Application of the Teaching Learning Optimization Algorithm to an Analytical Model of Thunderstorm Outflows to Analyze the Variability of the Downburst Kinematic and Geometric Parameters

Andi Xhelaj [1], Massimiliano Burlando [1]

[1] *Department of Civil, Chemical and Environmental Engineering*
Polytechnic School, University of Genoa, Via Montallegro 1, 16145 Genoa, Italy

Correspondence to: Andi Xhelaj (andi.xhelaj@edu.unige.it)

Abstract. Downbursts winds, characterized by strong, localized downdrafts and subsequent horizontal straight-line winds, presents significant risk to civil structures. The transient nature and limited spatial extent present measurement challenges, necessitating analytical models for accurate understanding and predicting their action on structures. This study analyzes the Sânnicolau Mare downburst event in Romania, from June 25, 2021, using a bi-dimensional analytical model coupled with the Teaching Learning Optimization Algorithm (TLBO). The intent is to understand the distinct solutions generated by the optimization algorithm and assess their physical validity. Supporting this examination is a damage survey and wind speed data recorded during the downburst event. Employed techniques include agglomerative hierarchical clustering with the K-means algorithm (AHK-MC) and principal component analysis (PCA) to categorize and interpret the solutions. Three main clusters emerge, each displaying different storm characteristics. Comparing the simulated maximum velocity with hail damage trajectories indicates that the optimal solution offers the best overlap, affirming its effectiveness in reconstructing downburst wind fields. However, these findings are specific to the Sânnicolau Mare event, underlining the need for a similar examination of multiple downburst events for broader validity.

KEYWORDS: Downburst analytical model, Metaheuristic optimization algorithm, Multivariate data analysis, Downburst kinematic and geometric parameters, Damage survey.

1 Introduction

The wind climatology of Europe and several mid-latitude countries are primarily dominated by the presence of extra-tropical cyclones and thunderstorms. The understanding of the formation and evolution of extra-tropical cyclones dates back to the 1920s (Bjerknes and Solberg, 1922). The atmospheric boundary layer (ABL) winds generated during such systems are well recognized, and their influence on structures has been extensively studied and coded starting from the 1960s (Davenport, 1961). These established models continue to be employed in contemporary engineering practice (Solari, 2019).

Thunderstorm winds known as “downburst” consists of a strong and localized downdraft of air generated within a convective cell. These downdrafts after reaching the ground begin to spread horizontally, resulting in the formation of the downburst gust front, also known as the downburst outflow. The presence of strong turbulent wind within the downburst outflow poses significant risk to civil structures. Given their high frequency of occurrence, downburst events are among the most severe meteorological phenomena in mid latitudes. Downbursts, often generated by isolated thunderstorms,

typically exhibit scales of less than few kilometers in extent, distinguishing them from the larger scale of thunderstorms themselves. Additionally, they can be originated from more complex convective systems such as squall lines and bow echoes, in this case the spatial length scale which can potentially be affected by downbursts or downburst clusters is in the order of hundreds of kilometers (Fujita, 1978, Hjelmfelt, 2007). The size of the downburst outflow area of strong winds exhibits variability, leading to the classification of this phenomenon as either a microburst or macrobust. A microburst is characterized by a strong outflow size that is less than 4 km, whereas a macroburst corresponds to an outflow size of intense wind greater than 4 km (Fujita, 1985).

For over four decades, intense downburst winds and their impact on the built environment have been key research topics in the field of Wind Engineering (Letchford, 2002). These winds, resulting from nonstationary behaviours in mesoscale thunderstorms, create a distinct horizontal wind profile. This profile, marked by a nose-shape with peak wind speed near the ground level, sharply contrast with the typical wind profiles in the ABL and significantly endangers structures, particularly those of low and medium height.

From a statistical point of view, wind velocities, characterized by a mean return period greater than 10 or 20 years, are often due to these phenomena (Solari, 2014). The lack of a unified model for downburst outflows and their actions on structures, similar to Davenport's (1961) model for extra-tropical cyclones, is primarily due to significant uncertainties arising by the inherent complexity of downburst winds. Indeed, the transient nature and limited spatial extent of downbursts presents challenges in their measurements and restrict the availability of an adequate number of test cases.

The early analytical models for downburst wind velocities stemmed from Glauert's (1956) impinging wall jet model and Ivan's (1986) ring vortex model. Glauert focused on radial jets, while Ivan developed an axisymmetric downburst model validated by the Joint Airport Weather Studies Project (Fujita, 1985; McCarthy et al., 1982), incorporating a primary and mirror vortex above the ground. Oseguera and Bowles (1988) developed the first three-dimensional downburst model, later refined by Vicroy (1991, 1992). This model, simpler yet comparable in effectiveness to Ivan's (1986) ring vortex model, was based on axisymmetric flow equations and empirical data from the TASS model (Proctor 1987a, b), and NIMROD Project (Fujita, 1978; Fujita, 1985). Holmes and Oliver (2000) revised the impinging jet model, simplifying the expression for radial mean wind velocity and integrating it with the downburst's translational speed. However, their model did not clearly distinguish between the low-level environmental flow in the ABL and the thunderstorm cell's motion. Abd-Elaal et al. (2013) used a parametric-CFD model coupled with an optimization algorithm to determine that downburst characteristics are significantly influenced by factors such as the touchdown location, time, and the downdraft's speed and direction. An essential aspect already highlighted with regard to the Holmes and Oliver model (2000), and then repeated in other subsequent papers (Chay at al. 2006, Abd-Elaal et al., 2013, and Hoa Le and Caracoglia (2017), is the lack of a clear distinction between the translational movement of the thunderstorm cell and the boundary layer wind in which the thunderstorm outflow is immersed at the ground. Hjelmfelt's (1988) study through radar measurements highlighted this problem's importance by examining two downbursts. The first case depicted a nearly stationary downburst in strong low-level environmental winds, while the second described a fast-moving downburst in a setting with little or no ABL flow. This lack of distinction in models hinders their ability to accurately describe such diverse real-world cases.

Based on these foundational insights provided by Hjelmfelt (1988), the authors of this paper introduced in 2020 a novel bi-dimensional analytical model to simulate the horizontal mean wind velocity at a specific height from a moving downburst (Xhelaj et al. 2020). This model conceptualizes the combined wind velocity at any given point during a downburst as the vector sum of three distinct components: the radial impinging jet velocity characteristic of a stationary

downburst, the translational velocity of the storm cell, and the ambient low-level ABL wind velocity, which encompasses the downburst winds near the surface. The model relies on 11 parameters, which are determined using a global metaheuristic optimization algorithm outlined in Xhelaj et al. (2022). This optimization process combines the analytical model with the Teaching Learning Based Optimization (TLBO) algorithm. TLBO operates with a population of solutions and employs iterative teaching and learning to find the best solution within the population (Rao et al., 2011). Due to the stochastic nature of TLBO, when integrated with the analytical model, the procedure can yield different optimal solutions each time it is executed. This variability arises from the initial random population of solutions and the intermediate transformations carried out by the algorithm to converge towards the best solution.

This study aims to examine the characteristics of the optimal solutions obtained through multiple runs of the optimization procedure, which integrates the Xhelaj et. al. (2020) model with the TLBO algorithm. It seeks to investigate the variability of the best solutions when applying the optimization algorithm to reconstruct the wind field during an intense downburst event. The main objective is to assess the extent to which the solutions differ from each other and from the solution with the lowest objective function value. Additionally, the study explores whether these alternative solutions can be considered physically valid, particularly when additional data describing the downburst event is incorporated.

The selected downburst event occurred in western Timis region of Romania on 25 June 2021 and was produced during the passage of an intense mesoscale convective system in the form of a bow echo over the town of Sânnicolau Mare. This event was recorded by a bi-axial anemometer and temperature sensor, both placed on a telecommunication tower 50 m above the ground level. The telecommunication tower lies approximately 1 km south of Sânnicolau Mare. The downburst that occurred in Sânnicolau Mare was of significant magnitude, resulting in extensive hail damage of the facades of numerous buildings within the city. In response to this event, a comprehensive damage survey was conducted through a collaborative partnership between University of Genoa (Italy) and the University of Bucharest (Romania). The survey (Calotescu et al., 2022 and Calotescu et al., 2024) pinpoints the GPS position of the buildings within the city that were predominantly impacted by the downburst. Moreover, a comprehensive map illustrating the hail damage of the building facades was generated. The map provides important information regarding the wind velocity experienced at urban scale, which has been used to validate the reconstruction/simulation of the downburst by the optimization procedure.

The analysis of the different optimal solutions (i.e., the data set) generated by the optimization algorithm was conducted through multivariate data analysis (MDA). This involved the joint application of cluster analysis and principal component analysis to effectively examine and interpret the dataset. Cluster analysis (CA) is a data mining technique that groups similar solutions together, aiming to identify patterns in the data. It is commonly used in fields like meteorology and climatology to identify clusters of weather phenomena or geographical regions with similar weather patterns (Burlando et al., 2008; Burlando et al., 2009). Principal component analysis (PCA) is a mathematical technique used to decrease the dimensionality of a dataset while minimizing the loss of information within the data. This analysis is commonly used in meteorology and climatology to decrease the number of variables required for representing weather pattern or climate trends and to identify regions with similar weather patterns (Amato et al., (2020); Jiang et al., (2020)). Principal component analysis is utilized in this context to enhance the interpretation of the different optimal solutions.

The present work is structured in 6 Sections. Following the introduction, Section 2 provides a description of the monitoring system that acquired the full-scale measurement employed in this research. Section 3 provides a brief meteorological description of the downburst event in Sânnicolau Mare (Romania). Section 4 describes the data set employed for performing cluster analysis and principal component analysis as well as the implementation of these

analyses. Section 5 presents an in-depth account of the main results derived from the CA and PCA. In conclusion, Section 6 offers a summary of the principal findings derived from this research.

2 Monitoring system and data acquisition

The complete set of measurements employed in this research were obtained through a monitoring system installed in Romania. Relevant information of this monitoring network can be accessed in the publications by Calotescu et al., (2021) and Calotescu and Repetto, (2022). The monitoring network received funding from the THUNDERR Project (Solari et al., 2020), which was conducted by the "Giovanni Solari – Wind Engineering and Structural Dynamics" Research (GS-Windyn) Group at the Department of Civil, Chemical, and Environmental Engineering (DICCA) of the University of Genoa. GS-Windyn, with a keen interest in monitoring poles and towers exposed to thunderstorm actions worldwide, secured funding for the acquisition of a full-scale structural monitoring network. This monitoring system was deployed on top of a 50 m lattice tower. The primary focus of this project revolves around three key objectives: first, the detection of thunderstorms; second, the analysis of wind parameters associated with these phenomena; and third, the experimental assessment of the structural response of telecommunication lattice towers to the forces generated by both synoptic and thunderstorm winds. The monitoring tower, named TM_424, is property of the SC TELEKOM ROMANIA SRL and is located in the western part of Romania, Timis county, at approximately 1 km south of Sânnicolau Mare (Figure 1). The site is an open field, the terrain is flat with low grass vegetation.

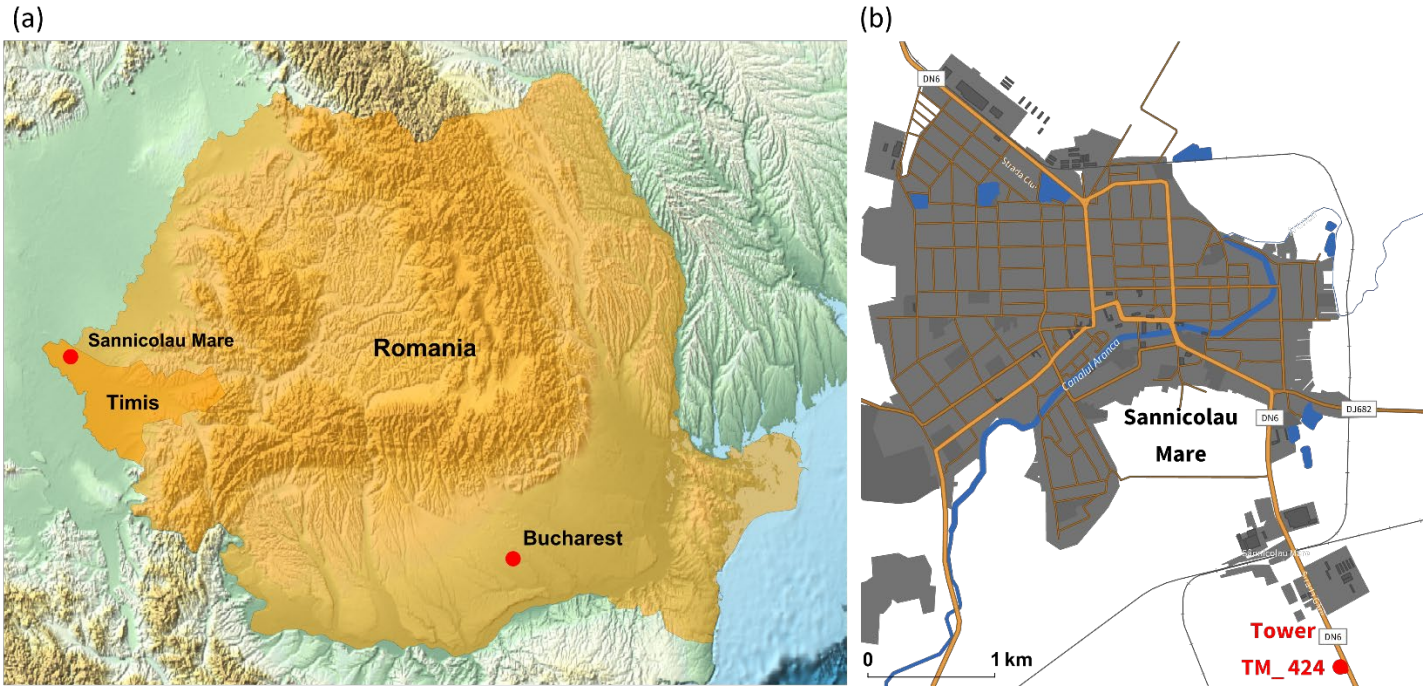
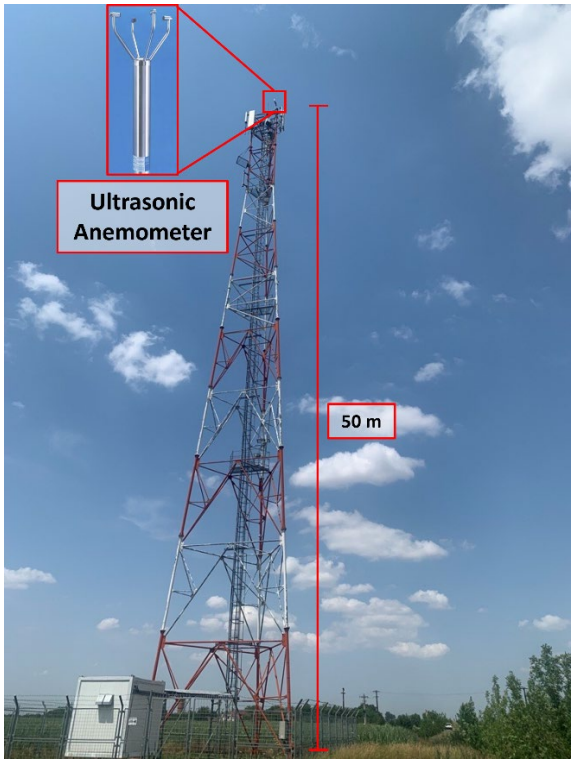


