Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Meccanica

Tesi di Laurea Magistrale

Unconventional analysis of turbulent plumes

Relatori
Ing. Giovanni Iacobello
Prof. Luca Ridolfi
Prof. Pietro Salizzoni (École centrale de Lyon)
Prof.ssa Stefania Scarsoglio

Candidato
Capello Luigi

Luglio 2019

Unconventional analysis of turbulent plumes

Luigi Capello

"Dobbiamo buttarci nel gioco senza nemmeno chiederci quale sarà la partita perchè quasi sempre non è dato saperlo in anticipo. In tutto questo, l'unica cosa che possiamo stabilire sono gli obiettivi. Lo abbiamo fatto in una maniera chiara e rigorosa a prescindere da quelle che sarebbero state le condizioni del mercato e dell'economia mondiale." Sergio Marchionne, A.D. Fiat Chrysler Automobiles

Dedicato a mia mamma ed a mio papà, le persone che mi hanno insegnato l'orgoglio del lavoro, che è e resterà l'insegnamento più grande che io abbia mai ricevuto. Grazie di cuore per tutto!

Contents

Introduction

Turbulence is a very common phenomenon, present both in nature and in technical applications. The smoke plume generated by a cigarette, the swirling motion of a river rather than the turbulent areas that generate vibrations of planes are all examples of turbulent natural phenomena. At the industrial level turbulence plays a key role. There are several examples that could be mentioned. In foundries, for example, it is possible to observe a relationship between the turbulence of the molten metal (normally generated during the transports) and the final mechanical properties. This because, the more turbulent is the liquid metal, the more it will enter in contact with the air, generating oxides and impurities. These latters, in the finished product, may be triggers of cracks and breakage at low mechanical loads.

Although turbulence has a fundamental role in many sectors, there are no approaches which allow to fully describe a turbulent flow. Turbulence still remains a great field for theoretical and industrial research. The turbulent motions are unpredictable and chaotic. Therefore, the spatial-temporal characterization and the streams control are still particularly complex. The turbulence nature makes it such a useful phenomenon, but at the same time, so difficult to describe and predict. For this reason, statistical treatments are often used for the study of turbulent flows.

The modern computational techniques, such as the DNS-LES-RANS, do not stop the continuous need of new methodologies for the turbulence study. New techniques that take into account the phenomena from several points of view. This because the turbulence is generated by the combined action of multiple causes. In order to understand turbulent flows dynamics and to get information about their spatial characterization, the complex networks theory may be used. This theory, combining elements of graph theory and statistical physics, allows the study of systems formed by a large quantity of elements interacting each others. It is therefore suitable for the study of turbulent flows, which are composed of a large number of elements that interact dynamically over time.

The aim of this thesis is studing the diffusion inside a turbulent boundary layer. This particular case of turbulence has been analysed starting from time series measured within a boundary layer suitably created inside a wind tunnel.

What is analyzed here must be considered as a "case of study" respect to a large quantity of natural, environmental and industrial phenomena. For example, the possibility of predicting presence, concentration and evolution of a pollutant inside an atmospheric flow is fundamental for the human health point of view. For chronic risks, mean concentrations estimations are sufficient. Conversely, when considering the accidental risk due to releases of pollutant, flammable and explosive substances, what matters is the instantaneous concentrations.

For the turbulence study there are different approaches that exploit complex networks theory. Most of theme are based on time series transformation in corresponding complex networks. In this thesis, through the time series measured in the wind tunnel, the complex networks analyzed here have been defined. This means that the experimental signals have been appropriately geometrized. There are several ways to perform time series geometrization in networks. This work aims to be a cognitive thesis of the Recurrence Networks method applied to turbulent time series. Networks generated using this method are based on the presence of "similar parts" within the signals themselves. In fact, these are divided into sub-signals, called embedded vectors, which contain m consecutive measurements coming from the considered series. Each embedded vector will be a network node. Two nodes will be connected together if the corresponding vectors present a recurrence, so if the two vectors contain similar measures. Physically, this similarity is to be understood as nodes distance in the phase space. Therefore, to define links between the various nodes of a network, the similarities among all possible pairs of nodes must be evaluated. To do this, given two embedded vectors, two different ways can be followed. The first uses the Euclidean norm of the difference, calculated element by element, of the two vectors. If this norm is less than a certain threshold, the connection between the two nodes analyzed can be activated. The second method is based on the Pearson correlation coefficients. Given two vectors, the absolute value of this coefficient can go from 0 to 1. The more it will be close to 1, the more the two vectors will be alike. Also for this case will be necessary defining a discriminating threshold for the connections definition.

Complex networks are a versatile tool, applicable in many fields of science and technology. Whatever the application, using the Recurrence Networks method, the mathematical model that transforms time series into networks will always be the same. What varies according to the specific problem analyzed are the optimal embedding parameters. So, the constructive parameters that allow to obtain networks that are significant for the considered signals. Therefore, for the geometrization of time series, the model calibration plays a fundamental role. In fact, as will be illustrated, even small variations of the embedding parameters with which networks are built may generate large differences on the networks structure and on their topology, having therefore an high effect on the results significance.

Once the optimal networks are defined, through the study of their metrics, so through the study of indicators describing the structure and the topology of the networks, it is possible to obtain information about the time series of interest. Therefore, it is possible to identify the behaviour of the considered physical phenomenon.

The thesis is organized as follows:

- *Chapter 1 is a general description of turbulent flows. Their main properties are introduced using appropriate mathematical-statistical models. The assumptions behind the K41 theory, the energy cascade concept and the modern computational techniques are illustrated.*

- *Chapter 2 describes the experimental apparatus used for the time series measurement: the wind tunnel and the sensors settings are described in order to illustrate the signals meaning. Then a first description of data is given through some statistical analyses carried out for the velocity and the concentration signals. For the latter it is also described the filtering necessary to clean up the signal.*

- *Chapter 3 gives ample space to complex networks theory. Starting from its historical evolution, the networks broad application fields and their main properties are illustrated. The introductory part concludes with a description of the metrics used in this work. At the chapter end there is a description of the Recurrence Networks (RN). The RN mathematical model and the steps to obtain a network starting from time series are illustrated.*

- *Chapter 4 describes the obtained networks, both from the point of view of the adjacency matrices and from the graphic point of view. In this chapter the effects of some embedding parameters on the networks are also analyzed.*

- *Chapter 5 describes the model calibration. Here, by comparing the metrics obtained with different parameters, the optimal ones are identified. These parameters allow to study the considered signals in the best way in terms of results significance. The metrics obtained with these optimal parameters are analyzed in chapter 6. Here there is a fundamental passage: from the networks properties the characteristics of the time series are inferred. So the characteristics of the considered turbulent plume are identified. Chapter 6 illustrates also a study related to the recurrences frequencies. As will be seen, these frequencies are directly connected to the network metrics, and thus to the system dynamics.*

- *Chapter 7 shows the main results identified.*

Chapter 1

Turbulent flows: characteristics and main properties

In the first part of this chapter the turbolent flows are defined. Their main characteristics and their properties are analyzed both from the scientific and the applicative point of view. In fact, two non conventional examples of turbulence use in industries are described. Then, the Reynolds average technique, which is, the decomposition of turbulent phenomena into a transport average flow and into a fluctuating flow, is described. Towards the end of the chapter, the concept of turbulent scales is defined. The mechanisms of the energy exchanges between the different scales, so the energy cascade, are described according to the hypotheses introduced by the so-called K41 theory. The chapter concludes with a description of techniques used for turbulent flows simulations: DNS, RANS and LES.

1.1 Turbulence: definition and main features

It is not easy to define a turbulent flow because of the turbulence nature. In fact, this may be considered as a rapid and irregular movement that occurs in fluids when they reach high Reynolds number or when they are flowing in contact with solid walls. The main characteristic of a turbulent flow is the irregularity of each fluid particle property. For example, the velocity of a particle in a turbulent flow varies irregularly in time and space. Therefore, it is not determinable beforehand with precision and continuity. The fast and irregular fluctuation of the velocity overlaps the mean value, so that the instantaneous velocity results the sum of an average speed and the corresponding fluctuation. Obviously, what is indicated here for the velocity is valid for any quantity relative to a turbulent flow.
In nature, most of the flows are turbulent: from atmospheric and oceanic currents to the arteries bloodstream. Even in the industry the turbulence plays a

fundamental role. This because the turbulent flows have characteristics of interest for the technique and for the mechanics. For example, turbulence produces an increase in the friction coefficient, less wake resistance, higher values of the lift coefficient for high-incidence wings and higher values of the heat exchange coefficient. Mentioning some examples of industrial use of turbulence, shell and tube heat exchangers may be considered. They are constructed having inside some plates that deviate and mix the flow inside the shell. This in order to increase the turbulence of the flow and, with it, its thermal exchange capacity. Another example of turbulence use allows the great mobility of modern times. An engine, gasoline or diesel, has a combustion phase of about 60 crank angles. If the flow inside the combustion chamber were laminar, the combustion speed would be so small to not allow combustion in the required time [2][1]. Using techniques such as squish, swirl and tumble the flow in the chamber is made turbulent, and so the flame front becomes corrugated, increasing its surface. This allows an increase in the combustion velocity that guarantees the well-known thermal engines performances.

By defining the turbulence, its stochastic nature is immediately perceved. Many turbulent analyses are in fact based on the use of statistical techniques. This aspect is in contrast with the ability of the Navier Stokes equations to describe every type of fluid and flow. In fact, the nature of these equations is deterministic. It is therefore interesting to ask whether it is possible to describe a stochastic phenomenon with deterministic equations. The answer to this question lies in the mathematical structure of NS equations. In fact, they are extremely sensitive to the initial conditions and to the numerical approximations. Thus, describing a turbulent flow through these equations, the solutions variability does not come from the equations themselves but from their high dependence on the minimal variations of the boundary conditions, i.e. the minimal variations of the environment in which the flow moves.

A turbulent flow can therefore be defined as irregular, stochastic and subject to a large amount of degrees of freedom. Hence, statistical methods are needed for the study of such flows. Note that, describing the turbulence, the term flow is always used instead of fluid. This because turbulence is a property of the flow and not of the fluid that composes it. In fact, using the same fluid, the Reynolds experiment shows that are the different efflux conditions to define the motion state. The latter is usually identified using the Reynolds number (Re). This acts as an index of the turbulence that is present. For low Re values, there is a laminar flow, which turns into turbulent for high Reynolds numbers. Intermediate Re values identify the transition zone.

A physical meaning can be attribute to the Reynolds number: it is the ratio between the inertial forces, from which the instability of the flow derives, and the

viscous stresses, responsible of energy dissipation. These two physical quantities, to which correspond as many terms in the NS equations, varying their relative weight, change the flow from laminar to turbulent and viceversa. Being the viscous term effect an energy dissipation, the turbulence needs therefore a "power supply". In fact, transforming the turbulent kinetic energy into internal energy, so warming the fluid, in order to not cease the turbulent effect itself, the viscous stresses always need new incoming energy.
Compared to the a laminar flow, of course, a turbulent one is characterized by a greater mass, momentum and heat mixing effectiveness. The effect of the turbolence on thermal and mass diffusivity can be observed considering the Prandtl and Schmidt numbers increase while the Re increases.

1.2 Decomposition of time signals

Considering a certain point in the space where measuring a fluid dynamic quantity: the fig. 1.1 illustrates the difference between a measured laminar signal and a turbulent one. Starting from the instantaneous values it is possible to obtain,

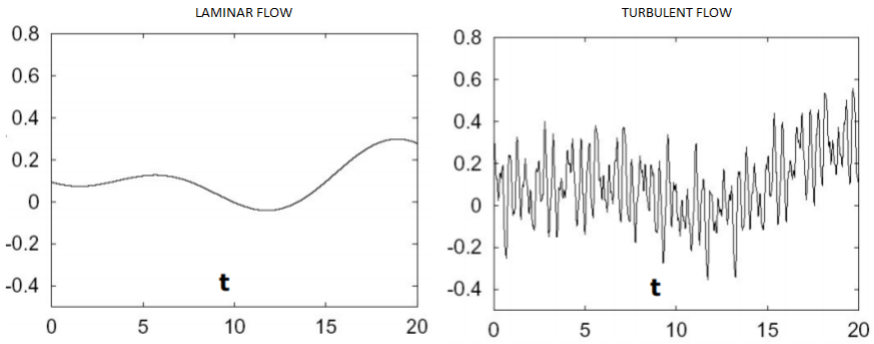

Figure 1.1: Difference between a laminar and a turbulent signal

through a temporal average, a mean value of the analysed quantity. Consider for example a velocity signal. The average velocity characterizes the transport that determines the overall displacement of the fluid mass. If this average speed is time-independent, the flow is stationary, otherwise it is non-stationary. Instant by instant, subtracting the average value from the instantaneous values, the turbulent fluctuation can be obtained. This shaking motion involves only an irregular oscillation of the fluid-dynamic quantities around the average transport values. Therefore, the turbulence vigour will be as much greater as these fluctuations are intense. But, in practice, more than the instantaneous fluctuations, one is interested in a single value that indicates with what force the instantaneous values vary from the mean. This parameter can be calculated as the fluctuations root mean

square. For the velocity, it is indicated as turbulence intensity.
This way of proceeding is called Reynolds decomposition [1] [2]. Therefore, considering a generic velocity component u, operating as above, the turbulent signal is so decomposed: $u(t) = \overline{u} + u'$, where $\overline{u}$ is the average velocity and u' is the fluctuation.

1.3 Scales of turbulence

One of the main turbulence features is being strongly rotational. A turbulent flow has within it a large number of vortices which are maintained thanks to the fact that there are three-dimensional fluctuations of the fluid dynamics quantities.
Being the turbulent fluctuations formed by a series of vortices, it is interesting to characterize these latters. In fact, each eddy has its own size, its speed and its rotation time. The metrics that characterize the eddies are called scale. For each eddy, it is possible to define three metrics, so three scales: a spatial scale to evaluate the eddy size, a velocity scale to evaluate the rotation speed and a time scale. Of course, for a given eddy, these three scales are not independent each other.
Given the large amount of vortices that characterize a turbulent flow, it is possible to define two extremes. That are, the larger space scale, so the larger eddy, and the smaller. The larger is called integral scale and is defined using geometric constrains. Considering for example the flow inside a engine piston, the integral scale will not be larger than the piston itself. However, does not count only the size. It must be considered the probability that a so big eddy will be actually present in the analyzed motion field. In fact, the integral scale must be representative of the kinetic energy actually present in the flow. It is therefore defined using a function called two-point correlation. This larger scale is indicated with L_i.
The smaller scale, called the Kolmogorov scale and indicated with η, is defined using kinetic energy and the viscosity. There aren't scales smaller than η because, at this scale, the kinetic energy is completely dissipated by viscosity.
The intermediate scales, i.e. those between the integral scale and the Kolmogorov scale, are part of the inertial subrange.
For each scale it is possible to define a Reynolds number as $Re_x = \dfrac{u_x \cdot l_x}{\nu}$.
The larger eddies are the first ones to be formed and derive their energy directly from the average motion of the flow. So, the turbulence energy "enters" at the larger scales. As noted by Richardson in the twenties of the last century, the large eddies are unstable, therefore, through a continuous process of destabilization and fragmentation (stretching) the energy is trasferred at the smaller scales. Within the inertial subrange, this energy is transferred without a viscose action, i.e. there

is an energy transfer without dissipation. All the scales, except for the Kolmogorov one, have a Reynolds number bigger than one. So, for them, inertial forces are predominant. Instead, considering the Reynolds number at the smaller scale, it can be obtained $Re_\eta = 1$. Therefore, once this scale is reached, due to the viscosity effects, the kinetic energy is dissipated. The rate of turbulent kinetic energy dissipated per volume unit, indicated with ϵ, may be defined. This also represents the energy entering inside the system, since it is not subject to dissipation in the inertial subrange. This process, known as the energy cascade, does not allow to answer to some questions. For example, what is the size of the smallest eddies that dissipate energy. Or, how the velocity and the time scales vary moving along the inertial subrange towards smaller spatial scale. These and others questions were answered by Kolmogorov's theory of turbulence in 1941. This theory is often indicated with the pseudonym of "K41 theory". It is based on three important hypotheses combined with dimensional arguments and experimental observations [3].

The first Kolmogorov's hypothesis is related to the local isotropy. The directional biases of the large scales are lost in the chaotic scale-reduction process as energy is transferred to successively smaller eddies. Therefore, the Kolmogorov's hypothesis of local isotropy states that at sufficiently high Reynolds numbers, the small-scale turbulent motions are statistically isotropic.

Not only the directional information are lost in the energy cascade, but also the information about the geometry of the eddies gets lost. So, the statistics of the small-scale motions are universal: they are similar in every high Reynolds number turbulent flow, independent of the mean flow field and the boundary conditions. These statistics depend just on ν and ϵ. This is called first similarity hypothesis. From this derive the definitions of η, u_η and t_η. They will only be a function of ν and ϵ appropriately combined.

Once the smaller scale is defined, it is possible to represent the inertial subrange, using the turbulent kinetic energy definition and the epsilon. As mentioned above, all mathematical formulations are not reported. However, can be proved that, in a bilogarithmic plane which has the spatial scales as abscissa axis and the velocity scale on the ordinates axis, the inertial subrange is a line with slope equal to $1/3$, which, starting from the Kolmogorov scale, reaches the integral one.

The last hypothesis is relative to the intermediate scales. Their Reynolds number is relatively large, therefore they will not be affected by the viscosity. Starting from this, the Kolmogorov's second similarity hypothesis states that in every turbulent flow at sufficiently high Reynolds number, the statistics of the inertial subrange scales have a universal form that is uniquely determined by ν and independent of ϵ.

1.4 Numerical simulations of turbulent flows

As indicated, a turbulent flow can be described through the Navier-Stokes equations. Therefore, the most immediate method for its numerical simulation is the NS discretization. This technique is called "Direct Numerical Simulation", DNS. Referring to the turbulence description shown in the previous section, it is easy to understand that the computational domain will have dimensions comparable to the integral scale. Instead, to describe in detail the effect of viscosity at the Kolmogorov scale, i.e. the transformation of kinetic energy into internal energy, the grid must have dimensions comparable to the scale η.
Obviously, a so fine grid has a direct effect on the calculation time: it is proportional to the third power of the Reynolds number. In modern engineering systems, the Reynolds numbers that are reached are extremely high. For example, inside a medium-small heat exchanger, the Reynold number is around 200000. So, even using the modern computational techniques, it is a effort of time and costs that is completely unsustainable. Therefore, at the industrial level, DNS is never used. Instead, its is used in the scientific research. In fact, directly solving the equations of motion without the use of any model, the results produced are comparable to the experimental ones. Compared to the latter, the numerical results are also easier to obtain. Think for example at the motion field within an engine. As long and complex as it is, the DNS is still easier than direct measurement. Just imagine the large amount of modifications to be made to the pistons in order to insert a sensor. Than, the sensors presence can change the considered motion field. Therefore, DNS is a fundamental technique for the research. The computational cost, however, binds the use to moderate Reynolds numbers, making it unusable for flows of engineering interest.
There are two other techniques that are used at the industrial level to model the turbulent flows [4]. The most used is the RANS technique (Reynolds Averaged Navier-Stokes). This is a model that consider an average in time of NS motion equations. Therefore, only the average quantities are explicitly calculated while the turbulent ones are modeled appropriately. This technique is widely used because the flows of interest present, normally, a stationary configuration of the motion fields. That is, the quantities of interest are usually the average ones and not the instantaneous.
The purpose of this method is decomposing, using the Reynolds average described above, all the terms present in the NS equations. The problem is that the equations system obtained is not solvable. In fact, the Reynold tensor depends only on the turbulent fluctuations values, and not on the average quantities. So, in face of 4 equations, 10 unknowns are obtained. It is therefore necessary to model the tensor components with respect to the average quantities. There are several techniques that allow to do this. The technique choice depends on the results accuracy

that is needed. In fact, the Reynolds tensor has an important influence on the flow dynamics. There are the first order models, having different number equations. Probably, the most used method is the two-equation model K-Epsilon. Increasing the computational cost, it is possible to use second order models or probabilistic models. It is interesting to note that, these models often have calibration constants that vary from case to case. This, together with the fact of modelling the term that most of all influences the dynamics of the flow, are the main criticalities of this method.
When more detailed results are required, a technique called LES (Large Eddy Simulation) can be used. As indicated previously, the larger eddies depend on the flow geometry while the smaller ones have a universal character. So, in the LES the large scales are calculated, while the effects of the smaller are modeled. The LES technique is conceptually more advanced of the RANS. This beacuse is based on the two hypotheses of similarity and universality that the K41 theory describes.
The LES technique requires less computational efforts than the direct numerical simulation, but more than the RANS methods. The main advantage is the greater level of detail than the RANS techniques: the LES is able to predict instantaneous features of the flow. This is particularly important in those simulations that include chemical reactions, such as combustion in engines and transporting phenomena.
To conclude, it is advisable to use RANS technique for stationary, two or three dimensional problems, while LES is advisable for non-stationary flow in which the temporal evolution plays an important role.

Chapter 2

Experimental activity: velocity and concentration measurements

This chapter describes the experimental activities performed to obtain the time series investigated in this work. The type of wind tunnel used and its setting are described in the first part of the chapter. Then, ample space is dedicated to the description of the measuring points choice, to the type of sensoristics and to their operating conditions. The middle part of the chapter describes the choices made in order to organize the experimental data optimally. The last part of the chapter is instead dedicated to the statistical analyses carried out on the different signals. For the concentration, the only series to require a filtering, is shown a study of the intermittency resulting from the filtering itself. To confirm what has been achieved, some parts of concentration time series are graphically represented. The average concentration profiles are then shown in order to identify the variation of this parameter within the plume. The chapter concludes with the statistical analysis of the velocity signals that allows to identify two important aspects: the effect of the friction between plume and wall on the average velocity u profile and the absence of convective phenomena inside the plume.

2.1 Purpose of the experimental activity

In recent years attention has been paid to environmental issues. Among these, air quality plays a key role. There are several standards that allow to assess the level of air pullution. Atmospheric pollution means the alteration of natural air conditions, due to emissions of exhaust gases from motor vehicles, boilers, power plants, factories, incineration plants. The possibility of predicting presence, concentration and evolution of a pollutant in an atmospheric flow is therefore fundamental. Different substances can, of course, have different effects on the environment and on

human health.

As far as human health is concerned, the statistical characterization of concentration is necessary. For chronic risks, estimates of mean concentrations are sufficient. Conversely, when considering the accidental risk due to releases of pollutant, flammable and explosive substances, what matters is the instantaneous concentrations. For this reason, the need to define the concentration probability density function (PDF) arises [5].

As shown in the literature, there are many ways to characterize emissions in the lower atmosphere. The dependence of concentration on the parameters controlling the emissions is also shown in some papers. The dependence of the dispersion process on the source configuration, including source size, elevation and emission velocity was analyzed at the the Laboratoire de Mécanique des Fluides et d' Acoustique at the Ecole Centrale de Lyon in France. A series of wind-tunnel experiments on the dispersion of a passive scalar emitted by a source of varying size and height, within a turbulent boundary layer, have been performed. The experimental dataset is composed by velocity and concentration time series measured at different points in the wind tunnel. These time series are the basis of the study carried out in this thesis. In fact, they have been transformed into networks, with the aim of getting statistical information about speed and concentration fields.

2.2 Wind-Tunnel set up

There are two ways to classify wind tunnels. Firstly, it is possible to evaluate the presence or not of fluid recirculation. Open loop and closed loop tunnels can be defined in this way. While open loop tunnels expel the fluid in the external environment after the test, the closed ones have a recirculation and reconditioning system of the fluid that is reintroduced into the test chamber. Wind tunnels are also classified by the amount of speed they can produce. There are four basic types of wind tunnels which are: low subsonic, transonic, supersonic and hypersonic.

The time series were obtained in the in the atmospheric wind tunnel of the Laboratoire de Mécanique des Fluides et d' Acoustique. This is a recirculating wind tunnel with a working section measuring 14 m long and 3.7 m wide.

The wind tunnel set up is sketched in the fig. 2.1.
When working in wind tunnels, both for the analysis of plumes and for models testing, it is very important to control the longitudinal pressure gradient. In case of models testing, the pressure gradient produces an effect of buoyant force acting

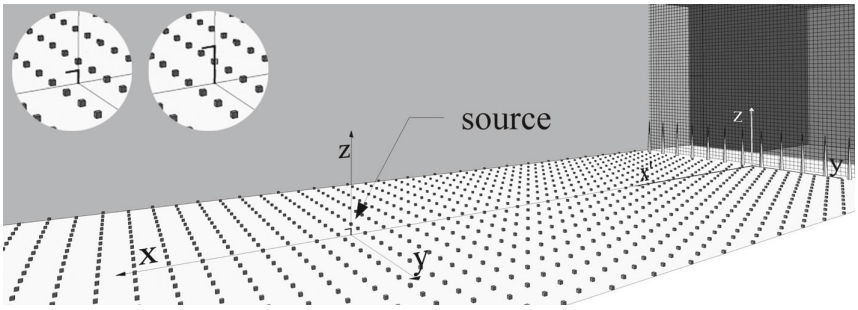

Figure 2.1: Sketch of the wind tunnel set up [5]

on the model in the direction of decreasing pressure. As a result, the longitudinal force acting on the model is determined with an error. While in the case of plumes structure study, the velocity field may be affected by the gradient.

The air temperature in the wind tunnel was regulated so that its variations during a 1-day experiment can be maintained in the range 0.5°C. The experiment can therefore be considered performed in isothermal conditions.

The experiment consists in the atmospheric dispersion of a passive scalar, emitted by a localized source, placed at a height h_s from the ground, with a diameter σ_0 and an ejection velocity u_s . Ethane (C_2H_6) was used as tracer in the experiments, since it has a density similar to air. It was continuously discharged from a source of varying diameter and elevation. The release takes place within a turbulent boundary layer with free-stream velocity u_∞. It is assumed to be fully characterized by self-similar relations, obtained by rescaling profiles of velocity statistics on the friction velocity u^* and the boundary-layer depth δ. A neutrally-stratified boundary layer was generated by combining the effect of a grid turbulence and a row of spires, placed at the beginning of the test section, and roughness elements on the floor. The presence of a turbulence grid is used here since it assists in minimizing the inhomogeneities of the flow. The spires were of the Irwin type[1]. The entire working section floor was overlaid with cubes of side $h = 0.02$ m acting as roughness elements. The cubes were placed in a staggered array and covered approximately 1.8% of the tunnel floor surface. This experimental set-up allow to reproduce a boundary layer of depth $\delta = 0.314$ m. Imposing a free-stream velocity

[1]Spires used to simulate the planetary boundary layer are so called. There are expressions for the height and base length of triangular spires which produce given values of the boundary layer thickness.

$u_\infty = 4.94$ m/s and $\nu = 1.5 \cdot 10^5$, the Reynolds number

$$Re = \frac{\delta \cdot u_\infty}{\nu} = 2.6 \cdot 10^5$$

was sufficiently high to ensure the adequate simulation of a fully turbulent flow [6]. In this calculation ν is the air kinematic viscosity.

The tracer was emitted by an L-shaped nozzle. This was positioned so that it can be considered in the area where the boundary layer was fully developed. The nozzle horizontal part was almost 30 times the internal diameter, so as to eliminate the effect of the vertical stretch and of the elbow.

As indicated before, measures were carried out considering two values of the internal diameter σ_0, 3 mm and 6 mm, while the emission height h_s was fixed equal to 75 mm.

In the wind tunnel, obviously, at different points the air assumes different speeds. In the experiments, ethane was injected with the air velocity at the injection point. This condition is know as Isokinetic Condition.

A reference system has been created to define the spatial position of the measuring points. Coordinates y and z denote the transversal and vertical direction, respectively. The x coordinate has its origin at the source location.

2.3 Measurements

2.3.1 Contemporaneousness and output of measurements

As indicated, this thesis uses as starting point the velocity and the concentration time series measured in the wind tunnel. The measurements were performed at different positions in the tunnel. When the measuring point is fixed, it is possible to consider the measurements of velocity and the measurements of concentration as contemporary. Considering the entire measure domain, however, all the values obtained cannot be considered contemporaries. It is possible to imagine taking measurements in a specific point and then moving to the next one. In doing so, the flow of time is inevitable. Therefore, all the measures cannot be considered contemporary.

The result of the measurements performed in a point is a six columns file. The two speed components (m/s), the concentration (ppm) and the air and ethane mass

flow rate at the source (l/h) are indicated for each time instant (s). An example is shown in the figure below:

1.7000000e-02	3.4325000e+00	5.4091000e-01	3.2333000e+01	3.6352000e+00	3.4077000e+02
1.8000000e-02	3.5116000e+00	3.0043000e-01	2.9866000e+01	3.6352000e+00	3.4077000e+02
1.9000000e-02	3.6066000e+00	3.9479000e-01	1.9125000e+01	3.6352000e+00	3.4077000e+02
2.0000000e-02	3.4394000e+00	3.1736000e-01	9.9451000e+00	3.6352000e+00	3.4077000e+02
2.1000000e-02	3.3895000e+00	2.0649000e-01	8.8565000e+00	3.6352000e+00	3.4077000e+02
2.2000000e-02	3.3718000e+00	1.9226000e-01	3.4909000e+01	3.6352000e+00	3.4077000e+02
2.3000000e-02	3.3648000e+00	2.7985000e-01	1.3905000e+02	3.6352000e+00	3.4077000e+02

Figure 2.2: Part of a measures file

Considering the two diameter values, 469 points were measured, and so, 469 files of this type were defined. Each file has a number of lines relative to a 3-minute measurement time with a 1000 Hz sampling. Therefore, for each point, about 180 000 time instants have been considered.

2.3.2 Concentration measurements

Concentration measurements were performed with a fast flame ionization detector (FID). This is an instrument used in gas chromatography. Using a hydrogen-based flame and electrons produced with an anode and a cathode, it derives the mass of a substance present in an efflux and thus derive its concentration. The concentration is then transformed into an electrical signal as the instrument output is an electrical output. For the FID used, the relation between ethane concentration and tension response was linear.

Concentration statistics extracted from each time series include the mean, the standard deviation, the third and the fourth moments.

It is important to note that, all measurements were made by keeping constant the volume flow rate through the nozzle. So, the outlet velocity u^* was also kept constant. This has been possible by monitoring and recording the ambient pressure and temperature.

It is necessary to consider the effects of a recirculated tunnel. In this case, in fact, the concentration will tend to grow with time. To take into account the contribution of this drift, the background concentrations were recorded before and after acquiring any of the concentration time series. The background concentration, which was assumed to evolve linearly with time from its initial to its final value, was then subtracted from the signals.

2.3.3 Velocity measurements

Measurement techniques

To assess the dynamics of the flows, speed measurements were carried out. Two different types of measures have been performed. Firstly the hot-wire anemometry, providing information about the spectral characteristics of the velocity.

The hot-wire anemometry is based on a heater placed in a fluid current. When an electrically heated wire is placed in a gas stream, heat is transferred from the wire to the gas and hence the temperature of the wire reduces, and due to this, the resistance of the wire also changes. The change of the wire resistance becomes a measure of flow rate. There are different operating modes depending on what parameters are kept constant or variable.

The hot-wire constant temperature anemometer was equipped with a X-wire probe with a velocity-vector acceptance angle of 45, allowing for the simultaneous measurements of two velocity components.

The calibration operations ware performed in the tunnel using a Pitot tube and measuring a reference speed.

Measurements were made in points at different coordinates and for each of them a sampling frequency of 1000 Hz was used.

The experimental error, estimated by repeating the measurements in a fixed reference location, was approximately 2% for the mean and the standard deviation.

2.4 Operational organization of data

In order to manage in the most appropriate way the large amount of data obtained from the experimental activity, it was necessary to organise them in a suitable way. The data were provided in the form of text files. It was therefore decided to operate as follows:

1. The data were transformed from text files to Matlab files

2. The concept of section was created: a section represents a specific x-coordinate along the wind tunnel axis. Each section has been numbered with progressive numbers, from number 1 to number 6. The x-coordinates are all referred

to the boundary layer depth, i.e. $x_d = \frac{x}{\delta}$. Therefore, the sections obtained with the relative coordinates are:

- Section 1: x=51mm, $x_d = 0.16$;
- Section 2: x=102mm, $x_d = 0.32$;
- Section 3: x=204mm, $x_d = 0.64$;
- Section 4: x=409mm, $x_d = 1.30$;
- Section 5: x=817mm, $x_d = 2.60$;
- Section 6: x=1226mm, $x_d = 3.90$;

The sections are the same for whatever the diameter of the source considered. As can be seen, section 1 has not been considered because it is characterised by a series of criticalities due to the measurement methods.

