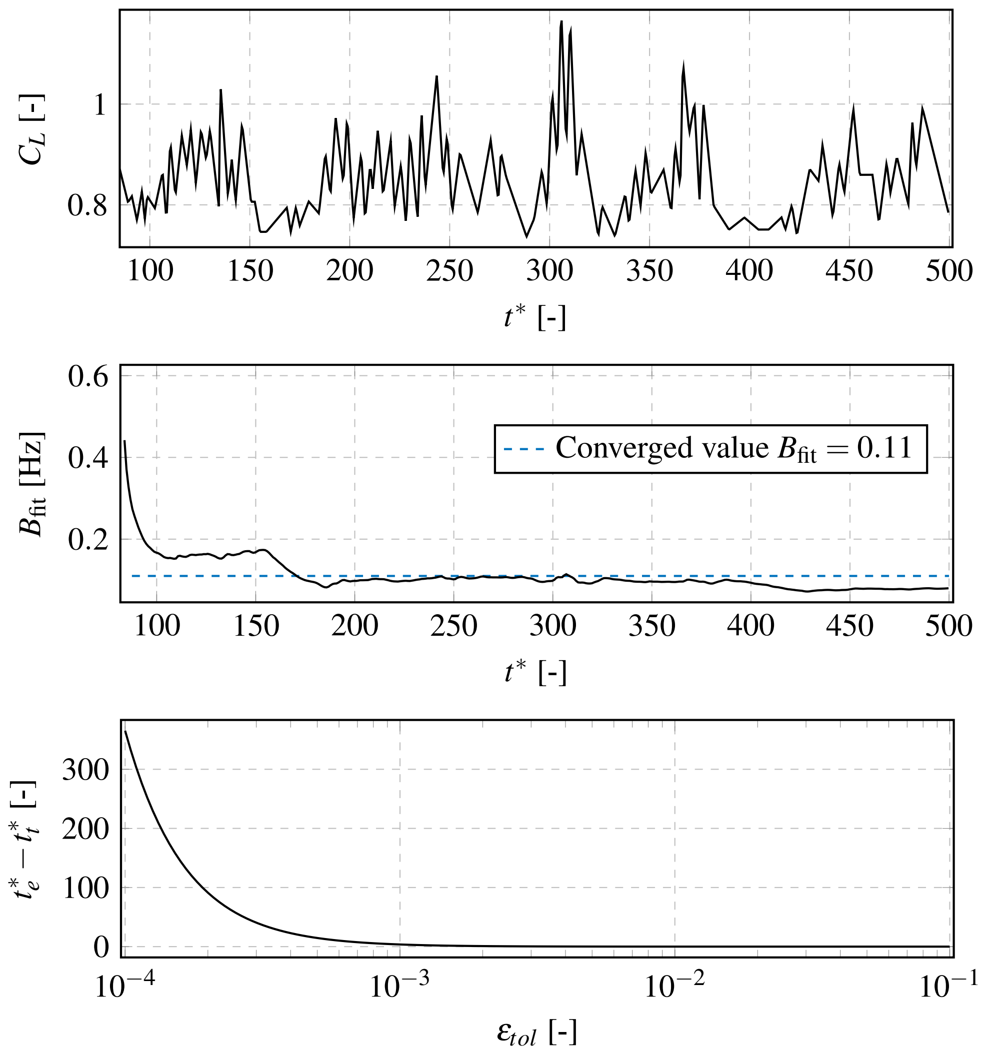

Figure 9.6: Time trace of the integrated sectional lift coefficient C_L (above) as a function of the non-dimensionalised as $t^* = t(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity. Variation of the product $\Pi(t_0^*)$ highlighting the minimum at t_t^* (mid). Prediction of the transient estimate as a function of the sample length (below). The data set is truncated as $0 \leq t^* \leq 500$.