Figure 1. (a) Location of the telecommunication tower TM_424, situated 1 km south of Sânnicolau Mare in Timis County, Romania. (b) Expanded view of the Sânnicolau Mare town with the telecommunication tower TM_424 represented by the red dot on the map. Maps generated using Mathematica (Wolfram Research, Inc., Version 13.3, 2023, https://www.wolfram.com/mathematica).



132

Figure 2. TM_424 Telecommunication tower and sensors position at the top of the tower. On the horizon, approximately 1 km from the tower lies the small city of Sânnicolau Mare. Image courtesy of Google Street View, 2022 (https://www.google.com/maps).

133 Figure 2 shows the dimension of the tower. Among the various networks for the monitoring systems, the tower is
134 equipped with a GILL WindObserver 70 ultrasonic anemometer at the top (Figure 2). The anemometer has a data
135 acquisition rate of 4 Hz, can measure the wind speed up to 70 m/s. In addition to the anemometer sensor, the tower is
136 equipped with a temperature sensor installed near the location of the anemometer. The working temperature range for this
137 sensor is between -55 and 70 °C.

138

139 **3 The Sânnicolau Mare (Romania) downburst event of 25 June 2021**

140 In this section, a brief overview of the meteorological aspects pertaining to the downburst event in Sânnicolau Mare on
141 25 June 2021 is provided. In the late afternoon of 25 June 2021, a severe downburst event affected the extreme western
142 region of Romania. The downburst event took place in the Timis county (Figure 1a) between 18:00 and 19:00 UTC and
143 struck the little town of Sânnicolau Mare (Figure 1b). At 17:30 UTC, a strong mesoscale convective system moving
144 toward the east was approaching the town of Sânnicolau Mare. Figure 3a, acquired from Eumetsat, captures an image of

a deep convective cell at 18:30 UTC. This weather phenomenon exhibits cloud tops ascending over 12 km above mean sea level, signifying the mature stage of the convection cycle. This mature storm cell was observed to have directly impacted the town under study. Figure 3b presents composite radar reflectivity data, indicating that this meteorological phenomenon can be classified as a mesoscale convective system known as bow echo. Radar reflectivity values at or above 60 dBZ, as seen in this event, are typically indicative of severe weather conditions. Such conditions are often associated with the production of hailstones, with an average diameter of approximately 2.5 cm.

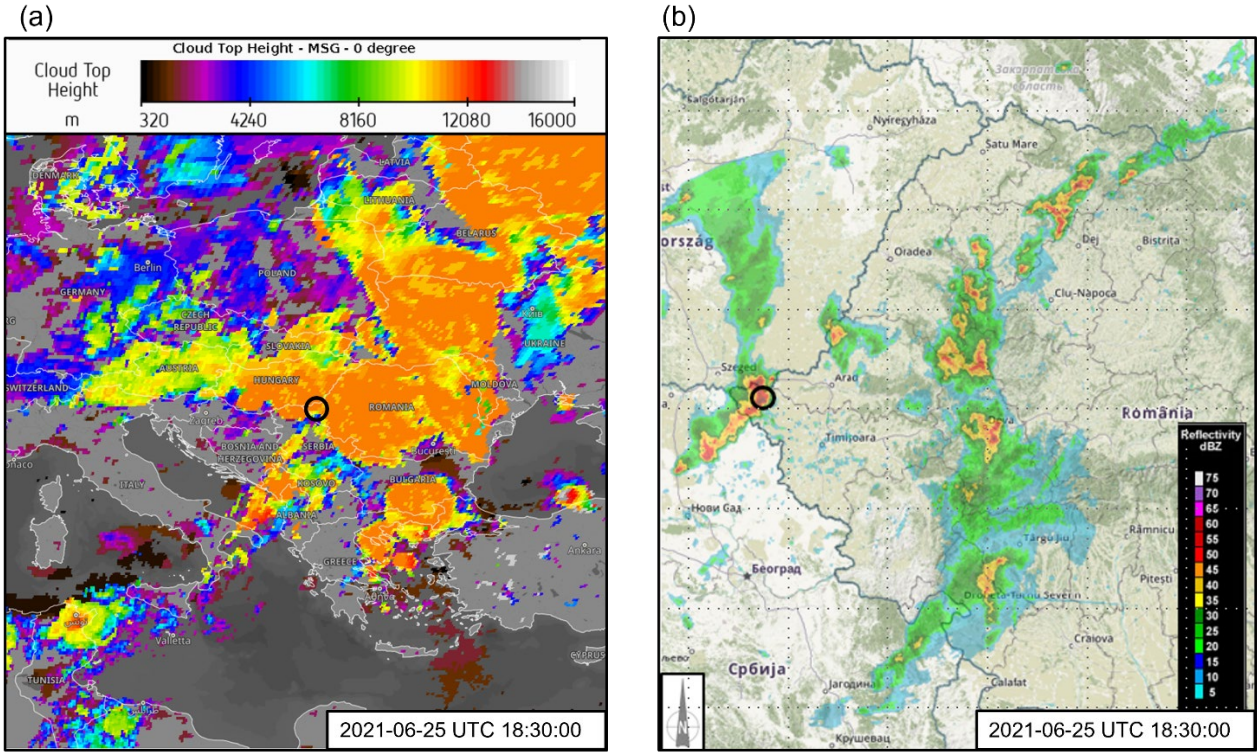
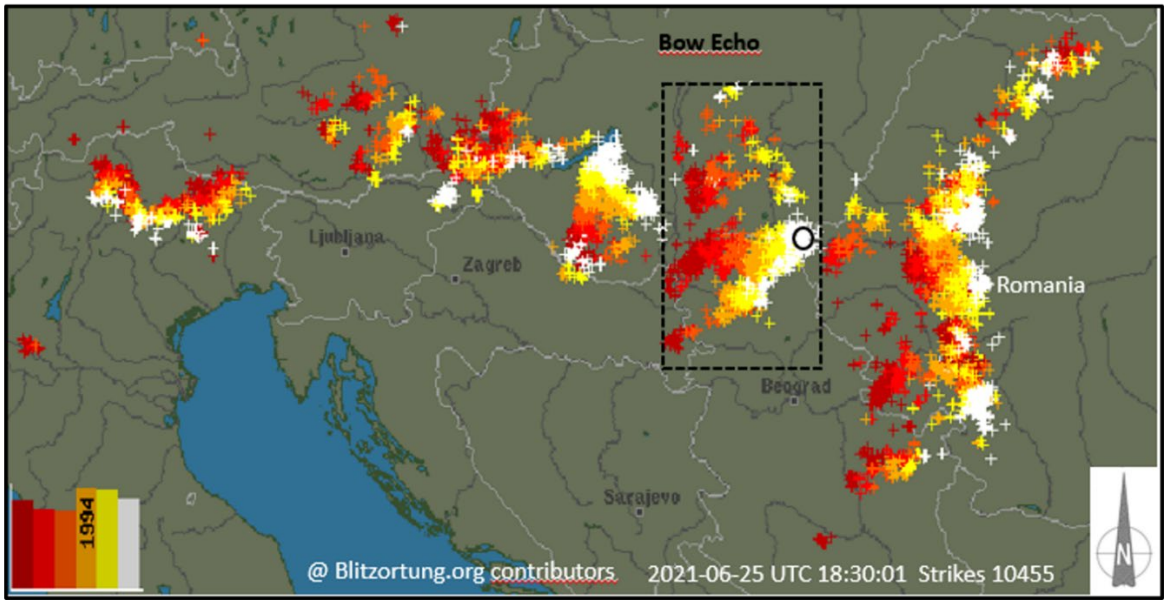


Figure 3. (a) Distribution of cloud top heights derived from Meteosat Second Generation (MSG) valid for 25 June 2021 at 18:30 UTC. Data and map obtained from ©EUMETSAT 2022 (https://view.eumetsat.int). (b) Composite radar reflectivity (dBZ) for June 25, 2021, at 18:30 UTC. The geographical location of Sânnicolau Mare and the apex of the bow echo are indicated by the black circle. Data and map obtained by ©2018 Administratia Nationala de Meteorologie (https://www.meteoromania.ro).

The existence of a robust convective motion, indicative of the typical kinematic structure of a bow echo, is distinctly portrayed through the distribution of intensive lightning activity, as displayed in Figure 4a. As the figure illustrates, an approximate total of 10455 lightning strikes were recorded by the Blitzortung.org network across Eastern Europe between 16:30 to 18:30 UTC. A significant concentration of these strikes correlates with the bow echo structure near the western Timis County in Romania. The color gradient in Figure 4a, ranging from red, orange, yellow and white serves as a temporal marker, with white color indicating the most recent strikes and with red color denoting older ones. This color coding effectively illustrates the temporal and spatial evolution of the lightning activity during the severe weather event, providing insight into the progression of the storm system. Bow echoes are a prevalent form of severe convective organization. These mesoscale convective systems can generate straight-lines surface winds that lead to extensive damage associated with downbursts. On occasion, they may also give rise to tornadoes. Interestingly, the observed bow echo seems to display a stratiform parallel structure, a rarer variety of squall lines (Parker and Johnson, 2004; Markowski and Richardson, 2010).

(a)



(b)

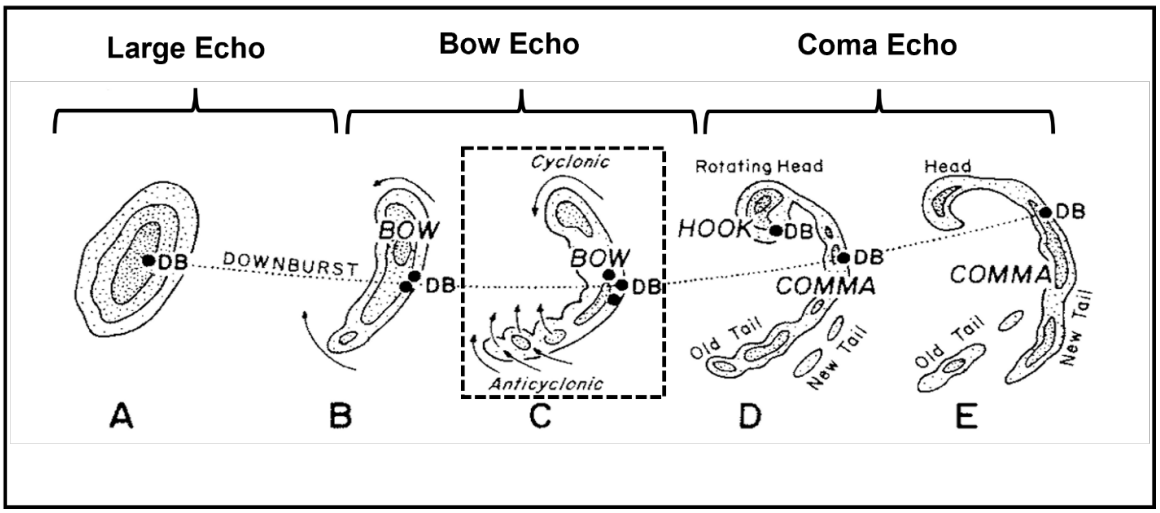


Figure 4. (a) Lightning strikes recorded between 16:30 to 18:30 UTC on June 25, 2021, sourced from the Blitzortung.org network archive for lightning and thunderstorms (www.blitzortung.org). The black circle marks the geographic location of Sânnicolau Mare, situated near the apex of the observed bow echo. (b) Typical radar echo morphology commonly observed in bow echoes, characterized by the generation of strong downbursts at the bow apex, denoted as DB. Adapted from Fujita (1978).