3. Considering the 3mm source diameter case (D3), starting from section 1 to section 6, a progressive number has been assigned to each measurement point. For the D3 case, 239 measurement points have been identified and therefore as many files.

4. The operation is repeated for the case of source diameter equal to 6mm (D6). The progressive numbers were entered starting from the value 500. Here, 260 measurement points within the six sections have been considered.

2.5 Concentration time series statistical analysis

2.5.1 Concentration measurements filtering

The data analyzed here have been elaborated, under different points of view, in several theses. In one of these, an appropriate filtering technique was implemented. As it is known, before operating on data deriving from experimental procedures, appropriate filtering operations should be carried out. For example to eliminate any noise or to compensate a certain behavior of the instruments.
From the studies carried out, it was observed the necessity to filter only the concentration signal, by applying a high pass filter. That is, all concentration values below a given threshold must be considered zero. This threshold, in the D3 case, is equal to 3.320 ppm. In the D6 case, instead, it is equal to 2.754 ppm. Therefore, before any analysis was performed, the data were filtered.

2.5.2 Intermittency of concentration signals

When a phenomenon that does not happen regularly is considered, it can be described using the adjective "intermittent". The presence of null values within the concentration time series can be defined as an intermittent phenomenon. Starting from this idea, the intermittency of concentration signals can be calculated. It is defined as:

$$Intermittency = \frac{\text{Time instants with a non-zero signal}}{\text{Total time instants measured}}$$

Therefore, with the time series considered here, the denominator will always be 179985. This quantity has been calculated at each measuring point. Then, for each section, the vertical and the horizontal profile were represented. The first profile is formed by all measuring points with y=0 and different z.
Vice versa for the horizontal profile. This was done considering both the diameters of the sources. Therefore, there are four profiles obtained in each section. Two for the case D3 and as many for the case D6. Results obtained in case D3 are shown in fig. 2.3, while those obtained in case D6 are shown in fig. 2.4.

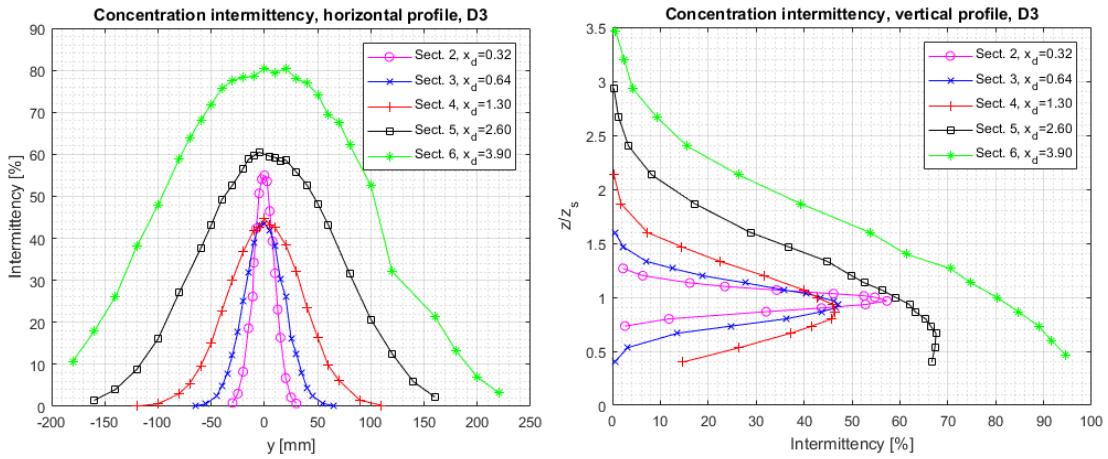

Figure 2.3: Intermittency profiles, source diameter 3 mm

For both diameters, profiles exhibit common characteristics:

- **Horizontal profile**
  These profiles should have, theoretically, a bell shape. Close to the nozzle axis, the intermittency should be close to one. While, moving towards the periphery of the plume, it should tend asymptotically to zero. The horizontal profiles obtained are reminiscent of the theoretical bell shape. This means that, at least for the shape, the floor does not have an effect on these trends.

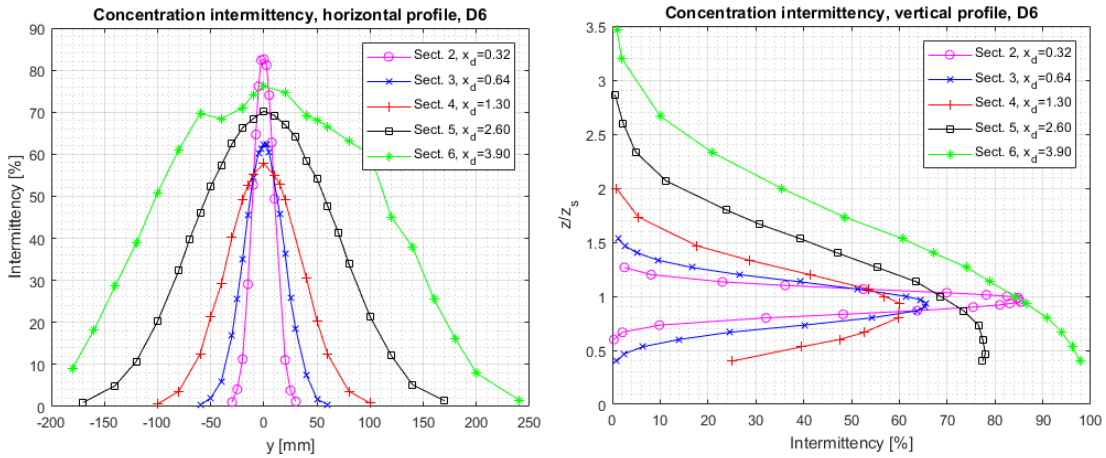

Figure 2.4: Intermittency profiles, source diameter 6 mm

However, in the theoretical case the central peak is around 100%. In the cases analysed here, instead, this central value is less than 100%. It tends to decrease from section 2 to section 4, while in the last two increase. This behaviour, probably, might be also attributed to the interaction between plume and wall.

- **Vertical profile**
  The shape of vertical profiles resembles the bell shape only in sections 2 and 3. From section 4 onwards, the vertical profile became not symmetrical. High intermittency characterise the areas under the nozzle while, above it, the intermittency tends to zero. This effect is certainly due to the interaction between wall and jet. In fact, similar trends are also found in the average concentration profiles. The maximum intermittency, even in vertical profiles, tends to decrease from section 2 to section 4, increasing afterwards. Note that, the maximum intermittency value is never at nozzle level, $z/z_s = 1$. It's always below it. Moving away from the source, the coordinate of the maximum value decreases: in section 2 it is around $z/z_s = 0.95$ while, in section 6, $z/z_s = 0.45$. Table 2.1 illustrate this concept. It shows, for both diameters, the maximum intermittency and its vertical coordinate.

Using the point with coordinates z=75 mm and y=0 mm, is possible to verify the correctness of the calculations shown here and the measurements repeatability. In fact, in each section, this point is measured twice: as part of vertical and horizontal profiles. Considering therefore this point as belonging first to a profile, and then to the other, similar results must be found. The table 2.2 illustrates the results obtained for D6 case. As can be seen, the results are comparable. This means that the same spatial point, measured at different times, is characterized by comparable

Table 2.1: Intermittency, vertical profiles: maximum values coordinates

	D3		D6	
	$(z_d)_{max}$	max int. [%]	$(z_d)_{max}$	max int. [%]
sect.2	0,97	57,30	0,95	85,31
sect.3	0,93	47,08	0,93	65,64
sect.4	0,87	46,48	0,93	60,05
sect.5	0,67	67,54	0,47	77,99
sect.6	0,47	94,65	0,40	97,78

measurements.

Table 2.2: Intermittency: common measuring points

D6	VERTICAL PROFILE	HORIZONTAL PROFILE
	Central point z=75	Central point y=0
	Intermit. [%]	Intermit. [%]
sect.2	82,52	82,61
sect.3	61,32	62,20
sect.4	56,90	57,78
sect.5	68,57	70,26
sect.6	84,13	76,12

2.5.3 Graphic representation of concentration time series

What has been said so far on intermittency can also be evaluated graphically. In fact, the time series can be represented using dots. When the value is zero, no dot is reported. Instead, when the value is different from zero, a dot is drawn. In the analysis carried out here, the first 2500 values of the concentration time series have been considered. For both diameters, section by section, vertical and horizontal profiles representation has been obtained. The results for the sections 2, 4 and 6 in the D3 case are shown in fig. 2.5, fig. 2.6 and fig. 2.7.

Looking at the horizontal profiles, so z=75 and y variable, it can be notice that:

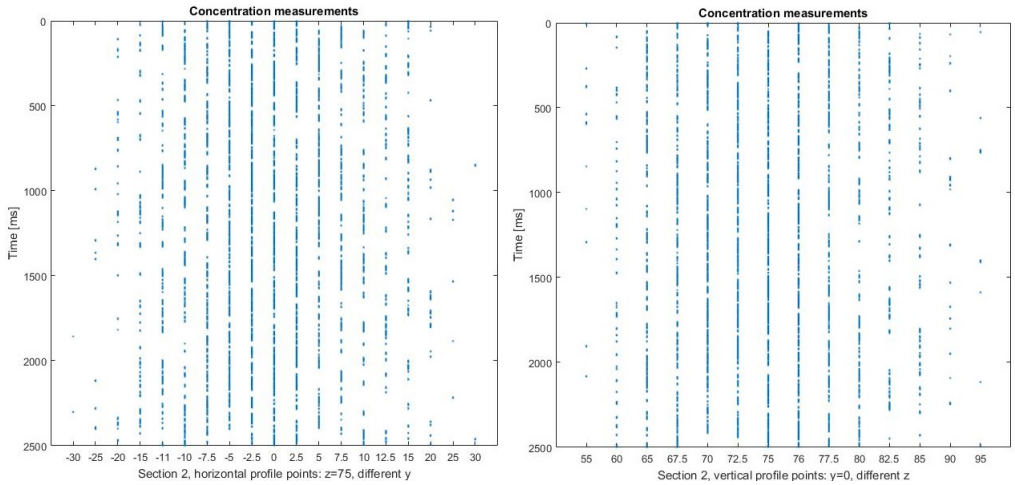

Figure 2.5: D3, section 2, $x_d = 0.32$, concentration time series representation (0-2500 ms)

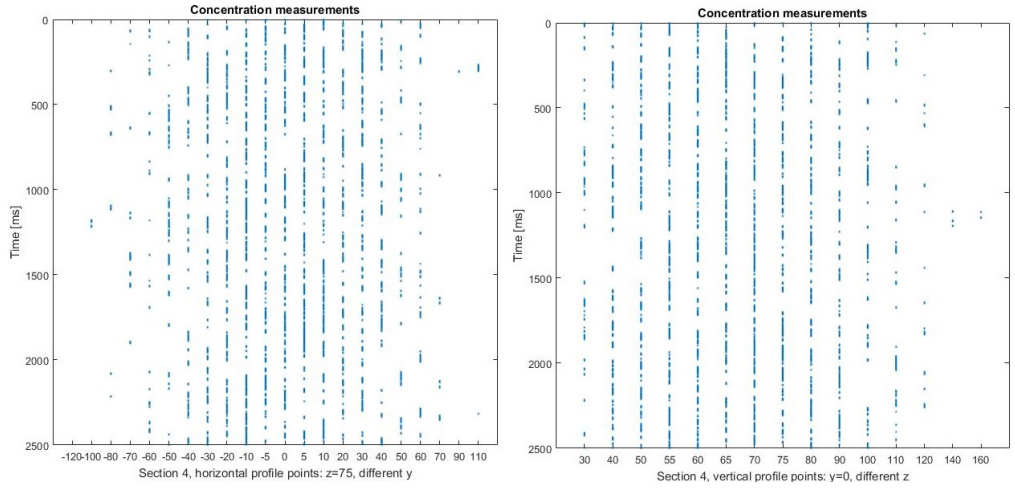

Figure 2.6: D3, section 4, $x_d = 1.30$, concentration time series representation (0-2500 ms)

- All sections have the same behaviour on horizontal direction. Around the central point, y=0, there are measurement points that present a large quantity of instants with non null concentration.
- Moving away from the centre, the quantity of points tends to decrease. This means that, for the points at the plume border, the time series has a large quantity of zeros. This occurs both on the right than on the left with respect to the central point, confirming the already observed symmetry present in the horizontal intermittency profiles.

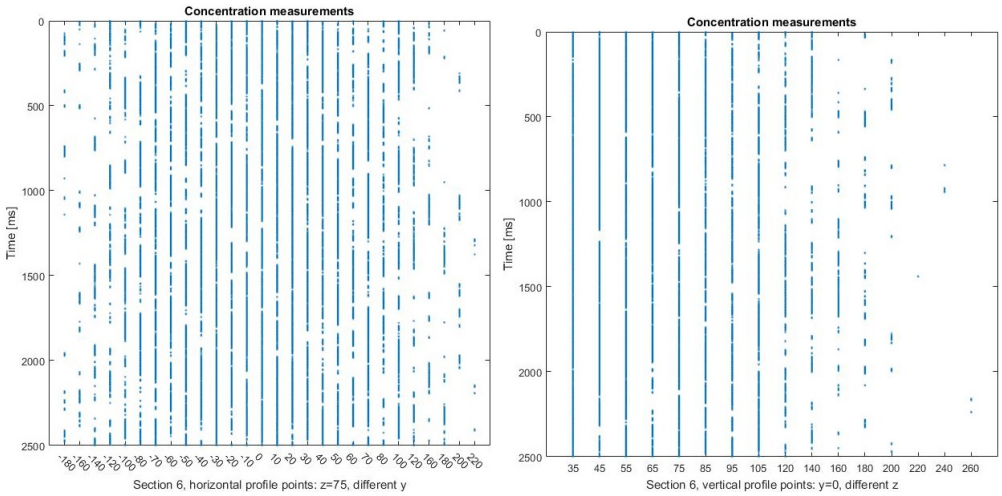

Figure 2.7: D3, section 6, $x_d = 3.90$, concentration time series representation (0-2500 ms)

Instead, vertical profiles show a very different behaviour:

- The maximum number of dots is not in the centre, $z/z_s = 1$, but at lower coordinates.
- In sections 2, 3 and 4 the profiles are still quite symmetrical. The greater quantity of dots is however located on the left side of the graph ($z/z_s < 1$)
- In sections 5 and 6 the effect of the floor is clearly present . A large amount of dots is located on the left side of the graph. Therefore, a large quantities of non-zero elements characterise the time series of points closest to the floor.

2.5.4 Average concentration profiles

In order to study the evolution of the plume, an important statistical parameter is the average concentration. That is, fixed a certain spatial point, the average temporal value that the concentration assumes. To facilitate the comparison of the results, it can be referred to the figures fig. 2.8 and fig. 2.9.
The average concentration trends are very reminiscent of intermittency:

- **Horizontal profile**
  The horizontal profiles tend to a fairly symmetrical bell distribution. So, the maximum value is in centre (y=0), while the concentration tends to zero at the maximum and the minimum y measured respectively.

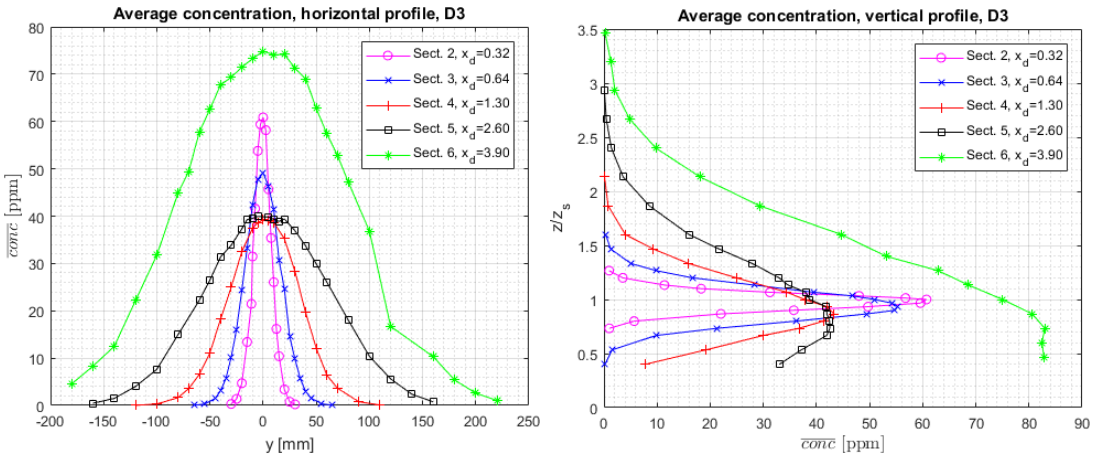

Figure 2.8: Source diameter D3, average concentration profiles.

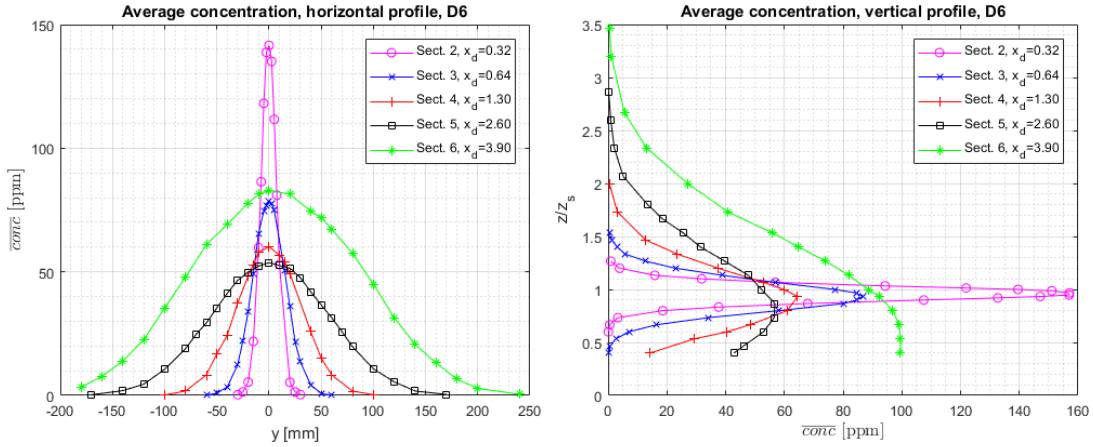

Figure 2.9: Source diameter D6, average concentration profiles.

- **Vertical profile**
  Two different types of vertical profiles can be identified: before the interaction between the jet and the wall and after it. In the first case, so in sections 2 and 3, the vertical profile is symmetrical. From section 4 onwards, instead, the effect of the wall tends to accumulate ethane in the lower part of the jet. The profiles are therefore very asymmetrical: they tend to zero for coordinates above the nozzle axis while they have a maximum for coordinates far below the axis.

Observing the maximum values, for both diameters and profiles, it can be seen how these decrease from section 2 to section 5, while going to the last section it tends to increase again. This trend, at first sight, is unexpected. However, it points out what was obtained in the case of intermittency: maximum values decreasing from

section 2 to 4 and increasing from section 5 to 6.
In order to evaluate the correctness of the data obtained, as was also done for the intermittence, in each section the two values relating to the central point were compared. This, as indicated, is measured twice. As expected, the two measurements show totally negligible differences.
The x-coordinate effect is substantially connected to the wall effect. Instead, it is interesting to evaluate the diameter effect. Logically, in the same spatial coordinate, the average concentration is greater in the D6 case. This is even more true in the initial sections. Just consider section 2 peak values: in the case D6 there is an average concentrations that is more than double those of case D3. On the other hand, in section 6 the difference is much smaller. The horizontal profile in section 6 seems almost to be about the same in the two cases. This means that, in D6 case, it is the section closest to the nozzle which has the highest average concentrations. Which does not happen in the D3 case. Here, the effect of the proximity to the nozzle can not however counteract the accumulation effect of ethane due to the wall. Therefore, the higher average concentration values are obtained in the section furthest from the nozzle, the number 6.

2.5.5 Average normalized concentration profiles

The concentration profiles shown above represent what is actually measurable: from them it is possible to infer what are the geometrical parameters that most influence the diffusion phenomenon. However, from fluid dynamics, it is known that to compare two different physical phenomena, like for example the two plumes obtainable with the two different diameters, it is advisable to use dimensionless parameters and not actual physical parameters. In fact, even if the latters are different, if the physical phenomena have the same dimensionless parameters they are very similar from the fluid dynamics point of view. The differences in dimensional variables my be due, for example, to the scale effect. This is exactly what happens in the diffusion phenomenon considered here. In fact, concentration series, at each time instant, can be normalized with respect to the ethane flow emitted from the source. Each instantaneous concentration can be normalized as $conc_n = \frac{conc}{\frac{Q_e}{u_\infty \cdot \delta^2}}$, where Q_e is the ethane volume flow at the source, which changes both with the diameter and the spatial coordinates.
In each measuring point a normalized concentration time series is obtained. Their average value may be represented exactly as done previously with the measured concentration. What have been obtained in the D3 case is shown in fig. 2.10, while fig. 2.11 shows the D6 case.
From the normalization it is possible to infer that, whatever is the diameter considered, the diffusion phenomena that are generated are completely comparable.

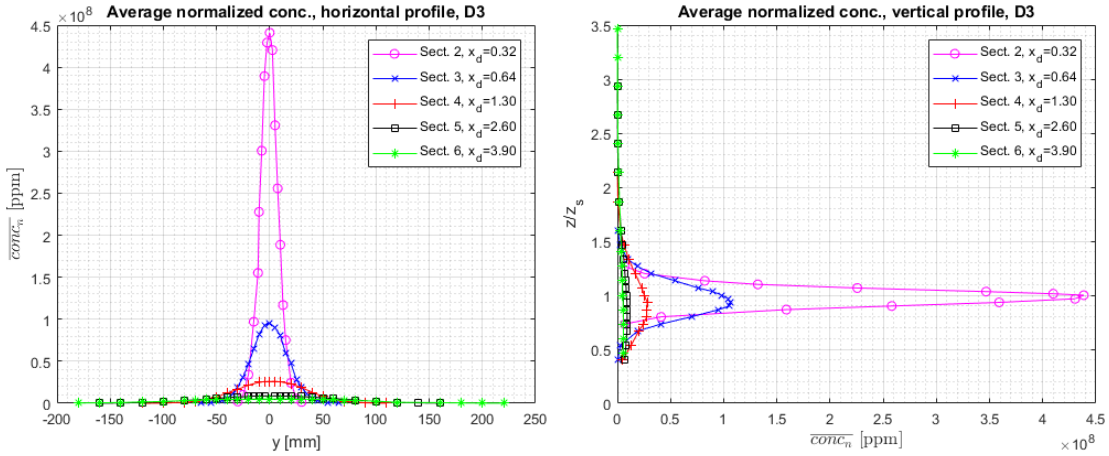

Figure 2.10: Source diameter D3, average normalized conc. profiles.

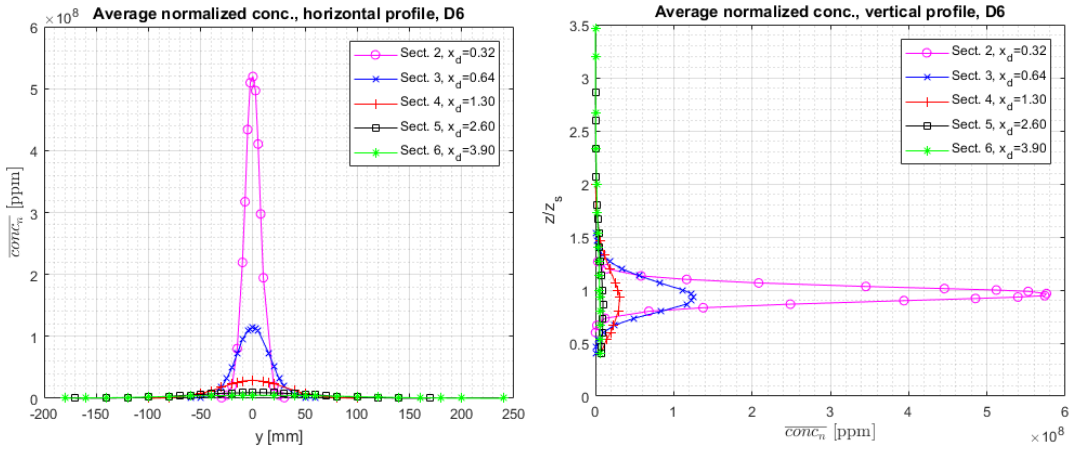

Figure 2.11: Source diameter D6, average normalized conc. profiles.

In fact, the profiles for D3 case and those relating to D6 case are very similar. It is also interesting to note that, in this representation, the effect of the longitudinal coordinate x, and thus the effect of the interaction between plume and the wall, is lost. The profiles are in fact arranged in order of section: higher values are found in the areas close to the nozzle while moving away from this the average normalized concentration tends to decrease.

2.6 Velocity time series statistical analysis

2.6.1 Average velocity u profiles

As for the concentration time series, also for the series of the velocity along the wind tunnel axis, so the velocity u, the average profiles can be built. The figures fig. 2.12 and fig. 2.13 illustrate what has been achieved.

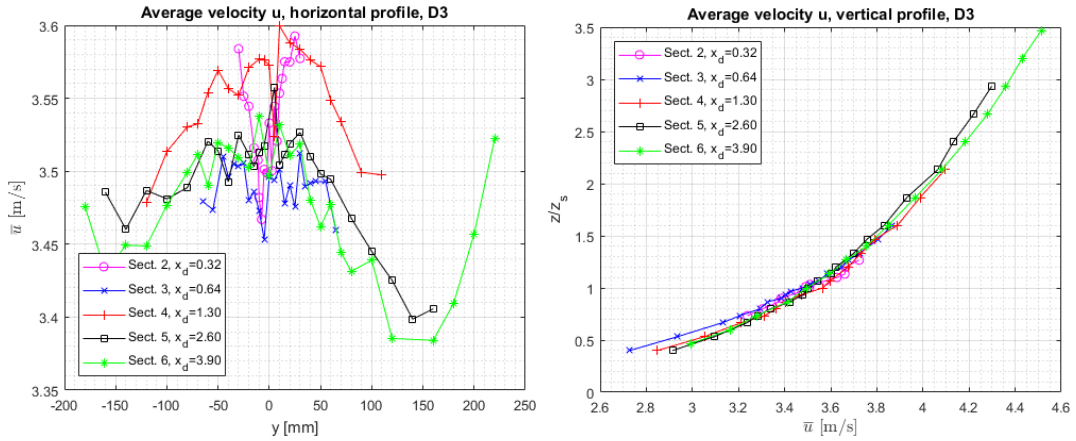

Figure 2.12: Source diameter D3, average velocity u profiles.

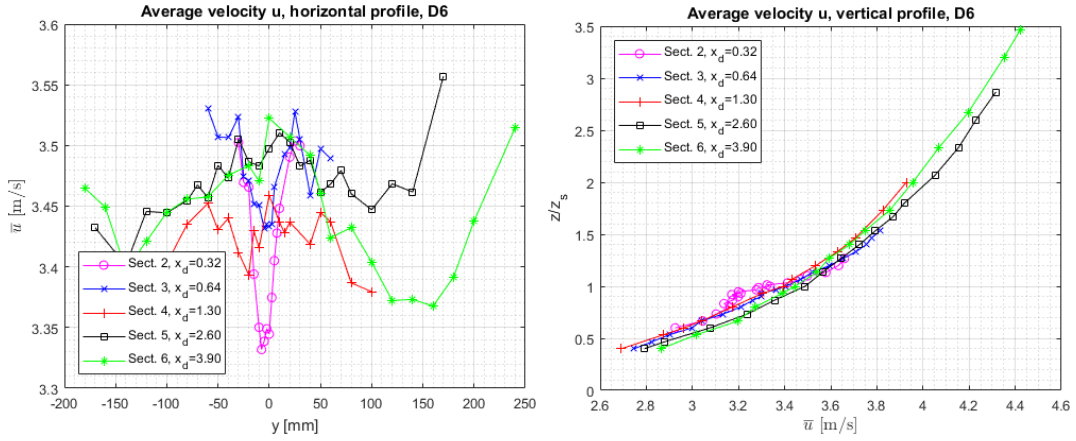

Figure 2.13: Source diameter D6, average velocity u profiles.

As is evident, fixed a certain diameter, there is a big difference between the vertical and the horizontal profile. The last one is characterized, in all the sections, by a velocity that oscillates around the 3.5 m/s. That is, the velocity u can be considered approximately constant at each point of the horizontal profile of each section. Instead, there is a very different situation for vertical profiles. Here the

braking effect of the wall is evident. In fact, the profiles are parabola-like profiles and the speeds vary considerably according to the z-coordinate.
It is interesting to observe a great difference between the average concentration and the average velocity u profiles. The former have a strong dependence on the diameter of the source. In the case of u instead, the vertical and the horizontal profiles, for both diameters, have very similar numerical values.

2.6.2 Average velocity w profiles

Also for the velocity w is possible to represent, for both the diameters, the vertical and the horizontal profiles of the average values. The fig. 2.14 illustrates what has been obtained in the D3 case while the fig. 2.15 refers to the D6 case.

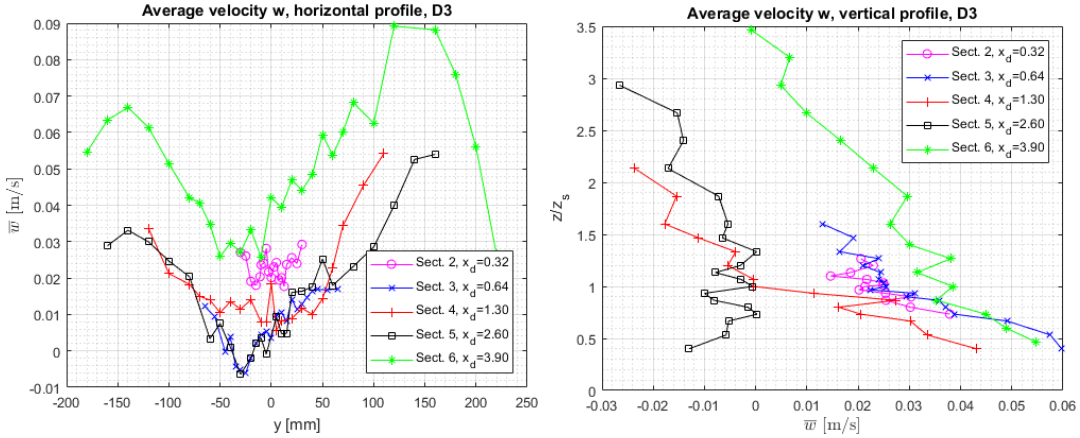

Figure 2.14: Source diameter D3, average velocity w profiles.

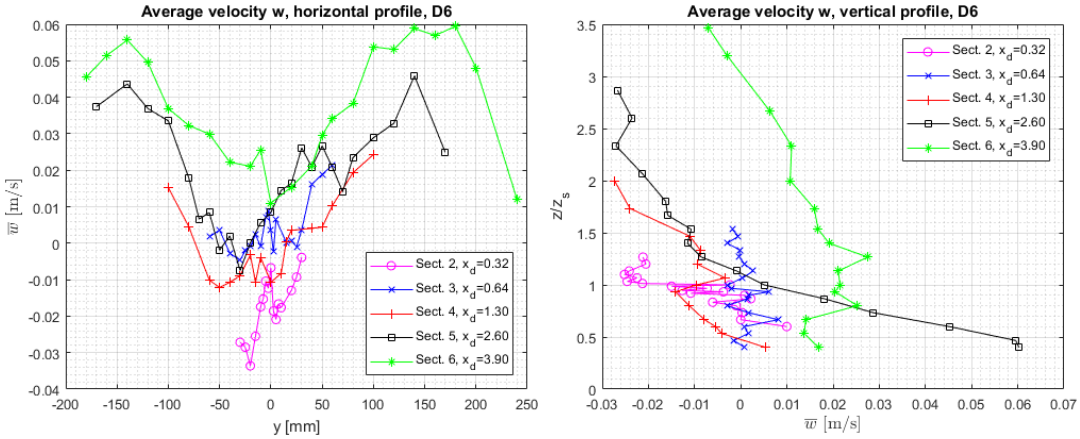

Figure 2.15: Source diameter D6, average velocity w profiles.