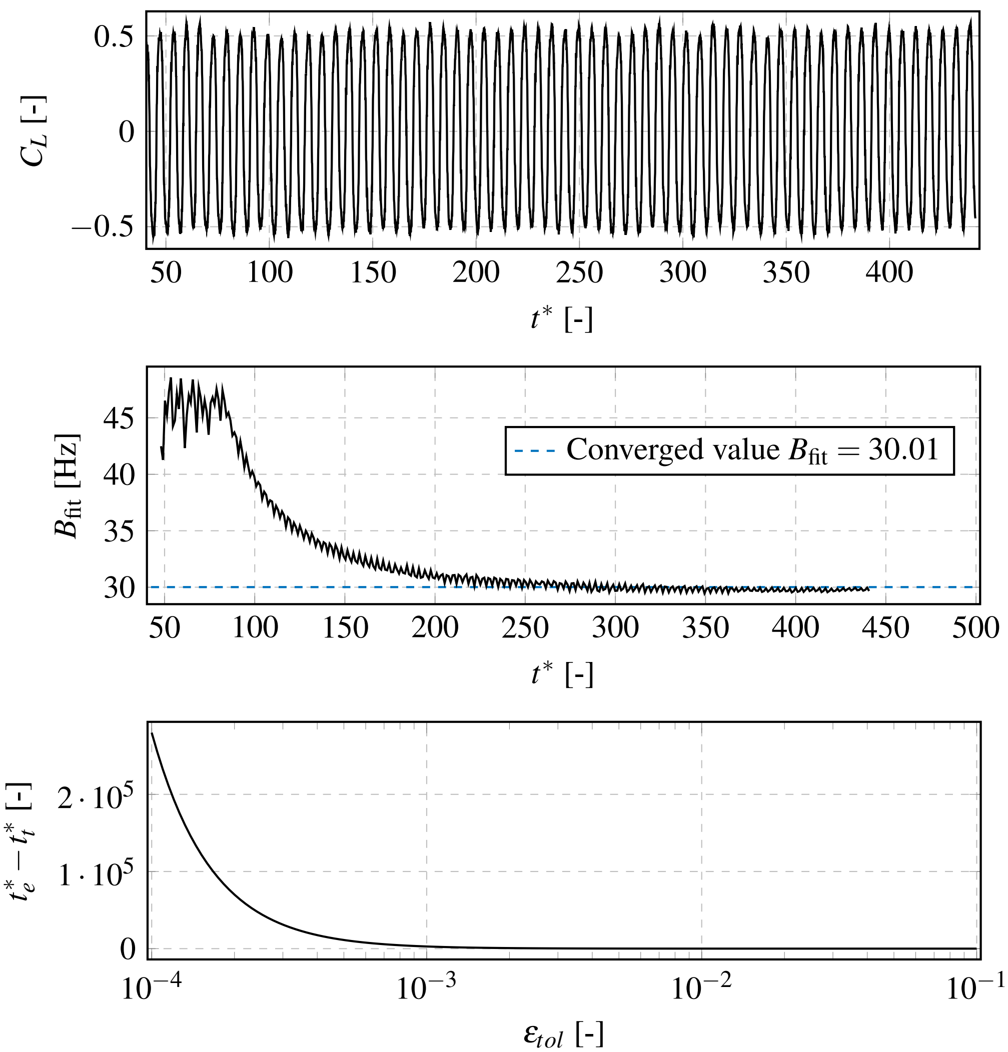

Figure 9.7: Time trace of the sectional lift coefficient C_l (above) as a function of the non-dimensionalised as $t^* = t(|U_\infty|/c)$ with c the chord and $|U_\infty|$ the free stream velocity. Variation of the product $\Pi(t_0^*)$ highlighting the minimum at t_t^* (mid). Prediction of the transient estimate as a function of the sample length (below).

provements that are made due to the new mathematical model introduced in Chapter 5. Indeed, the statistical error for the triangular cylinder is firstly influenced by the periodic contribution and then by the additional noise generated by the DDES model. The scaling with the parameters $f_0 = 14$ [Hz] and $\beta^* = 0.1388$ provides a good agreement between the mathematical model and the statistical error. As expected, the bandwidth ratio β^* is smaller than the unity which is characteristic of a strong periodic contribution.