Figure 4b illustrates the characteristic kinematic structure of a bow echo as outlined by Fujita (1978). Typically, the system originates as a singular, prominent convective cell, either isolated or embedded within a broader squall line system (Phase A). As the surface winds strengthen, the parent cell undergoes transformation, evolving into a line segment of cells with a bow-shaped configuration (Phase B). During the maximum intensity, the bow's center might develop a spearhead echo (Phase C), characterized by the occurrence of the most severe downburst winds at the apex of the spearhead. During the decay phase, the wind system frequently evolves into a comma-shaped echo (Phase E) (Weisman, 2001). The comparisons between Figures 3b, 4a, and 4b elucidate that the bow echo positioned above Sânnicolau Mare at 18:30 UTC is in its most intense stage (Phase C), as evidenced by the formation of the characteristic spearhead echo shape. The intense downburst event generated at the apex of the bow echo was recorded by the anemometer and temperature sensor

situated 50 meters above the ground on the TM_424 tower. The time histories of the moving average wind speed and direction (averaged over 30 seconds) (Solari et al., 2015; Burlando et al., 2017) for the recorded one-hour duration of the downburst event are given in Figure 5a and Figure 5b, respectively. At approximately 18:30 UTC the anemometer recorded an instantaneous maximum velocity (sampled at 4 Hz) of $\hat{V} = 40.8$ m/s while the maximum moving average wind velocity was $V_{\max}$, = 35.8 m/s. This notable high velocity clearly evidences of the occurrence of an intense downburst. The time interval spanning from 18:20 to 18:45 UTC represents the primary indicator of the downburst's occurrence in the proximity of the telecommunication tower. This period is characterized by a sudden surge in wind speed, commonly referred intensification stage followed by a subsequent decrease in velocity after 18:30 UTC. Throughout the initial phase of intensification, the wind direction exhibited a clockwise rotation, ranging from 235° and extending to approximately 360°. Additionally, Figure 5a also includes 1-hour time series of the recorded temperature data. The temperature sensor is positioned at the same location of the anemometer. Before the passage of the downburst, the environmental temperature was on average 27 °C, while at approximately 18:20 UTC the temperature dropped very sharply reaching the minimum value of 14.5 °C at approximately 18:30 UTC. After the sharp drop the temperature started to rise and eventually returned to its pre-storm level (not shown).

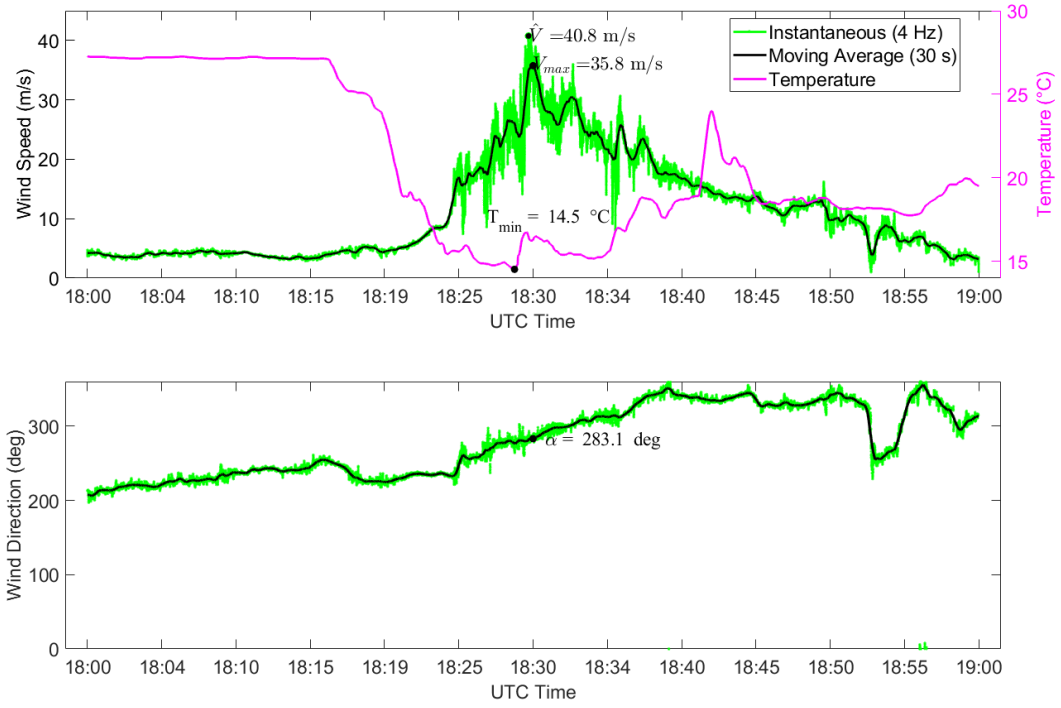


Figure 5. Telecommunication tower monitoring network measurements from 18:00 to 19:00 UTC on June 25, 2021: (a) Time history of the instantaneous wind speed (green), moving average mean wind speed (black) and temperature record (magenta); (b) Instantaneous (green) and moving average mean wind direction (black).

The downburst in Sânnicolau Mare was also marked by a substantial hail occurrence. The interaction between the high-velocity winds and hail, potentially influencing the trajectory and impact of the hailstones, contributed to extensive damage, especially to the facades of numerous buildings. To comprehensively assess this damage, a collaborative survey

was conducted by the University of Genoa (Italy) and the University of Bucharest (Romania) (Calotescu et al., 2022; Calotescu et al., 2024). The survey identified the affected buildings and produced a comprehensive map illustrating the hail damage. Figure 6 shows a schematic representation of the distribution of hail damage per area (600 x 600 m^2) and the position of the buildings that suffers hail damage in the town of Sânnicolau Mare. Correlating specific damages like hail impacts with near-surface wind velocities involves inherent uncertainties, which are extensively explored in the study by Calotescu et. al., 2024.

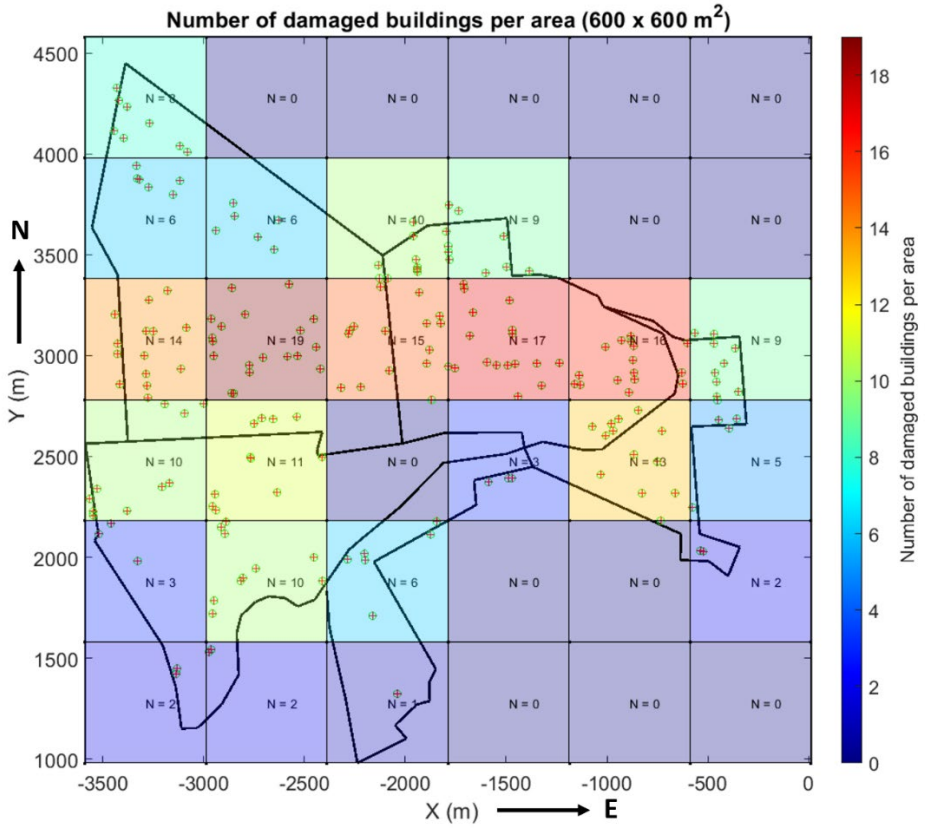


Figure 6. Spatial distribution of damaged buildings and locations of hail-damaged structures within 600 x 600 m^2 area in the town of Sânnicolau Mare during the downburst event on June 25, 2021. The city boundaries of Sânnicolau Mare are delimited by the black line.

4 Downburst reconstruction

This section focuses on the modeling, optimization, and reconstruction of the Sânnicolau Mare downburst event. Section 4.1 delves into the modeling and optimization approach used for downburst reconstruction. Section 4.2 introduces metaheuristic optimization and its application in the reconstruction of the specific downburst event under study. Finally, Section 4.3 outlines the multivariate data analysis employed to examine the solutions generated by the optimization algorithm.

4.1 Modeling and optimization approach for downburst reconstruction

In this study, the authors utilize the computational model developed in a previous work by Xhelaj et al. (2020) for the reconstruction and simulation of the Sânnicolau Mare downburst event discussed in Section 3. The Xhelaj et al. (2020) model can simulate the spatiotemporal evolution of the bi-dimensional moving average (30 second window) wind speed

and direction experienced during a typical downburst event at a specified height z above ground level (AGL). In general, the wind system simulated by the analytical model represents the outflow structure of a translating downburst, typically occurring in diverse meteorological conditions such as single cell thunderstorms, multicell thunderstorms, squall lines and bow echoes. For the specific case of the Sânnicolau Mare downburst, the analytical model operates under the hypothesis that the downburst occurs near the tip of the bow echo during its mature stage (Phase C, Figure 4b), in line with the studies of Fujita (1978) and Weisman (2001). It is worth noting that the model does not encompass the broader, complex mesoscale circulations, commonly associated with high winds in bow echoes. This represents a focused approach, considering the downburst evolution within a specific context, rather than attempting to model the entire spectrum of atmospheric phenomena related to bow echoes.

The analytical model comprises 11 variables that describe the kinematic structure of the downburst wind. Table 1 presents a short description of the 11 variables upon which the model relies. As a result, the model allows for the reconstruction of the time-evolving moving average wind speed and direction generated by the simulated downburst at every point within the simulation domain. The model simulates the downburst wind velocity field by combining three components, the stationary radial velocity from a jet impacting a flat surface, the downdraft's translation velocity (i.e., storm motion) and the low level ABL wind velocity. The virtual anemometer, situated at the center of the simulation domain, measures the simulated wind speed and direction generated by the model. By employing anemometric wind speed and direction data collected during the Sânnicolau Mare downburst event, an optimization procedure can be formulated to minimize the relative error (objective function F), which quantifies the discrepancy between the observed time series of the moving average wind speed and direction and the corresponding simulations generated by the model. Since the Sânnicolau Mare downburst event was recorded by an anemometer positioned at a height of 50 meters AGL, the analytical model will reconstruct the wind speed and direction at the corresponding height.

Table 1. Variables of the Xhelaj et, al. (2020) analytical model.

1	X-component touchdown location (at $t = 0$) (m)	X_{C0}
2	Y-component touchdown location (at $t = 0$) (m)	Y_{C0}
3	Downdraft radius (m)	R
4	Normalized radial distance from the center of the downburst where $V_{r,max}$ occurs (-)	$\rho = \frac{R_{max}}{R}$
5	Maximum radial velocity (m/s)	$V_{r,max}$
6	Duration of the intensification period (min)	T_{max}
7	Total duration of the downburst event (min)	T_{end}
8	Storm translational velocity (m/s)	V_t
9	Storm translational direction (deg)	α_t
10	ABL wind speed below the cloud base (m/s)	V_b
11	ABL wind direction below the cloud base (deg)	α_b

The reconstruction procedure gives rise to a mathematical optimization problem characterized by being single-objective, nonlinear, and bound constrained, as discussed in Xhelaj et al. (2022). To tackle this optimization problem, the analytical model is integrated with a global metaheuristic optimization algorithm. Specifically, the Teaching Learning Optimization Algorithm (TLBO) proposed by Rao et al. (2011) is employed. The details pertaining to the integration of the analytical

model with the optimization algorithm, as well as the estimation of the kinematic and geometric variables associated with the downburst event, are explained in detail in Xhelaj et al. (2022). The TLBO algorithm is an iterative, stochastic, and population-based algorithm comprising two distinct phases: the Teacher Phase and the Learner Phase. In the Teacher Phase, the best solution in the population (the teacher) shares its knowledge (objective function) with the other solutions (the students) to enhance their performance. In the Learner Phase, the students interact with each other to further improve their performance. TLBO requires only two user-specified parameters: the maximum number of iterations T and the population size N_p. When incorporating the objective function into a stochastic metaheuristic optimization algorithm, running the algorithm independently multiple times is crucial to reach the optimal solution. This iterative approach allows for a deeper exploration of the variable space, reducing the risk of getting trapped in local optima. However, it is important to note that in the context of metaheuristic optimization, there is no guarantee of attaining a globally optimal solution. As a result, the procedure can yield a range of solutions ordered based on the values assumed by the objective function, with some being better than others. In this study, the TLBO algorithm is executed 1024 times independently, with each run producing an optimal solution. Consequently, 1024 solutions are obtained. The reconstruction of the downburst event can be accomplished by selecting the solution with the lowest objective function value, as it is considered the best representation of the event based on the optimization process. This study aims to analyze and clarify the nature of all the solutions generated by means of the TLBO algorithm for the downburst outflow reconstruction. This choice was made for a twofold reason.