From these graphs a very important information may be derived. Observing the two vertical profiles, it is possible to see that in some spatial points the average velocity is positive, while in others it is negative. This means that an integral of these profiles would get a value near zero. This null value indicates that the considered phenomenon is non convective. That is, there is no a velocity w that tends to drive the plume in the vertical direction.

2.7 Graphical representation of the plume

It is interesting to give a graphical representation of the considered plume. This in order to identify the proportions between the different geometrical parameters and to identify, compared to the geometry, the concentration distribution. To this end, fig. 2.16 can be consider. It refers to the D6 case and it is represented in scale. Blue lines indicate the vertical coordinates in which, section by section, the sensors have been inserted. Red lines, instead, try to represent, using some kind of constraint, the plume. They identify the z-coordinates where the intermittency of the concentration is greater or equal to 45%. As can be seen, compared to the total number of measuring points, just few points, especially above the nozzle axis, respect this constraint.

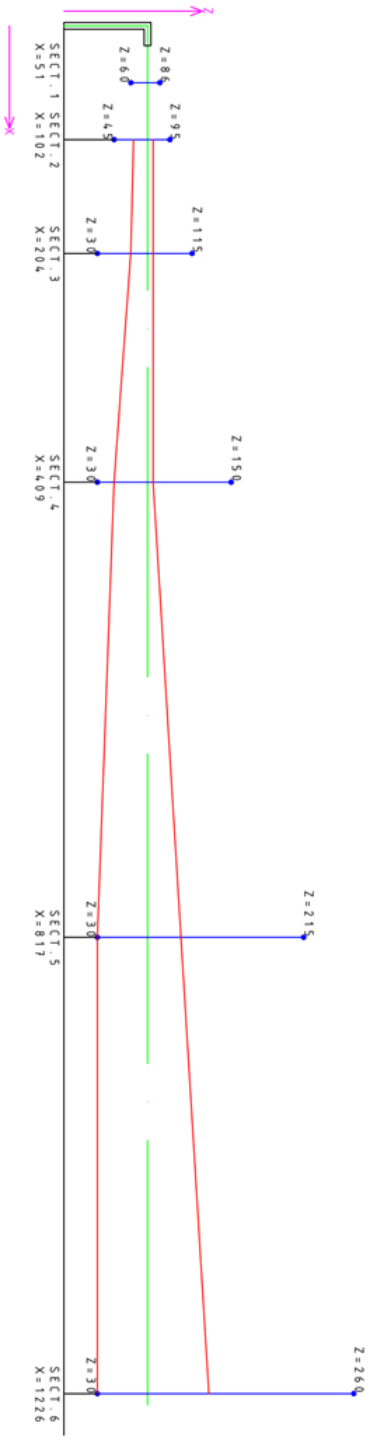

Figure 2.16: Intermittency: scale drawings. The blue lines identify the areas in which, section by section, the measuring sensors have been inserted. Instead, the red ones identify the z-coordinates where the concentration intermittency is greater or equal to 45%.

Chapter 3

Complex network analysis of time series

In the first part of this chapter ample space is given to the complex networks theory. Starting from its historical evolution, the main properties of networks, their topology and the metrics that identify their salient characteristics are described. Some examples of main modern fields in which complex networks are used defines their versatility and their wide development given by the use of computer techniques. Instead, the second part of this chapter is devoted to the techniques that allow to study the time series through the use of the complex networks theory. Among these, particular attention is paid to the Recurrence Networks (RN). It is through this technique that the experimental time series introduced in the second chapter are studied. The chapter concludes with a detailed description of the mathematical operations and the parameters needed to obtain complex networks starting from time series through the use of the recurrences, that is, through the distances between series fragments in the phase space.

3.1 Network theory: historical notes

The birth of graph theory is usually fixed in 1736, when L. Euler[1] tried to answer the famous question known as the "problem of the seven bridges of Königsberg" [7]. The city of Königsberg, located not far from St. Petersburg, had an islet enclosed between the two branches of the river Pregel, connected to the mainland by seven bridges as shown in fig. 3.1. The following question was then posed: "Is it possible to take a walk along the seven bridges without ever crossing the

[1]Leonhard Euler, known in Italy as Eulero (Basel,1707 - St. Petersburg,1783), was a Swiss mathematician and physicist.

same twice?". Tradition wants the wealthy citizens of Königsberg on Sundays to stroll around their city, trying in vain to solve the problem. Euler approached the problem in a different way: building a graph. Each of the four areas of land was represented using a node, each bridge was represented by a link. The graph thus generated, consisting of four nodes and seven links, allowed to give an answer to the problem. In fact, Euler discovered that on a graph with more than two nodes and with an odd number of links, such as the Königsberg graph, a path with the required characteristics could not be drawn.

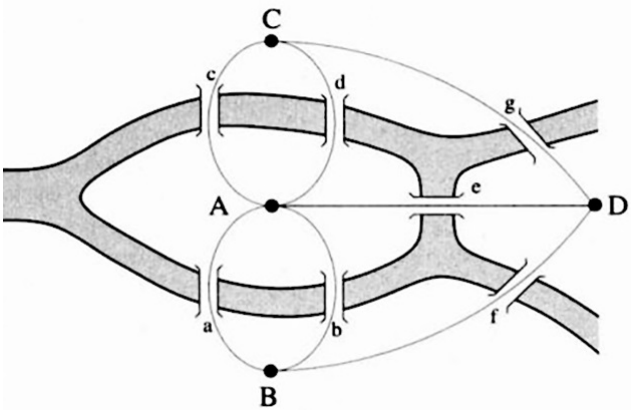

Figure 3.1: The seven bridges of Königsberg

The work on the Königsberg bridges is considered the birth of modern graph theory. Theory that, in the years after Euler, had a great development. In fact, it is extremely versatile to represent real problems in different fields. The basic idea is very simple: use nodes to represent certain entities and links to identify their connections. The graphs thus obtained are a very powerful tool in various fields [8]. Just think at the development that the theory had after Euler, with the studies of Kirchhof and Cayley. The first tried to describe the electrical networks using graphs. In doing so he identified some of the main electrotechnics laws: the junction rule and the mesh rule. Cayley faced the hydrocarbon isomers problem. There are molecules that have the same number of atoms, but they are bound in different ways. These molecules are called isomers. Cayley sensed the potential of the representation of these molecules using graphs and he defined two different chemical graphs typologies: the plerogram, containing all the atoms present in the molecule, and the kenogram containing only the carbon atoms. For clarity, fig. 3.2 illustrates a three-dimensional graph of two propanol isomers.

Interesting applications of the graph theory are also in the discussion of the "Four

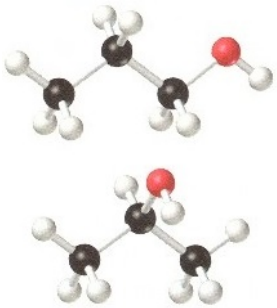

Figure 3.2: Three-dimensional graph of two propanol isomers

colour problem", formulated by Francis Guthrie[2] in 1852.
Paul Erdos and Alfred Renyi in 1959 proposed the theory of random graphs. In a random graph, the probability of connection between two nodes is the same for any pair of nodes in the network and it is randomly distributed. In this model, the network is almost completely interconnected. So, you can reach any other node in the network in few steps. Starting from this model, Stanley Milgram[3] defined his theory of the "Little World". With an experiment based on parcel mailing, he showed that two people are separated by a chain of six links. In this way the average distance between two randomly selected people was defined. This social experiment, as well as the theory of random networks, did not consider many factors. In fact, geographic distances, being part of a tied group of friends or working for a given company are aspects that may create social networks characterized by clusters presence. A cluster is a group of nodes very connected with each other but with poor connections with other external nodes. Starting from this concept, Mark Granovetter[4] in 1973 realized that the weakest links, such as those between nodes belonging to different clusters, are the most important links in a social network. Indeed, these connections prevent networks from fragmenting.
Starting from these preliminary studies the theory is still evolved at the end of the last century. The definition of "small-world" networks and "scale-free" networks has allowed more detailed analyses. These new theories, together with the large amount of data relating to social networks, deriving from the World Wide Web and mobile communication, and together with an ever-increasing computing capacity guaranteed by more and more performing computer tools have allowed the analysis of networks, often dynamic, containing millions of nodes and links.

[2]Francis Guthrie (London,1831 - Cape Town,1899) was a South African mathematician and botanist.

[3]Stanley Milgram (New York,1933 - New York,1984) was an American psychologist.

[4]Mark Granovetter (Jersey City,1943) is an American sociologist.

Thanks to modern information technologies, the ever-increasing amount of data and the "easy way" to describe the reality using complex networks theory, it can be considered that in the future it will still be subject of studies and researches. So, despite its application for about 280 years, it can be considered a young and ever-growing theory.

3.2 Network theory: applications

The graph theory is a completely general tool: this is definitely his big advantage. In fact, it is applicable in different fields, also very different from each other.
Studying the historical evolution of the theory, it is possible to observe that one of the fields in which it has been widely applied is the study of social relations. Now, due to large amount of traceable social relations, for example Internet and mobile telephony, the analyses on social dynamics can be based on a greater amount of data, often easily detectable, that are not affected by geographic and time barriers.
In addition to social relations field, there are countless other application. Given the vastness, here are shown three examples of applications that want to make clear the idea of general applicability of this theory.
In the financial field [9], the theory of complex networks, and in particular its integration with the study of time series, has allowed to correlate catastrophic events that have characterized the market in the past. In fact, a correlation between the changes in network structures and stock market critical issues has been defined.
The networks theory has been successfully applied also in the seismology [10]. With it, studies to describe the seismic phenomena that occurred in the last decade in Italy have been carried out.

3.3 Network theory: Definitions and main properties

3.3.1 Definitions

The term graph means a set of entities, in which some of these entities are "related". In fact, a graph is formed by nodes and links. The nodes represent the entities to study, while the links identify the connections that exist. To analyse a real problem with a graph, the first thing to do is to find out which entities will form the nodes of the networks. In a social network, for example, the nodes will represent people. Instead, if you want to study a district heating system, the nodes will be the thermal utilities distributed in a city. Once you define the "objects of the study", so the nodes, you must identify the connections between them, that

are the links of the graph. Referring to the two previous examples, in the first case you will have links between people who know each other. While, for a hydraulic network, the connections can represent the pipelines.
Note that, in this thesis, the terms graph and network will be used as synonyms. As well as links and connections.
Mathematically, a graph is defined as $G = (V, E)$. V denotes the set of nodes and E the set of arcs, which is 2-element subsets of V [11]. The number of nodes is usually indicated with N. Obviously, V should not be an empty set. Instated, this condition is not required to E. If $E = 0$, the minimum number of links in the graph is identified. Two nodes connected by a link are called adjacent. Each node can be identified by an integer value $i = 1,2, ..., N$ while each edge can be identified by a pair (i, j).
There are different types of graphs:

1. **Undirected graph**
   An undirected graph is a graph in which links have no orientation. The connection (i, j) is identical to (j, i). The maximum number of edges in an undirected graph without loops is $\frac{N(N-1)}{2}$.

2. **Directed graph**
   A directed graph is a graph in which edges have orientations. If G is oriented, then the set E consists of ordered pairs of nodes: the connection (i, j) is called oriented link and is characterised by a direction from i to j. The maximum number of edges in an undirected graph without loops is $N(N - 1)$.

3. **Weighted graph**
   A weighted graph is a graph in which a number (the weight) is assigned to each edge. Such weights might represent for example costs or lengths, depending on the problem. For instance, consider a graph describing an urban district heating system, the weight of each link can be equal to the pressure drop that occurs in that pipe. In this way, working on mathematical quantities of the graph, it is possible to define the head of the pumps needed to guarantee the circulation of the heat transfer fluid.
   The weight of each link can be represented using a scalar value $w(i, j)$. In matrix notation, a weighted graph can be completely represented by its weight matrix W, where $w_{ij} = w(i, j)$.

4. **Unweighted graph**
   The graphs in which all the links have weight $w(i, j) = 1$ are define as unweighted or binary graphs. In this case, a link that connects two nodes can just be present or not. It can not have different characteristics from the other links.

For understanding the first two types of graphs, fig. 3.3 can be considered. Two graphs are represented in it. On the left, the links are characterised by arrows that establish a direction. The graph is therefore a directed graph. Conversely, in the graph on the right, there are no links with defined direction. It is therefore an undirected graph.

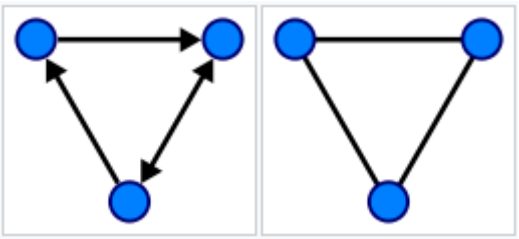

Figure 3.3: Comparison between undirected and directed graph

The type of graph to be used depends on the relationships that exist between the nodes to be connected. For example, if the vertices represent people at a party, and there is an edge between two people if they shake hands, then this graph is undirected because any person A can shake hands with a person B only if B also shakes hands with A. In contrast, if any edge from a person A to a person B corresponds to A's admiring B, then this graph is directed, because admiration is not necessarily reciprocated.
A graph $G = (V, E)$ is said complete when all nodes have a link that connects them to all the others nodes. A network is called completely disconnected if $E = 0$. Usually it is assumed that there are no self-connections (or loops), i.e. each vertex has not a link with itself. That's why, given a network of N nodes, the maximum number of connections is $N(N-1)$ and not N^2. Obviously, this in the case of directed networks. In fact, in this case, the connection from node i to node j may be active or not regardless of the presence or not of node j and node i connection. In the undirected networks, instead, these two connections coincide. Therefore, a split two appears into the maximum number of links for the undirected graphs.

An important definition in graph theory is the path. Two nodes that are not adjacent may be reachable through an sequence of nodes and edges that begins and finish with the two nodes. A path is a walk in which no node is visited more than once. The shortest path is therefore the walk of minimal length between two nodes. When graph theory is applied to real problems, graphs with a high number of nodes are often obtained. Therefore, the graphical representation is not recommended. As done in this thesis, once the graph is generated, its properties, its metrics and its characteristics must be evaluated, in order to obtain information about the real system under examination. It is clear that, if the graph is only in

graphical form, these operations are impossible.
A more compact form to represent the network is the adjacency matrix. The elements of the matrix indicate if two nodes are adjacent or not. The use of a matrix allows to generate computer algorithms to calculate the main properties of the graph under examination.
The adjacency matrix definition depends on the graph type. Given an unweighted graph G, the adjacency matrix A consists of a square binary matrix NxN. The row and column indexes are the names of graph nodes. The element a_{ij} of the matrix will be 1 if there is a link that goes from node i to j. Otherwise $a_{ij} = 0$. It is easy to see that, if the graph is undirected, the matrix A will be symmetric, i.e. $a_{ij} = a_{ji}$. Instead, if the graph is directed, the matrix may not be symmetric. Fig. 3.4 shows an example related to an undirected graph. Node G is connected only to node F. Accordingly, in row G of the matrix there is a 1 in correspondence of column F. Instead, a 0 is in all the others columns. So, the matrix is symmetric, and on the diagonal there are only 0 because no self-loops are present in the graph.

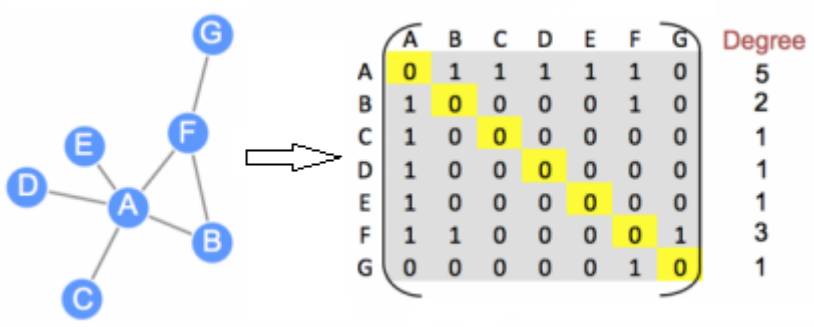

Figure 3.4: Example of graphical and matrix representation of a network

3.3.2 Metrics to describe the network structure

As indicated, the use of networks based models is a very powerful tool. In fact, it is possible to identify characteristics of the system under consideration evaluating the topology of the networks appropriately obtained. To do this, network metrics are necessary.
The metrics are indicators describing the topology, and therefore the structure, of the networks. In literature, a large amount of metrics and parameters are described . This is due to the fact that each single metric highlights different aspects of the network. For example, for the study that is taking place, may be appropriate determining the importance of each node, so the amount of links that are connected to that node. Or, it may be important evaluating a metric that describes the distances between nodes. That is, according to the dynamical

system under study and according to the properties of the networks that should be highlighted, the most appropriate metrics must be chosen. Therefore, there is not a set of metrics to be applied in bulk in order to describe a certain system: the choice varies from case to case. Below, a brief description of the metrics used in this thesis is reported.

Degree centrality

The degree centrality of a node i, k_i, is defined as the number of neighbours, i.e. the number of nodes that are directly connected with i. Considering a directed network, there are two different degree: one relative to the links entering the node i and the other relative to the links coming out from the node i:

$$k_i^{out} = \sum_{j=1}^{N} a_{ij}$$

$$k_i^{in} = \sum_{j=1}^{N} a_{ji}$$

Obviously, the overall degree will be given by the sum of the two contributions:

$$k_i = k_i^{out} + k_i^{in}$$

For the undirected networks, i.e. the ones analyzed in this work, it is not necessary to separate the degree into two contributions, since there are not two different types of links connected to each node.
The degree centrality estimates the ability of the nodes to have a direct relationship with other nodes. From this point of view, the concept of node importance can be defined as the amount of connections that this node presents within the network. Starting from the concept of degree centrality, it is possible to define two other metrics connected to it:

- **Average degree**
  The average degree is the average of the degrees of each node in the network.

$$\overline{k} = \frac{1}{N} \sum_{i=1}^{N} k_i$$

  From what has been indicated, it is easy to understand that the degree is a metric that is definable for each node of the network. So, it can be considered as a local metric. While the average degree is a metric that is defined globally for the network. It is therefore considered as a global metric.

- **Local connectivity**
  To identify the importance of each node, in the sense defined above, the degree may not be sufficient. In fact, it is interesting to evaluate the number of effective connections with respect to the maximum number of possible connections. The latter is defined as $(N-1)$. The presence of the minus one indicates that each node is at most connectable with each other node of the network but not with itself, thus avoiding the presence of self loops. It is therefore possible to define, for each node, the local connectivity as the ratio between the degree of that node and the maximum number of possible connections:
  $$LC_i = \frac{1}{(N-1)} \sum_j^N a_{ij}$$

Edge density

This metric describe the amount of links actually present. Given a network formed by N nodes, the number of links present between these nodes can vary with respect to many causes. The concept of the threshold that leads to the definition or not of a link between two nodes will be extensively treated in the following . However, defined the number of nodes N, the maximum number of possible links is easily defined as $N(N-1)$. So, comparing the number of links actually present with the maximum number of connections, the edge density is defined as:

$$\rho = \frac{E}{\frac{N(N-1)}{2}}$$

This metric leads to two interesting aspects. Firstly, it can be shown that, for an undirected network, the average local connectivity and the edge density are proportional. Therefore, in the remainder, only the density will be considered. Secondly, it is important to observe that this metric refers directly to the study of the adjacency matrix. In fact, for a certain adjacency matrix, the density represents the percentage of values equal to one with respect to the matrix elements (always without considering diagonal values).

Closeness centrality

The closeness centrality provides the measure of the distance of a node from all the other nodes. The idea is that an important node is typically close to the other nodes in the network. So, the closeness centrality is defined as the reciprocal of the total distance from a node to all the other nodes:

$$c_i = \frac{N-1}{\sum_{i=1}^N l_{v,i}}$$

If there is no path connecting two nodes, the maximum shortest path length in the graph, $(N-1)$, is used in the sum by definition [12].

3.4 Recurrence network analysis

3.4.1 Complex network analysis of time series

Characterizing dynamical processes in a time-dependent complex system, from observed time series of just one or more variables, is a fundamental problem of significant importance in many fields. Thanks to modern measurement and data processing techniques, in order to characterize a time-dependent complex system, it is possible to use measurements "directly carried out in the field". These measurements, characterized by an increasing accuracy, repetitiveness and sampling frequency, lead to time series. A time series is a series of data listed in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Two different types of time series may be defined:

- **Univariate time series**: consists of a single scalar observation, recorded sequentially over equal time increments. For instance, the monthly CO_2 concentration.

- **Multivariate time series**: in this case, more than one quantity is measured simultaneously. An example is used to illustrate a bivariate time series. Inside a gas furnace, air and methane are combined in order to obtain a mixture of gases containing CO_2. The input series x_t is the methane gas feedrate and the CO_2 concentration is the output series y_t. Successive pairs of observations (x_t, y_t) can be collected obteining a bivariate time series.

Time series have great potential for characterizing important properties of complex dynamical systems [13]. In order to extract meaningful statistics and other characteristics, time series analysis may be implemented. These analysis has been broadly adopted in engineering applications. In fact, many theoretical developments for time series analysis have significantly contributed to the understanding of complex systems. Different methods have been developed: chaos analysis, trend estimation, curve fitting, fractal analysis, complexity measure, multiscale entropy, etcetera. As normally happens in the case of dynamic analyses, also time series analysis can be done in two different domains. Methods developed in time or in frequency domain may be chosen, depending on the purposes.
However, when system complexity increases, it becomes difficult to describe the dynamical behaviour from time series. Traditional analysis methods have difficulty to deal with this increased complexity. So, recently, the complex network

theory has been incorporated into the analysis of time series. As described before, a complex network is shorthand for real entities characterised by a large number of components, that interact with each other in complex way. Therefore, in addition to this, complex networks can be used to analyse univariate or multivariate time series.
The literature presents different methods to implement complex network analysis of time series. Among these, the most important are:

- Visibility graph networks, VG
- Cycle networks, CN
- Transition networks, TN
- Recurrence networks, RN
- Synchronization networks, SN

In this thesis, the approach used to analyzing time series is the recurrence networks method.

3.4.2 Recurrence Networks

Many dynamical processes exhibit recurrences in phase space, as showed in the pioneering work of Poincaré in the late 19th century. In time series analysis, the quantification of recurrence properties in phase space has attracted interest. In fact, recurrences can be easily visualized by means of the so-called recurrence plots. They are automatically manageable with computer codes and, thanks to appropriate operations, they can be represented by the use of binary matrices.
Consider a time series $x(t)$, with $t = 1,2,...,N$, where N represents the number instants considered. A state, in the m-dimensional phase-space, may be define has [12]:

$$\mathbf{x}^m(t_i) = [x(t_i), x(t_i + \tau), ..., x(t_i + (m-1)\tau)]$$

where t_i is the point in time associated with the ith observation recorded in the time series. In the previous formulation, m represents the vector size while τ the interval between the temporal instants forming the vector. In order to simplify the notation, $\mathbf{x}^m(t_i)$ will be indicated as $\mathbf{x}_i$ in the following. Observing the definition of a state in the phase space, it is possible to notice the concept of recurrence plots does not require observations that are equally spaced in time. This is not true in many other methods of time series analysis.
A state in the phase space, i.e. the vector define before, is also called embedded

vector. So, the words "state in the phase space" and "embedded vector", will be considered synonyms.
In the context of time series analysis, there is a recurrence in phase-space when two state, for example $\mathbf{x}_i$ at time t_i and $\mathbf{x}_j$ at time t_j, are similar, i.e. $\mathbf{x}_i \approx \mathbf{x}_j$.
For the visualization of recurrences in phase space, one of the most powerful way is the use recurrence plots (RPs). A RP represents all recurrences in form of a binary matrix R. This is a square and symmetric matrix, with a number of rows and columns equal to the number of embedded vectors obtained from the time series in question. So, at each state in the phase space corresponds a row and a column of the matrix R. R_{ij} is 1 if the state $\mathbf{x}_i$ is similar to $\mathbf{x}_j$, R_{ij} is 0 otherwise. Appropriate mathematical tools must be applied to identify the closeness between two different states in phase space. One of these tools can be the Pearson's correlation coefficient (PCC). Secondly, a norm of the two vectors difference can be a way to quantify their closeness. Considering latter case, the generic element R_{ij} will be the norm described before. In doing so, R will not be a binary matrix. To do that, a threshold ϵ must be applied:

$$R_{ij} = 1 \text{ if } R_{ij} > \epsilon$$

$$R_{ij} = 0 \text{ if } R_{ij} < \epsilon$$

The same concept can be expressed in a more compact way through the use of the Heaviside function θ:

$$R_{ij} = \theta(\epsilon - ||\mathbf{x}_i - \mathbf{x}_j||)$$

RPs of dynamical systems with different types of dynamics exhibit distinct structural properties. In other words, experimental time series often yield a recurrence plot displaying complex structures, forming diagonal or vertical "line" structures. A variety of statistical characteristics of the length distributions of these lines can be used for defining additional quantitative measures, that characterise different aspects of the dynamic complexity of the studied system. This is known as recurrence quantification analysis (RQA). For example, a periodic regime is reflected by long and non-interrupted diagonal lines. The vertical distance between these lines corresponds to the oscillation period. A chaotic dynamics also leads to diagonals, which are however clearly shorter. However, most of these RQA measures are sensitive to the choice of embedding parameters, which are found to sometimes induce spurious correlations in a recurrence plot.
As showed, the structure of R allows to identify information on the dynamics of the system under study. But, this is not the only potential of RPs. It is possible to reconsider the concept of recurrences in phase space for defining complex network structures directly based on time series. For this purpose, it is straightforward to interpret the recurrence matrix R as the adjacency matrix A of an unweighted

and undirected complex network, which is called the "Recurrence network" RN, associated with a given time series. The adjacency matrix can be defined from R easily:

$$A = R - \delta_{ij}$$

where δ_{ij} is the Kronecker delta introduced to avoid artificial self-loops.
In comparison with similar network-based techniques, the RN has important conceptual advantages, and can be considered as a unifying framework for transforming time series into complex networks that also includes other existing methods as special cases. Quantitative characteristics of the recurrence matrix, such as average path length, clustering coefficient, and the centrality are measures of the RN. But, they are directly related to the dynamical complexity of a time series. This allows to study the time series using the complex networks theory.
Therefore, the fundamental concepts of a RN are:

1. Starting from the concept of recurrences in phase space, the recurrence matrix of a time series is interpreted as the adjacency matrix of an associated complex network, which links different points in time if the considered states are closely neighboured in phase space.

2. The consideration of recurrence plots as graphical representations of complex networks allows a reinterpretation of many network-theoretic measures in terms of characteristic phase space properties of a dynamical system. So, quantitative descriptors of the topological features of recurrence networks can be considered as novel and complementary measures to describe systems.

3. This method, introduced through the use of a binary and symmetrical matrix, i.e. introduced with undirected and unweighted graph, can be easily used even in the case of weighted networks. A generalization to weighted networks is straightforward if the recurrence matrix is replace by the associated distance matrix between pairs of states.

3.4.3 Network construction, states in phase space

The previous theoretical description clarified the method used in this thesis for the study of time series. As indicated in the preceding chapter, by means of experimental procedures, for each measuring point inside the wind tunnel, three different time series have been measured. Two represent the velocities, u and w, while the third is the concentration. Each series is built from about 180 000 elements.
In order to create RNs, the first step is to generate the states in phase space. The formula by which the embedded vectors (EV) are generated from the time series has been illustrated above. It highlights that the embedding parameters are two:

1. m: is the EVs size;
2. τ: is defined as delay and indicates the distance between the elements that are selected in the time series.

Actually, another important parameter does not appear inside the EVs formulation: the overlap. To understand in detail the meaning of these three parameters, fig. 3.5 can be considered. It shows the creation of two EVs by imposing m=20, τ=5, both in the case of 0% overlap than 50%. For easier comprehension, the time series consists of progressive numbers. Starting from the first element of the time

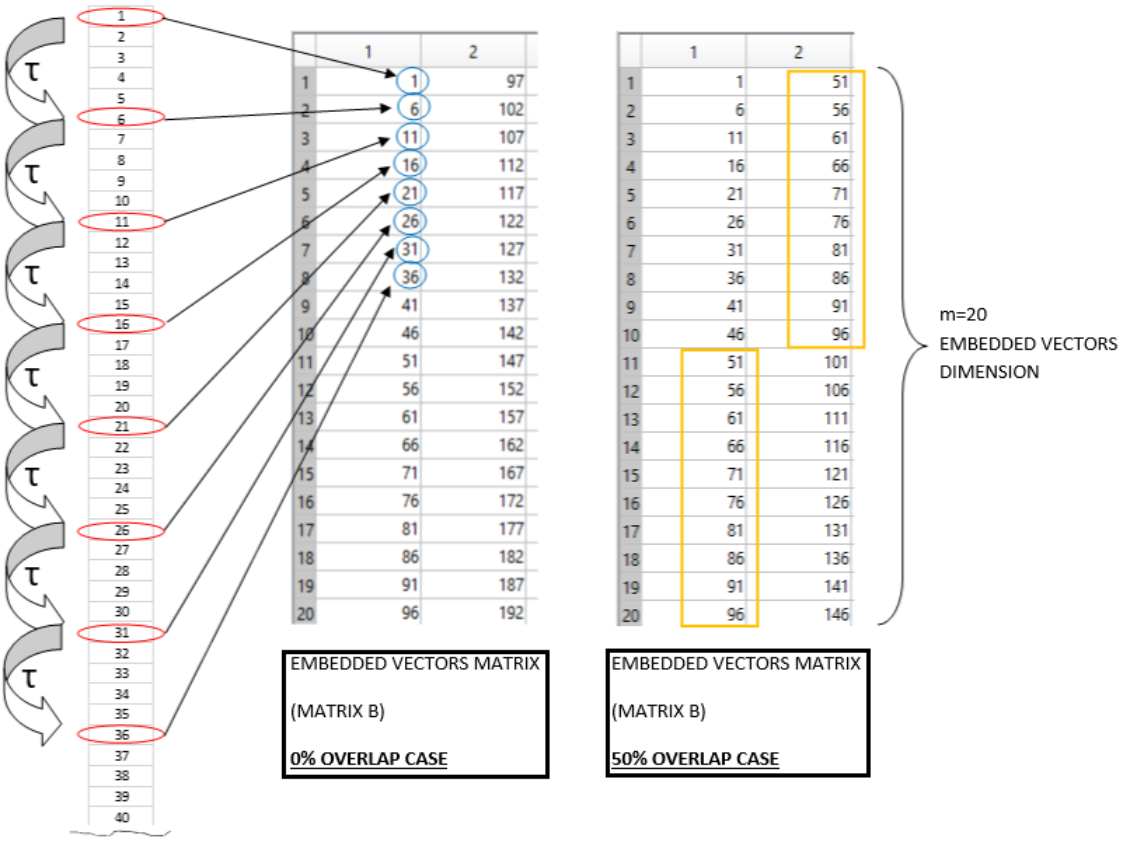

Figure 3.5: Embedding parameters

series, through τ, the m elements that will form the 1st EV can be identify. In the figure, they are circled in red. The definition of the second EV depends on which overlap is considered. In the 0% case, the first elements of the 2nd EV is the next element of the last component of the 1st EV. From this element, again through τ, other m elements are identified in order to fill the 2nd EV. The procedure continues in the same way in order to define all the others EV. In the case 50% of

overlap, the first m/2 elements of the 2nd EV are equal to the last m/2 elements of the 1st EV. This is the meaning of overlap: the first half of an EV is equal to the last half of the previous EV. Then, the last m/2 elements are again identified with the τ parameter. Continuing this operation, all the EVs of the case 50% of overlap will be defined. Therefore, through this description two things are clear:

- Considering same values for m and τ, in case of 50% of overlap, the number of EVs is greater than in the 0% case. For example, with the embedding parameters considered, 3598 EVs are obtained in the case with overlap and 1874 in the case without overlap. Therefore, it can be said that the overlap generates a number of EVs that is twice the one generated in the other case.

- It is clear that some elements in the terminal part of the time series, even if corresponding to the identification criteria outlined above, will not constitute any EV. This is due to the fact that even the last EV must be composed of m elements: it can not be smaller. The generated code automatically evaluates which is the last EV that can contain m elements and this vector will be the last state in the phase space considered for the time series analised.