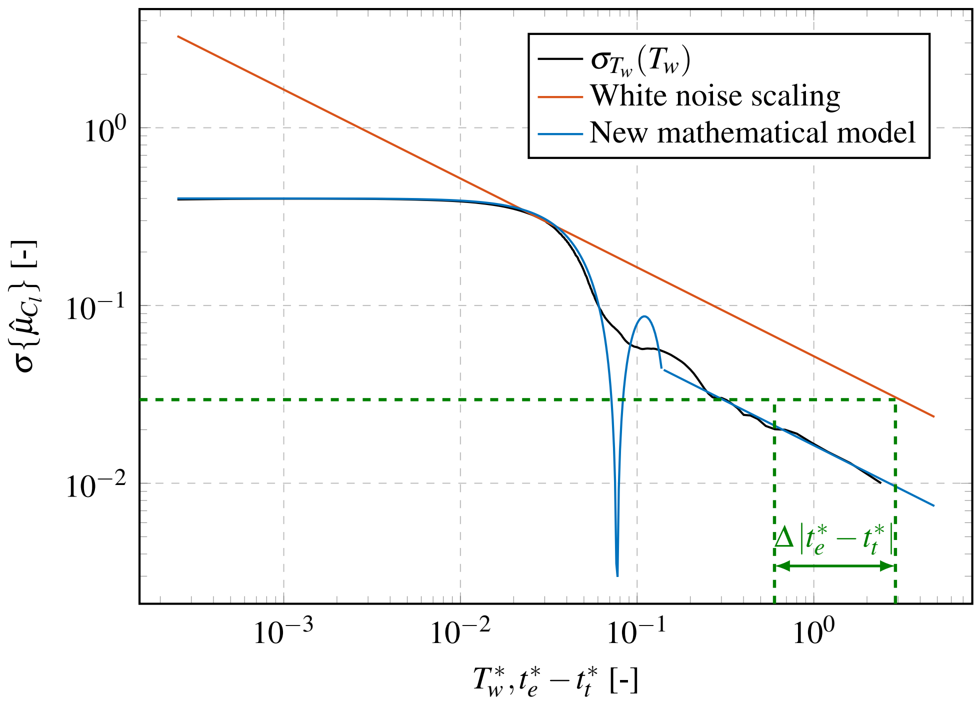

Figure 9.8: Time evolution of the statistical error on the mean estimator of the lift coefficient for the triangular cylinder test case (black). The signal is freed of any initial transient. Both white noise scaling (orange) and the new mathematical model (blue) are represented.

It can then be seen that a huge gain of accuracy is achieved. To quantity this improvement, the difference between the estimated required simulation time for both scaling can be computed as :

$$\begin{aligned} \Delta|t_e^* - t_t^*| &= |t_e^* - t_t^*|_{\mathrm{wn}} - |t_e^* - t_t^*|_{\mathrm{new}} \\ &= \hat{\sigma}_{C_L}^2 \frac{1}{2\varepsilon_{\mathrm{tol}}^2} \left(\left| \frac{1}{B_{\mathrm{wn}}} - \frac{1}{B_{\mathrm{new}}} \right| \right) \end{aligned} \tag{9.4}$$

where B_{wn} and B_{new} are respectively the curve fitting parameter of the white noise scaling and of the new mathematical model. For the white noise scaling, the parameter is equal to 29.89 while for the new mathematical model it is equal to 297. For instance, considering a pre-defined error of 10^{-2}, the difference in required simulation time is thus in the order of 25 convective units i.e more or less half of the initial transient.