- The first reason is to determine the best possible solution among the 1024 totals, where best solution is the one that minimizes the objective function F, and allows to reconstruct the Sânnicolau Mare downburst event.
- The second reason, which is the primary objective of this study, is to analyze these 1024 solutions using multivariate data analysis (MDA). The method used in MDA are the Agglomerative hierarchical clustering (AHC) coupled with the K-Means algorithm and principal component analysis (PCA).

The objective is to investigate the distinct characteristics of the different solutions provided by the TLBO algorithm, enabling an understanding of their divergence from the optimal solution. If alternative solutions do exist, it signifies that the algorithm's solution is not unique. This highlights the challenge in accurately reconstructing downburst wind field form just one anemometric time series, underlining the problem's inherent complexity and underdetermined nature.As such, a more comprehensive definition of the objective function is necessary to accurately discern between the optimal solution and its alternatives.

4.2 Metaheuristic optimization and reconstruction of the Sânnicolau Mare downburst

In metaheuristic optimization, a commonly used guideline suggests setting the population size N_p as ten times the number of variables to estimate D (Storn, 1996). In this study, where D corresponds to 11 variables, a population size of $N_p = 110$ has been chosen. Additionally, considering the reported fast convergence rate of the TLBO algorithm (as mentioned in Xhelaj et al., 2022), the maximum number of iterations T for this study has been set to $T = 100$. Table 2 displays the lower and upper bounds of the optimization problem pertaining to the reconstruction of the Sânnicolau Mare downburst. These parameter values have been determined based on a comprehensive literature review, available in Xhelaj et al. (2022).

Table 2. Lower and upper bound of the decision variable parameters for the reconstruction of the Sânnicolau Mare downburst. Table form Xhelaj et al. (2022).

	Parameters/Variables	Lower Bound	Upper Bound
1	X_{C0} (m)	-10000	-10000
2	Y_{C0} (m)	-10000	-10000
3	R (m)	200	2000
4	$\rho = \frac{R_{max}}{R}$ (-)	1.6	2.6
5	$V_{r,max}$ (m/s)	0	40
6	T_{max} (min)	2	15
7	T_{end} (min)	15	60
8	V_t (m/s)	0	40
9	α_t (deg)	0	359.9
10	V_b (m/s)	0	40
11	α_b (deg)	0	359.9

The spatial domain of the downburst simulation covers an area of 20 x 20 km² while the grid resolution in both the X and Y directions is set at 50 m. This approach employs a comprehensive simulation approach, primarily using anemometric data, due to its common availability. The methodology entails numerous simulations to extract downburst's kinematic and geometric parameters. However, when additional data like Radar or Lidar is available, this information can be used to bound some variables and restrict the model variables domain (Table 2) to enhance model accuracy. Figure 7 illustrates the "performance chart" depicting the convergence pattern of the objective functions during the reconstruction of the Sânnicolau Mare downburst using the TLBO algorithm. The performance chart in Figure 7 illustrates the convergence pattern of the objective functions as iterations progress. It shows the upper and lower envelopes that encapsulate all 1024 independent runs. The region within the envelopes represents the objective function values' trend for all runs. At the end of the 100 iterations, the lower envelope represents to the best objective function value obtained, while the upper envelope corresponds to the worst objective function value obtained by the TLBO algorithm. The performance chart in Figure 7 includes additional visual representations: a dashed line representing the mean convergence curve, and dotted lines representing the mean plus/minus one standard deviation curves. These curves provide insights into the average behavior and deviation of the objective function values across the 1024 runs. The performance chart demonstrates that after approximately 70 iterations, the TLBO algorithm ceases to find significantly better or worse solutions. This is evidenced by the convergence of both the upper and lower envelope curves. Concurrently, the mean curve appears to plateau, although it exhibits a slight yet continuous improvement beyond the 70th iteration. This suggests that the algorithm is still optimizing, albeit at a reduced rate. The increasing spread between the mean and the plus/minus one standard deviation curves as iterations progress indicates a complex solution landscape. This complexity is manifested in the algorithm's convergence to various local minima, maintaining steady average performance while increasing the variability of solutions. In this study's context, such expanding spread represents a deeper and more intricate exploration of the solution space, a desirable characteristic to ensure a comprehensive search across the objective function domain. . At the conclusion of 100 iterations, the best and worst objective function values correspond to F_{min} = 0.730 and F_{max} =

1.062, respectively. The mean and standard deviation of the objective function values are determined as m_F = 0.893 and s_F = 0.080, respectively.

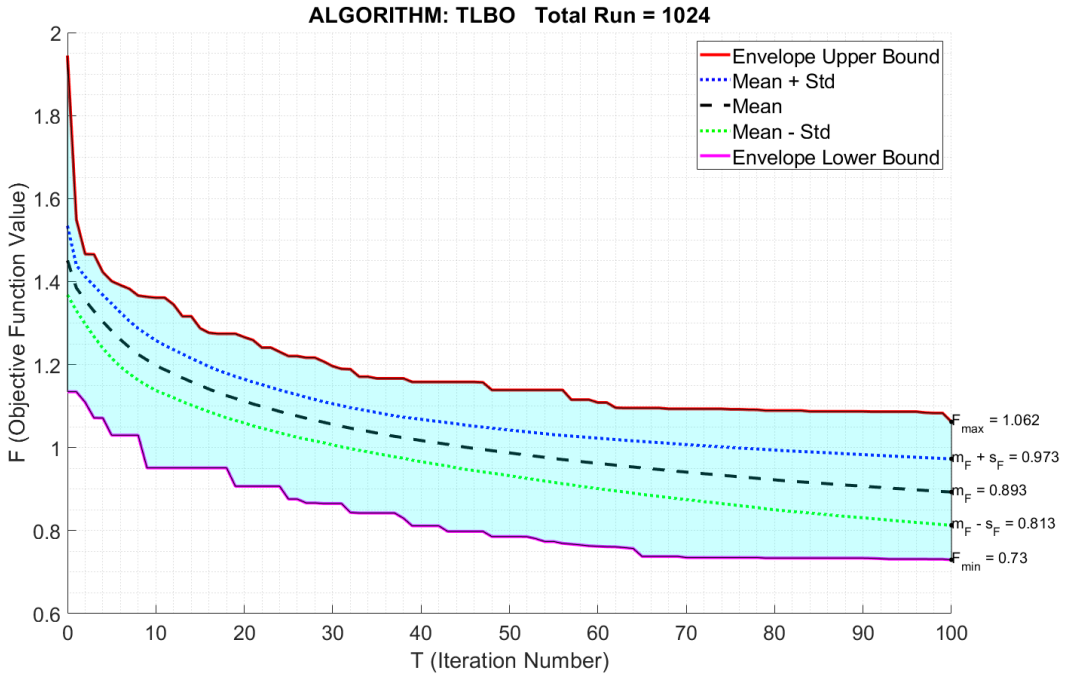


Figure 7. Performance chart for the reconstruction/simulation of the Sânnicolau Mare downburst using the TLBO algorithm.

4.3 Multivariate data analysis of solutions for the Sânnicolau Mare downburst reconstruction

The optimization algorithm provides in output a data table, where each row of the table is a solution of the optimization problem. Therefore, the data table is composed of 1024 rows (solutions). The table has 12 columns, where 11 columns represent the 11 variables/parameters of the analytical model, while the last column contains the values assumed by the objective function F of each solution (i.e., each row). Although the objective function F, is not a variable of the analytical model, it is treated in Section 5 as a variable from the point of view of the multivariate data analysis. The solutions are sorted in descending order based on their objective function value F. This means that the best overall solution among the 1024 lies in the last row of the data table. The analysis of the data table indicates that most variables exhibit multimodal histograms, with two or more peaks. However, only the variables V_b and α_b are characterized by a unimodal histogram. Since the aim of this document is to conduct a multivariate data analysis (MDA), the variables of the data table are split into primary and secondary variables. Primary variables participate in the analysis of multivariate data (i.e., AHC + K-Means and PCA), as opposed to secondary variables, which have no role in the calculation. However, secondary variables can indeed assist in the interpretation of the data table. In the present study, V_b, α_b and α_t are considered as secondary variables. This choice is primarily driven by the observation that V_b, and α_b exhibit unimodal histograms, suggesting that they may not significantly contribute to distinguishing different cluster solutions. However, the choice of α_t as a secondary variable is purely practical, since it makes it possible to carry out a multivariate statistical analysis, avoiding the problem of circular statistics and, hence, simplifying the calculation and the interpretations of the results.

Let's define the data table that contains only primary variables by a matrix $\mathbb{X}$. Each row i of the matrix represents a solution vector $\boldsymbol{X}_i$, encompassing the values associated with the nine primary variables. Therefore the solution vector can be expressed as $\boldsymbol{X}_i = \left(X_{C0\,i}, Y_{C0\,i}, R_i,\ \rho_i, V_{r,max_i}, T_{max\,i}, T_{f_i}, V_{t_i}, F_i\right)^T$ with i ranging from 1 to I, where I represents the total number of solutions, in this case I = 1024. Since the solution vector $\boldsymbol{X}_i$ contains $K = 9$ primary variables, the resulting data matrix $\mathbb{X}$ is an I-by-K matrix with 1024 rows and 9 columns. For the sake of simplicity, in order to shorten the notation, let X_{ik} be the value of the k-th primary variable in the i-th solution. Henceforth, the term "variable" will refer to primary variables, unless explicitly specified. Consequently, the dataset within the matrix $\mathbb{X}$ can be regarded either as a collection of rows representing solutions to the optimization problem or as a collection of columns representing variables of the analytical model. The focus of MDA is to apply statistical clustering to identify similar analytical solutions. Since a generic solution $\boldsymbol{X}$, is a set of $K = 9$ numerical values, $\boldsymbol{X}$ evolves within a space $\mathbb{R}^K$(a space with 9 dimensions), called "the solution's space". Defining in the solution's space the usual Euclidean metric (i.e., the l_2 norm $\|\cdot\|_2$), then, the squared distance between two solutions $\boldsymbol{X}_i$ and $\boldsymbol{X}_l$ can be expressed by the Euclidean distance d_{il}:

$$d_{il}{}^2 = d^2(\boldsymbol{X}_i, \boldsymbol{X}_l) = \|\boldsymbol{X}_i - \boldsymbol{X}_l\|_2^2 = \sum_{k=1}^{K} (X_{ik} - X_{lk})^2 \qquad (1)$$

The distance d possesses the following metric properties:

$$\begin{cases} d(\widehat{\boldsymbol{X}}_i, \widehat{\boldsymbol{X}}_l) = 0 \quad \Leftrightarrow i = l \\ d(\widehat{\boldsymbol{X}}_i, \widehat{\boldsymbol{X}}_l) = d(\widehat{\boldsymbol{X}}_l, \widehat{\boldsymbol{X}}_i) \quad \text{(symmetry)} \\ d(\widehat{\boldsymbol{X}}_i, \widehat{\boldsymbol{X}}_l) \leq d(\widehat{\boldsymbol{X}}_i, \widehat{\boldsymbol{X}}_j) + d(\widehat{\boldsymbol{X}}_j, \widehat{\boldsymbol{X}}_l) \quad \text{(tirangle inequlity)} \end{cases}$$

Variables in the data matrix $\mathbb{X}$ are standardized to account for different units and scales. This common practice in statistical modelling neutralizes scale effects allowing for meaningful comparisons across variables. Therefore, the variables are standardized according to the following equation:

$$\hat{X}_{ik} = \frac{X_{ik} - \bar{X}_k}{S_k},\ \forall\, i = 1, \dots, I = 1024 \ \text{ and } \ \forall\, k = 1, \dots K = 9 \qquad (2)$$

where $\bar{X}_k$ denotes the sample mean of the k-th variable calculated over all I solutions: $\bar{X}_k = \frac{1}{I}\sum_{I=1}^{I} X_{ik}$ and S_k is the sample standard deviation of k-th variable: $S_k = \sqrt{\frac{1}{(I-1)}\sum_{i=1}^{I}(X_{ik} - \bar{X}_k)^2}$.

Finally, the normalized data matrix $\widehat{\mathbb{X}}$ containing the set of vectors $\widehat{\boldsymbol{X}}_i$, $i = 1, \dots, I$, has been used in the MDA for the identification of different typology of solutions provided by the TLBO algorithm for the simulation/reconstruction of the Sânnicolau Mare downburst. Figure 8 showcase a summary statistic in the form of a box plot, illustrating the distribution of the standardized variables. Variables such $\widehat{\boldsymbol{R}}_{max}$ and $\widehat{\boldsymbol{T}}_{max}$ have a large number of outliers which indicates extreme values within the dataset. Therefore, even in the context of standardized data, outliers can still be informative and may hold important information for distinguishing distinct solution clusters.