To not weigh down the text, the case with 50% of overlap will be indicated as "the overlap case", i.e. OC. On the other hand, when indicating the case without overlap, i.e. WOC, of course, we will refer to the case with 0% of overlap.
Once calculated, the EVs are stored as columns of a matrix, indicated as matrix B. This matrix, therefore, will have a number of columns equal to the number of EVs and will be characterized by m rows. A matrix B will be obtained for each considered variable (u, w, concentration), for each tau and for each overlap. For example, assuming that the series taken into consideration in the previous example were the velocity u time series, the two matrix B that are obtained are so named:

$$B_{u,\tau 5,0}$$

$$B_{u,\tau 5,50}$$

where the last digit in the name indicates the overlap.
Particular attention must be paid to the concentration EVs. As indicated in the previous chapter, concentration measurements must be appropriately filtered using a threshold. So, especially for measurement points that are located in the peripheral areas of the jet, this involves the presence of a large amount of zeros within the concentration time series. To avoid the presence of EVs containing a large number of null elements, it is necessary to make a "correction" of the EVs obtained. Therefore, each columns of the matrices B_{conc} are analysed and the number of null elements contained is evaluated. If this number is greater than $m/2$, then the column, and so the EV, is discarded. So, for the concentration, the

matrices B are constructed and than they "must be corrected": columns containing more than $m/2$ null elements are deleted. New matrices, called $B_{corrected}$, are defined.

3.4.4 Network construction, steps

In the previous subparagraph, the procedure for identifying the EVs of the time series has been described. Obviously, this operation is one of the most important step to be performed to obtain a network starting from a time series. It is not the only operation, though. The steps to follow in order to obtain a graph, starting from a time series, are here showed in form of a numbered list. To this end, it is necessary to emphasize an important aspect. The embedding parameters are three: m, τ and overlap. It is clear that, every dynamic system, and therefore every time series, will have different dependencies on these parameters. So, different values for these parameters will highlight different properties of the system. Consequently, "optimal values" of these parameters may be defined according to the system and the aim of the study. The purpose of the this first part, in addition to defining the networks themselves, was understand what these "optimal values" were. Hence, the networks were built considering more values of m and τ. The following are the steps to generate the networks:

1. **Measurement point and the time series definition**
   A specific measurement point, so a given point in space, shall be fixed. Then, the variables to analyse must be chosen: time series of u, w and concentration have been directly measured. Vertical turbulent transport can also be considered, deriving the $\tilde{w} \cdot \tilde{conc}$ time series directly from the previous ones. Note that in this work, the tilde sign is used to indicate turbulent fluctuation from the mean value.

2. **Embedding parameters definition**
   Different values of these parameters have been used in order to identify the optimal setting. The EVs dimension, m, was set equal to 50, 100, ..., 450, 500. For each value of m, five different values of τ were considered. From $\tau = 1$ to $\tau = 5$. Each combination of m and tau was then apply both in the WOC and in the OC.

3. **EVs calculation: matrices B definition**
   For each variable, and for a considered m, ten different matrices B are obtained ($5\tau \cdot 2overlaps = 10$). When the VTT is considered, its matrices B can be obtained directly from the w and concentration matrices. As indicated previously, the concentration EVs must be corrected, thus avoiding the excessive presence of null values. Therefore, for the VTT, the B matrices of

the velocity w must be corrected in the same way as the concentration ones. So, discarding the same EVs eliminated in the concentration cases.

4. **Recurrences calculation: matrices R definition**
   Matrices R are the most important elements of the recurrence method. In fact, each matrix R contains the measure of the proximity, in the phase space, between the EVs contained in the corresponding matrix B. Therefore, one matrix R corresponds to each matrix B.
   These matrices are square, with dimension equal to the number of EVs stored in the corresponding B matrices.
   Consider a matrix B and its corresponding matrix R. At each EV of B corresponds a row and a column of R. Therefore, the generic element $R(i, j)$ must contain the proximity, in the phases space, between the EV i and the EV j of B. There are two ways in which this closeness can be assessed:

   (a) Euclidean norm
   The euclidean norm of difference between the two EVs is calculated. The greater the norm, the more the EVs are dissimilar to each other. So, they are not close in the phase space. Obviously, in this way, the diagonal positions will be occupied entirely by null values.

   (b) Pearson correlation coefficient (PCC)
   It is possible to measure the proximity using the linear correlation between the two EVs. The PCC can go from +1 to -1, where +1 is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation. By operating in this way, the diagonal positions will be formed by unitary values.

   Operationally, both methods for evaluating the distance were considered. Therefore, two matrices R have been obtained from each matrix B. One obtained in the case of euclidean norm and one in the case of PCC.

5. **Adjacency matrices A definition**
   Once the matrices R are defined, two more steps are needed to generate the adjacency matrices A, and so the networks.
   Firstly, the matrices R must be transformed into binary matrices applying a certain threshold. Using PCC, matrix A can be so defined:

$$R(i, j) \geq threshold \leftrightarrow A(i, j) = 1$$

$$R(i, j) < threshold \leftrightarrow A(i, j) = 0$$

   Instead, in the Euclidean case, it is necessary to reverse the inequalities. The more small is the norm of the difference, the more similar are the vectors

and therefore $A(i,j) = 1$. Note that, in the PCC case, it is convenient to consider the absolute value of the correlations. In fact, a correlation close to -1 is synonymous of two EVs formed by very similar time series elements, even if opposites. Consider for example the fig. 3.6. It refers to two velocity u EVs (source diameter 3 mm, section number 6, $z = 75$mm, $y = 0$mm). Their $m = 50$ components and the average value are represented for both the EVs. As it is possible to see, the two EVs have similar trends, albeit "opposites". Their PCC is therefore very high and negative:-0.9753. To confirm these characteristics, fig. 3.7 can be considered. It represents, for the EVs already shown in fig. 3.6, the absolute value of the difference between each component and the average value of the EV. As can be seen, the trends are quite similar.

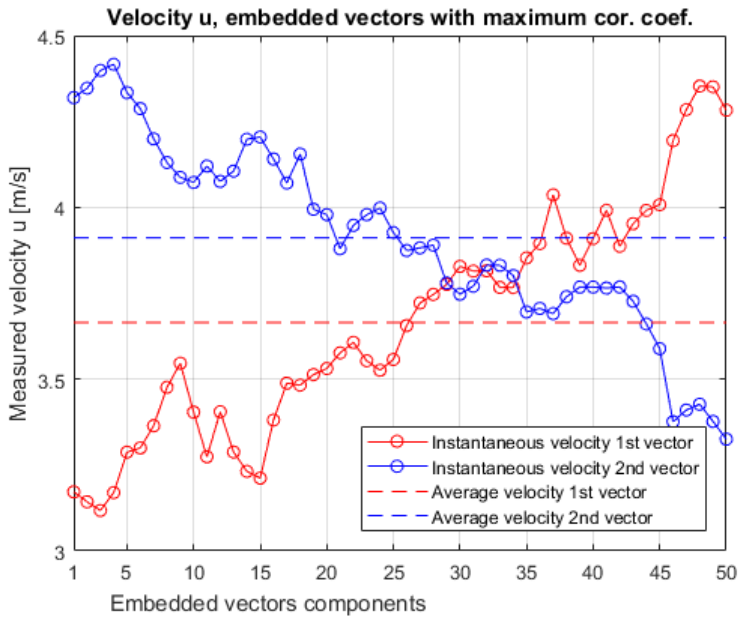

Figure 3.6: Velocity u EVs components and average values

The second action to get the networks is avoiding the presence of self loops, which is to set 0 any non-null diagonals. Therefore:

$$A(i,i) = 0$$

So far it has spoken in a generic way of the threshold so, now, it is necessary to define this parameter. There are two different ways to identify the value to be assigned to the threshold.
A first way is to impose arbitrarily a certain threshold value. This way of

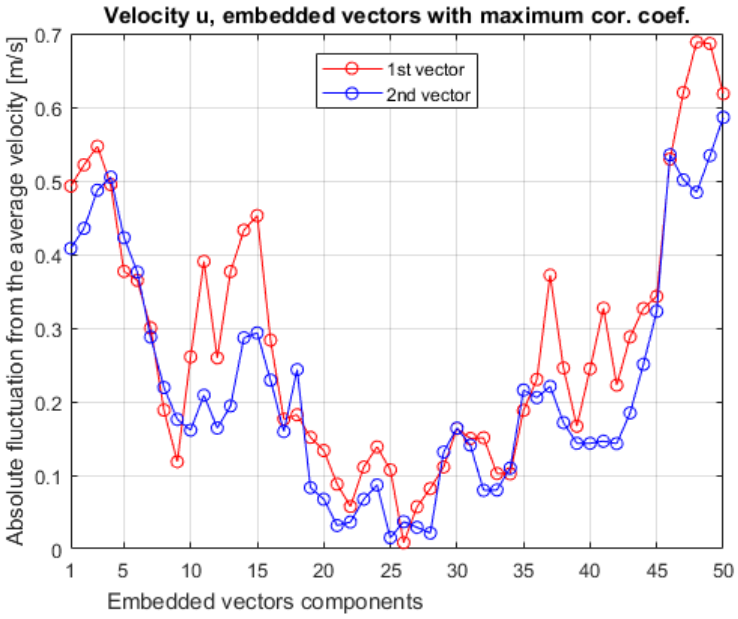

Figure 3.7: Velocity u EVs fluctuations absolute values

working is easier in the PCC case. In fact, the threshold value must be selected between 0 and 1, considering the absolute values of the correlations. This is whatever the variable analysed: velocity, concentration and VTT. In the case of the euclidean norm, the order of magnitude of the threshold will be different according to the variable considered. For the concentration, the norms are of the order of the hundreds of ppm, while for the velocity w, the norms under the unit are the most present.
The second way consists in adapting the threshold based on the values actually present in the R matrices. In fact, for each matrix it is possible to build the frequency distribution of its values . The obtained distributions have forms quite traceable to Gaussian distributions, centred around the value 0 in the PCC case or around a certain average value in the case of euclidean norm. An example of the velocity u in the PCC case (source diameter 3 mm, section number 2, $z = 75$mm, $y = 0$mm, $m = 30$) is shown in fig. 3.8. Once these distributions have been defined, the value that defines a certain quantile of the distributions can be used as a threshold. In the work presented here, the value of the quantile which identifies 95% of the distribution was used as the threshold value. This second method can be considered as an adaptive method. That is, according to the specifications of each single matrix, a certain threshold is identified. As will be more detailed below,

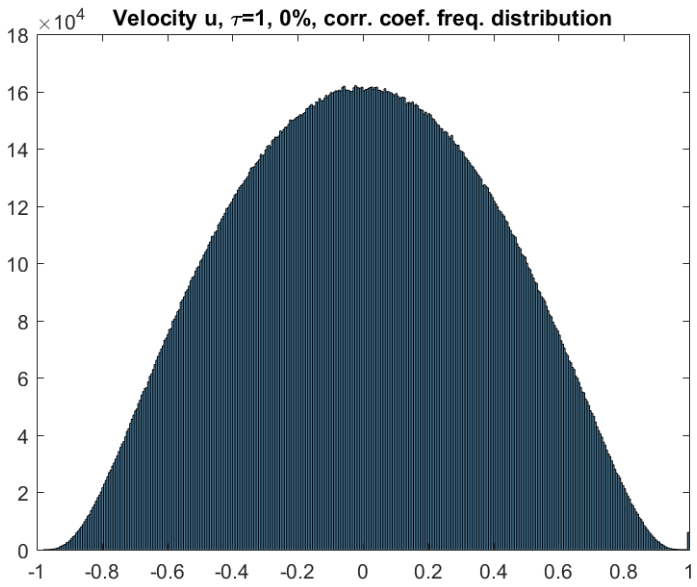

Figure 3.8: Velocity u correlation coefficients distribution

this way of operating has advantages and disadvantages. An important advantage is that the networks are all built in the same way with respect to the distribution, and this will give at the spatial distributions of the metrics certain constant trends. However, the fact of using an adaptive method from case to case does not allow a clear comparison of the networks structures.

Therefore, following these steps, the networks of interest can be obtained for each variable analyzed. Now analyzing the structure of these networks, their metrics and their properties observable in different cases, so with different combinations of parameters, it will be possible obtain information about the plume and the movement of the pollutant, in relation to the motion field and boundary conditions present.

Chapter 4

Characterization of obtained recurrence networks

In this chapter the main properties of the obtained networks are discusses. In the first part a detailed description of the adjacency matrices structures is given. The main differences between the matrices obtained using the Euclidean norm and the PCC are illustrated depending on the the considered measuring points and the setted embedding parameters. The analysis of the PCC optimisation is then executed through the study of their variability and defining the parameters that most influence the correlation coefficients. Obviously, the 0/1 threshold used for the adjacency matrices definition is directly connected to the obtained the phases space distances distributions. For this reason, a comparison between the main results deriving from the PCC and the values of the quantiles used as threshold is performed towards the end of this chapter. This chapter concludes with the description of the main results related to the graphical representation of the networks using the Gephi software and the OpenOrd algorithm.

4.1 Structure of the adjacency matrices

As indicated in the previous chapter, starting from time series, it is possible to generate networks that allow to model the series themselves. Through the study of these networks, the behaviour and the characteristics of the series may be obtained. For the signals discussed here, the study of the networks through their graphical representation is very difficult because the number of nodes and links is too high. So, the networks must be analysed using their adjacency matrices and their metrics. The purpose of this section is to analyse the structure of the adjacency matrices for the networks of interest. In fact, as indicated in the literature, some properties of the considered dynamic system may be identified by observing

the structure of the matrices (i.e. RQA, recurrence quantification analysis).
For this purpose, the graphical representation of matrices A, so the RPs, is needed. To do that, the following convention was adopted:

- in positions that have a connection, that is, where $A(i, j) = 1$, a blu point is reported;
- in positions with a null value, that is, where $A(i, j) = 0$, nothing is reported (it is just a white space).

In the following the structure of the matrices calculated both using the Euclidean norm and the PCC will be illustrated. All the matrices analysed here have been identified using a threshold 0/1 defined with the quantile relative to the 95% of obtained recurrences distributions. Note that, in each measuring point, considering the three signals and having to identify the optimum embedding parameters for the study of the signals themselves, 300 different networks have been generated.

4.1.1 Structure of the matrices obtained in the Euclidean case

The fig. 4.1 represents four different concentration adjacency matrices, obtained in the measuring point on the nozzle axis in section 2, i.e. $z_d = 1$ and $x_d = 0.32$, for the case of 3 mm source diameter. The figure has a twofold purpose: illustrating the matrices structures and evaluate graphically what is the effect of the embedding parameters used for the networks construction. In fact, it is to possible see what is the effects of delay tau and the overlap effect on the number of nodes and on the number of connections. So, the effect on the EVs in which the time series are split and on their recurrences. Fixed m, in fig. 4.1 m=450, it is noted that:

- The number of nodes is directly proportional to the considered delay tau. The number of nodes in the $\tau = 4$ case is about a quarter of the nodes present in the same case but with $\tau = 1$. In the case $\tau = 2$ is half of those in $\tau = 1$ and so on.
- Fixing specific parameters (m/τ), the number of nodes present in the OC (overlap case) is about twice of those in the WOC (without overlap case).

Observing fig. 4.1 in terms of matrices structure in the Euclidean case, it can be observed that:

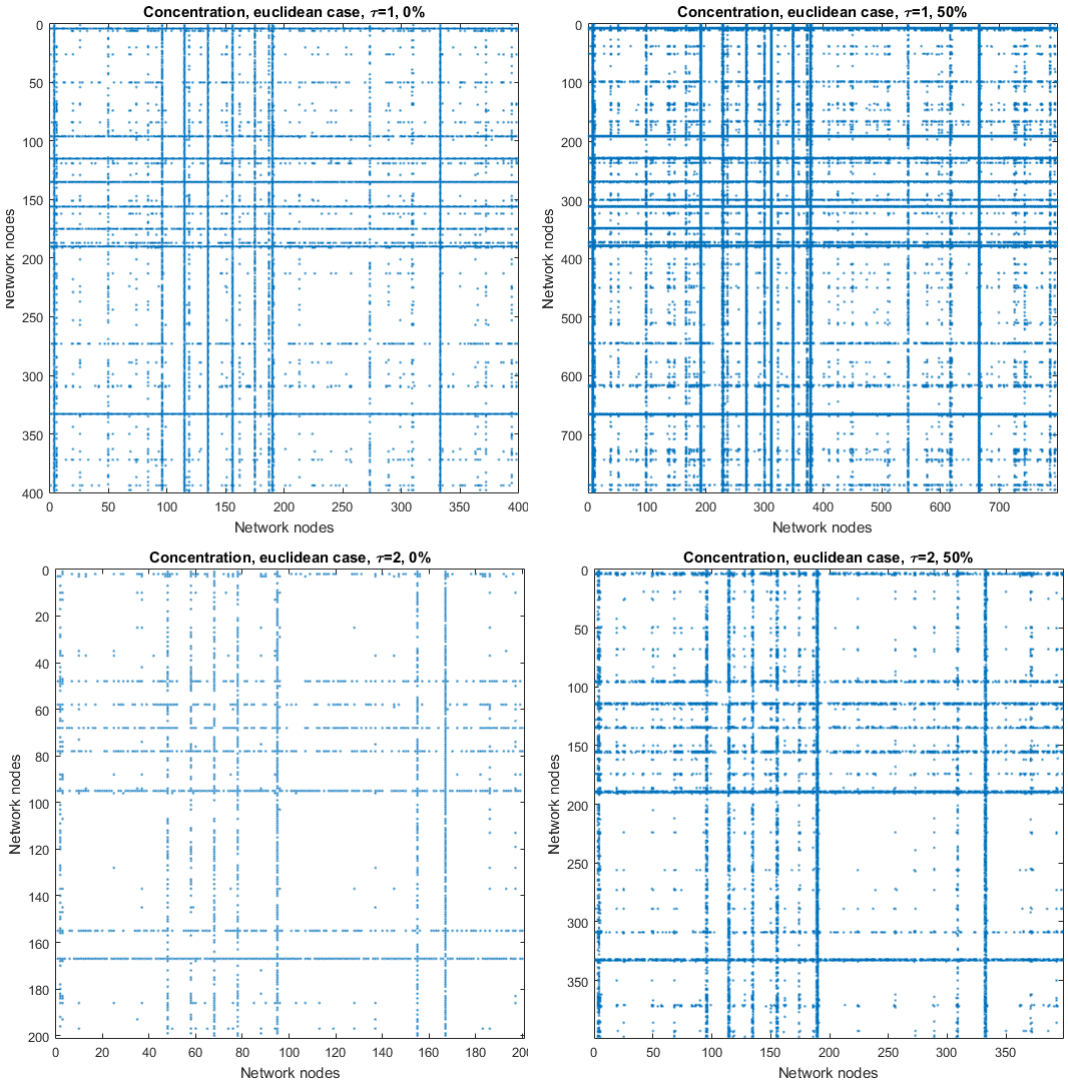

Figure 4.1: Concentration, D3, section 2, $z_d = 1$, $m = 450$: adjacency matrices in the Euclidean case

1. The structure could be defined as non-diagonal. In fact, there are no diagonal marked lines, but rather vertical marked lines. So, lines formed by many dots, indicating the presence of some extremely central nodes within the network. Being symmetric matrices, at these vertical lines correspond horizontal lines. Among these more dense vertical and horizontal lines, there are other more sporadic dots. However, even these more sparse points tend to be organized on rows and columns.

2. Considering the same embedding parameters (m, tau) but with overlap at 50% (right column in the figure), the vertical lines visible already in the 0% case are still present but:

   - They are even more dense of points. This is because, with an overlap at 50%, each node represents a time series fragment that is equal, for an half, to that represented by the adjacent node.
   - They are in a double position. For example, if the WOC has a fairly dense line at the node number 30, the same line will be found in the OC, but at node number 60.

3. Consider the same m and the same overlap, but with different tau, it is possible to see that the densest lines:

   - Tend, as tau increases, to "fade". So, to become less dense.
   - Keep their position in the graphical representation. But note that are the "axis" that change. Therefore, the denser lines, increasing tau, refer to different nodes, despite their position remains about unchanged in the graphical representation.

4. As mentioned above, for each measuring point, several values of m were considered: 50, 150, 250, 350 and 450. What is observed for corresponding networks but with different m, is basically the same structure but with the links density smaller as much as m increases.
   It can therefore be said that the EVs dimension m is one of the most important parameters from the network point of view. A greater amount of links, fixing the others parameters, indicates a greater ability to detect recurrences in the time series. That is, a greater ability to study the series. This is the reason why part of this chapter will be devoted to the study of this parameter and its effects on the adjacency matrices.

4.1.2 Structure of the matrices obtained in the PCC case

Considering now the adjacency matrices obtained by evaluating the distances in the phase space through the use of PCC, some similarities and some differences, with respect to the Euclidean case, may be identify. The first big difference is in the matrices structure. The fig. 4.2 illustrates, for the signal and the measuring point already considered in fig. 4.1, the adjacency matrices obtained using the PCC. Also here the threshold 0/1 was identified by the use of the 95% quantile. The structure of these matrices is less rigid. In fact, the connections within the matrices are distributed more evenly: there are no more a small number of nodes

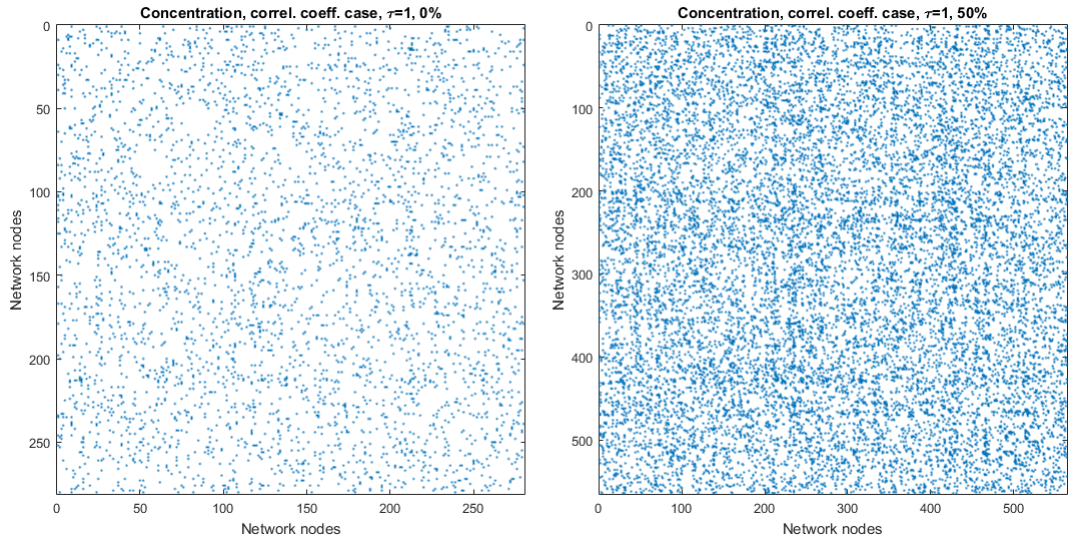

Figure 4.2: Concentration, D3, section 2, $z_d = 1$, $m = 450$: adjacency matrices in the PCC case

with a high centrality in the network. Thus, the generated networks will not present excessively interconnected nodes or isolated clusters.
The structure of these matrices can therefore be defined as "random", compared to the rigid structure of Euclidean matrices.
Regarding the variation of these matrices, according to the embedding parameters of the method (m, tau and overlap), all the considerations already exposed for the Euclidean case are still valid. From this point of view, there is a full similarity between the two different ways of working. For example, fig. 4.2 illustrates the increase in the links density due to the overlap. This behaviour is completely similar to the Euclidean case.
Consider now the case of adjacency matrices obtained by imposing an arbitrary threshold value. That is, through a method that is not adaptive respect to the distribution of the phase space distances. Imposing an ever higher threshold value, the density of the network will tend to decrease more and more. Therfore, when the threshold value imposed will be greater than the maximum PCC, no more unitary values will be present in the adjacency matrices. So, the network will be completely free of links.
In order to identify how a matrix completely devoid of connection can be obtained, so to have a magnitude on the thresholds that can be inposed, three measuring points along the section 3 ($x_d = 0.64$) in the D3 case have been considered. The points were the number 78 ($z_d = 0.87$), the number 80 ($z_d = 0.93$) and the number 82 ($z_d = 1$). For each measuring point, three values of m were considered : 50,

250 and 450. It was therefore studied the threshold value for which is obtained, with respect to the z and the m considered, an adjacency matrix of only zeros. So, network with nodes that do not have links between them. Two important results have been obtained from this analysis:

- Fixed a certain value of m, the threshold that bring to the null density is about the same for the three z considered. In fact, fixed a certain m, the density does not vary much with the z-coordinate.

- Instead, the thresholds that identify the first null densities have a variability with the parameter m. Fixing a specific z and considering different values of m, it is observed that these thresholds are greater in the case of small m (50) while decreasing with the increase of m (450). The table below shows the thresholds that identify matrices with zero density, for the point 78, varying m, for the three signals of interest.

Table 4.1: D3, section 3, measuring point at $z_d = 0.87$: thresholds that identify completely disconnected networks

m=50		**m=250**		**m=450**	
Velocity u	0,9	Velocity u	0,7	Velocity u	0,5
Velocity w	0,8	Velocity w	0,4	Velocity w	0,3
Concentration	0,85	Concentration	0,25	Concentration	0,15

4.2 Effect of the EVs dimension on the correlation coefficients

4.2.1 Purpose of the analysis

The parameter m represents the embedded vectors size, so the number of elements contained within the vectors generated starting from the time series. The purpose of the following is to analyze the effects of this embedding parameter with respect to the correlation coefficients obtained in order to generate the network adjacency matrices. Having interest in the study of the PCC, all the analyses listed herein refer to their use and not at the Euclidean norms case.

4.2.2 Measuring points and considered cases

Any calculation has been done considering only two delay tau: $\tau = 1$ and $\tau = 2$. This choice is due to the fact that, in this first phase, the signals behaviour is unknown. Using small tau values, multiple elements of time series are analysed, achieving more meaningful results.
The analyses were carried out at the measuring points shown in fig. 4.3

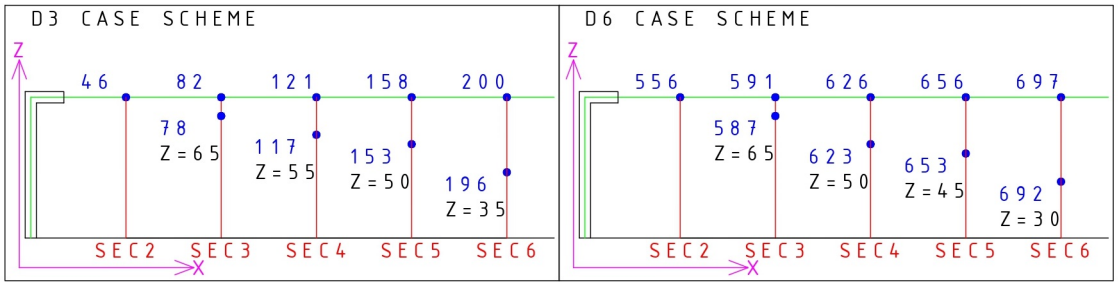

Figure 4.3: Measuring points considered

The points on the nozzle axis have been considered for each section and for both the source diameters. In addition, from section 3 onwards, a point below the nozzle axis has been analysed, with vertical coordinate descending from section to section. This choice is due to the desire to follow "in Lagrangian way" a particle of mixture air-ethane. In fact, once out of the injector, statistically, a singol particle will tend to move towards increasing x but also to gradually decreasing z.
Therefore, in these section, for different values of m, the PCC obtained for the two velocity signals and for the concentration have been analysed. Both in the WOC than in the OC.

4.2.3 Analysis in a single measuring point

Generated graphs and description of the obtained curves

Considering now a single measuring point, for example the point number 692, at the base ($z_d = 0.4$) of section 6 ($x_d = 3.90$), in the D6 case. Then, consider a certain signal, for example the velocity u. For the point and for the signal chosen, the graph shown in fig. 4.4 has been generated. The graph has different values of the m parameter, from $m = 20$ to $m = 500$, on the abscissa axis. The ordinates represent the percentage of PCC greater or equal than a certain threshold, defined in the following as the curve-threshold, with respect to the total number of correlations.
If the EVs number obtained from a time series is N, an NXN matrix will define to

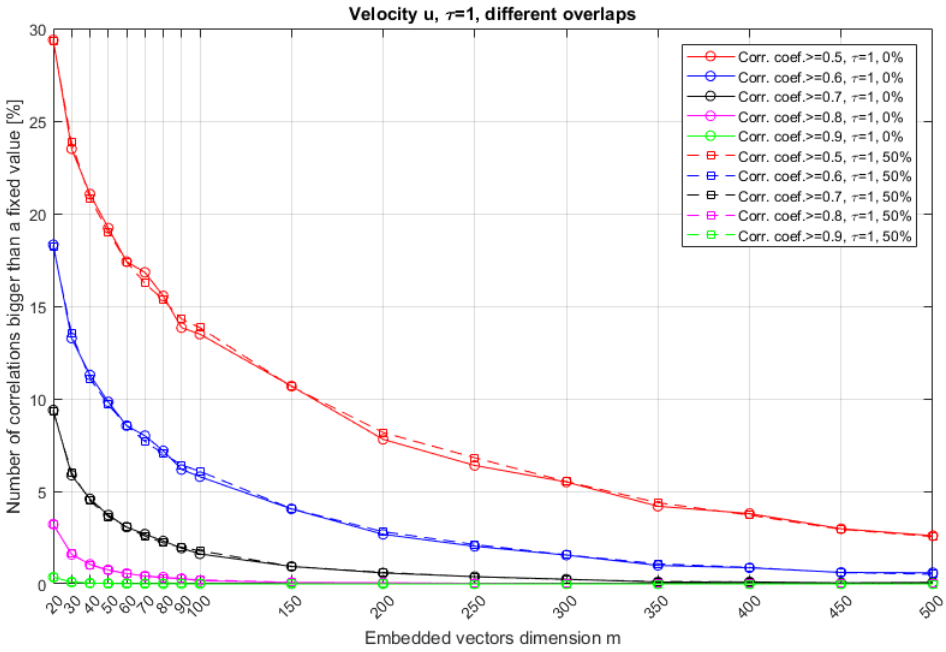

Figure 4.4: Velocity u, D6, section 6, $z_d = 0.40$, $\tau = 1$, WOC/OC: percentage of PCC greater or equal of a certain threshold, depending on the m parameter.

contain the correlation coefficients. The matrix will be symmetrical and will have unitary values on the diagonal. The total number of correlations represents the number of elements contained in the upper triangular part or, equivalently, in the lower triangular part, excluding diagonal elements.
In this analysis five different curve-threshold values were considered: from 0.5 to 0.9.
The obtained curves have trends similar to negative exponential. In fact, each curve has a maximum for $m = 20$ and, as the abscissa increases, it has an asymptote represented by the null percentage. The percentages are all the lower the more the curve-threshold increases. In fact, by imposing the value 0.9, the curves tend to degenerate into a line with percentages all close to 0%.
From the fig. 4.4 it is possible to derive an important information: curves relative to the same curve-threshold have similar trends and similar numerical values, regardless of overlap. This can be observed whatever the considered signal (u, w and concentration) and for both the analysed tau. What is observed here is not inconsistent with the links increase shown in the fig. 4.2 due to the overlap which refers to the case of thresholds defined in an adaptive way. Here, instead, the threshold is arbitrarily imposed. Therefore, the overlap plays a different role depending on the method considered.

The graphs obtained using $\tau = 2$ are not reported because they do not add information. In fact, they have the same trends already observed using $\tau = 1$, just with lower percentages.
The results shown here refer to a specific measuring point and a specific signal. Obviously, these analyses were carried out for each of the three considered signals and in each of the above mentioned measuring points.