10 Conclusions and perspectives

This work aimed to present a robust algorithm to monitor unsteady CFD simulations. In particular, a methodology to estimate the initial transient has been introduced as well as an estimation of the required simulation time to reach reliable statistics. The algorithm is based on statistical considerations and especially on the statistical error encountered by a point estimator. In the present thesis, the statistical error has been evaluated on the mean and standard deviation estimators of the lift coefficient. This statistical error has been evaluated as a function of the running time to be able to monitor a numerical simulation. In particular, the algorithm has been implemented as a User-Defined function in the CFD software ANSYS Fluent. Moreover, by means of analytic formulas, a prognosis of this error has been derived. For that purpose, the first part of the thesis has been dedicated to the literature review. It aimed to provide a complete summary of the main tools and concepts used throughout the manuscript. In the part 2 has been described the complete algorithm to detect the initial transient and estimate the required simulation time. Finally, the third part was dedicated to the discussions and the validations of the implementation for several test cases. Beyond the numerical implementation of the algorithm, this thesis has studied the influence of the frequency content on the statistical error and has open the doors to a wide field of investigations. Therefore, in the following are summarized the main findings and outcomes of this thesis.

10.1 Summary of findings and major outcomes

The major outcomes and findings of the thesis are wanted to be summarized in a concise way. References to the different relevant sections of the thesis are made to ease the reader to obtain more details.

In Chapter 4 has been introduced and reproduced a previous work to compute/prognosis the statistical error and estimate the initial transient. This methodology is based on a combination of windows averaging of a time series and a white-noise scaling to prognosis the statistical error (section 4.2.1). Furthermore, the initial transient of a numerical simulation has been estimated by shortening the sample to make the transient totally disappear. By means of an adequate criterion, it has yielded a robust algorithm of initial transient detection (section 4.2.2). However, this methodology has presented some limitations regarding the periodic flows. In particular, the white-noise scaling appeared to strongly overestimate the statistical error of this particular kind of flow. Indeed, while the statistical error of a white noise varies as $1/\sqrt{t}$, the one of a periodic flow varies as $1/t$ (section 4.3.2).

Directly influenced by the above outcomes, the Chapter 5 has been dedicated to an analysis of the influence of the frequency content on the statistical error. The starting point of this idea has been to notice that the white-noise and the sine wave were totally opposed regarding their spectral content (section 5.3). Indeed, a white-noise is characterised by a uniform power spectral density while the sine wave has a Dirac impulsion as frequency spectrum. Based on this observation a mathematical model has been introduced allowing to model, in the limit, both ends of the frequency domain. That is, a rational function has been considered as power spectral density whose shape presents a peak centred in a frequency f_0. In particular, a band-width parameter β has been defined to control the acuity of the peak. In that way, a large β leads to a pseudo white-noise while a small β leads to a pseudo sine wave (section 5.4.2). In the end, this mathematical model has been able to reproduce the statistical error of signal with both a periodic and a broadband contributions.

In the Chapter 9 the discussions and validations of the implementation have been collated for different test cases . The methodology has provided satisfactory results and an interesting robustness. In particular, the methodology to estimate the initial transient has allowed to handle noisy as well as periodic flows, which was not a certitude regarding the limitations of the statistical error estimation. Regarding the required simulation time, a non-negligible improvement has been made on the estimation by introducing the new mathematical model.

10.2 Closing comments and perspectives

The perspectives for this thesis are multiple, challenging and above all, motivating. In the one hand, the idea to obtain a robust, efficient and more complete algorithm to detect the initial transient should definitely be further investigated given its great utility. In particular, the implementation of this kind of methodology in an industrial software such as ANSYS Fluent could be highly valuable for any user which performs unsteady simulation. On the other hand, this master thesis has introduced a beginning of analysis regarding the intrinsic relationship between frequency content and statistical error. This piece of insight suggests that a more complete and rigorous analyse can be made and that rigorous mathematical links exist between both entities.

As far as the transient detection algorithm is concerned, the latter is obviously perfectible. In the section 9.1.2 the issue of a first plateau in the evolution of the transient estimation has been highlighted . This issue has been overcome numerically by imposing an additional constraint taking into account the throughflow time required for the turbulence to pass one time through the fluid domain. This constraint is a rule of thumb and not a rigorous mathematical

condition. Therefore, an other option should be considered to assess the relevant convergence of the transient duration.
Several tracks of answer can be enumerated :

i. The initial transient algorithm could be applied continuously during the simulation to allow a reset of the statistics if the transient duration considered appears not to be the asymptotic value ;

ii. Other quantities of interest could be used to check and validate or note the estimated transient duration. In particular, in the case of an airfoil for instance, the spanwise forces are expected to be null for a RANS simulation and are thus dependent on the scale resolving model. Furthermore, they should oscillate around zero. Hence, they could provide an additional information to estimate the initial transient. Monitoring points could also be considered to investigate different variables along the time at specified relevant position e.g the analysis of the instantaneous pressure or velocity in the wake downstream of a bluff body.