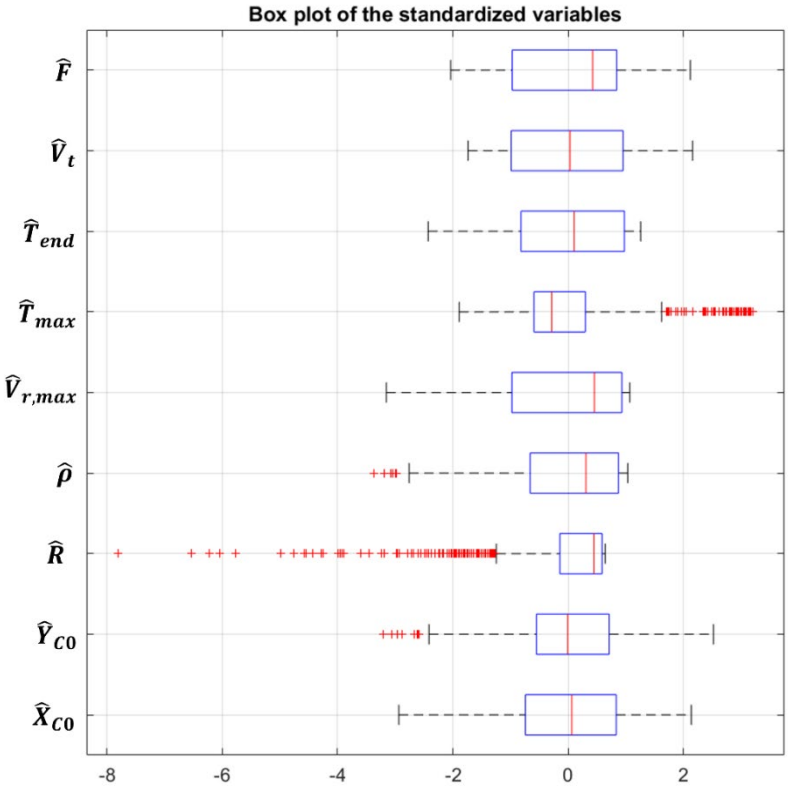


Figure 8. Box plot of the distributions of the standardized variables. Outliers in the data are plotted individually using the red marker symbol + .

5 Results

In the following section the results of multivariate data analysis (MDA) including cluster analysis and principal component analysis applied to the data matrix $\widehat{\mathbb{X}}$ is presented. After the clusters have been established a comprehensive description of each of them is provided. This involves examining the variables that contribute to each cluster's composition as well as identifying specific representative solutions within each cluster. Such an analysis allows for a deeper understanding of the cluster characteristics and facilitates the interpretation of meaningful patterns and insights within the data. Sections 5.1 to 5.3 provide an in-depth analysis of data matrix $\widehat{\mathbb{X}}$ from the variable's perspective, employing agglomerative hierarchical K – Means clustering and principal component analysis. In Section 5.4 the clusters are analyzed from the point of view of the specific solutions which are the most representative of the clusters. Finally, these representative solutions are compared with the best overall solution founded from the TLBO algorithm. The comparisons of the representative solution for each cluster and the best overall solution with the full-scale data is therefore enriched considering the data from the damage survey that was carried out after the Sânnicolau Mare downburst event.

In order to identify the appropriate number of clusters for grouping the solutions, the agglomerative hierarchical clustering (AHC) is firstly employed (Hartigan, (1975), Kaufman and Rousseuw (1990)). In AHC, each individual solution is initially treated as an independent cluster (leaf). Through a series of iterative steps, the most similar clusters are progressively merged, forming a hierarchical tree structure known as a dendrogram. This merging process continues until all the individual clusters are combined into a single cluster (root).

Subsequently, the hierarchical tree is analysed, and a suitable level is chosen to cut the tree, leading to distinct and meaningful clusters. The number of clusters obtained from the AHC forms a partition of the data set. To refine and optimize this partition, a partitioning clustering algorithm called K-means (MacQueen, 1967, Hartigan and Wong, 1979) is subsequently applied. Partitioning algorithms, like K-Means, subdivides the data sets into distinct clusters, ensuring that solutions within each cluster are similar to one another while exhibiting noticeable differences between clusters. Hence the two steps clustering procedure is called agglomerative hierarchical K – means clustering (AHK-MC) and is employed to analyse the standardized data matrix $\widehat{\mathbb{X}}$. By combining the strengths of both algorithms AHC and K-means, AHK-MC aims to provide a comprehensive and improved clustering algorithm of the data, enabling a more accurate identification of distinct solution groups.

The hierarchical tree in Figure 9 (i.e., dendrogram) is constructed following the Wards' method (Ward, 1963). Since the total solutions of the optimization problem are $I = 1024$, the dendrogram is very dense at the bottom level (i.e., at the leaf level, where each solution is considered as a cluster by itself). The hierarchical tree is composed therefore by $I - 1 = 1023$ nodes, the points where two clusters (solutions or set of solutions) are merged. The level (height) of each node in the tree is described by the within-cluster variance. The level of a node in the agglomeration process, when examined from top to bottom, indicates the reduction in within-cluster variance achieved by merging two connected clusters. This reduction in variance can be visualized using a bar graph, as depicted in Figure 10.

From Figure 10 it is possible to establish the level where to cut the dendrogram and consequently to establish the number of clusters for partitioning the data set. The choice of the number of clusters is important because partitioning with too few clusters risk leaving groups which are not at all homogeneous. On the other hand, partitioning with too many clusters' risks creating classes that are not very different from each other. Being $\sum_{s=1}^{I-1} \Delta_s = K = 9$ (the total variance contained in the standardized data), the separation into two groups is able to describe $\Delta(1,2)/K = 4.314/9 = 0.4793$ (47.93 %) of the total variance. Considering the partitioning into three groups, the explained variance by the three clusters is equal to $[\Delta(1,2) + \Delta(2,3)]/K = [4.314 + 1.044]/9 = 0.5954$ (59.54 %) of the total variance, while for four clusters the "explained variance" is equal to $[\Delta(1,2) + \Delta(2,3) + \Delta(3,4)]/K = [4.314 + 1.044 + 0.406]/9 = 0.6404$ (64.05 %) of the total variance.

Therefore, considering more than three clusters (refer to Figure 10) is going to have a very little impact on the explained variance since very little information is gained and is no longer useful to group together any more classes. For this reason, the dendrogram in this work is partitioned in 3 clusters (refer to Figure 9) and therefore they can explain approximately 60% of the total variance present in the data.

The three-cluster solution's ability to explain about 60% of the total variance is significant, especially considering the single-point (anemometric) measurement nature of the downburst data. This inherent limitation often leads to high variability, making the extraction of consistent patterns challenging. As noted in related studies, such as those by Bogensperger and Fabel (2021), benchmarks for acceptable levels of explained variance in clustering are not universally applicable but rather depend on the specific context and data characteristics. The present study's level of variance explanation, given the complexity and variability of the downburst captured from one location, is therefore robust. This is further supported by the observation in Figure 10 that additional clusters contribute minimally to the total variance explained, suggesting that the primary structural patterns in the data are adequately captured with three clusters.

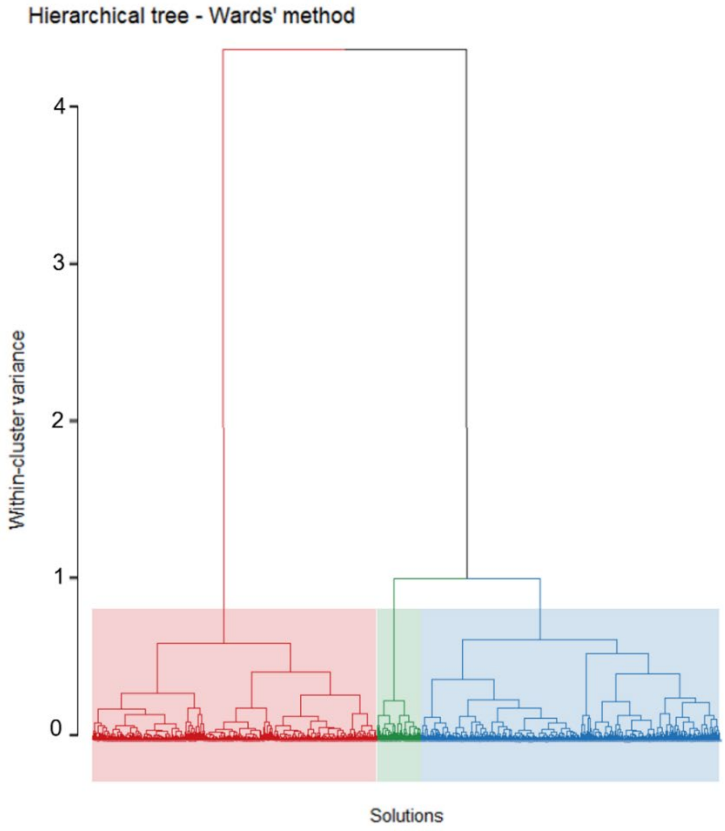
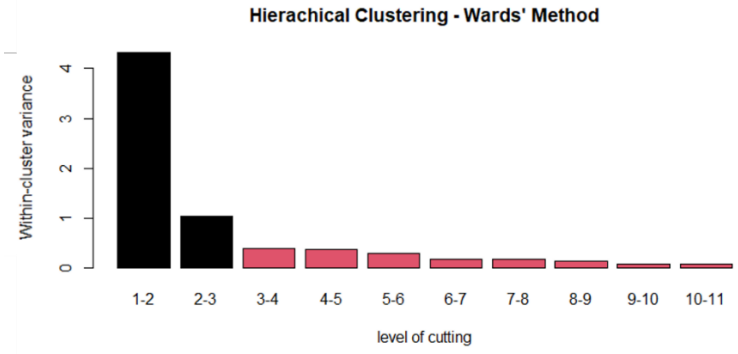


Figure 9. Hierarchical tree (dendrogram) created with Ward's Method, categorizing the optimization problem solutions for the Sânnicolau Mare downburst into three clusters, each marked by a distinct color: red for cluster 1, green for cluster 2, and blue for cluster 3.

408



409 **Figure 10. Bar graph of the relation between the number of merged clusters and the within-cluster variance.**

410 **5.2 Clusters interpretation via PCA and Optimization with K-Means**

411 The three clusters of solutions are analyzed using principal component analysis (PCA) to identify the key variables that
412 drive the system's behavior. By extracting the principal components, which captures the most significant variation in the
413 data, the complexity of the system can be reduced. In particular, the eigenvalues of the correlation matrix $\mathbb{S} = \frac{1}{(I-1)} \widehat{\mathbb{X}}^T \widehat{\mathbb{X}}$
414 quantify the amount of variance accounted by each principal component (Kassambara, 2017). The eigenvalues shows that
415 the first components have larger values, indicating that they capture the most significant variation in the data set. In
416 contrast, the subsequent components have lower eigenvalues, representing a diminishing level of variation. Table 3

presents displays the eigenvalues, the percentage of variance explained by each component, and the cumulative percentage of variance.

Table 3. PCA results in term of the eigenvalues, percentage of variance and cumulative percentage of variance.

	Dim-1	Dim-2	Dim-3	Dim-4	Dim-5	Dim-6	Dim-7	Dim-8	Dim-9
Eigenvalues (λ_s) or variance	5.278	1.458	0.884	0.499	0.378	0.195	0.167	0.093	0.048
Percentage of variance	58.645	16.204	9.825	5.542	4.204	2.170	1.852	1.028	0.530
Cumulative perc. of variance	58.645	74.849	84.674	90.216	94.420	96.589	98.441	99.470	100.000

The first two principal components capture 74.85% of the total variance in the dataset. These components define a plane that provides significant insights into the underlying patterns and structure of the data. Eigenvalues greater than 1 (Table 3) signify that the respective principal components explain more variance in the data compared to any single standardized variable. In contrast, eigenvalues less than 1, starting from the third principal component (Table 3) indicate that the associated principal components explain less variance than individual standardized variables, suggesting they have relatively less influence on the overall variability in the data. Therefore, it is probably not useful to interpret the next dimensions and better focusing on the first two principal dimensions for a more meaningful analysis. It is worth mentioning that the percentage of variance explained by the first principal component (58.65 %) is very close to the variance explained by the hierarchical tree when is partitioned into three clusters (59.54 %).

The three clusters, found using the Wards' method only, are represented in terms of solutions in the principal component map (Figure 11a). This figure shows how solutions are grouped together into three clusters when the overall cloud of solutions is projected into the first two principal components. Here, cluster 1 is not very well separated from cluster 3, which means that both clusters share similar solutions.

To enhance the distinctiveness of the cluster partitioning, the K-Means algorithm is subsequently applied. This refinement step adjusts the initial partitioning obtained through Ward's method. The K-Means algorithm optimizes cluster separation by iteratively recalculating the centroids for each cluster and reassigning solutions according to their proximity in Euclidean space. This procedure incrementally increases the ratio of between-cluster variance to the total variance, which results in the reduction of overlap and a clearer delineation of clusters. The process continues until the improvement in this variance ratio does not exceed a certain threshold, thus solidifying the partitioning. The iterative optimization by the K-Means algorithm is what transforms the initial, less distinct clusters arrangement (Figure 11a) into a final partitioning where clusters are well-separated and more compact (Figure 11b). This refined partitioning is not only more visually apparent but also statistically significant, and it is this final configuration that is retained for further analysis within the paper.