Comparison between velocity u, velocity w and concentration at the same measuring point

The aim now is to make a comparison between the three considered signals always in the measuring point indicated as 692. The graphs obtained for the velocity w and for the concentration are shown in fig. 4.5 and in fig. 4.6

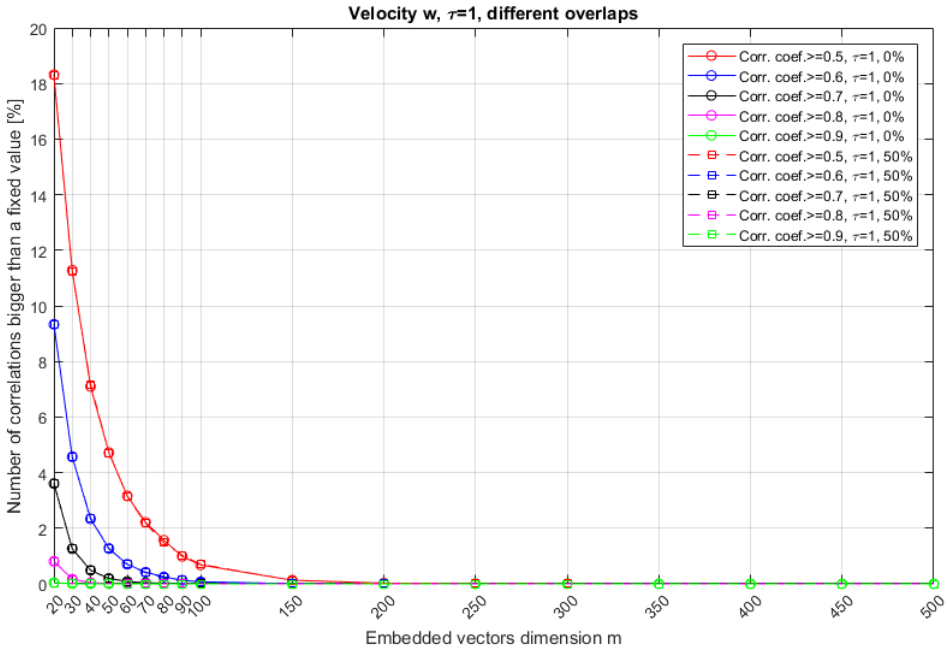

Figure 4.5: Velocity w, D6, section 6, $z_d = 0.40$, $\tau = 1$, WOC/OC: percentage of PCC greater or equal of a certain threshold, depending on the m parameter.

Comparing the figures, it is possible to observe that:

- Considering small m values, like $m = 20$, the concentration has the highest percentages. The velocity u has intermediate values, while the w has the lower. This means that, for a given m and for a certain curve-threshold, the signal that has the greater quantity of recurrence, is the concentration.

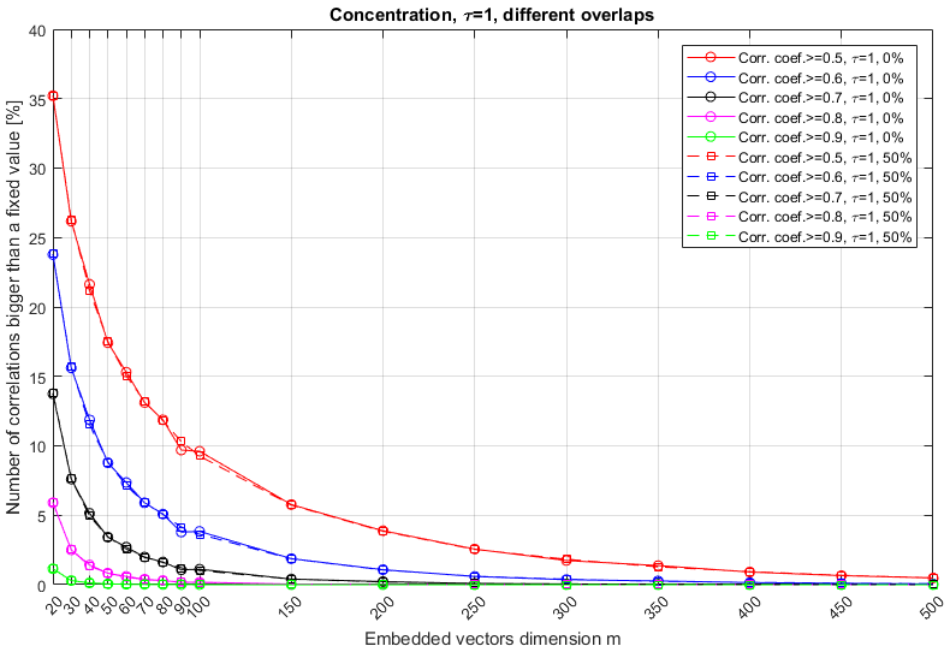

Figure 4.6: Concentration, D6, section 6, $z_d = 0.40$, $\tau = 1$, WOC/OC: percentage of PCC greater or equal of a certain threshold, depending on the m parameter.

- Comparing the velocity u and the concentration, it is true that the percentages for $m = 20$ are greater for the velocity, but it is also true that the u curves have a lighter fall. Therefore, the percentages that are obtained for $m = 20$ are greater for the velocity u.
  So, each signal has a different rate of decay for the curves in exam. This must be taken into account when analyses that require the use of different values of m are performed considering all the measured signals. The velocity w is the signal characterized by the most pronounced fall of the curves. For m greater than 150, the curves are practically all around 0%, meaning that, in these cases, the correlation coefficients calculated are all lower 0.5. This feature of the w signal do not particularly amaze. In fact, the phenomenon in question is not convective. Therefore, the velocity w basically has a random behaviour, keeping an average near zero. It is clear that, a parameter that measures the "similarity" between the embedded vectors can not achieve particularly high values.

4.2.4 Analysis varying the measuring point

Effect of x coordinate on the curves (source diameter and z coordinate fixed)

The purpose of this part is to analyse the joint effect of the EVs dimension m and the x coordinate on the PCC. To do this, the 6 mm source diameter was fixed and, in each section, the measuring point on the nozzle axis was considered. Fixed then a certain curve-threshold value, chosen equal to 0.60, on the same plot the curves for the points on the nozzle axis present in the five different sections were represented. The curves obtained for the signal u are shown in fig. 4.7. All the

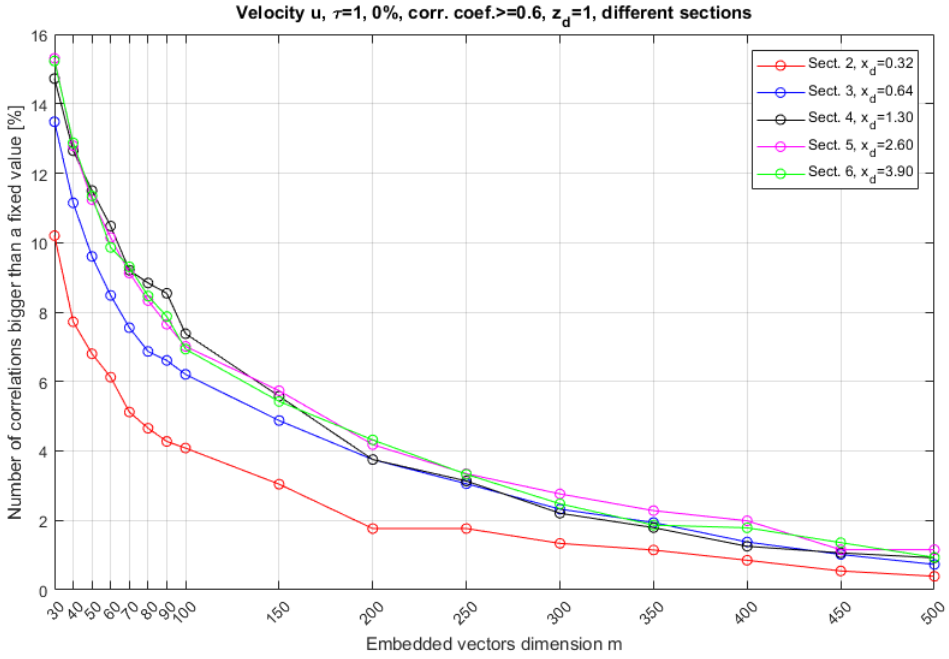

Figure 4.7: Velocity u, D6, $z_d = 1$, $\tau = 1$, WOC: effect of the x coordinate on the curves

curves refer to $\tau = 1$, which is the tau that allows the larger percentages. For ease of reading, only the curves related to the WOC were represented, given the little dependence from the overlap discovered before.
The x effect of x turns out to be a curves vertical displacement. In fact, increasing the x-coordinate, slightly greater percentages are obtained and so curves shifted higher than the red curve relative to section 2, the closest to the nozzle. Therefore, the considered time series have a longitudinal coordinate dependence: as this increases, the time series elements give rise to more correlated EVs. That is, the elements of time series tend to be more uniform and with a lower variability.

Even the velocity w and the concentration have the same dependence on the x coordinate.

Effect of the x coordinate and of the source diameter on curves (z coordinate fixed)

The previous analysis refers to the points on the nozzle axis, in the D6 case. Similar graphs were also obtained for the D3 case. The structure of these graphs is completely similar to the one illustrated in fig. 4.7.
In order to compare, in a compact way, the results obtained with the two different diameters, table 4.2 can be considered. It shows the percentages obtained for the two diameters, for the curve-threshold equal to 0.60, in the case of $m = 30$ and $m = 100$, for all points on the axis in the five different sections.

Table 4.2: $z_d = 1$, $\tau = 1$, WOC: the table shows the joint effect of the x coordinate and diameters on the percentage calculated using 0.60 as curve-threshold

	VELOCITY u				VELOCITY w				CONCENTRATION			
$z_d = 1$	D6		D3		D6		D3		D6		D3	
	m=30	m=100	m=30	m=100	m=30	m=100	m=30	m=100	m=30	m=100	m=30	m=100
sect.2	10,20	4,08	11,92	6,21	3,19	0,06	4,30	0,17	8,24	0,25	6,91	0,38
sect.3	13,48	6,21	14,16	7,55	5,67	0,27	6,35	0,35	8,27	0,39	7,53	0,55
sect.4	14,72	7,38	14,78	7,99	6,95	0,38	6,68	0,42	10,02	0,98	9,38	0,98
sect.5	15,31	7,01	15,28	7,71	7,36	0,49	7,16	0,46	13,26	1,60	12,85	1,62
sect.6	15,23	6,93	15,30	7,75	7,22	0,47	7,58	0,50	15,61	2,31	15,70	2,61

The most important result that can be obtained from the table is the independence of the percentage from the source diameter. The numerical variations that are present varying the diameter, are very small. These variations are a little bit greater in the first three sections.

Lagrangian study of an ethane particle

In addition to the points on the nozzle axis, other measuring points were also considered. These, shown in fig. 4.3, may be considered as positions subsequently occupied by a particle which, once ejected from the injector, will tend to move both in the verse of the increasing x and in the verse of the decreasing z. They are therefore positions that would be considered by studying the motion of the particles in Lagrangian way.
In each of these points, the PCC were analysed with respect to different m, imposing a curve-threshold of 0.6. The results obtained for the concentration are reported in fig. 4.8 This figure reflects the behavior already illustrated for the

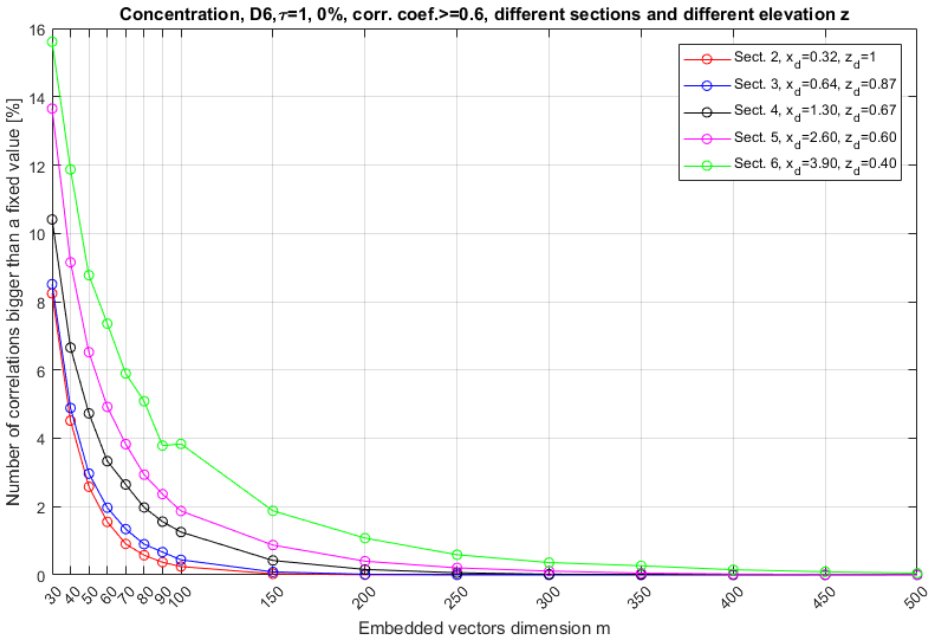

Figure 4.8: Concentration, D6,$\tau = 1$, WOC: analysis of Lagrangian plume positions

points on the nozzle axis. So, percentages that increase considering points gradually farther away from the source and at smaller vertical coordinates.
The same graphs were also constructed for the points identified in the D3 case. The percentages that are found, for corresponding points but with different diameters of the source, are in fact completely comparable. They are slightly larger in the D6 case but just of few percentage points.

4.3 Trends of the threshold to generate the adjacency matrices

So far, the PCC obtained from the EVs generated by the time series of interest have been analyzed. Once these correlation coefficients are computed, they are used to generate the networks adjacency matrices. For this purpose, an adaptive method can be used: the threshold can be variable according to the PCC distribution, using for example the quantile at 95%.
Fig. 4.9 illustrates, for the points on the nozzle axis, section by section, the value of the threshold corresponding to the quantile at 95%, both in D3 than in D6 case. Each line of the image refers to a certain value of m: $m = 30$, $m = 50$ and $m = 100$. Each column refers to a certain signal: u, w and concentration. Even if

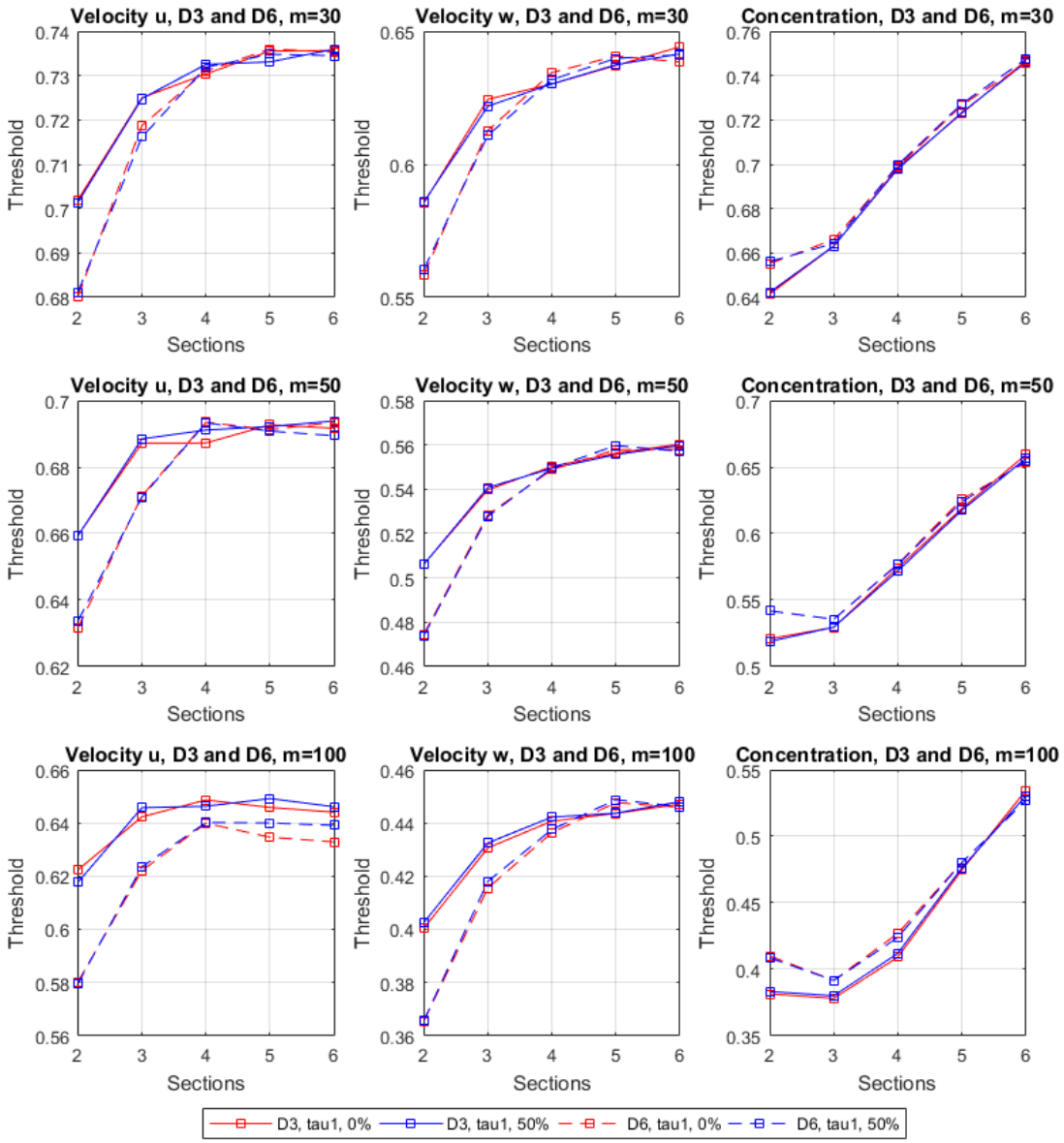

Figure 4.9: Distribution of the adjacency matrices thresholds

the analysis carried out in the previous section is more akin to an arbitrary defined threshold, of course, it must be consistent with the quantile values. If the PCC increase or decrease as a function of a certain embedding or geometric parameters, their distribution suffers, and, with it, the value of the quantile.
The behaviour of this threshold is therefore similar to the behaviour of the percentages analysed above. In fact, looking the fig. 4.9, it si possible to see that:

- The overlap effect is minimal. In fig. 4.9, the two continuous curves (D3) are practically overlapping, as are the two dashed curves (D6).
- Increasing m, the threshold decreases. In fact, observing corresponding graphs with different m the values on the ordinates axis tend to decrease.
- The effect of the source diameter on the threshold value, as well as on the percentages, is minimal. The curves relating to the D3 case and those relating to the D6 case are substantially the same. A "greater" difference between these two curves is presented in sections 2 and 3.
- The percentages and the threshold, fixed all the parameters except the x coordinate (i.e. varying only the section), tend to increase moving away from the source. It is interesting to note that, considering the threshold, for the two speeds, the trend is monotonously increasing. For the concentration, instead, in some cases, presents a minimum.

4.4 Graphic representation of the obtained networks

This paragraph shows what has been achieved with regard to the networks graphic representation. In fact, the velocities and the concentration networks obtained in four different measuring points, in the D6 case, defined using both the Euclidean norm and the PCC have been represented. In this analysis, different values of the embedding parameters have been setted: two different values of m, $m = 100$ and $m = 350$, two different values of tau, $\tau = 1$ and $\tau = 5$, both in the OC than in the WOC.
Once the networks are obtained, the two-dimensions representations can be performed using the Gephi software.
Regarding the settings of this software, it is useful to report two aspects:

- OpenOrd was chosen as mathematical algorithm for the two-dimensional networks representation [14]. This choice was motivated by the fact that it is usable for undirected networks with a large number of nodes. The algorithm behaviour is based on a similarity between the networks and a dynamic system: each node represents a mass while each links a spring. The springs stiffness is defined by the weight of each links while the masses are defined by the properties set for each node. The algorithm, iterating, define the equilibrium position of the system and returns its graphical representation. The iterations are composed of five phases: liquid, expansion, cooldown, crunch and simmer. The first two act on the representation "enlarging the

network", or "pulling the springs" and evaluating possible relative balances, while the last two tend to "reassemble" the network. In order to avoid a contrast between these phases it is useful to indicate different time percentages. Therefore, for the first two a percentage of 30% was set, instead for the last two a percentage of 0%.

- Two distinct graphical representations have been done for each network in order to highlight different aspects. The first shows a specific networks metric, the degree. The nodes are in fact coloured according to the degree value, from light white in case of small degree up to intense red for elevated degree. The nodes size is also variable depending on the degree. In the second representation the parameter used to classify the nodes is the time. Each time series has been subdivided into five sub-intervals representing five different macro-times. At each macro-time a colour was attributed. In this way it is possible to observe whether the nodes related to different temporal moments of the signal tended to distribute themselves evenly or not, generating for example clusters of nodes related to contiguous times, within the network.

It was observed that, varying the measuring point, the networks generated with the same embedding parameters have similar representations. Therefore, there isn't a connection between the geometry, that is the coordinates of the measuring points considered, and the graphical representation. Just as, for different m and tau, the networks tend to vary, in their graphical representation, only their density rather than their structure. In the sequel what was achieved for the measuring point 697 (D6, $z_d = 1$, $x_d = 3.90$) using m=100 is reported.

Representation as a function of time

Whatever the signals and the embedding parameters, performing the representation as a function of time, the same results are always obtained. That are, networks in which the nodes related to different times are uniformly distributed. Therefore, there are not clusters related to particular temporal moments. An example is reported in fig. 4.10.

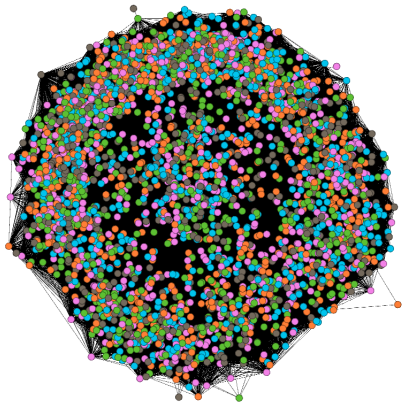

Figure 4.10: Velocity w, D6, $z_d = 1$, $x_d = 3.90$, $\tau = 1$, OC: representation as a function of time

Representation of the velocity u networks as a function of the degree centrality

Also in the graphical representation there are, of course, the same differences between the Euclidean case and the PCC case already observed through the study of the adjacency matrices structure. In fact, in the Euclidean case, it is possible to observe nodes with low degrees all connected to a central node, with a bigger degree. Instead, using PCC, nodes with small degree are more connected to each other forming a much thicker network. Going to consider different embedding parameters, τ and the overlap in this case, a networks thinning can be observed. Therefore, these have no effect on the characteristics and the structure of the networks but solely on its size. This aspect will also be highlighted by the study of the metrics in the next chapter. The variation of some parameters will only have an effect on the metrics numerical values and not on their trends.

Representation of the velocity w networks as a function of the degree centrality

From the point of view of the graphic representation, there are no particular differences between the two speeds. As can been seen, the structure of the velocity w networks is quite similar to that of the u.

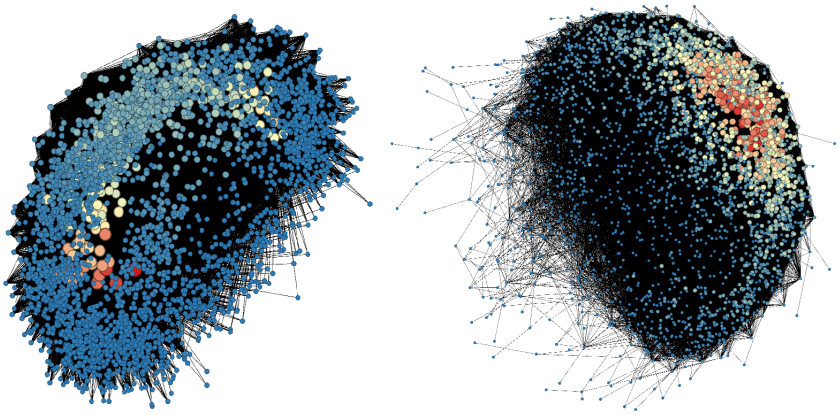

Figure 4.11: Velocity u, $\tau = 1$, OC: Euclidean case (left) and PCC case (right)

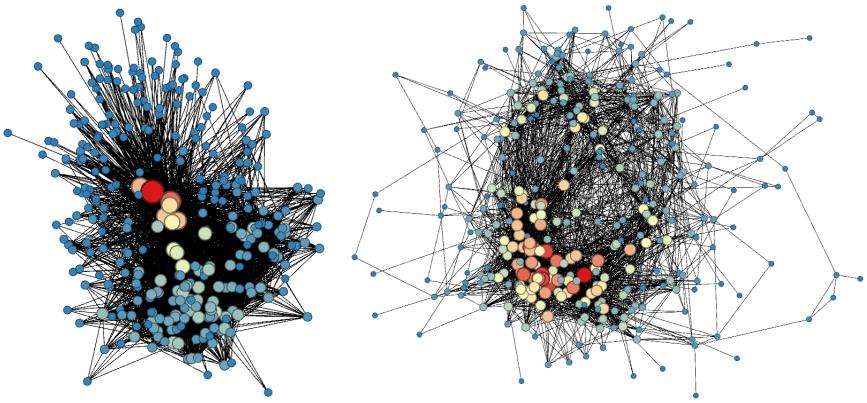

Figure 4.12: Velocity v, $\tau = 5$, WOC: Euclidean case (left) and PCC case (right)

Representation of the concentration networks as a function of the degree centrality

The concentration signals presents, in the PCC case, a difference compared to the speeds. Both in the measuring point here considered than in all the others analysed, the concentration representation has nodes with an intermediate degree, that are, light coloured nodes, distributed in a much more uniform way than the velocities. For example, it is possible to compare the right column of the fig. 4.11 with the right column of fig. 4.15. The structure of the latter is less "roundish" than the one of the velocity u and has a more uniform distribution of nodes with medium degree.

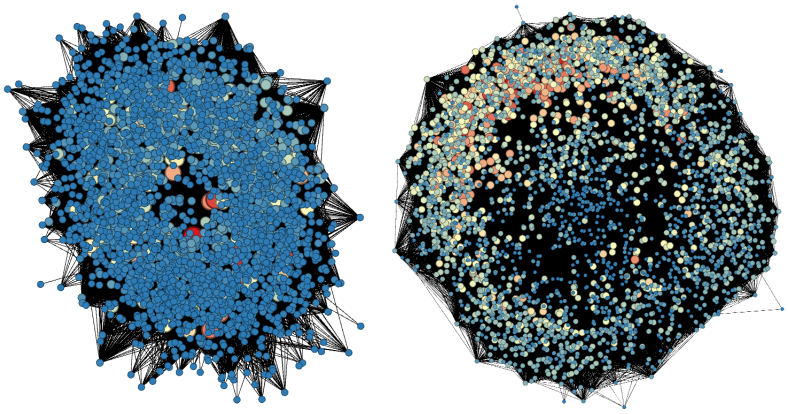

Figure 4.13: Velocity w, $\tau = 1$, OC: Euclidean case (left) and PCC case (right)

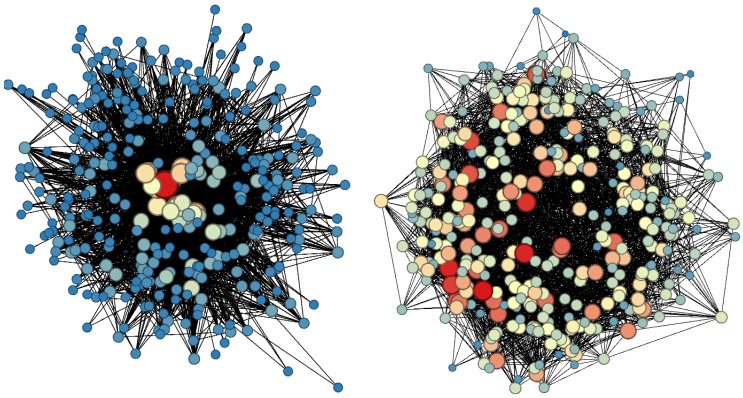

Figure 4.14: Velocity w, $\tau = 5$, WOC: Euclidean case (left) and PCC case (right)

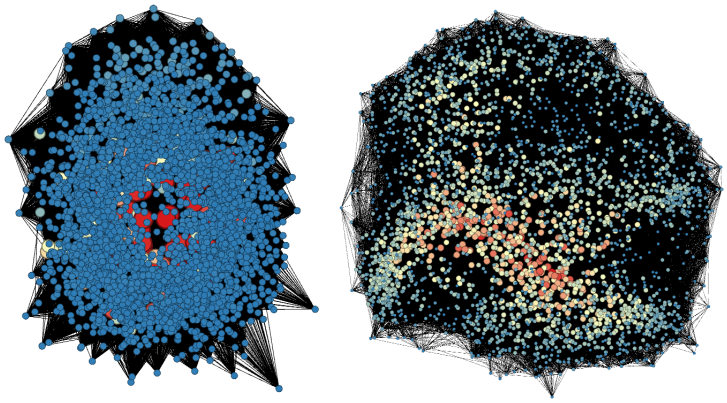

Figure 4.15: Concentration, $\tau = 1$, OC: Euclidean case (left) and PCC case (right)

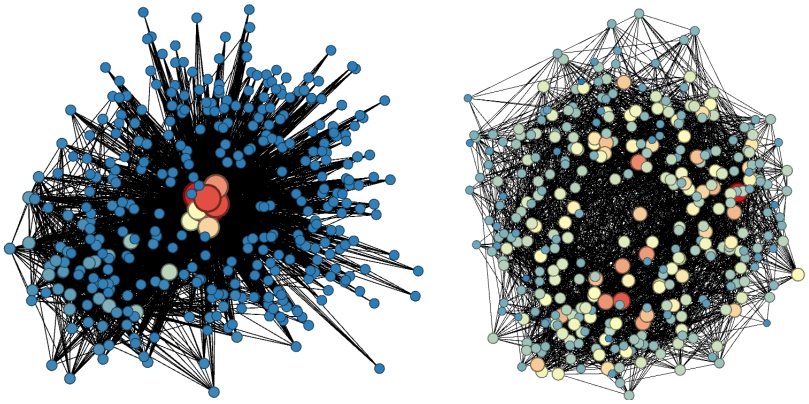

Figure 4.16: Concentration, $\tau = 5$, WOC: Euclidean case (left) and PCC case (right)

Chapter 5

Calibration of the networks embedding parameters

This chapter shows the main results deriving from the study of the network metrics. The first part describes the analyses carried out, using both the Euclidean norms and the PCC, to evaluate the metrics dependencies with respect to the geometric and the embedding parameters. Once the metrics are defined, a series of similarities between them and the statistical analyses carried out in the second chapter is presented for the concentration signals. Combining the main results deriving from the PCC obtimisation illustrated in the fourth chapter with the metrics obtained in the cases analyzed in this chapter, the optimal embedding parameters are identified.

5.1 Purpose and measurement points considered

Starting from a time series, using the techniques illustrated in the third chapter, the corresponding network can be created . This network contains numerous information about the signal used to generate it. So, by analysing its tolopology and its characteristics, information about the time series of departure may be obtained. To this end, it is therefore necessary the study of the network metrics, that are the parameters that identify the main properties and characteristics of the network, and so of the time series.
The purpose of this part is to identify, for the metrics, the main dependencies with respect to the plume parameters and the embedding parameters. So, evaluate the dependence with respect to:

- Source diameter;
- Geometric spatial parameters: the effect of the distance from the source,

identified by the parameter x_d (section), and the vertical position, identified by the z_d coordinate;

- Embedding parameters. In these analyzes $m = 50$ and $m = 100$ were the values considered. Regarding τ and the overlap, two cases were considered: the case $\tau = 1$ OC and the case $\tau = 5$ WOC. These two cases are indeed as "extreme" cases. Considering time series with the same number of elements, they generate respectively the network with the greatest number of nodes and the one with the least number of nodes;

- Recurrences definition: using the Euclidean norm or the PCC. In the latter case, it is possible to distinguish between the threshold set arbitrarily or as a quantile.

As indicated, there are many dependencies that must be analyzed. Among these, one of the most significant is the dependence on coordinate z. Therefore, in each analysis carried out, graphs have always been generated whit the z coordinate on the abscissas axis. This z-coordinate is always compared to the vertical coordinate of the nozzle. That is, a dimensionless coordinate will always be shown in the graphs.
Being the vertical coordinate of considerable interest, particular attention has been paid to the choice of the considered points within the vertical profiles in each section. In fact, having to make a correction for the concentration in the PCC case, the minimum number of non-zero elements in each embedded vector was set equal to $m/2$. However, not all the measuring points are able to satisfy this constrain. In fact, the most extreme points do not present vectors able to comply this rule. This means that not all measuring points in a given section took part to this analysis. Only the points that present a quantity of corrected EVs bigger than 20%, compared to the maximum number of generable embedded vectors, have been considered.

5.2 Metrics trends using Euclidean norm

The trends and the dependencies of the network metrics obtained through the use of the Euclidean norm are now described. All the cases reported in this section have been obtained by generating the thresholds, necessary to create the adjacency matrices, using of the quantile that identifies the 95% of the obtained phase space distances distributions.
In order to analyse the metrics in each measuring point, varying the embedding parameters appropriately, many analyses have been carried out. Two of the graphs obtained are reported as examples. Fig. 5.1 represents the trend of the average

closeness for the velocity u, while fig. 5.2 shows the links trend for the concentration. Both refer to section number 2,$x_d = 0.32$, in the case of 3 mm source diameter.