In general, a methodology gains in credibility with the number of cases investigated. Hence, additional test cases should be considered to better highlight the capabilities and the limitations of the implementation.

Regarding the analysis of the relationship between frequency content and statistical error, the field of possibilities seems infinite given the links between every quantity : power spectral density, autocorrelation function, statistical error, time series etc. The mathematical model introduced in Chapter 5 should therefore be considered as a small piece of a much larger puzzle. In particular, a better understanding of the statistical error could perhaps, in the end, help to express the input time series as a rigorous sum of a pseudo-periodic signal and an additional noise. Indeed, it has been shown that the shape of statistical error is directly influenced by the presence or not of a periodic content. Therefore, by means of a scaling with the parameters β^* and f_0 it should be possible to clearly highlight the two contributions. Following the same philosophy, an investigation of the noise due to the numerical integration could be feasible. In particular, an expression for the signal-to-noise ration could be deduced.

Bibliography

[1] Veena G Adlakha and George S Fishman. Starting and stopping rules for simulations using a priori information. *European Journal of Operational Research*, 10(4):379–394, 1982.

[2] Mohamed H Ahmed and Thomas J Barber. Fast fourier transform convergence criterion for numerical simulations of periodic fluid flows. *AIAA journal*, 43(5):1042–1052, 2005.

[3] Søren Asmussen, Peter W Glynn, and Hermann Thorisson. Stationarity detection in the initial transient problem. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2(2):130–157, 1992.

[4] Julius S Bendat and Allan G Piersol. *Random data: analysis and measurement procedures*, volume 729. John Wiley & Sons, 2011.

[5] Farid Benyoucef. *Amélioration de la prévision des écoulements turbulents par une approche URANS avancees*. PhD thesis, Toulouse, ISAE, 2013.

[6] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[7] Barbara K Butland, Ben Armstrong, Richard W Atkinson, Paul Wilkinson, Mathew R Heal, Ruth M Doherty, and Massimo Vieno. Measurement error in time-series analysis: a simulation study comparing modelled and monitored data. *BMC medical research methodology*, 13(1):136, 2013.

[8] Pijush K. Kundu Ira M. Cohen and David R. Dowling. *Fluid Mechanics*. Academic Press, 2016.

[9] Alan G Davenport. The spectrum of horizontal gustiness near the ground in high winds. *Quarterly Journal of the Royal Meteorological Society*, 87(372):194–211, 1961.

[10] Eric J.M. Delhez. *Analyse Mathematiques*, volume 1. 2015.

[11] Eric J.M. Delhez. *Analyse Mathematiques*, volume 2. 2015.

[12] Eric J.M. Delhez. *Analyse Mathematiques*, volume 3. 2015.

[13] V. Denoel. *Elements de processus stochastiques*. 2015.

[14] Paul A Durbin and BA Pettersson Reif. *Statistical theory and modeling for turbulent flows*. John Wiley & Sons, 2011.

[15] Ionut Florescu. *Probability and stochastic processes*. John Wiley & Sons, 2014.

[16] Fluent. Fluent udf manual. *Theory Guide*, 2009.

[17] A Garbaruk, S Leicher, C Mockett, P Spalart, M Strelets, and F Thiele. Evaluation of time sample and span size effects in des of nominally 2d airfoils beyond stall. pages 87–99, 2010.

[18] Michel Géradin and Daniel J Rixen. *Mechanical vibrations: theory and application to structural dynamics*. John Wiley & Sons, 2014.

[19] ANSYS FLUENT Tutorial Guide. Ansys tutorial. *ANSYS Inc*.