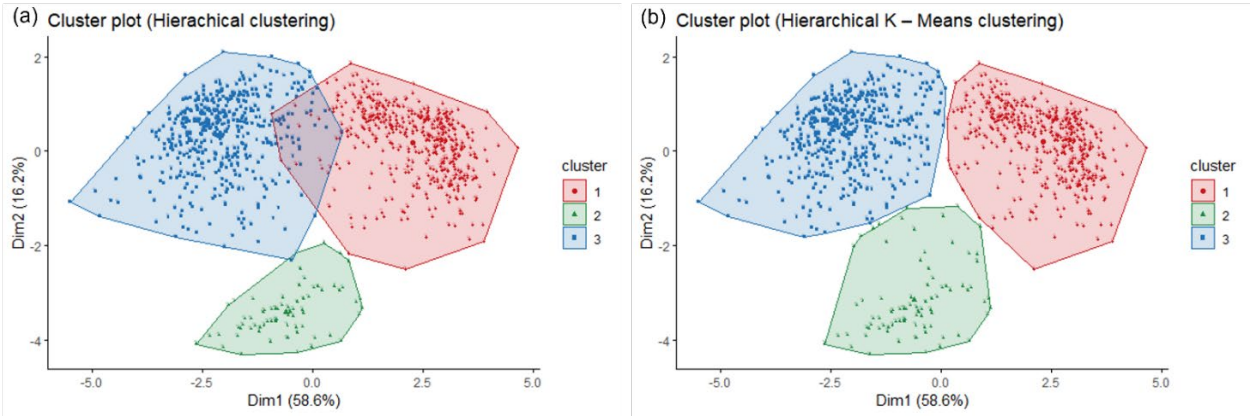


Figure 11. (a). Solutions' clusters partitioning on the principal component map, using the Ward's method only. (b) Solutions' clusters partitioning using the Hierarchical K-Means method.

5.3 Further considerations on the model's parameters

In Table 4, each standardized variable $\widehat{\boldsymbol{V}}_k$ is presented as a vector, summarizing observations form the 1024 solutions. This forms the basis for the analysis focusing on the first two principal components, denoted as $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$. The table displays the correlations $r_s = r(\widehat{\boldsymbol{V}}_k, \boldsymbol{p}_s)$ (where s = 1,2) between the variables and these components (column 1 and 4). Additionally, the table includes the quality of the representation qlt_s of each variable (columns 2 and 5), and the weight of each variable qtr_s in the construction of these components (columns 3 and 6). The quality of representation, $qlt_s = {r_s}^2$, measures the extent to which a variable is accurately projected onto a principal component. The weight of a variable, $qtr_s = \frac{{r_s}^2}{\lambda_s} \cdot 100\%$, quantifies the variable's relative contribution to the variance explained by the principal component, with λ_s being the eigenvalue corresponding to that component (Husson and Pagès, 2017).

Table 4. Principal component analysis results for variables in terms of correlations (r), quality of representation (qlt), and contribution to the construction (qtr) relative to the first two principal components. $\widehat{\boldsymbol{V}}_{\boldsymbol{k}}$ represents the k-th standardized variable; $\boldsymbol{p}_1$ and $\boldsymbol{p}_1$ denotes the first and the second principal components, respectively.

Variables $\widehat{\boldsymbol{V}}_k$	Dim-1 $r(\widehat{\boldsymbol{V}}_k, \boldsymbol{p}_1)$	Dim-1 $qlt_1(\widehat{\boldsymbol{V}}_k)$	Dim-1 $qtr_1(\widehat{\boldsymbol{V}}_k)$	Dim-2 $r(\widehat{\boldsymbol{V}}_k, \boldsymbol{p}_2)$	Dim-2 $qlt_2(\widehat{\boldsymbol{V}}_k)$	Dim-2 $qtr_2(\widehat{\boldsymbol{V}}_k)$
$\widehat{\boldsymbol{X}}_{\boldsymbol{C0}}$	-0.831	0.691	13.094	-0.443	0.196	13.441
$\widehat{\boldsymbol{Y}}_{\boldsymbol{C0}}$	0.723	0.523	9.912	-0.489	0.239	16.377
$\widehat{\boldsymbol{R}}$	0.578	0.334	6.326	-0.256	0.066	4.504
$\widehat{\boldsymbol{\rho}}$	0.715	0.512	9.699	-0.216	0.047	3.200
$\widehat{\boldsymbol{V}}_{\boldsymbol{r,max}}$	-0.909	0.827	15.664	0.079	0.006	0.424
$\widehat{\boldsymbol{T}}_{\boldsymbol{max}}$	0.182	0.033	0.628	0.916	0.839	57.502
$\widehat{\boldsymbol{T}}_{\boldsymbol{end}}$	-0.823	0.678	12.847	0.117	0.014	0.942
$\widehat{\boldsymbol{V}}_{\boldsymbol{t}}$	0.969	0.939	17.789	0.132	0.017	1.189
$\widehat{\boldsymbol{F}}$	0.861	0.741	14.042	0.188	0.035	2.421
Secondary variable						
$\widehat{\boldsymbol{V}}_{\boldsymbol{b}}$	0.299	0.089	-	-0.073	0.005	-

Table 4 also presents the secondary variable $\widehat{\boldsymbol{V}}_{\boldsymbol{b}}$. The other two variables $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{t}}$ and $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{b}}$ are non-considered in the PCA due to their circular nature, which does not align well with the linear interpretation framework of principal component analysis. Despite $\widehat{\boldsymbol{V}}_{\boldsymbol{b}}$ not being involved in the construction of the principal components, it is still possible to evaluate the correlation and the quality of the representation of this variable using the two principal components. To facilitate the interpretation of Table 4, a correlation circle plot (Abdi and Williams, 2010) can be used to visually represent the variables. This plot represents each variable as a point in a two-dimensional space, where the coordinates of each point correspond to the correlation coefficients between the variable and the two principal components (i.e., $r(\widehat{\boldsymbol{V}}_k, \boldsymbol{p}_1), r(\widehat{\boldsymbol{V}}_k, \boldsymbol{p}_2)$). Figure 12a illustrates the correlation circle plot.

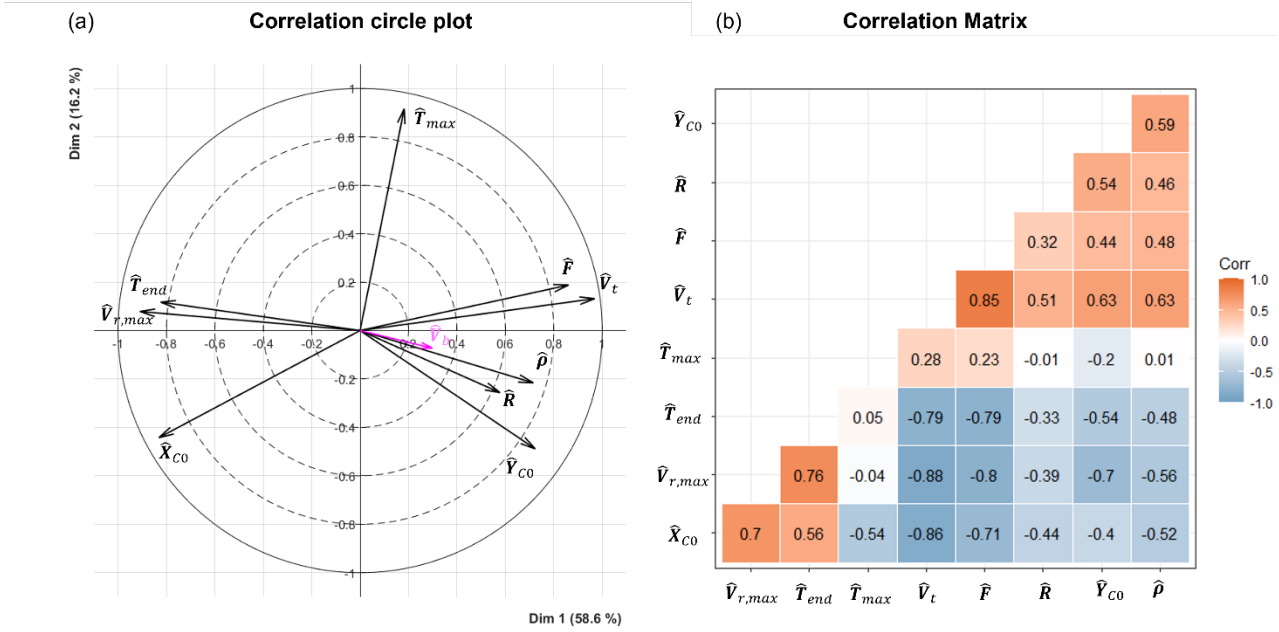


Figure 12. (a) Correlation circle plot. The variables in black are considered as primary variables whereas the variable in magenta is a secondary variable (b) Correlation matrix plot.

The plot geometrically represents variable correlations: the angles between the variables indicate the level of correlations between variables, with acute angles suggesting positive correlation and obtuse angles indicating negative correlation. Each variable's total contribution across all principal components equals 1. Variables fully explained by the first two components will be located on the circle's circumference (radius 1) in the correlation circle. Variables not well represented by these components will be near the center, indicating that only those near the circumference are significantly represented. Except for the variables $\widehat{\boldsymbol{V}}_{\boldsymbol{b}}$ and $\widehat{\boldsymbol{R}}$ which are not very well represented by the first two principal components, the remaining variables are very well represented since their tip is close to the circle of radius 1. The set of variables $\{\widehat{\boldsymbol{V}}_{\boldsymbol{t}}, \widehat{\boldsymbol{F}}, \widehat{\boldsymbol{Y}}_{\boldsymbol{C0}}, \widehat{\boldsymbol{\rho}}\}$ are positively correlated, increasing together, similarly to $\{\widehat{\boldsymbol{V}}_{\boldsymbol{r,max}}, \widehat{\boldsymbol{X}}_{\boldsymbol{C0}}, \widehat{\boldsymbol{T}}_{\boldsymbol{end}}\}$. The variable $\widehat{\boldsymbol{V}}_{\boldsymbol{t}}$ is highly correlated with the first component (correlation of 0.97). Essentially $\widehat{\boldsymbol{V}}_{\boldsymbol{t}}$ can be viewed as a representative summary of the first principal component axis. Figure 12a indicates that the variable $\widehat{\boldsymbol{V}}_{\boldsymbol{t}}$ has a strong negative correlation with the variables $\{\widehat{\boldsymbol{V}}_{\boldsymbol{r,max}}, \widehat{\boldsymbol{X}}_{\boldsymbol{C0}}, \widehat{\boldsymbol{T}}_{\boldsymbol{end}}\}$. This suggests that high storm motion values V_t correspond with lower maximum radial velocities $V_{r,max}$, position closer to the station for positive values of X_{C0}, farther for negative values, and shorted downburst duration T_{end}.

Since $\widehat{\boldsymbol{V}}_{\boldsymbol{t}}$ is positively correlated with the variables $\{\widehat{\boldsymbol{F}}, \widehat{\boldsymbol{Y}}_{\boldsymbol{C0}}, \widehat{\boldsymbol{\rho}}\}$, what is true for $\widehat{\boldsymbol{V}}_{t}$ with respect to the group of variables $\{\widehat{\boldsymbol{V}}_{\boldsymbol{r,max}}, \widehat{\boldsymbol{X}}_{\boldsymbol{C0}}, \widehat{\boldsymbol{T}}_{\boldsymbol{end}}\}$, will also remain true for the variables $\{\widehat{\boldsymbol{F}}, \widehat{\boldsymbol{Y}}_{\boldsymbol{C0}}, \widehat{\boldsymbol{\rho}}\}$. Finally, from the correlation circle plot, it seems that the variable $\widehat{\boldsymbol{T}}_{\boldsymbol{max}}$ is not very well "linearly" correlated with the groups of variables $\{\widehat{\boldsymbol{V}}_{\boldsymbol{r,max}}, \widehat{\boldsymbol{T}}_{\boldsymbol{end}}, \widehat{\boldsymbol{\rho}}\}$ since it is nearly orthogonal with these variables. From a quantitative point of view the values of the correlation coefficients between all the pairs of variables are plotted in Figure 12b. Table 4 lists each variable's contributions to the first and second principal components (columns 3 and 6 respectively). Figure 13a and 13b graph these contributions in percentages, showing which variables have the most impact on these two components.

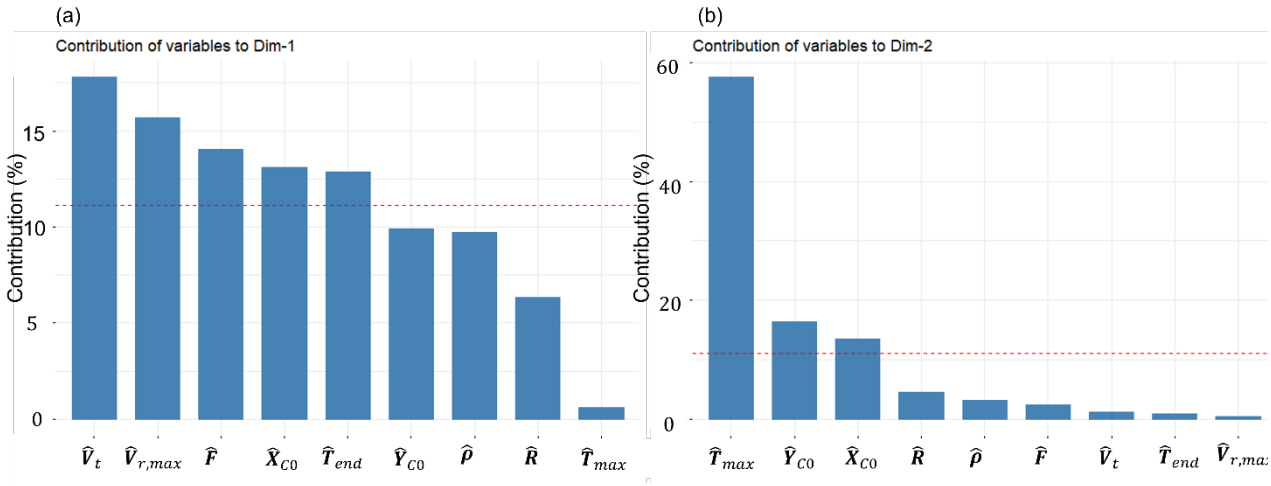


Figure 13. (a) Contribution of the variables in the reconstruction of the first principal component (Dim-1). (b) Contribution of the variables in the reconstruction of the second principal component (Dim-2). Variables are sorted from the strongest to the weakest. The red dashed line indicates the expected average contribution.