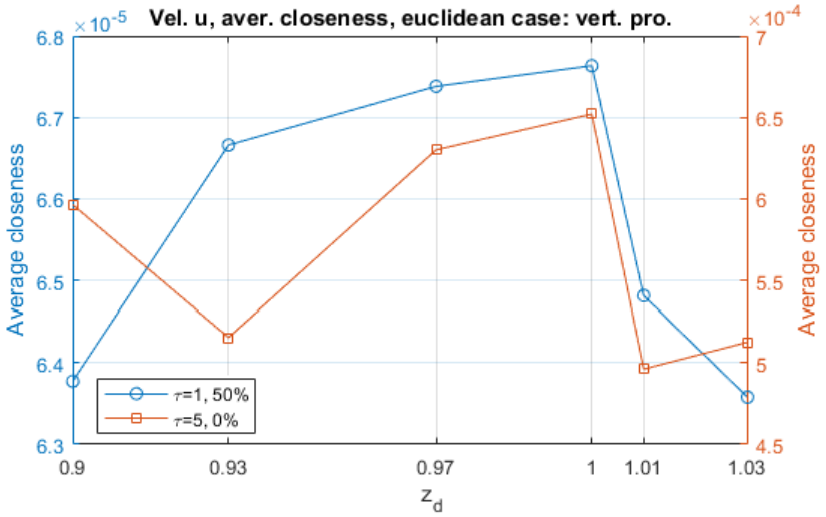

Figure 5.1: Euclidean case, velocity u, section 2, D3: trend of the average closeness

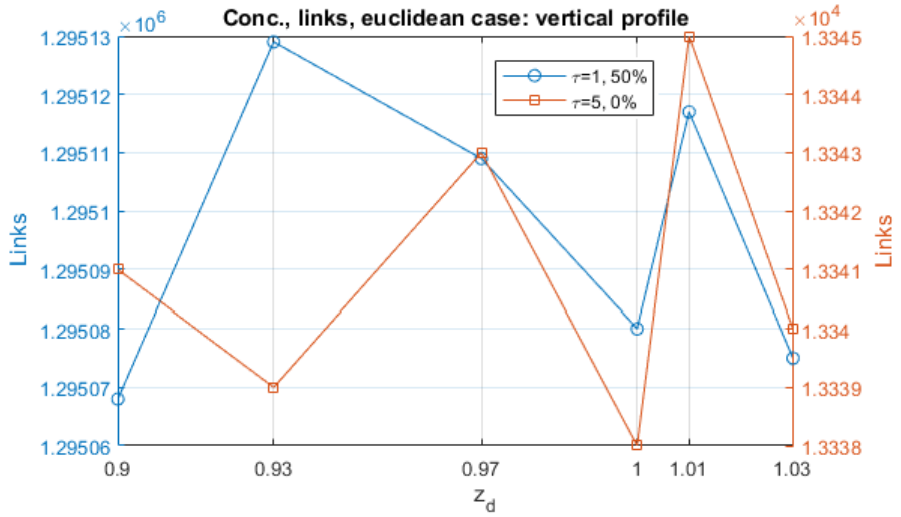

Figure 5.2: Euclidean case, concentration, section 2, D3: trend of the links number

The two main properties of the metrics in the Euclidean case are:

- All the considered metrics remain almost constant when the vertical coordinate changes. Of course, from point to point, a minimal variability,

considered as noise deriving from the experimental activity, is present.

- The behaviour of the metrics is completely general: both speeds, as well as the concentration, have similar and comparable trends. Therefore, it is possible to analyse just one of these three signals to obtain information and trends.

As indicated, the metrics oscillate around an "average" value. For ease of reasoning, it is possible to table these mean values in order to study their variability according to the embedding parameters and the spatial geometry. Table 5.1 and table 5.2, can be considered. They refer to the velocity and report the metrics changing the most important settings.

Table 5.1: Euclidean case, velocity u, $m = 50$: metrics average values

m=50		NODES		LINKS		AVERAGE CLOSENESS	
		$\tau = 1$, 50%	$\tau = 5$, 0%	$\tau = 1$, 50%	$\tau = 5$, 0%	$\tau = 1$, 50%	$\tau = 5$, 0%
SECT.2	D3	7198	731	$1{,}29{\cdot}10^{6}$	$1{,}32{\cdot}10^{4}$	$6{,}5{\cdot}10^{-5}$	$5{,}5{\cdot}10^{-4}$
	D6	7198	731	$1{,}29{\cdot}10^{6}$	$1{,}32{\cdot}10^{4}$	$6{,}4{\cdot}10^{-5}$	$5{,}5{\cdot}10^{-4}$
SECT.4	D3	7198	731	$1{,}29{\cdot}10^{6}$	$1{,}32{\cdot}10^{4}$	$6{,}5{\cdot}10^{-5}$	$6{\cdot}10^{-4}$
	D6	7198	731	$1{,}29{\cdot}10^{6}$	$1{,}32{\cdot}10^{4}$	$6{,}3{\cdot}10^{-5}$	$6{\cdot}10^{-4}$
SECT.6	D3	7198	731	$1{,}29{\cdot}10^{6}$	$1{,}32{\cdot}10^{4}$	$6{,}5{\cdot}10^{-5}$	$5{,}75{\cdot}10^{-4}$
	D6	7198	731	$1{,}29{\cdot}10^{6}$	$1{,}32{\cdot}10^{4}$	$6{,}5{\cdot}10^{-5}$	$5{,}25{\cdot}10^{-4}$

Table 5.2: Euclidean case, velocity u, $m = 100$: metrics average values

m=100		NODES		LINKS		AVERAGE CLOSENESS	
		$\tau = 1$, 50%	$\tau = 5$, 0%	$\tau = 1$, 50%	$\tau = 5$, 0%	$\tau = 1$, 50%	$\tau = 5$, 0%
SECT.2	D3	3598	362	$3.23{\cdot}10^{5}$	$3.23{\cdot}10^{3}$	$1.28{\cdot}10^{-4}$	$1.05{\cdot}10^{-3}$
	D6	3598	362	$3.23{\cdot}10^{5}$	$3.23{\cdot}10^{3}$	$1.25{\cdot}10^{-4}$	$1{\cdot}10^{-3}$
SECT.4	D3	3598	362	$3.23{\cdot}10^{5}$	$3.26{\cdot}10^{3}$	$1.30{\cdot}10^{-4}$	$1.05{\cdot}10^{-3}$
	D6	3598	362	$3.23{\cdot}10^{5}$	$3.23{\cdot}10^{3}$	$1.28{\cdot}10^{-4}$	$1.15{\cdot}10^{-3}$
SECT.6	D3	3598	362	$3.23{\cdot}10^{5}$	$3.26{\cdot}10^{3}$	$1.25{\cdot}10^{-4}$	$9{\cdot}10^{-4}$
	D6	3598	362	$3.23{\cdot}10^{5}$	$3.23{\cdot}10^{3}$	$1.26{\cdot}10^{-4}$	$1{\cdot}10^{-3}$

Each table refers to a specific value of m: $m = 50$ and $m = 100$. In each table, for every metric, two different values of tau and overlap are reported. That are

$\tau = 1$ OC and $\tau = 5$ WOC. In order to analyse the effect of the distance from the nozzle, the metrics obtained in three different sections, number 2-4-6, have been reported, while the presence of the data relative to both the diameters allows to quantify the effect of the nozzle geometry.
Analyzing the dependence on the geometrical parameters, it is possible to observe:

- Effect of the vertical coordinate z: as indicated above, the metrics tend to be constant when the z varies. This parameter is therefore irrelevant compared to the study of metrics in the Euclidean case.

- Effect of the longitudinal coordinate x: as can be seen from both tables, considering different sections there are minimal variations.

- Effect of the nozzle geometry: also in this case the variations are minimal: when the diameter varies, the percentage variation of the metrics is practically negligible.

As regards the dependence on the constructive parameters of the networks, it is possible to observe:

- Effect of the dimension m: doubling the size of the EVs, the number of nodes is reduced by about an half. The number of links also varies with m: doubling m the number of links increases by an order of magnitude. It goes from 10^5 to 10^6. The same consideration can be done for the average closeness, in which a decrease is observed: from 10^{-4} to 10^{-5}.

- Effect of delay τ and the overlap: Fixing a certain value of m, a certain section and a certain diameter of the nozzle, it is possible to see that, varying the value of the delay and the overlap from 1 OC to 5 WOC, nodes and closeness vary by a factor of 10 while the number of links by a factor of 100. As indicated previously, the cases $\tau = 1$ OC and $\tau = 5$ WOC are considered as "extremes". That is, with the highest and lowest number of generable EVs. Considering intermediate cases to the previous ones, the order of magnitude of the metrics variations will obviously be smaller.
  The only metric that never changes is the density: it remains constant whatever parameter is changed.

5.3 Metrics trends using PCC

Now, a similar analysis is proposed, but here the recurrences between the EVs were defined using the PCC. As in the previous case, the thresholds for the adjacency matrices definition were obtained by imposing the quantile relative to 95% of each

correlation coefficient distribution.
In this section, the metrics trends of the speeds and of the concentration will be discussed separately. This because, as will be shown, they have different behaviour.

5.3.1 Speed metrics trends

As regards the velocity, both longitudinal and transverse, the trends and the numerical values are quite similar to those of the Euclidean case. In fact, for a fixed diameter and a given section, imposing a certain value of m, delay and overlap, the metrics turn out to be constant with the z-coordinate. Representing the five sections, it is noted that the variability of these metrics with the x-coordinate, that is, changing section, is minimal also in this case.

5.3.2 Concentration metrics trends

Very different trends are obtained considering the concentration metrics in the PCC case. Fig. 5.3 illustrates the trends obtained in the section 2 ($x_d = 0.32$), calculated with $m = 50$ and relative to the D6 case.

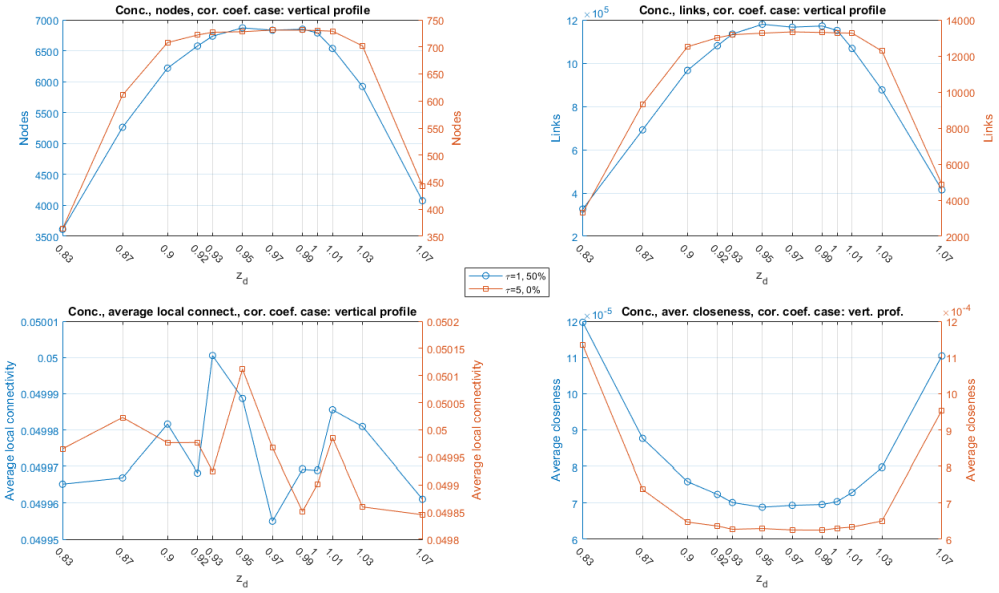

Figure 5.3: PCC case, concentration, section 2, m=50: metrics trends

The first important difference from Euclidean cases is the variability of the number of nodes. In fact, this metric has a parabolic shape. It is necessary to underline that not all the networks analysed in this chapter have the same number of nodes.

This fact is due to two different reasons. As regards the velocities and the concentration analysed with Euclidean norms, as well as for the velocities in the PCC case, the number of nodes depends only on the chosen embedding parameters (m, tau and overlap). When these parameters are fixed, the number of nodes will be the same in every measuring point. For the concentration in the PCC case, in addition to the dependence, the number of nodes also depends on the considered measuring point and the correction chosen to take into account the large number of null measures. A minimum number of non-zero elements, $m/2$, that each vector must contain is imposed (EVs correction). So, setting the constructive parameters, the number of nodes will be variable considering different measuring points.
Having networks with different number of nodes is the reason why it is necessary to identify dimensionless metrics, like for example the density. Otherwise, a direct comparison of non-dimensionless metrics would not be possible, for example for the average degree.
The trend of the links number is quite similar to the nodes one.
The average local connectivity (or density), as in the Euclidean cases, is constant and equal to 5%.
The average closeness shows a pot shape. This trend is justifiable by referring to its definition. In fact, it is defined as the reciprocal of an average path. Where the nodes and links are in smaller numbers, so at the sections edges, the average path will be statistically lower and therefore its reciprocal will be a high number. Conversely, in the case of a high number of nodes and links, the path will be greater and its reciprocal a smaller number. This is how, taking advantage of the definition, the average closeness trend can be described starting from the nodes and links trends.
The concentration metrics have a direct fluid-dynamic interpretation. The number of nodes and links is greater where the average concentration is also greater. That is, just below the nozzle axis or in the accumulation areas. This indicates the presence of a greater number of recurrences within the time series measured in those points.
What was shown here allows to describe the metrics obtained with certain embedding parameters and in a specific x coordinate along the axis of the wind tunnel. The following shows the dependencies of the metrics compared to the geometry and the networks embedding parameters.

5.3.3 Effect of the x coordinate on the concentration metrics

As indicated above, neither the velocities in the Euclidean and in the PCC cases, nor the concentration in the Euclidean case, have a marked dependence by the x-coordinate. This is due to the fact that the boundary layer is fixed. That is, the motion field is similar at different sections. Instead, the concentration

metrics obtained using the PCC have an important dependence on the section in which they are calculated. In order to represent the trends of the metrics in the different sections, the following 3 figures, formed by four sub-figures each, have been generated. Each figure refers to a specific metric: nodes, links and average closeness. Each sub-figure refers to a certain source diameter (D6 and D3) and to a certain m ($m = 50$ and $m = 100$). For simplicity, the curves refer all to the case $\tau = 1$ OC.

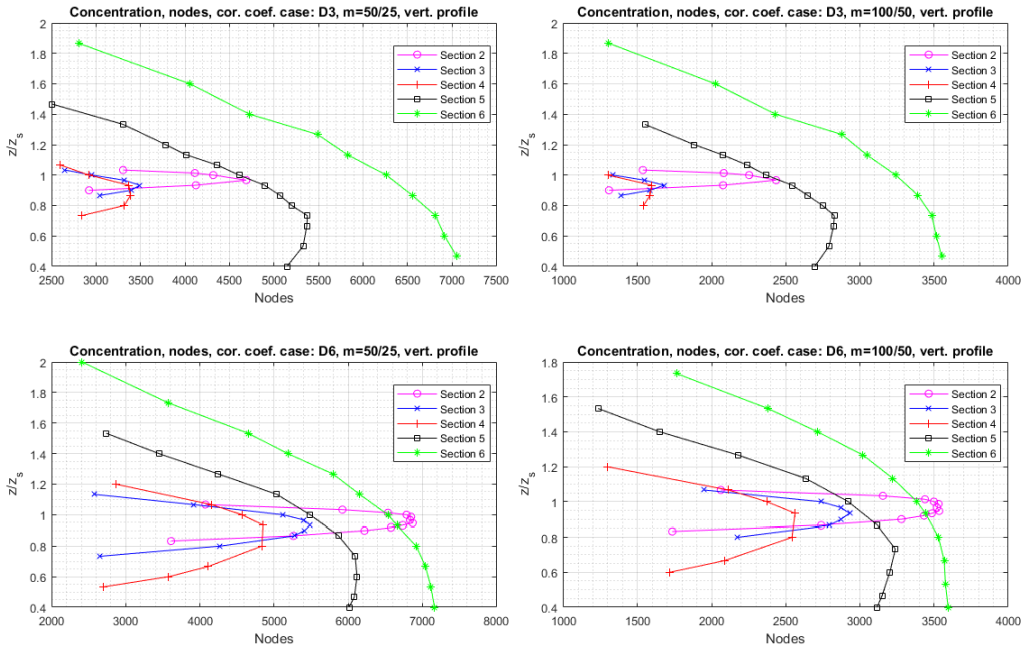

Figure 5.4: PCC case, concentration, nodes: profile trend as x-coordinate varies

As regards the number of nodes and links, observing fig. 5.4 and fig. 5.5, it is evident that:

- The vertex of the parabolic move to lower vertical coordinates increasing the x. In section 2, the parabolic profile vertex is located at $z_d \approx 1$. Considering section 5, it is located at $z_d \approx 0.7$.

- The profile, as the x-coordinate increases, is no more a symmetrical parabola. It becomes asymmetric. It retains its parabolic shape only for $z_d > 1$, i.e. for points above the nozzle axis. The more the x-coordinate increases, the more the part of the parabola present for $z_d < 1$ tends to have a less pronounced fall. Until, in section 6 ($x_d = 3.90$), the trend has no more a vertex and the previous parabolic shape is completely lost.

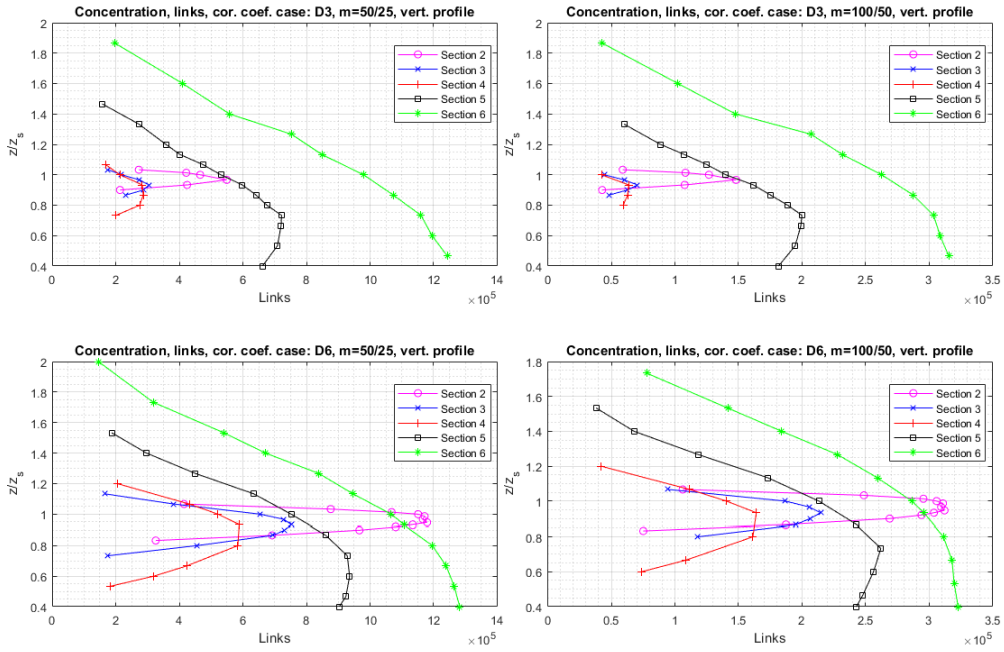

Figure 5.5: PCC case, concentration, links: profile trend as x-coordinate varies

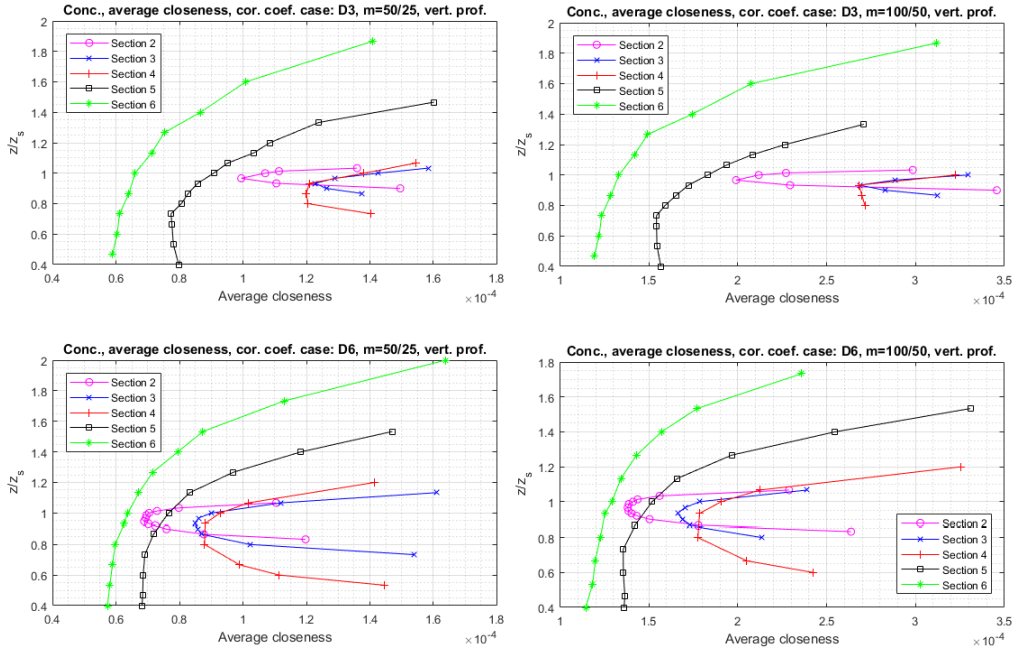

Figure 5.6: PCC case, concentration, aver. closen.: profile trend as x-coordinate varies

- Fixing a certain m, the green curve (section 6) is essentially the same, both in case D3 and in the D6 case. The purple curve (section 2) changes considerably as the diameter varies. For the intermediate sections the variations are much less pronounced the more the x-coordinate is elevated.

The average closeness varies its shape in a opposite way compared to the nodes and links.
The density is always about 5%, regardless of the considered section. This is why the figure with the plots of this metric has not been reported .

5.3.4 Effect of the nozzle geometry and of the embedding parameters on the concentration metrics

The aim now is to evaluate the effect of the nozzle diameter, of the EVs dimension m, of the delay tau and of the overlap on the concentration metric trends in the PCC case. To do this, fig. 5.7 and fig. 5.8 can be considered.
They represent the number of nodes and links, both in the D3 case (left column) and in the D6 case (right column), for three different sections (2-4-6). Each graph shows the metrics obtained with $tau = 1$ and $tau = 3$, in the OC and in the WOC, by imposing two different values of m, $m = 50$ and $m = 100$. The choice of using an intermediate delay ($\tau = 3$) gives the possibility to analyze what is the variation with respect to the extreme case, i.e. $\tau = 5$. Therefore, there are eight different curves in each chart.
They are represented using only one ordinates axis. This is due to the fact that the trends are already known, what is investigate here is the parameters effect on the order of magnitude of the metrics.
Referring to fig. 5.7 and fig. 5.8, it can be inferred that:

- Fixing a certain m and a certain τ, by changing the overlap from 0% to 50% the number of nodes doubles. The variation of the links, instead, is non linear with the overlap. That is, if superimposing the vectors between them is obtained an increase of nodes equal to the double, the increase of the links is about the quadruple of those present in the WOC.

- Fixing a certain m and an overlap, the tau effect is a vertical translation of the curves. Obviously depending on the overlap, the translation will be different. By setting the OC and passing from $\tau = 1$ to $\tau = 3$ the number of nodes decreases much more than in the WOC. The same applies for the links trends.

- Fixing a certain tau and a certain overlap, the m effect is also a curves vertical translation. From m=50 to m=100, the number of nodes is halved

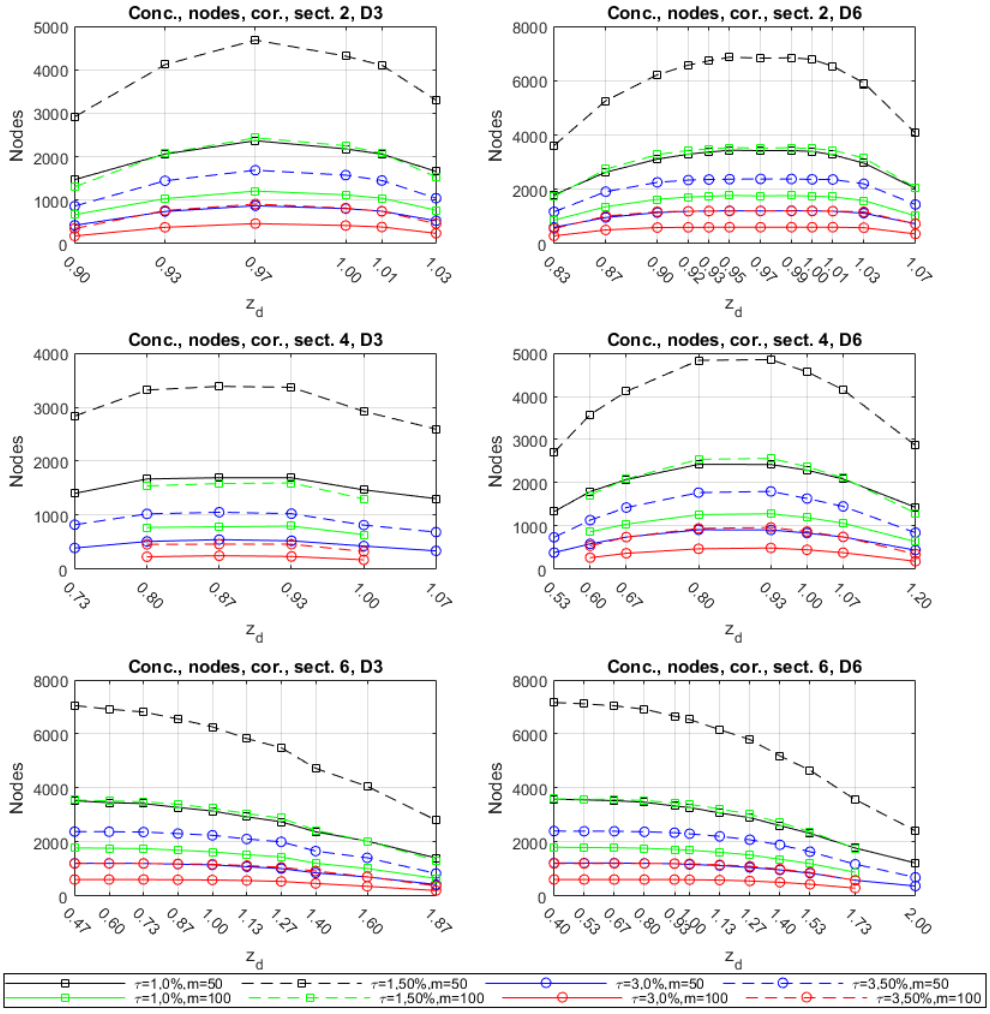

Figure 5.7: PCC case, concentration, nodes: magnitude and dependence on parameters

while the links are pariah to a quarter of the previous.
It is easy to understand that doubling the overlap, from 0% to 50%, and also doubling m, from $m = 50$ to $m = 100$, have opposite effects that delete each other out. In fact, for example, the curves for the case $\tau = 1$,$m = 50$, 0% (continuous black) and $\tau = 1$,$m = 100$, 50% (green with dashes) are practically overlapping.

- The effect of the nozzle diameter is located more in the initial sections. In section 2 there is a greater difference between the D3 and the D6 case. Conversely, in section 6, the metrics are practically the same irrespective of the

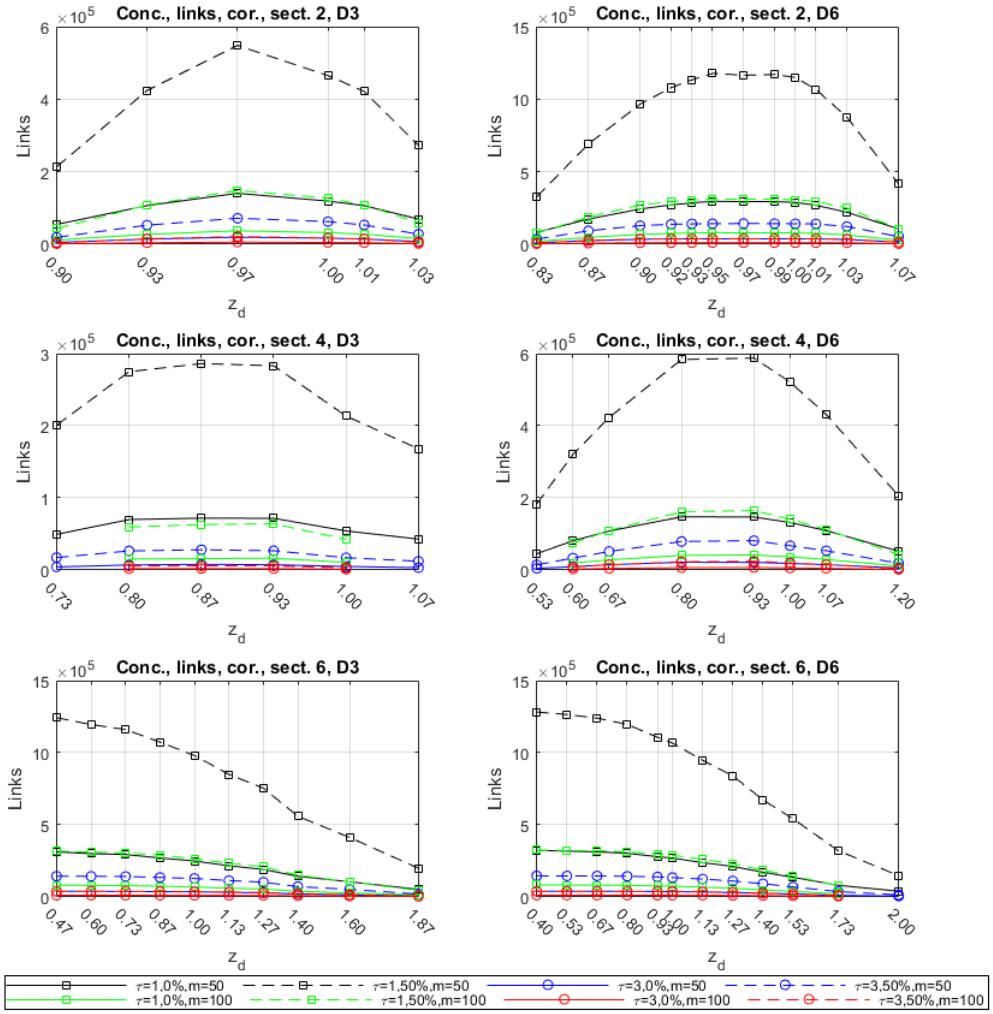

Figure 5.8: PCC case, concentration, links: magnitude and dependence on parameters

considered diameter. By imposing for example $m = 50$, $\tau = 1$ and null overlap, in the second section the maximum number of nodes is about 5000 for the D3 case while it turns out to be about 7000 in the D6 case. With the same embedding parameters, a greater number of nodes, and therefore of EVs, is indicative of less intermittent time series. From this is possible to conclude that the time series relative to the two diameters will have more similar behaviour in the sections farthest from the source. In the first sections, instead, their behaviour is more dissimilar.

The density is independent of the parameters considered and is always about 5%.

The average closeness presents trends opposite to those of the number of nodes and links, coherently to its definition.

5.3.5 Similarities between the average concentration and the concentration networks metrics using the PCC

The metrics of the networks obtained for the concentration signals using the PCC have important properties in common with the average concentration profiles. These profiles, shown in fig. 2.8 and in fig. 2.9, have trends similar to those of the nodes and links illustrated in fig. 5.4 and fig. 5.5. In fact, as the x-coordinate increases, the z coordinate where the maximum of the average concentration occurs tends to decrease. Than, thanks to the wall effect, the average concentration tends to increase in the area around the wall in the last two sections. Therefore, the average concentration presents similar trend that is observed for the nodes and for the links when the x-coordinate varies.
To underline how close is the relationship between these quantities it is possible to consider that in the D3 case, neither in the average concentration profile nor in the metrics, the curve relative to section 2 exceeds that of section 6. Conversely, in the D6 case, in both trends the curve of section 2 exceeds that of section 6.
Another important similarity between the metrics and the statistical study of the concentration signals executed in the second chapter may be identify. Considering a certain curve in the graphs of fig. 5.4, for example the black dashes curve corresponding to the case $\tau = 1$ OC, with $m = 50$, the numerical values corresponding to this curve moving from section 2 to section 4 decrease. Then, moving from section 4 to section 6, increase. A similar behavior has already been identified in a previously analyzed quantity: the intermittence. Calculating the intermittency in each measuring point of every section and considering then, in each section, the maximum value of the intermittence, a decrease up to the fifth section was observed then it increased again in the sixth.