[20] Charles Hirsch. *Numerical computation of internal and external flows: The fundamentals of computational fluid dynamics*. Elsevier, 2007.

[21] Edwin T Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003.

[22] Steen Krenk. Dynamic response to pedestrian loads with statistical frequency distribution. *Journal of engineering mechanics*, 138(10):1275–1281, 2012.

[23] LH Lee, ME Kuhl, JW Fowler, and S Robinson. The problem of the initial triansient, or why mser works.

[24] Georg Lindgren, Holger Rootzén, and Maria Sandsten. *Stationary stochastic processes for scientists and engineers*. CRC press, 2013.

[25] Florian R Menter. Best practice: scale-resolving simulations in ansys cfd. *ANSYS Germany GmbH*, 1, 2012.

[26] Charles Mockett. *A Comprehensive Study of Detached Eddy Simulation*. Univerlagtuberlin, 2009.

[27] Charles Mockett, Thilo Knacke, and Frank Thiele. Detection of initial transient and estimation of statistical error in time-resolved turbulent flow data. pages 9–11, 2010.

[28] Alan V Oppenheim, John R Buck, and Ronald W Schafer. *Discrete-time signal processing. Vol. 2*. Upper Saddle River, NJ: Prentice Hall, 2001.

[29] Raghu Pasupathy and Bruce Schmeiser. The initial transient in steady-state point estimation: Contexts, a bibliography, the mse criterion, and the mser statistic. In *Proceedings of the 2010 Winter Simulation Conference*, pages 184–197. IEEE, 2010.

[30] Charles L Phillips, John M Parr, and Eve A Riskin. *Signals, systems, and transforms*. Prentice Hall Upper Saddle River, 2003.

[31] RR Rekha and P Sumathi. Stochastic modeling of irrigation system in tamilnadu paddy fields–a study. *International Journal of Pure and Applied Mathematics*, 119(16):679–695, 2018.

[32] Stewart Robinson. A statistical process control approach for estimating the warm-up period. In *Proceedings of the Winter Simulation Conference*, volume 1, pages 439–446. IEEE, 2002.

[33] Gilbert Saporta. *Probabilités, analyse des données et statistique*. Editions Technip, 2006.

[34] R. Sepulchre and G. Drion. *Modelisation et analyse des systemes*. 2016.

[35] M Shur, PR Spalart, M Strelets, and A Travin. Detached-eddy simulation of an airfoil at high angle of attack. In *Engineering turbulence modelling and experiments 4*, pages 669–678. Elsevier, 1999.

[36] Mikhail L Shur, Philippe R Spalart, Mikhail Kh Strelets, and Andrey K Travin. A hybrid rans-les approach with delayed-des and wall-modelled les capabilities. *International Journal of Heat and Fluid Flow*, 29(6):1638–1649, 2008.

[37] Oleg Sokolsky and Serdar Tasiran. *Runtime Verification: 7th International Workshop, RV 2007, Vancover, Canada, March 13, 2007, Revised Selected Papers*, volume 4839. Springer, 2007.

[38] PR Spalart. Strategies for turbulence modelling and simulations. In *Engineering Turbulence Modelling and Experiments 4*, pages 3–17. Elsevier, 1999.

[39] Katrina Elise Swalwell. *The effect of turbulence on stall of horizontal axis wind turbines*. PhD thesis, Monash University, 2005.

[40] Paul G Tucker. Novel miles computations for jet flows and noise. *International Journal of Heat and Fluid Flow*, 25(4):625–635, 2004.

[41] Paul G Tucker. *Advanced computational fluid and aerodynamics*, volume 54. Cambridge University Press, 2016.

[42] Theodore Von Karman. Progress in the statistical theory of turbulence. *Proceedings of the National Academy of Sciences of the United States of America*, 34(11):530, 1948.

[43] Rob J Wang and Peter W Glynn. On the marginal standard error rule and the testing of initial transient deletion methods. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 27(1):1–30, 2016.

[44] L. Wehenkel. *Elements du calcul des probabilites*. 2016.

[45] David Williams. *Weighing the odds: a course in probability and statistics*. Cambridge University Press, 2001.