The graph shows a red dashed line indicating the average expected variable contribution at 11.11 %, based on 9 variables. Variables with contributions over 11.11% significantly construct a principal component. For the first component, $\{\widehat{\boldsymbol{V}}_{\boldsymbol{t}}, \widehat{\boldsymbol{V}}_{\boldsymbol{rmax}}, \widehat{\boldsymbol{F}}, \widehat{\boldsymbol{X}}_{\boldsymbol{C0}}, \widehat{\boldsymbol{T}}_{\boldsymbol{end}}\}$ are key contributors. For the second, $\{\widehat{\boldsymbol{T}}_{\boldsymbol{max}}, \widehat{\boldsymbol{Y}}_{\boldsymbol{C0}}, \widehat{\boldsymbol{X}}_{\boldsymbol{C0}}\}$ are most influential. The leading contributors for both components combined, ranked by importance in building the first two principal components are $\{\widehat{\boldsymbol{V}}_{\boldsymbol{t}}, \widehat{\boldsymbol{X}}_{\boldsymbol{C0}}, \widehat{\boldsymbol{T}}_{\boldsymbol{max}}, \widehat{\boldsymbol{V}}_{\boldsymbol{rmax}}, \widehat{\boldsymbol{F}}, \widehat{\boldsymbol{Y}}_{\boldsymbol{C0}}\}$. The remaining variables $\{\widehat{\boldsymbol{T}}_{\boldsymbol{end}}, \widehat{\boldsymbol{\rho}}, \widehat{\boldsymbol{R}}\}$ fell below the average contribution of 11,11 %. It is worth mentioning that the categorization of variables from stronger to weaker is not universal since the partitioning might depend on the downburst case under investigation.

5.4 Physical description of the solutions corresponding to clusters 1-3

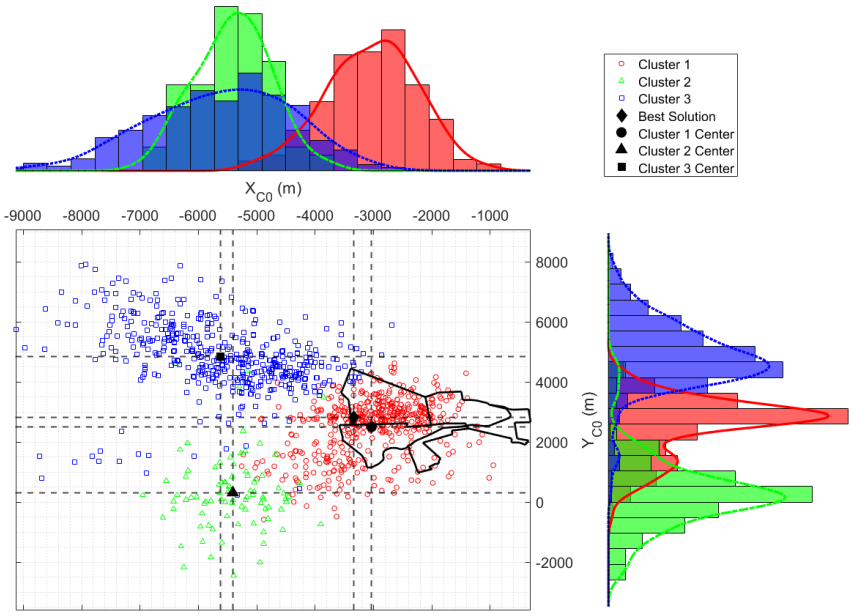
Once the partitioning of the solutions of the optimization problems in three cluster is completed, it is important to have a closer look at them and describe common features of solutions which belong to the same cluster. From the partition analysis, it is found that cluster 1 is made up of 481 solutions, cluster 2 of 85 and cluster 3 of 458 solutions. Table 5 summarizes a few key statistics related to the three clusters. This table includes primary and secondary (i.e., not used for clustering) variables, which are no longer standardized to investigate their physical meaning.

508 **Table 5. Description of the partition by the mean and standard deviation of all the variables.**

Variables $\boldsymbol{V_k}$	Overall Mean	Overall Std	Cluster 1 Mean	Cluster 1 Std	Cluster 2 Mean	Cluster 2 Std	Cluster 3 Mean	Cluster 3 Std
$\boldsymbol{V_t}$ (m/s)	6.025	3.371	2.811	1.042	6.527	1.407	9.307	1.492
$\boldsymbol{X_{C0}}$ (m)	-4386.350	1613.337	-3034.079	789.682	-5410.461	629.282	-5616.465	1209.346
$\boldsymbol{T_{max}}$ (min)	6.954	2.517	5.860	1.172	13.336	1.910	6.919	1.797
$\boldsymbol{V_{r,max}}$ (m/s)	24.293	5.356	28.639	1.465	28.182	1.793	19.006	3.266
$\boldsymbol{F}$ (-)	0.893	0.080	0.823	0.058	0.919	0.043	0.961	0.021
$\boldsymbol{Y_{C0}}$ (m)	3363.669	1809.316	2499.896	975.450	313.553	1257.946	4836.890	1160.613
$\boldsymbol{T_{end}}$ (min)	26.035	3.167	28.269	1.895	27.622	2.295	23.394	2.235
$\boldsymbol{\rho}$ (-)	2.189	0.108	2.126	0.104	2.134	0.100	2.265	0.050
$\boldsymbol{R}$ (m)	1334.478	102.519	1289.518	124.661	1301.969	90.475	1387.728	23.115
Secondary variables								
$\boldsymbol{\alpha_t}$ (deg)	290.383	0.480	276.439	0.416	253.518	0.217	310.868	0.229
$\boldsymbol{V_b}$ (m/s)	6.811	0.670	6.648	0.774	6.705	0.768	7.002	0.449
$\boldsymbol{\alpha_b}$ (deg)	268.218	0.118	264.854	0.138	273.055	0.074	270.827	0.055

509

510 In columns 2-3, the overall mean, and the overall standard deviation (std) are calculated with respect to each variable
511 (primary and secondary). In the other columns, the same calculation was repeated taking into consideration the three
512 clusters. Mean and the std of the secondary variables $\boldsymbol{\alpha_t}$ and $\boldsymbol{\alpha_b}$ have been calculated using circular statistics (Rao and
513 Sengupta, 2001). To start clarifying the characteristics of the different clusters, Figure 14 shows the scatter plot and
514 distribution of the touchdown components $(\boldsymbol{X_{C0}}, \boldsymbol{Y_{C0}})$ for all the solutions, partitioned into three clusters. In this figure it
515 is shown the center (namely the mean) of each cluster and the location of the touchdown position of the best overall
516 solution. The figure shows also with a black line the position of the city of Sânnicolau Mare. Also, on the left and on the
517 top of this figure is possible to show the histograms of the variable $(\boldsymbol{X_{C0}}, \boldsymbol{Y_{C0}})$ relative to each cluster.



518 **Figure 14. Scatter plot and histogram density distribution for the variables (X_{C0}, Y_{C0}). The dark black line shows the**
519 **contours of the city Sânnicolau Mare.**

520

521 The three clusters appear well separated in terms of touchdown position ($\boldsymbol{X_{C0}, Y_{C0}}$). Since it is very unlikely that the
522 cluster means coincide with one of the solutions present in the data set, let's define as "cluster solution", the solution
523 which is the closest to the mean of the cluster across all variables. Accordingly, the cluster solutions, reported in Table 6,
524 will be used to interpret the average features of each cluster. The first row of this table is dedicated to the best solution
525 founded by the optimization algorithm (i.e., the one that have the lowest objective function F among all the solutions);
526 the best solution belongs to Cluster 1.

527 **Table 6. Overall best solution and clusters representative solutions.**

528

Solutions	V_t (m/s)	X_{C0} (m)	T_{max} (min)	$V_{r,max}$ (m/s)	F (-)	Y_{C0} (m)	T_{end} (min)	ρ (-)	R (m)	α_t (deg)	V_b (m/s)	α_b (deg)
Best solution	2.76	-3339.53	6.50	29.80	0.73	2826.55	29.89	2.15	1381.38	271.74	5.49	58.35
Cluster 1	2.51	-2944.15	6.05	29.54	0.81	2769.36	27.23	2.09	1287.53	278.25	7.15	268.19
Cluster 2	6.14	-5105.66	14.05	27.07	0.91	383.39	28.18	2.14	1295.33	255.36	7.13	272.82
Cluster 3	9.25	-5930.81	7.15	17.36	0.97	4575.50	22.95	2.27	1392.86	307.61	6.15	272.71

529

530 Figure 15 shows the time histories produced by the best solution and the three cluster solutions, in terms of wind velocity
531 (Figure 15a) and direction (Figure 15b), compared with the moving averaged recorded data. The figure provides a
532 qualitative representation of the goodness of fit between the simulations and the recorded data. The goodness of fit is
533 quantitatively measured by the objective function F. The simulations produced from the best solution and the Cluster 1
534 solution fit the data better than Cluster 2 and 3. This is quite obvious since the best solution have the lowest objective

function F and belongs to Cluster 1, whereas Cluster 2 and Cluster 3 solutions have slightly higher objective function values F (refer to column 5 in Table 6).

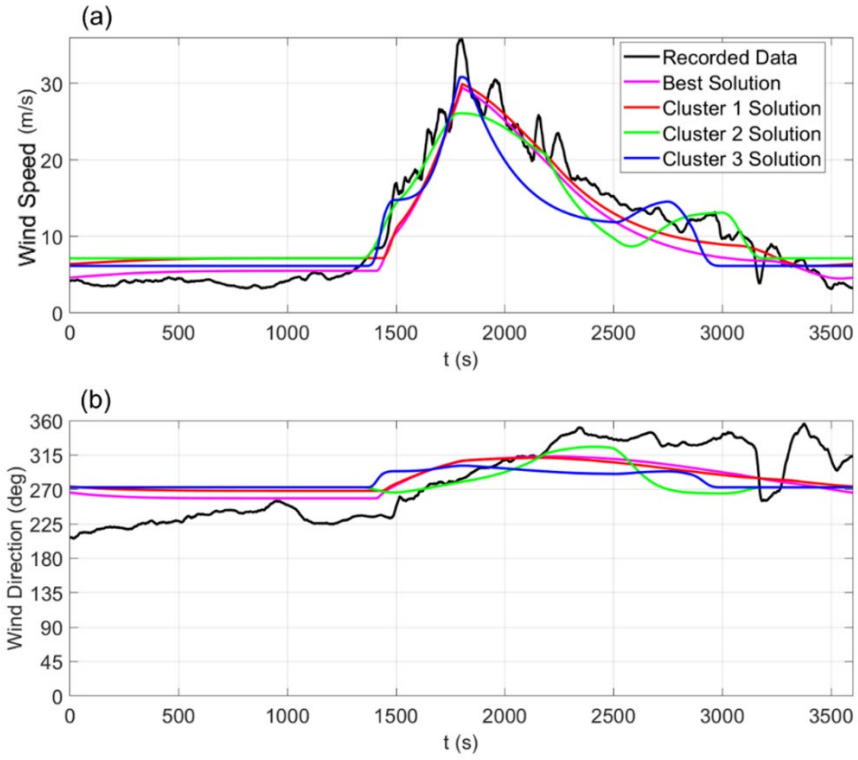
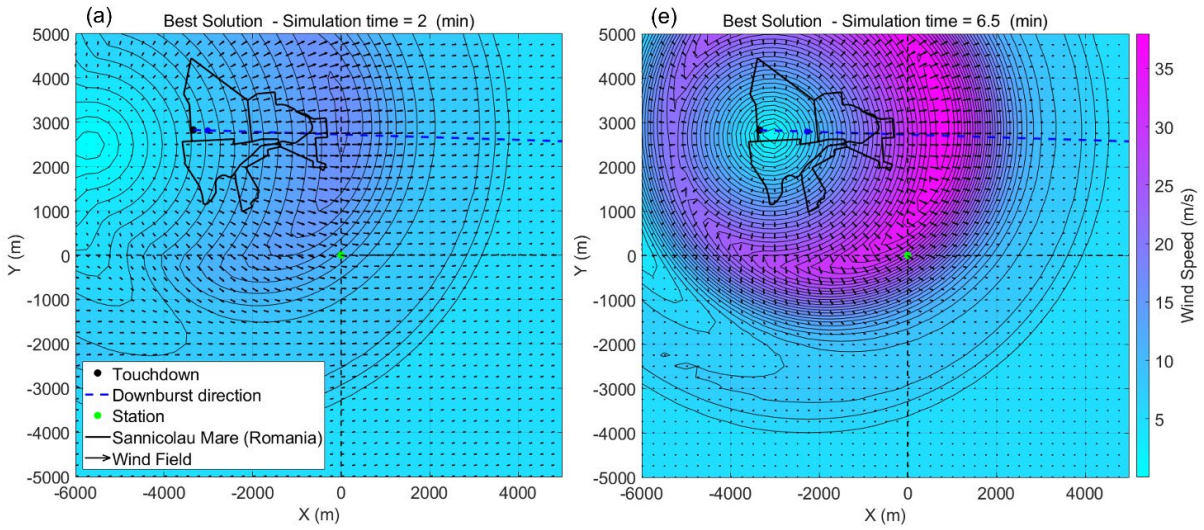
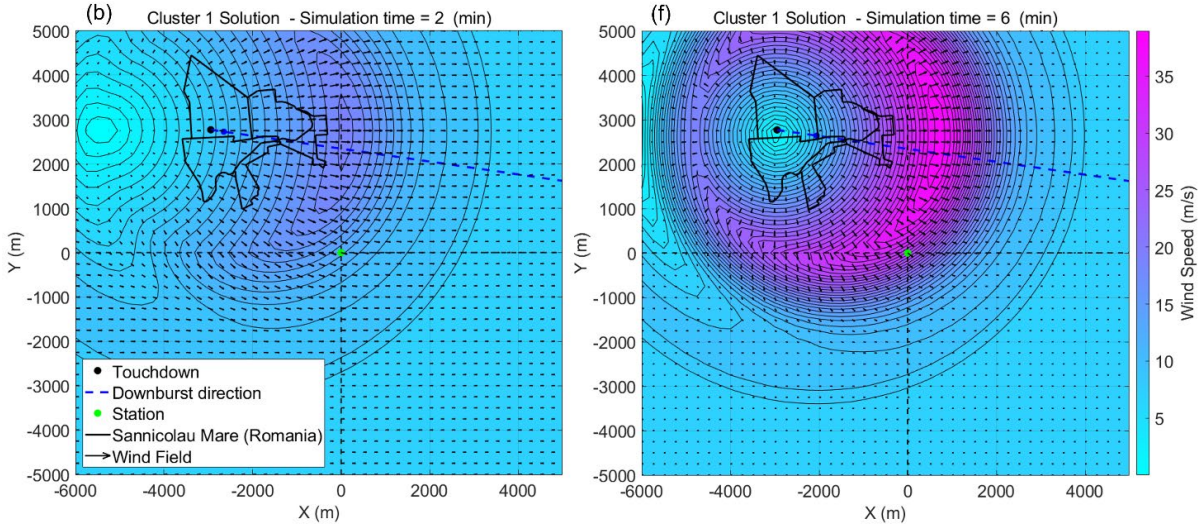


Figure 15. Comparison among the moving averaged wind speed (a) and direction (b) obtained from the measurements of the Sânnicolau Mare downburst, along with the best solution and the three cluster solutions.