5.4 Best method and optimal embedding parameters definition

The paragraph 3.4 illustrates the model that, from time series, generates the corresponding networks. It is necessary to underline two important aspects when time series are modelled through the use of networks:

1. There is a mathematical procedure that allows to generate networks. This procedure is characterised by some embedding parameters: the dimension m, the delay tau and the overlap. It is not possible to know beforehand

what are the optimal values for these parameters in order to obtain the "best networks". Best networks in terms of significance compared to the time series under exam. This is why, in this thesis, several cases have been analysed so far. All the analyses showed, both in the Euclidean case and in the PCC case, were performed using different m and tau values, both in the OC and in the WOC.

2. Once the networks are obtained, it is possible to calculate the corresponding metrics and evaluate any trends and dependencies. However, these information must be transformed into physical characteristics of the considered system. This step is the pivotal step that is needed every time that complex networks are used to study any phenomenon. In fact, networks can be generated in order to represent different systems: the time series of a turbulent flow, the social relationships between people, rather than the highway system of a nation. For each of these networks, it is possible to calculate the same metrics, but they will have a different physical meaning. For example, in the case of a network generated by a time series, the links will represent a "similarity" between groups of measures, then a "periodicity" in the measured values. While in the case of social networks, they will indicate the presence of a relationship between two people.

Here is the need to identify what is the best method and the what are the optimal embedding parameters in order to build networks, which, once analysed, allow to identify significant information on the fluid dynamics of the turbulent plume in exam.
To this end, it is possible to refer to two studies: the PCC optimization, illustrated in chapter 4, and the metrics trends shown earlier in this chapter.
From the first of these two analyses important information, applicable to obtain a large number of high PCC, can be inferred:

- An exponential degrowth in the number of elevated PCC is connected to parameter m. Therefore, it is advisable to use small m values, for example $m = 50$.
- The more the tau used is low, the more the correlation coefficients are elevated. Therefore, it is advisable to use $\tau = 1$.
- The overlap, using the PCC, is influential. Fixed a certain m and a certain tau, the percentage of PCC greater than a certain value is practically identical whatever is the overlap considered. In order to optimize the calculation time, it is advisable to choose the case without overlap.

With regard to the metrics trends study, it can be observed that:

- The type of method used to define recurrences, the Euclidean norm rather than the PCC, leads to metrics trends that have different shapes. The advantage of the PCC is that the trends obtained have forms directly connectible to the fluid dynamics statistical analysis of the plume.
- If the shape of the metrics depends on the chosen method, the orders of magnitude depend only on the embedding parameters. Fixing a certain m, a certain tau and an overlap, the metrics will have different shape, depending on the method, but will have comparable orders of magnitude.

Therefore, starting from the results of these two analyses, it is possible to identify the best methodology for the construction of the networks related to the plume under consideration.
The PCC use is preferable compared to the Euclidean norm. It allows to obtain metrics that better reflect the fluidynamics of the considered phenomena, thus providing a better description of the time series behaviour.
Regarding the definition of optimal embedding parameters, compared to the three results obtained in the study of the PCC optimisation, they can be fixed equal to:

- $m = 50$;
- $\tau = 1$;
- $overlap = 0\%(WOC)$.

In addition to the method and the optimal parameters, the methodology used to define the 0/1 threshold for the adjacency matrices construction must be defined. In the metrics trends mentioned previously, this was always defined with the quantile representing 95% of the norms or of the PCC obtained. In the Euclidean case, this way can be intuitive. In fact, the norms variability is much higher than the PCC case. The use of the quantile can be considered an "adaptive method". So, a way that, from point to point, using a threshold always different, leads to obtain networks with a density that is maintained about constant. This is evident from the metrics seen above. Whatever the method and the parameters used, the density was always about 5%.
Having now identified, as best calculation method, the use of correlation coefficients, a fixed 0/1 threshold may be used. This way, not being more adaptive with the measuring point considered, allow a greater comparability between the networks obtained in different spatial points. By analysing the graphs shown in chapter 4, with respect to the optimal parameters identified, it was decided to set this threshold equal to 0.70.
Once the method and the optimal parameters have been identified, it is possible to evaluate the metrics trends obtained by using them. In the next section the

main metrics and their characteristics will be presented. These metrics will then be used to describe the behaviour of the considered time series.
Note that, as the purpose of the analyses carried out so far was the determination of the best calculation methods, the vertical turbulent transport (VTT) has not been taken into account. This signal, extremely important for the study of the plume, will be analysed in the next sections.

Chapter 6

Optimized networks metrics analysis

The optimal embedding parameters allow to build the networks that provide the greatest amount of information about the signals of interest. Therfore, this chapter is the heart of the thesis because, from the topology of the networks, the characteristics related to the considered diffusion phenomenon are defined. Since the networks were calculated using the recurrence, in order to confirm what was achieved by the networks themselves, a study of the recurrences is presented at the end of the chapter. The recurrences average distances, both in terms of spatial positions within the adjacency matrices than in temporal terms, and the recurrences frequencies have been calculated. This study leads to results confirming what has been achieved through the networks.

6.1 Study of the time series behaviour using the networks metrics

6.1.1 Analysis of the velocity time series behaviour

In order to represent in a compact way the dependence of the velocities metrics, with respect to the main geometric parameters, for each section a single image has been generated. Each image consists of 4 sub-graphs in which the four considered metrics are shown: nodes, links, density and average closeness. What was obtained in section 2 ($x_d = 0.32$) and section 6 ($x_d = 3.90$) is illustrated in fig. 6.1 and in fig. 6.2.
In each sub-graph there are four curves that represent the metrics for the two speeds, both in the D3 case and in the D6 case.
Despite the different diameter and the different orientation of the velocity, the four curves are comparable, indicating that they have all the same order of magnitude. The number of nodes, whatever the diameter and the section considered, is fixed.

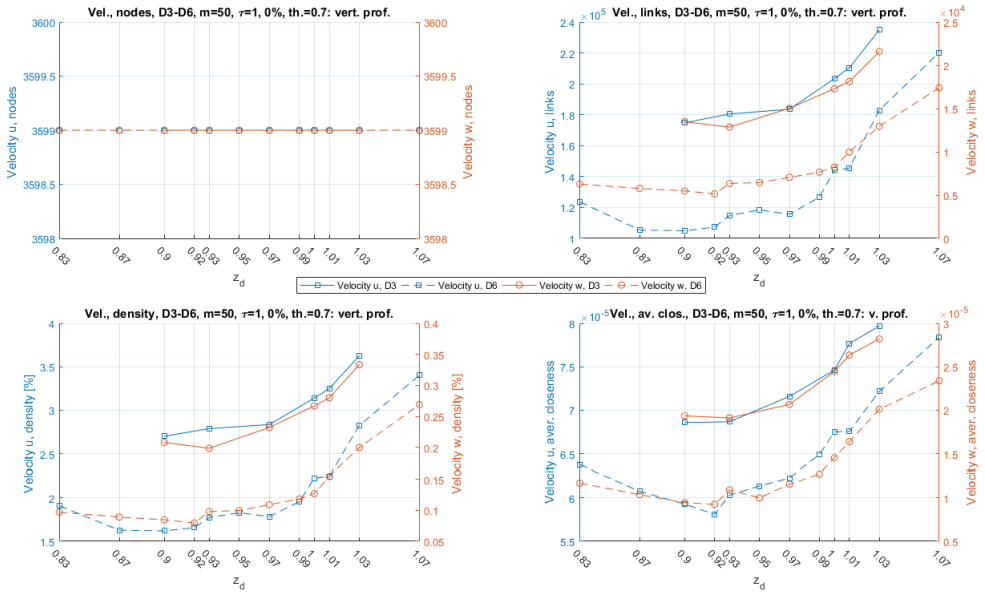

Figure 6.1: Velocities, sect.2, $x_d = 0.32$, metrics for the study of time series

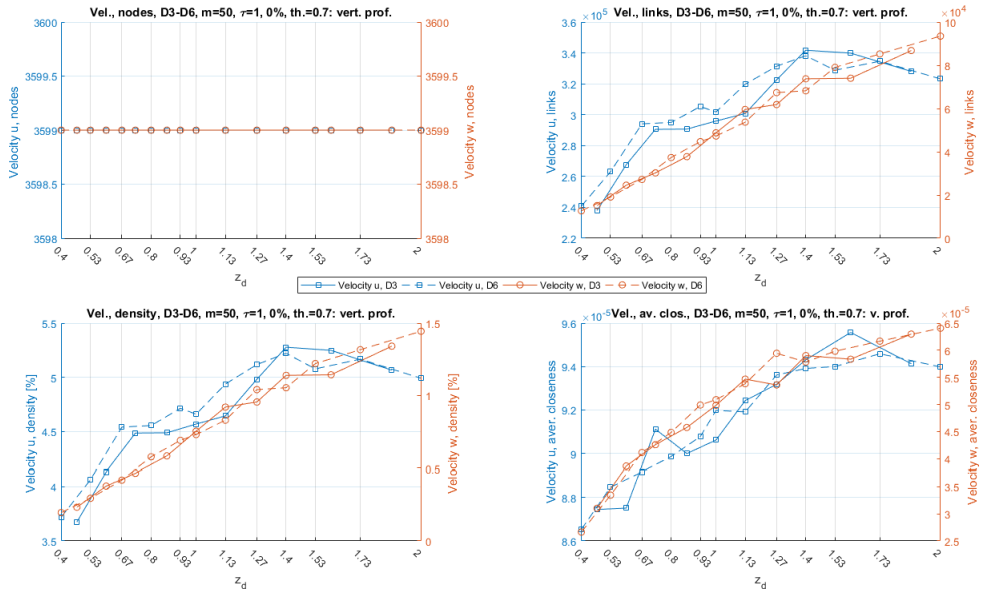

Figure 6.2: Velocities, sect.6, $x_d = 3.90$, metrics for the study of time series

It depends directly from the embedding parameters used. Physically, the number of nodes in a network has a very precise meaning: it represents the number of EVs, so, groups of time series elements. It is interesting to note that using $\tau = 1$ and without overlap, the EVs that are generate contain all the elements of the time

series. By imposing $\tau = 1$, no elements are skipped, while, having no overlap, the elements of two successive EVs are identified in the same order in which they occur in the time series. Therefore, these parameters allow a greater ease and a greater detail in the analysis of the series.
Fixing the x coordinate and moving in the direction of increasing z, the links numbers tends to increase. It tends be a line with positive slope. From the fluid dynamic point of view, this behaviour is very interesting. If the number of nodes remains constant, the size of each adjacency matrix will also remain unchanged. When the z-coordinate increases, the links increase, meaning that, as z increases, the elements equal to 1 present in the matrices increase. Knowing that, in matrices A, a 1 is generated when the correlation between two EVs is greater than 0.7, it can be inferred that, as the z-coordinate increases, the EVs tend to be more correlated. That is, they tend to be more similar. It can therefore be said that, in a given section, as the z-coordinate increases, the velocities time series have less variability.
Consider for example, in the D6 case, the velocity u in section 6. The metrics analysis identifies information about the variability of time series. In order to verify the obtained result, it is possible to calculate a quantity that allows to quantify the variability: the standard deviation. Fig. 6.3 illustrates the average velocity u profile and the standard deviation profile, calculated in section 6 measuring points.

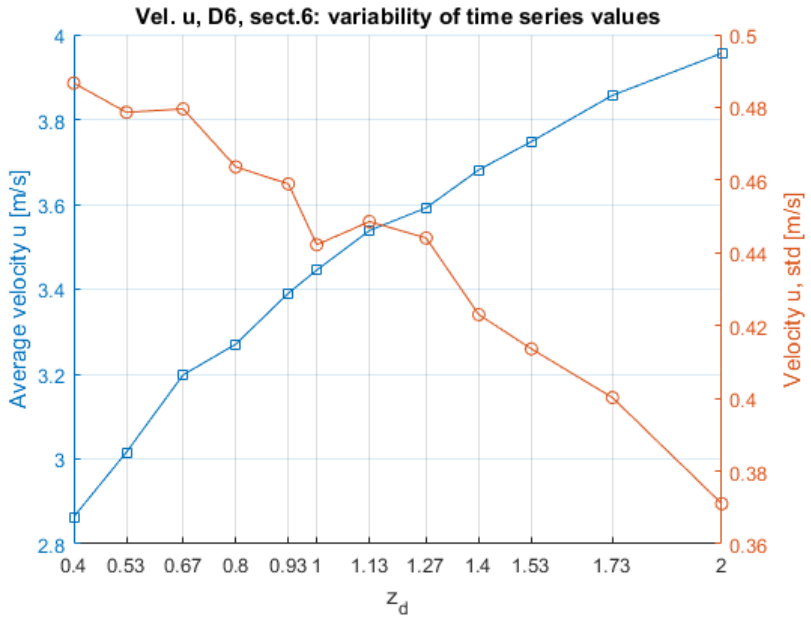

Figure 6.3: Vel. u, sect.6, $x_d = 3.90$, analysis of time series elements variability

This graph reinforces what has been inferred from the metrics. So, the decrease of time series variability for the velocity, in a given section, as the z coordinate increases.
The density trend, remaining constant the matrices A size and increasing the number of links, can only increase. This trend, considering the velocities, reflects the links behaviour.
The average closeness has a growing trend. This metric represents the centrality of the nodes. Considering a single node, i.e. a certain EV of the time series, increasing the z-coordinate, the number of links connected to it will increase. In this way, its centrality will increase, or equivalently, the average path that the separate it from the all the other nodes will decrease. If the considered networks were social networks, characterised by nodes representing people, it could be said that increasing z, and so increasing the average closeness, direct communications would be easier. This because each person would be directly connected to other members, i.e. nodes, of the network.
The links increase, with the same number of nodes, also leads to a decrease in the unconnected nodes number. So, the nodes with $degree = 0$. This metric, calculated in section 6, is illustrated in fig. 6.4

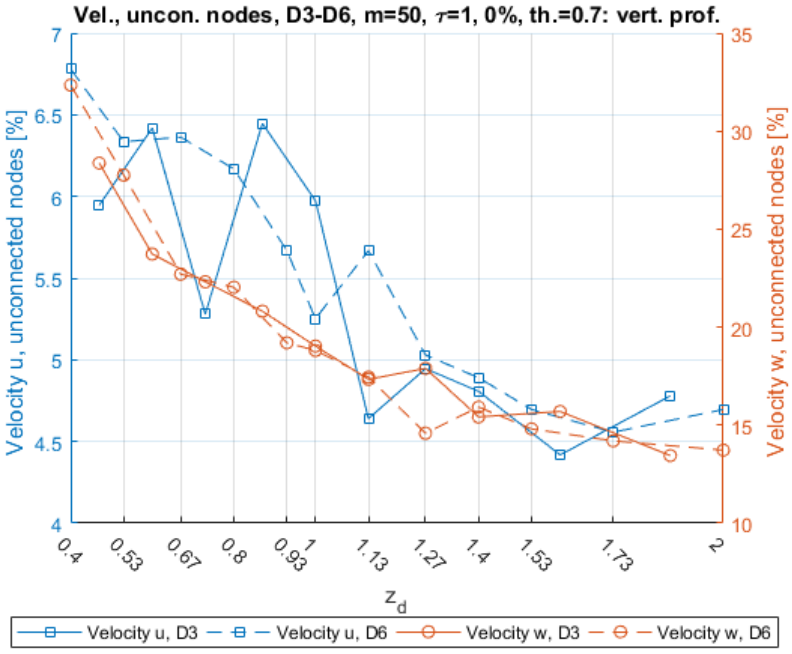

Figure 6.4: Velocities, sect.6, $x_d = 3.90$, unconnected nodes

To summarize the geometry effect on the metrics examined here, it can be said that:

- The effect of the vertical coordinate z is predominantly localizable in time series variability. As z increases, this variability tends to decrease.

- The longitudinal coordinate x effect is not related to the shape of the metrics profiles, but to the numerical value of the metrics. Just observe that, the maximum number of links shown in fig. 6.1 (section 2) is equal to the minimum number of links shown in fig. 6.2 (section 6).

- The effect of the source diameter, even here, is much greater near the nozzle. In section 2, whichever metric is considered, the curves of the D3 case and the D6 case are well separated. Numerically, the D3 case has higher values. In section 6, on the other hand, the curves relative to the two diameters tend to overlap, indicating a decay of the diameter effect moving away from the source.

It is interesting to observe a difference between the velocity u and the velocity w. The latter tends to have, with the same nodes number, metrics that are always lower than the ones of the u, except for the unconnected nodes. This behaviour derives from the fluid dynamics of the problem. What has been noted above on the variability and on the behaviour of the velocity u time series, it is obviously valid also for the w time series. But these will always have an intrinsic variability due to the fact that the phenomenon is non convective. Just think at the average velocity profiles. For the velocity u , these profiles have well-defined shapes, which enclose the effect of friction between the fluid and the wall. Considering, in a given section, the w profile, a "messy" profile is observed, with positive and negative mean values, indicating an integral of the profile near zero, typical of a non-convective problem. That's why, by setting an high threshold like 0.7, there is a lower number of links for the w than the u velocity. Contextually, an higher percentages of the velocity w unconnected nodes indicate that some nodes, so some time series EVs, don't have "any other similar EVs". This means that they don't tend to repeat themselves inside the series.

6.1.2 Comparison of concentration network metrics

Also for the concentration, in each section, a figure containing the graphs of the four considered metrics has been created. In each graph there are two curves, one relative to the concentration in the D3 case and the other relative to the D6.
The concentration metrics have a double behaviour. They have different trends, depending on whether or not there is an interaction between the plume and the wall in the section considered.
The first three sections (2-3-4) have a similar behaviour. Fig. 6.5 illustrates the

graphs obtained in section 2. Different trends are in the last two sections, fig. 6.6 illustrates what has been calculated in section 6.

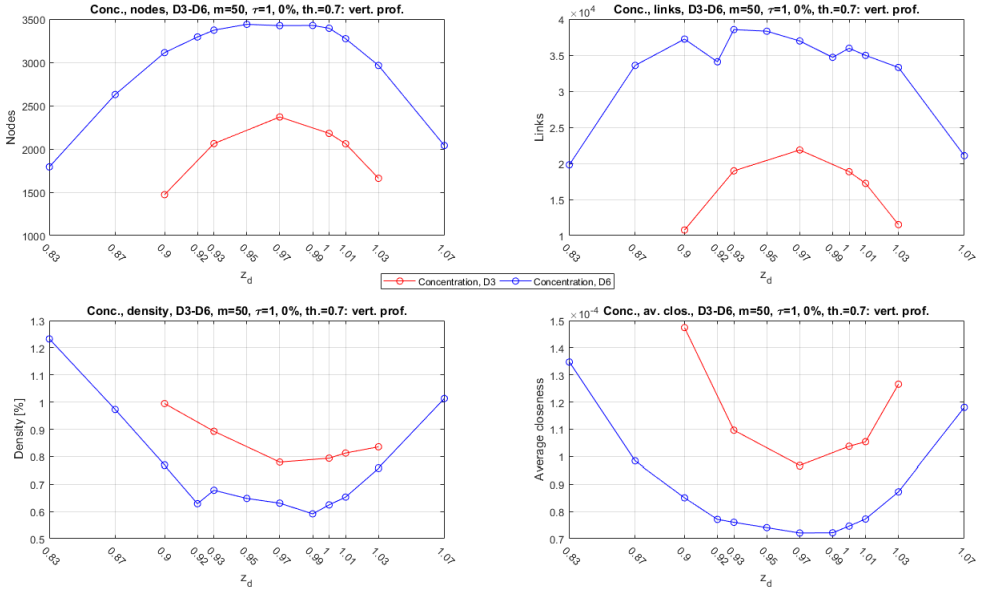

Figure 6.5: Concentration, sect.2, $x_d = 0.32$, metrics for the study of time series

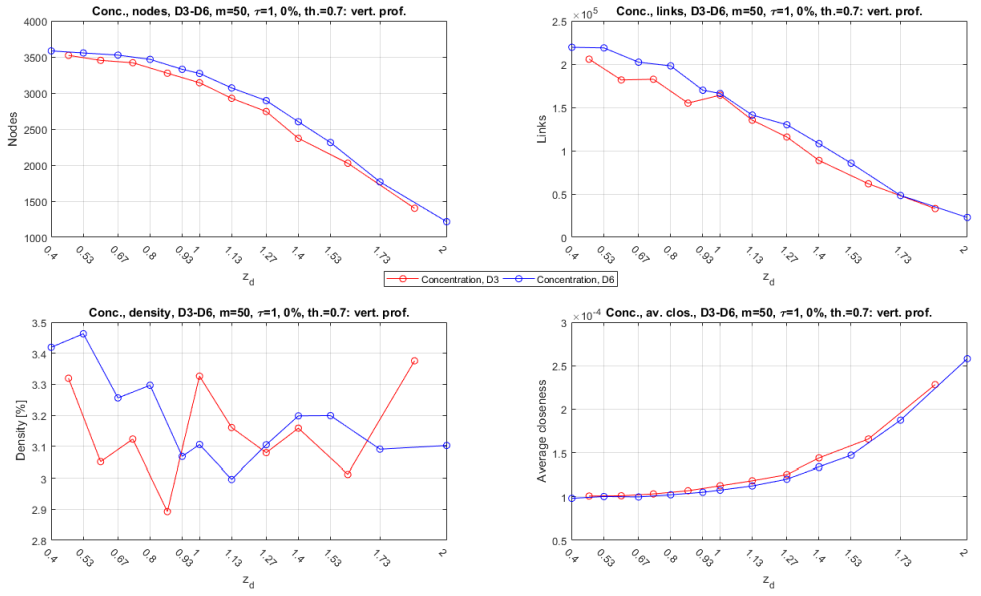

Figure 6.6: Concentration, sect.6, $x_d = 3.90$, metrics for the study of time series

The number of nodes, unlike the velocity case, is no more constant along the vertical coordinate. This is because a "correction" of the EVs with the purpose of

eliminate the insignificant ones, so those that contain too many null elements, is operated. The greater amount of EVs eliminated has been in the measuring points located near the edge areas of the plume. Therefore, in the first three sections, so where the effect of the wall is not yet perceived, the largest number of deleted EVs was in the measuring points most far from the nozzle axis. Instead, in the last two sections, where the effect of the wall generates a concentration accumulation, the largest number of EVs eliminated was in the measuring points farther away from the wall, so at the higher vertical coordinates. In the points where so many more EVs are eliminated, there are fewer nodes and vice versa.
So, this metric has an immediate physical meaning. Where the number of nodes is greater, there the concentration is stronger. That is, the concentration time series will be characterized by a higher average value, will present a lower variability and will be subjected to a less signal filtering. In this way, by evaluating the variation of the nodes, relative to the geometry, it is possible to identify the accumulation effect due to the wall.
In fig. 6.5 a parabolic shape is observed, with the maximum just below the nozzle axis. In the edge areas, i.e. above and below the axis, the number of nodes decreases. Index that, in those measuring points, the filtering of the time series has been more heavy, and that many generated EVs had lots of zeros.
Instead, observing fig. 6.5, the accumulation effect, due to the floor, is immediately visible. The maximum nodes number, that is, the less filtration of the concentration, is close to the wall. The more one moves away from the wall in the increasing z direction, the more the time series will be characterized by lower average concentration values and by a large amounts of zeros resulting from the experimental measurements filtering.
The number of links follow exactly the the number of nodes. Where this is greater, the time series are more stable, i.e. they will have elements slightly oscillating around the mean value. This means that the EVs tend to contain similar elements and, therefore, an higher correlations between them is observed. These correlations, exceeding the imposed threshold of 0.70, allow the formation of more links.
These considerations, which derive from the metrics, can be validated in terms of standard deviation, as done for the velocity u. Fig. 6.7 illustrates the trend of the mean concentration and the standard deviation for the measuring points on section 6, in the D6 case. Being the concentration a transported quantity, it is clear that its measurements will be characterised by a greater variability compared to other quantities, like for example the velocity. However, the figure shows that in the areas where the number of nodes and links is higher, this variability is comparable with respect to the average value. In the border zones, instead, the variability is greater than the average value.
For the density, there is an important difference between the concentration and

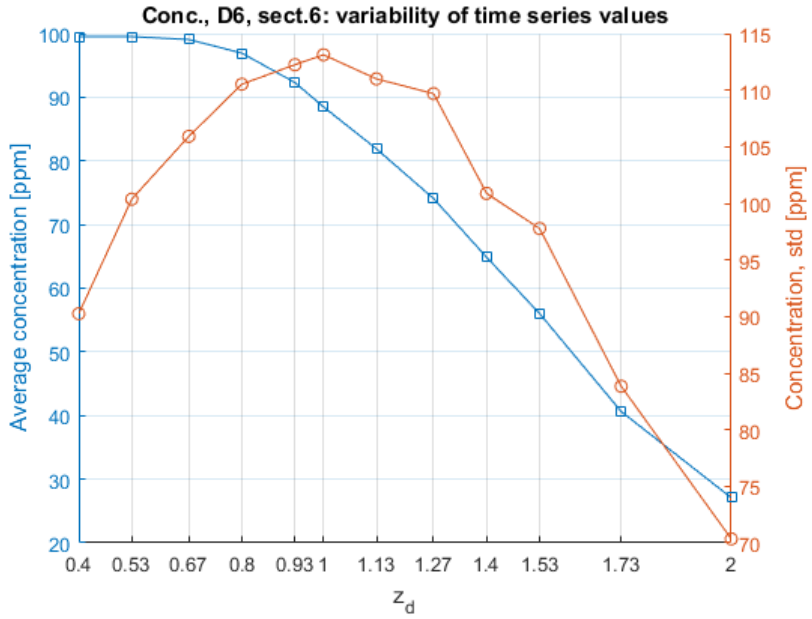

Figure 6.7: Concentration, sect.6, $x_d = 3.90$, analysis of time series elements variability

the velocity. For the latter, being the number of nodes constant, the increase of the links will also increase the density. For the concentration instead, since the number of links and the numbers of nodes vary with the same law, the density is kept about constant. In fact, observing both fig. 6.5 and fig. 6.6, it is noted that the variation in the density is less than 1%.
Wishing to summarize the effect of the geometrical parameters on the metrics examined here, it can be said that:

- The effect of the vertical coordinate z is predominantly localizable in the time series structure. The dependence on z allows to highlight the areas where the concentration has higher average values, less variability and less corrections of the measurements during the data filtering phase.
- The effect of the longitudinal coordinate x is clear. As the x coordinate increases, the shape of the trends changes: it passes from simmetric parabolic shapes to "half parabolas". This dependence is not present in the velocity metrics. One can therefore affirm that the dependence on the longitudinal coordinate is linked to the interaction between the plume and the wall.
- The effect of the diameter is greater close to the nozzle, as already noted.

It is interesting to note that the metrics shown here refer to the concentration

and not to the turbulent fluctuation of the concentration. So, the fluctuation with respect to its average value. Given the use of the PCC, the same metrics can be obtained for both the signals, measured signal and fluctuation signal. In fact, given two EVs, subtracting the average concentration to both and then calculating their PCC, the result is exactly equal to the one obtained without subtracting the mean value. For this reason, it is not necessary to consider the concentration and its turbulent fluctuation separately.

6.1.3 Analysis of the vertical turbulent transport time series

The vertical turbulent transport (VTT) metrics are now illustrated. These metrics have a peculiar characteristic, compared to the geometry of the wind tunnel. Fig. 6.8 and fig. 6.9 represent the metrics obtained in the first and last section examined.

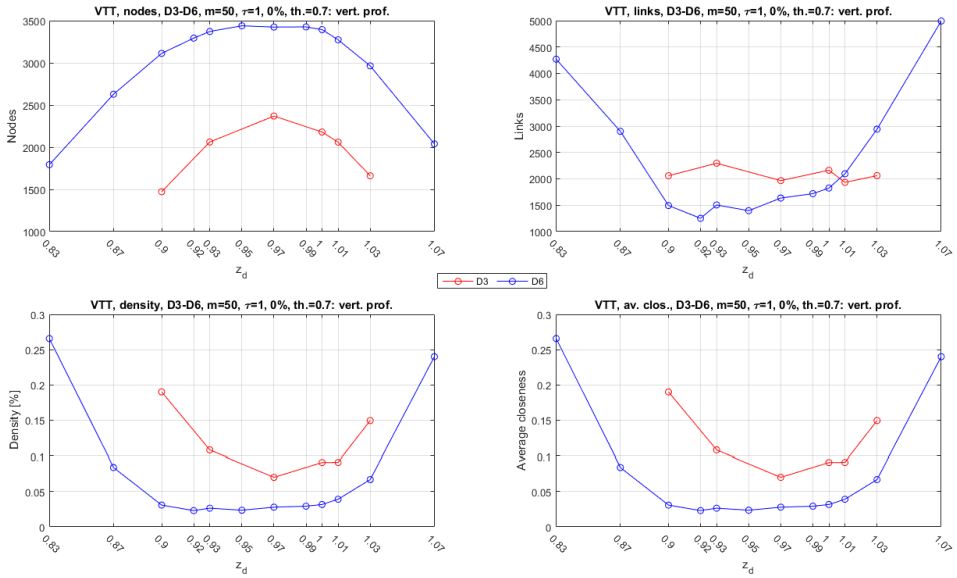

Figure 6.8: VTT, sect.2, $x_d = 0.32$, metrics for the study of time series

As shown in fig. 6.9, where the effect of the wall is perceptible, the metrics present the same trends described for the concentration. Therefore, also for the VTT time series, in section 5 and 6, it is possible to infer what has been described previously for the concentration series.
In the first three sections, so for the horizontal coordinates in which the plume has not yet interacted with the wall, the metrics present some differences with respect to the concentration. The number of nodes is obviously the same, because the EVs used for calculating the VTT networks are exactly those considered for the

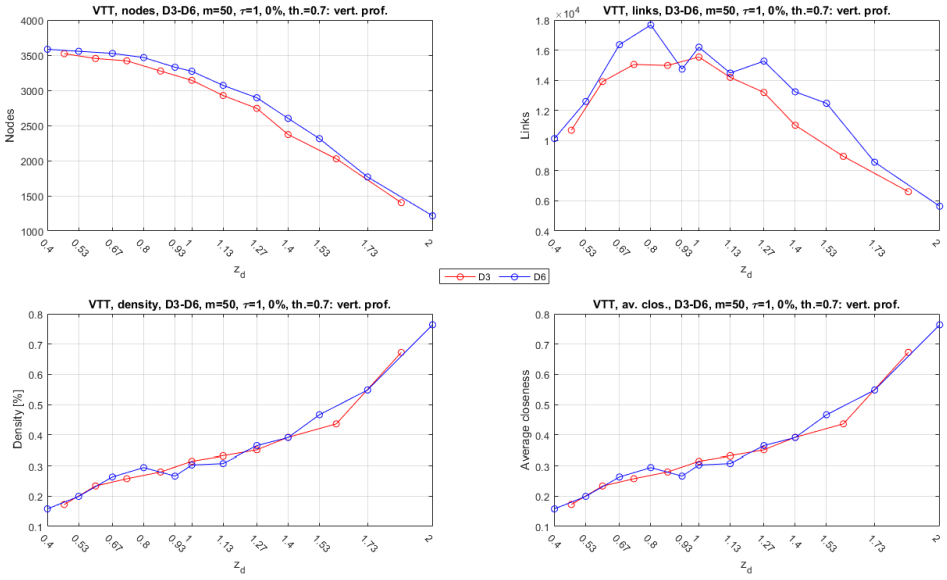

Figure 6.9: VTT, sect.6, $x_d = 3.90$, metrics for the study of time series

concentration. Instead, the number of links has an opposite trend: for the vertical coordinates close to the nozzle axis, the number of links is less than one present in the plume edge. Below the nozzle axis, having a number of nodes growing and a number of links decreasing, the density can only decrease. Opposite condition occurs above the axis.
As described, the phenomenon is not convective. Therfore, the speed w has an high variability compared to the average value, equal to zero. When the w fluctuations are multiplied by the concentration fluctuations, even where these are reduced (around the nozzle axis), the time series that is generated is extremely variable. It does not exhibit many recurrences and so, imposing a correlation threshold equal to 0.70, it is not possible to obtain a large number of connections within the network.
The dependence of these metrics from the longitudinal coordinate x is therefore double: it has an effect on the metrics shape and on their numerical values. It should be noted that, unlike the previous signals (velocity and concentration), for the VTT, the variations of the numerical values are larger. The number of links, from section 2 to section 6, increases by an order of magnitude. To represent these dependencies more clearly, fig. 6.10 and fig. 6.11 can be considered. They represent the metrics obtained in the five different sections. The left column refers to the D3 case, while the right one to the D6.

The previous images show that, also for the VTT, the diameter effect decrease moving away from the source.

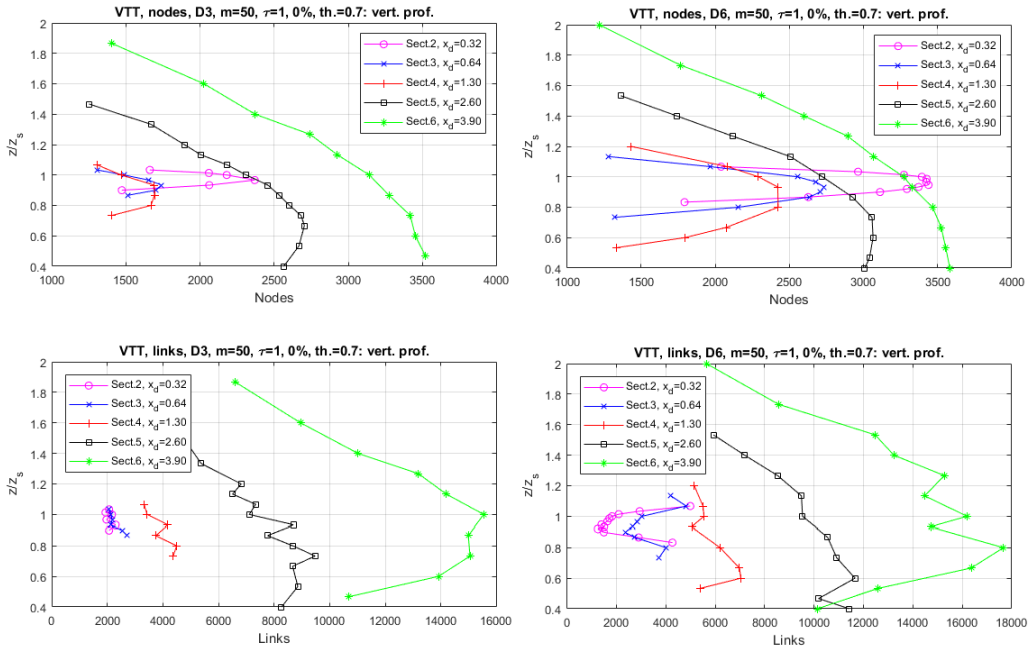

Figure 6.10: VTT, nodes and links, joint effect of the longitudinal and the vertical coordinates

To confirm the extreme variability of the VTT time series, it is possible to consider the unconnected nodes trends. As already indicated above, the more this metric is high, the more the EVs are dissimilar from each other. That is, the elements of the time series tend to be very different, thus lacking recurrences in the considered signal. As shown in fig. 6.12, of all the signals analysed, the VTT is the one that has the maximum number of unconnected nodes. Also for this metric the behaviour is very dependent on the x coordinate. In the first three sections, even near the nozzle axis where there is an high average concentration and where the plume is stable, the variability of the velocity w prevails, and so the percentages of unconnected nodes are very high. The 70% is exceeded in the second section. These percentages are partly reduced in the last two sections, where the effect of the wall occurs. The concentration accumulation due to the wall generates particularly stable concentration time series and with reduced variability. These characteristics of the concentration partially attenuate the VTT variability. This is visible both through the number of nodes and links, that increase in the lower part of the section, than through the percentage of unconnected nodes that is clearly reduced compared to previous sections. This indicates a greater presence of "recurrences" in the time series, or equivalently , EVs more similar to each other and then connected through a link.

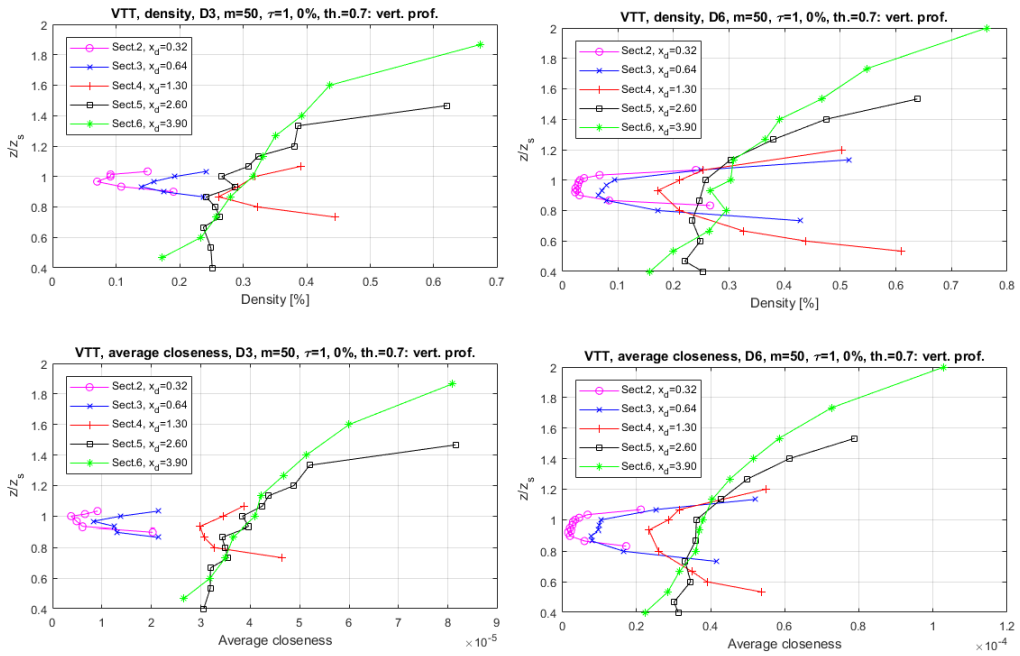

Figure 6.11: VTT, density and average closen., joint effect of the longitudinal and the vertical coordinates

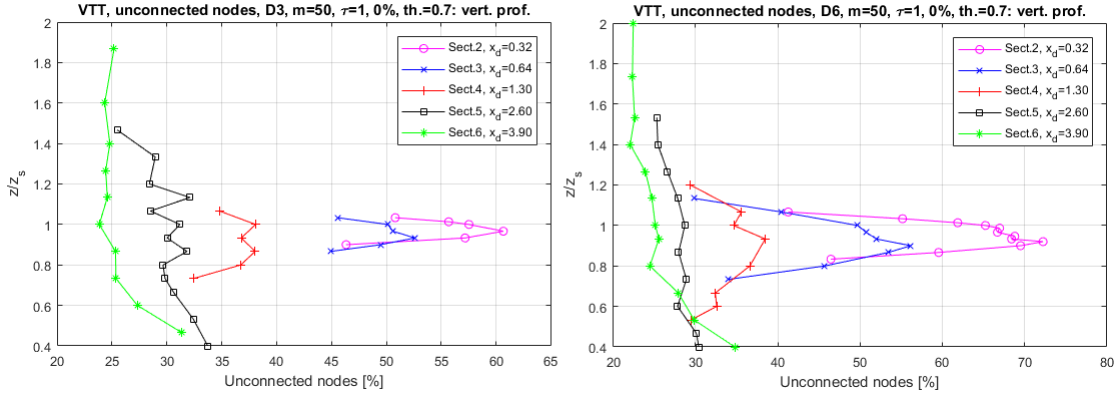

Figure 6.12: VTT, unconnected nodes, joint effect of the longitudinal and the vertical coordinates

6.2 Frequency of recurrences

As already described, the heart of the method used here consists in the study of time series through the recurrences that are within them. Until now, through two different analyses, the PCC optimisation and the network metrics, important

information about the time series of interest has been obtained. The main features illustrated are the series stability, to be understood as a limited variability of the instantaneous values with respect to the average value, the influence of the geometric parameters on the measured values, the interaction between the plume and the wall and the transport of the passive scalar inside the plume.
Being the recurrences inside the series the pivotal point of the method, it is interesting to evaluate how these are distributed. That is, considering these recurrence as similar parts inside the time series, it is interesting to evaluate when they repeat within the signals.
In order to evaluate the recurrences the following steps are considered:

1. Given a certain signal and fixing the embedding parameters equal to the optimum ones previously identified, the corresponding network and its adjacency matrix are generated.

2. As illustrated in fig. 6.13, for each row, the relative intervals δ_{ij} between each recurrences in that row is identified. Than, the average value of these distances is defined. This value, for notation consistency, is indicated with $\overline{\delta_{ij}}$.

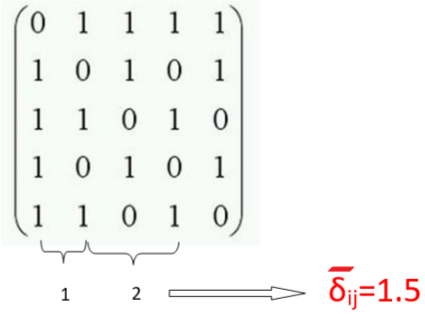

Figure 6.13: Distances between recurrences

   Therefore, for each generated network, a vector containing the average distances $\overline{\delta_{ij}}$ evaluated in each row is obtained. This vector will therefore have a number of elements equal to the network nodes.

3. To show the various distances obtained, so the vectors elements, their frequency distributions can be represented. The range between the maximum distance and the minimum (which is one, without considering the null elements) has been divided into small intervals. Then, the number of distances in each intervals, compared to the total number of calculated distances in the node, was assessed. In this way the percentage frequency has been defined. Thus, graphs that show the frequency on the ordinate axis and the values of the average distances $\overline{\delta_{ij}}$ on the abscissa have been created.

The following sections show what has been achieved for the four analysed signals.

6.2.1 Concentration recurrences

Fig. 6.14 illustrates what has been obtained in section number 2, in the D3 case. For each measuring point, that is, for each vertical coordinate, the corresponding frequency profile has been shown. As it is possible to see, in the first part of the graph, the curves tend to grow reaching a maximum, after which the curves tend to decrease having the abscissa axis as an asymptote. The peak corresponds to a distance of about 150-200 and is characterized by frequencies of about 15-18%. This small variability depends on the vertical coordinate considered. Moving

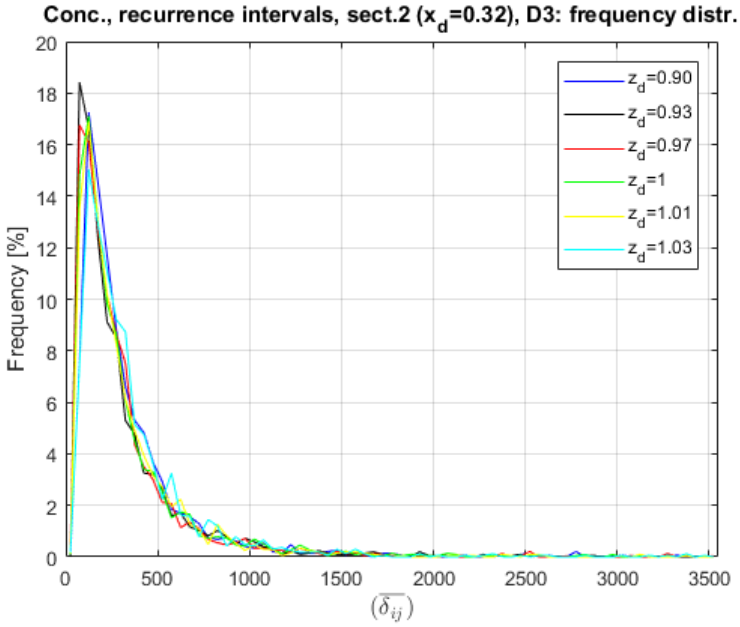

Figure 6.14: Concentration, section 2, D3: recurrences frequency analysis

towards larger x-coordinates, similar trends to that on fig. 6.14 are found up to section 4 included. But there are two important differences changing the section:

- the peak moves towards the left, so at smaller distances;
- the peak frequencies became bigger.

Instead, in sections 5 and 6 the curves have slightly different trends. The fig. 6.15 illustrates what has been obtained in section 6. The growing initial part disappears and the curves immediately have a peak followed by the decreasing part. The peaks frequency distances are of about 50 and percentages over 50%.

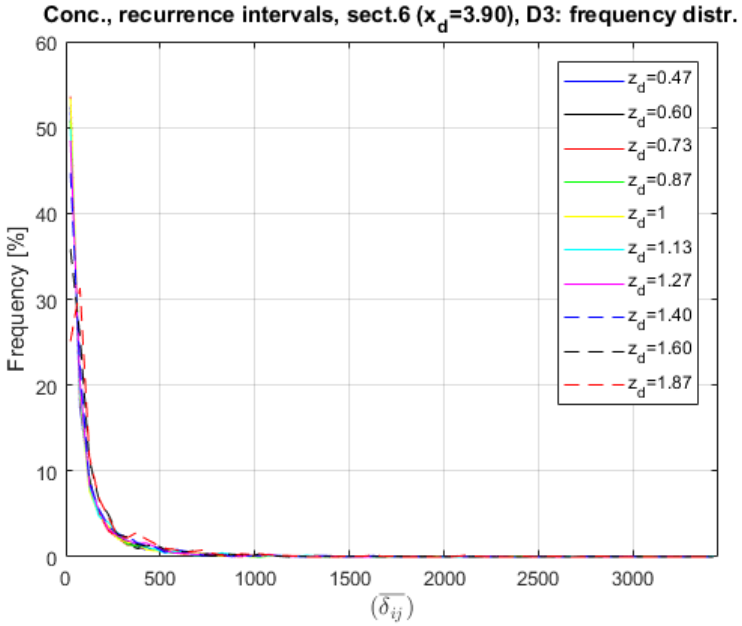

Figure 6.15: Concentration, section 6, D3: recurrences frequency analysis

The physical meaning of this analysis is completely in tune with was identified by the networks metrics studies. In fact, in the first sections, for vertical coordinates close to the nozzle axis, there are concentration time series stable compared to their average value. They are characterized by signal fragments that tend to repeat at medium distances between each other. Moving away from the nozzle, the intervals between these repetitions decrease. It can be said that, in sections 3 and 4, the effect of mixing ethane with air leads to a more homogeneous measured signals. So, signals characterized to more recurrences gradually closer together, that is, separated by a lower time interval.
In the last two sections, instead, the stabilizing effect is due to the wall: the concentration accumulation around it. In these sections most of the signals recurrences (peak values over 50%) are very close to each other. In fact, moving from section 2 to section 6, the peak distance is decreased by about 4 times.
The concentration metrics study had identified, in the accumulation zones, time series with high stability, with poor signals filtering and characterized by a small variability compared to the average value, this latter always increases moving towards the wall. All these characteristics are perfectly in agree with the information that is inferred from the recurrences study: series formed by very similar elements, which tend to repeat themselves at short distances. Distances that do not depend much on the vertical coordinate considered in a given section but, rather, from the x-coordinate considered. That is, from the distances from the source.

What has been reported here is the D3 case. By analyzing the 6 mm case, the same results are found. The only difference is slightly larger peak percentages in the first sections. Also here, as mentioned above, the effect of the diameter is perceptible only in the initial sections. In the last two sections this parameter is practically irrelevant.

6.2.2 Velocity u recurrences

In this section, the purpose is to analyze the velocity u signals with respect to the recurrences. To do this, fig. 6.16 can be considered. It shows what has been obtained in section 2, considering the 3 mm nozzle diameter. As it is possible to see, the curves have an initial peak followed by a decreasing part.

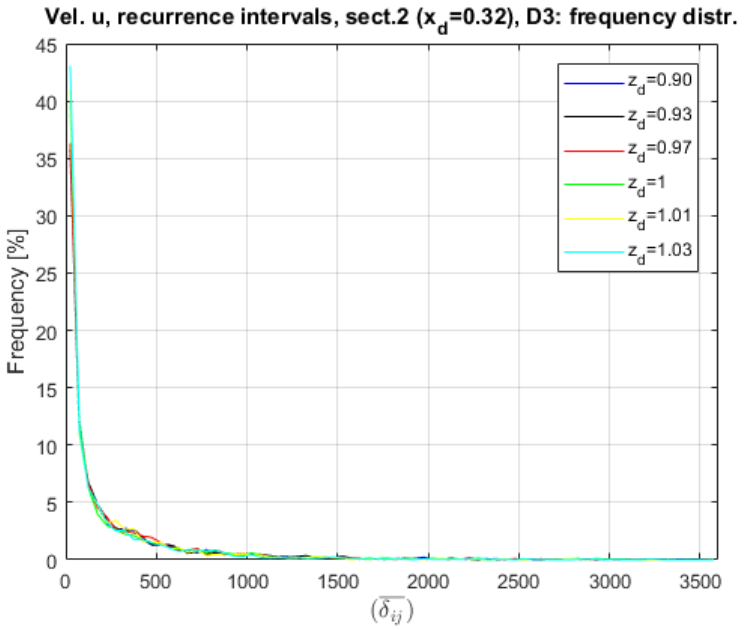

Figure 6.16: Velocity u, section 2, D3: recurrences frequency analysis

Considering different sections, very similar figures and numerical values are obtained. The effect of the diameter is also practically null. It can be inferred that the velocity u time series, whatever the diameter or the measuring point considered, will be characterized by a great quantity of recurences very close to each others. A high stability of these signals was also detected studying networks metrics.

6.2.3 Velocity w recurrences

As far as velocity w, the obtained trends are much more similar to the concentration ones than those of velocity u. Fig. 6.17 illustrates the curves obtained in section 2, for the D3 case. Even here it is possible to observe a first growing trend that leads to a peak, followed by a decreasing part going asymptotically to zero. The curves relative to different vertical coordinates z have all the same trends but are considerably less similar than in the case of velocity u. In fact, for this latter, the different curves were practically overlapping.

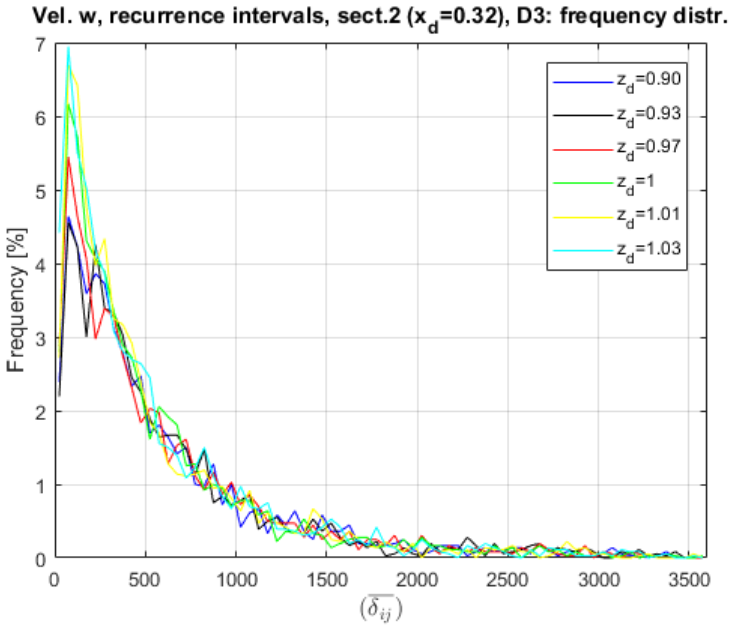

Figure 6.17: Velocity w, section 2, D3: recurrences frequency analysis

Moving from section 2 to section 4, the same trends are observed. The only difference is in the peaks that have bigger percentages.
Slightly different trends are observed in the last two sections: here, for measuring points with $z_d > 1$, the curves no longer have a maximum but have trends similar to those of the velocity u and higher percentages respect to the measuring points under the nozzle.
Even for this signal the diameter effect is a slight variation of the frequencies observable mainly in the initial sections.

6.2.4 VTT recurrences

As mentioned above, the VTT time series are calculated from the concentration and the velocity w signals in each measuring point. These series, therefore, reflect

in part the behaviour of the signals from which they derive. This is also observed in the case of the recurrences frequency: there are very similar trends to those obtained for the concentration and the velocity w. Fig. 6.18 illustrates the curves obtained in section 2, for the D3 case. Obviously, since this signal derives from

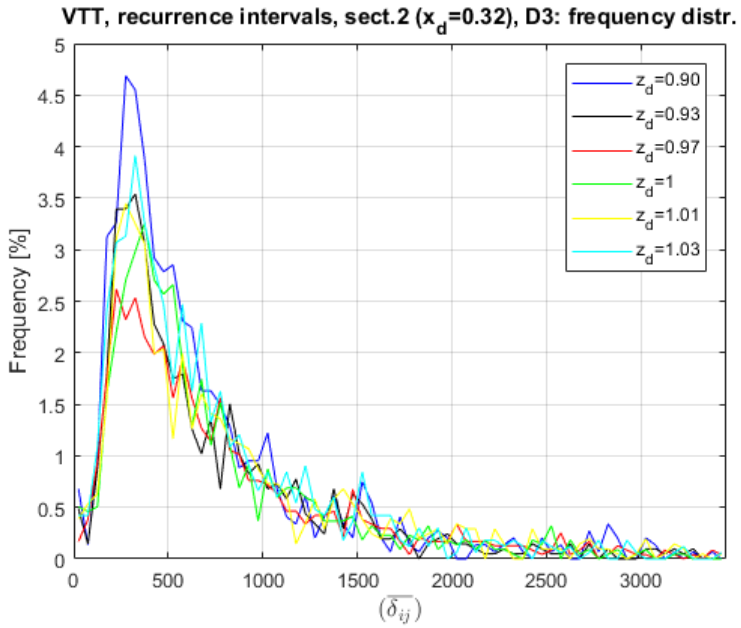

Figure 6.18: VTT, section 2, D3: recurrences frequency analysis

the combination of turbulent fluctuations, it will have an intrinsic randomness. This randomness can be observed in the peak that each curve present. In fact, the maximum percentages do not exceed 5% in the first two sections, while is below the 10% in the remaining sections. The peak distance is quite high, around 300 in section 2. As the x-coordinate increases, this distance tends to decrease, but much less than the signals from which the VTT derives, i.e. in section 6 it is around 100-120.
Therefore, there are similarities in the signal, but there are few (low percentages) and very far (high peak distances). What is deduced from this analysis is entirely appropriate with the randomness of this signal and in accordance with the features illustrated by the networks metrics.

6.2.5 Recurrences in the time domain

As shown in the previous sections, the δ_{ij} distance represents the distance between two recurrences, and so, between two similar EVs. Using the frequency with which the measurements have been made, 1000 Hz, it is possible to evaluate the distances

in temporal terms. In fact, dividing the distances obtained by the measurement frequency the temporal distances are obtained. Thus, in the preceding graphs, the values on the abscissa axis can also be interpreted as a time distance expressed in milli seconds.

Chapter 7

Conclusions

This thesis illustrates the application of complex network theory to the study of turbulent flows. In particular, through the analysis of experimental time series, the diffusion of a passive scalar within a completely developed turbulent boundary layer has been investigated. The time series, and so the signals that characterize the turbulent plume, have been firstly analysed from a statistical point of view. Then, these series have been transformed into complex networks using the Recurrence Networks approch, i.e. going to evaluate any similarities within the signals. So, in the obtained networks, the nodes represent signal contiguous measurements. Instead, the links represent the recurrences between nodes. The definition of the embedded vectors distances in phase space, i.e. the similarities between the signal parts, have been defined using two different calculation methods: the Euclidean norm and the Pearson correlation coefficients. This aspect is interesting in order to report the ways in which this thesis have been developed. The final elaborate does not report chronologically all the calculations and all the tests executed. This because, being a research thesis, to reach the results reported here, several ways of proceeding have been followed before choosing which one to take definitively. One example is the double calculation of the distance in the phases space, or the various experimentations carried out to identify the embedding parameters to use, rather than the tests to identify the most appropriate metrics. Once networks have been obatained, their dependence from the embedding parameters has been assessed in order to identify the optimum setting. The metrics of the "optimized" networks have been studied in order to identify the signals characteristics, so in order to characterize the diffusion within the analysed boundary layer.
The main results, divided in three macro-categories, are listed below:

1. **SIGNALS STATISTICAL ANALYSIS**
   This section shows the main results related to the system geometry effects on time series. So, their dependence on the nozzle size, on the distance from

the source, on the vertical z-coordinate and on the measuring points relative position respect to the wall. In fact, the geometry of the system has an important effect on the signals structure and on their properties. The effect of the wall on the concentration and on the longitudinal velocity u is emblematic. From section 4 ($x_d = 1.30$) onwards, the interaction between the plume and the wall creates a concentration accumulation adjacent to the wall itself. It is interesting to observe that the average concentrations found in this accumulation area are scarcely dependent on the source diameter value. Opposite behavior compared to the average concentrations found in the first sections. Here, doubling the diameter, the average concentrations are also doubled. The diameter effect on the average concentration must not be mislead: increasing the diameter does not increase the ethane diffusivity within the flow. The increase in the concentration signals that occurs doubling the diameter is due to the simultaneous increase of the ethane flow at the source. In fact, normalizing concentration signals with respect to the ethane flow at the source, it is possible to observe that, whatever is the diameter considered, the obtained plumes are characterised by comparable diffusive phenomena, even in the initial sections.
The velocity u does not present any particular dependencies on the considered diameter and on the longitudinal x-coordinate. Instead, it depends on the z-coordinate, so on the distance from the wall. Its effect is a friction effect that generates vertical profiles with parabolic shapes.
The statistical analysis carried out on the transverse velocity w signals confirms that the phenomenon is non convective. This because, integrating an average profile of this signal, a value close to zero can be obtained.

2. MODEL CALIBRATION AND OPTIMUM PARAMETERS DEFINITION

This section shows the main results related to the model setting and on its calibration, so the optimum embedding parameters definition. These latter, i.e. the parameters that allow to build significant networks compared to the considered signals, are not known a priori. Therefore, to define them, networks structures and their metrics must be evaluated varying both the constitutive parameters than the geometrical ones.
Using the same embedding parameters, the networks obtained with the Euclidean norms and those obtained with the Pearson correlation coefficients have a different structure. The former have adjacency matrices characterized by extremely connected nodes. So few nodes with very high degree centrality. Instead, through the use of the correlation coefficients, the unitary values inside the matrices are distributed in a much more uniform way.
Obviously, networks with different structures will also present metrics with

different trends. The networks obtained using the Euclidean norm have metrics that oscillate around an average value. In fact, section by section, considering the metrics vertical profiles, constant graphs have been obtained. Instead, the metrics profiles obtained using correlation coefficients tend to trace more the physics of the problem, so highlighting for example the effect of the nozzle size and of the wall for the concentration networks. From the analyzed metrics it is also possible to observe that the shape of the profiles depends on the calculation method used (Euclidean norm or correlation coefficients), while the numerical values depend on the embedding parameters used. Therefore, for the study of the diffusion phenomenon analysed here, it is advisable to use the Pearson correlation coefficients and not the Euclidean norm. Considering networks obtained using them, the embedded vectors size m has a direct effect on the correlation coefficients values. The more m is small, the more high coefficients values will be. Other parameters, such as the delay τ, have a smaller effect on the correlations, while the effect of the overlap is completely null. Then the embedding parameters can be set with $m = 50$ and $\tau = 1$, in order to optimize the Pearson correlation coefficients distributions. With regard to the overlap, it is preferable a null value of it. This in order to improve the calculation time. To increase networks comparability, it is advisable to use a fixed threshold for the determination of the adjacency matrix A and not an adaptive method such as the quantile at 95%.

3. **NETWORKS METRICS AND TIME SERIES BEHAVIOUR**

Once the optimum networks have been identified, through the study of their topology, information about the starting time series can be obtained. From the velocity networks it is possible to observe that, fixed a given section, as the z-coordinate increases, the number of nodes remains constant, while the number of links increases. This means that, as the z-coordinate increases, the velocities time series are more stable, so they will be characterized by less variability with respect to the average value. That is, as the z-coordinate increases, the standard deviation of the signals will decrease. Instead, the distance from the source does not have an effect on the time series behaviour, but on the numerical values of the metrics. This is due to the fact that the motion field is "frozen", so it is fixed by the wind tunnel settings.

Observing concentration networks it is possible to see that the nodes tends to follow the plume movement: in the first sections, the maximum of the nodes is obtained for vertical coordinates close to the nozzle coordinate, whereas, after the interaction between the plume and the wall, the maximum number of nodes is located near the wall. The same behaviour characterizes the links trends. These metrics have an immediate physical meaning. Where the

number of nodes and links are greater, there the concentration is stronger. So the concentration time series will be characterized by a higher average value, will present a lower variability (higher stability) and will be subjected to a less signal filtering. Metrics allow also to understand that the effects of the geometrical parameters on the concentration time series is particularly marked. The z-coordinate defines the stability of time series, to be understood as a reduced variability with respect to the mean value. The vertical coordinates at which the series are more stable depend on the longitudinal x-coordinate considered. Moving away from the nozzle, these coordinates tend to decrease. As already indicated for the average concentration, also for the metrics, and therefore also for the characteristics of the time series, the effect of the diameter is localized more in the first sections. Here the nodes and links are greater in the D6 case. Moving away from the nozzle, instead, the two diameters tend to have similar metrics. This means that, in the first sections, the D6 concentration time series is less intermittent and more stable.

For the study of the vertical turbulent transport it is necessary to observe that these time series derive directly from those of the concentration and of the w velocity, and so, from these series inherit its behaviour. As a result of the interaction between the plume and the wall, the concentration behaviour prevails over that of the velocity w and, in the accumulation zones, the vertical turbulent transport signal is similar to that of the concentration. In the first sections, instead, is the velocity w variability to characterize more the time series of the vertical turbulent transport. In fact, when the w fluctuations are multiplied by the concentration fluctuations, even where these are reduced (around the nozzle axis), the time series that are generated are extremely variable. They do not exhibit many recurrences and so, imposing a correlation threshold equal to 0.70, it is not possible to obtain a large number of connections within the network. Through the analysis of the distances between the recurrences, it is possible to obtain information about the considered signals. In particular, by plotting the distances frequencies, it is possible to perform a comparison between the turbulent vertical transport and the other signals. The frequence curves have a peak that represents the distance that is repeated more frequently within the signals. Obviously, since this signal derives from the combination of turbulent fluctuations, it will have an intrinsic randomness. This randomness can be observed in the peak that each curve present. In fact, the maximum percentages do not exceed 5% in the first two sections, while is below the 10% in the remaining sections. The peak distance is quite high, around 300 in section 2. As the x-coordinate increases, this distance tends to decrease, but much less than

the signals from which the vertical turbulent transport derives, i.e. in section 6 it is around 100-120.

List of acronyms

EV	Embedded vector
OC	Overlap case
PCC	Pearson correlation coefficient
RN	Recurrence network
RP	Recurrence plot
RQA	Recurrence quantication analysis
TVT	Turbolent vertical transport
WOC	without overlap case

Bibliography and sitography

[1] J.L. Lumley, Engines: An Introduction, Cambridge University Press, 1999;
[2] D. Citrini, G. Noseda, Idraulica, CEA, 1987;
[3] E. Spessa, Lezioni di Combustione e gasdinamica delle macchine, Politecnico di Torino, 2018;
[4] S.V. Patankar, Numerical heat transfer and fluid flow, Hemisphere Publishing Corporation, 1980;
[5] C. Nironi, P. Salizzoni, M. Marro, Boundary-Layer Meteorol 2015, pp. 156:415;
[6] J. Jimenez, Turbulent flows over rough walls, Annual Review of Fluid Mechanics 2004, pp. 173-196;
[7] Enciclopedia Treccani, Problema dei ponti di Konigsberg;
[8]https://www.isc.cnr.it/public-outreach/divulgazione/scienza-della
-complessita-e-teoria-delle-reti/;
[9] E. Zhuang, Physica A 2014, pp. 410-483;
[10] L. Telesca, M. Lovallo, EPL, Volume 97, Number 5, 2012;
[11] S. Boccaletti et al., Complex networks: structure and dynamics, Physics reports, Volume 424, Issues 4-5, 2006, pp. 175-308;
[12] Donner Reik V. et al., Recurrence networks as novel paradigm for non-linear time series analysis, New Journal of Physics, 12.3 2010;
[13] Donner Reik V. et al., Recurrence-based time series analysis by means of complex network methods, International Journal of Bifurcation and Chaos 21.04 2011, pp. 1019-1046;
[14] https://gephi.org/tutorials/gephi-tutorial-layouts.pdf.