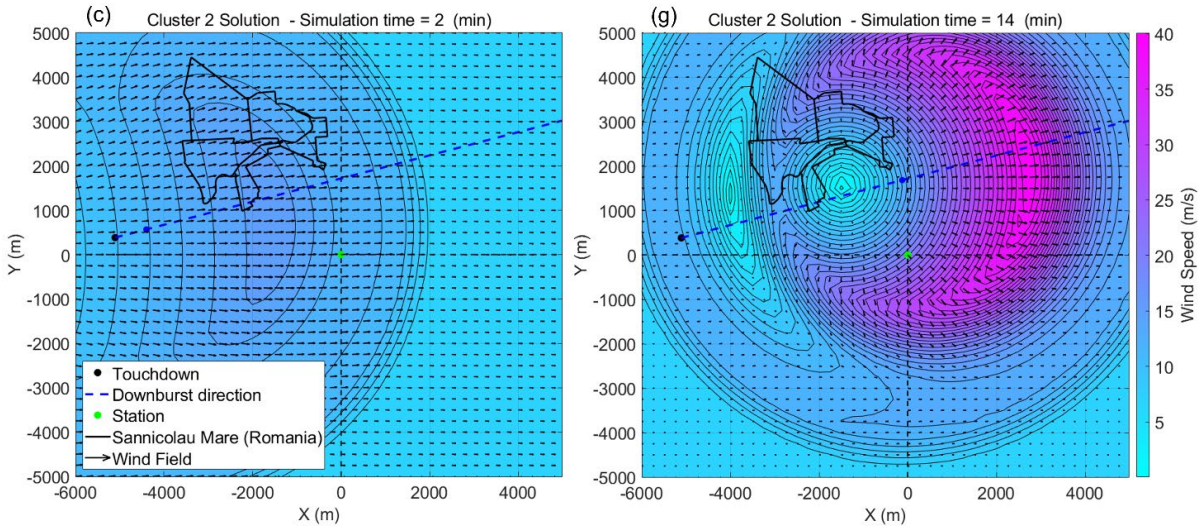
In order to better understand the nature of the different solutions relative to each cluster, for each solution present in Table 6, the downburst 2D wind velocity is evaluated at the same height of the anemometric station (i.e., at 50 m AGL). The left panels of Figure 16 (from (a) to (d)) show for each of the 4 solutions the wind field reconstruction during the intensification stage of the downburst, while the right panels (from (e) to (h)) describes the stage of maximum intensity. Note that the time of maximum intensity is different for each cluster according to the corresponding value of T_{max} reported in column 3 of Table 6.


(a)
Best Solution - Simulation time = 2 (min)
(e)
Best Solution - Simulation time = 6.5 (min)
Touchdown
Downburst direction
Station
Sannicolau Mare (Romania)
Wind Field
X (m)
Y (m)
Wind Speed (m/s)

547


(b)
Cluster 1 Solution - Simulation time = 2 (min)
(f)
Cluster 1 Solution - Simulation time = 6 (min)
Touchdown
Downburst direction
Station
Sannicolau Mare (Romania)
Wind Field
X (m)
Y (m)
Wind Speed (m/s)

548


(c)
Cluster 2 Solution - Simulation time = 2 (min)
(g)
Cluster 2 Solution - Simulation time = 14 (min)
Touchdown
Downburst direction
Station
Sannicolau Mare (Romania)
Wind Field
X (m)
Y (m)
Wind Speed (m/s)

549

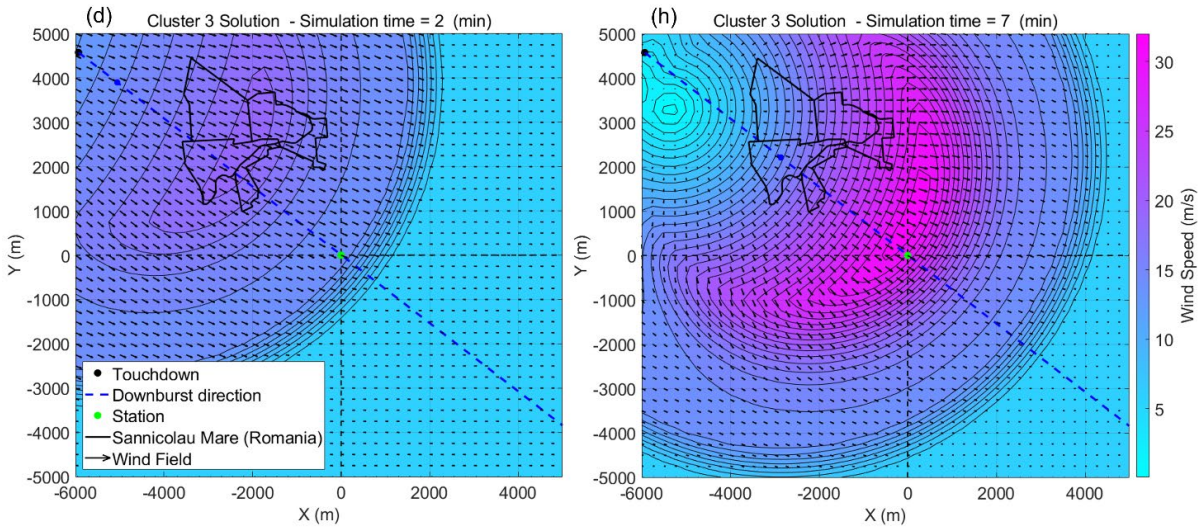
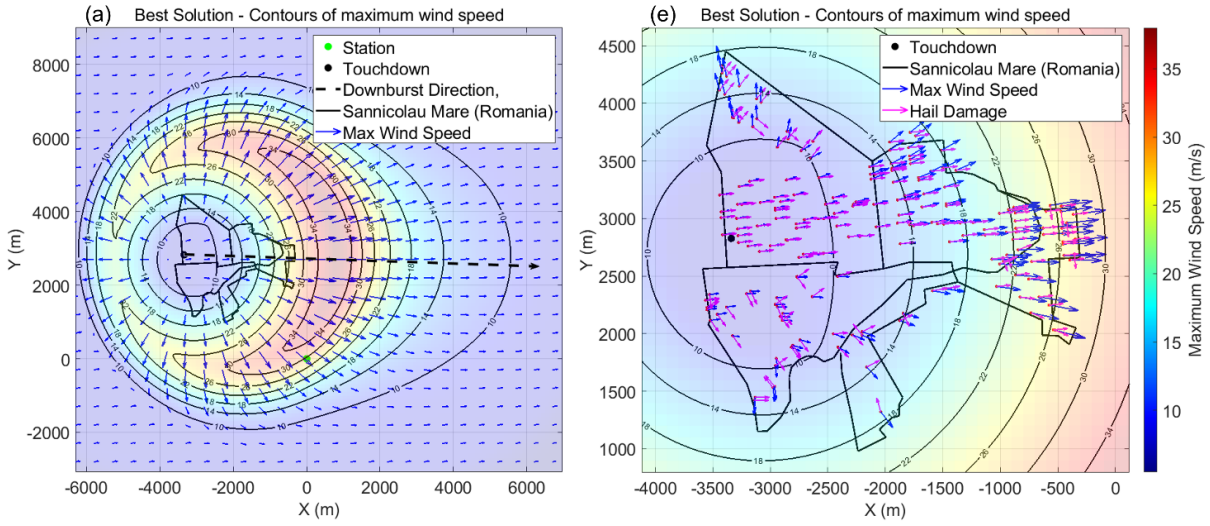


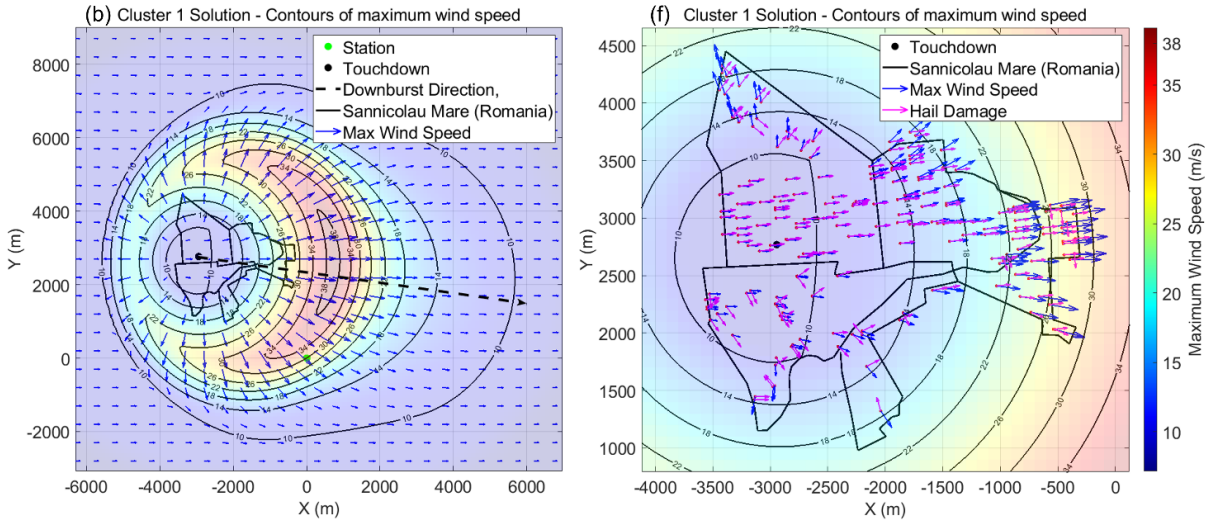
Figure 16. 2D wind field reconstruction at 50 m AGL. From (a) to (d), the best solution, Cluster 1, 2 and 3 solutions are depicted at the intensification stage of the downburst. From (e) to (h), the best solution, Cluster 1, 2 and 3 solutions are depicted at the maximum intensification stage of the downburst.

Cluster 1 touches down very close to the city center and moves slowly eastward, it is characterized by a low value of the downburst translation velocity V_t, with mean value 2.8 m/s against the overall mean among all clusters which is 6.0 m/s. In addition, it has maximum radial velocity $V_{r,max}$ higher and overall duration of the downburst event T_{end} longer with respect to the mean values of the other two clusters. The solutions belonging to Cluster 2 touch down around 2 km southwest of the city, they propagate northeastward with higher translation velocities compared to Cluster 1 and the longest intensification periods T_{max} overall. The solutions in Cluster 3 touch down about 3 km northwest of the city, they move southeastward with the highest values of downburst translation velocity V_t but they are the lowest-lasting as the duration of the downburst event T_{end} is on average 23.4 min while the overall mean is 26.0 min. They also have the lowest values of maximum radial velocity $V_{r,max}$ which compensate the high translation velocities. According to these descriptions, it is clear that in the solution's space of the model three different solutions exist that can describe similarly the time-series measured at TM_424. The existence of different plausible solutions means that the problem of finding the downburst wind field time-space evolution using a single time-series is an underdetermined problem.

The Sânnicolau Mare downburst had a strong impact, causing hail damage to numerous buildings in the town. A damage survey was conducted to assess the affected areas and identify buildings that experienced hail damage during the event. To estimate the extent of the damage, the simulated wind field generated by the analytical model was utilized. By analyzing the wind speeds at various locations, the "footprint" of the simulated damage was determined. This footprint represents the maximum wind speed recorded at different places during the downburst, providing valuable information on the areas most affected by the event. The left panels of Figure 17, labeled from (a) to (d), depict the complete footprint of the downburst potential damage area for the best solution and the three cluster solutions. In contrast, the right panels, labeled from (e) to (h), provide a closer view of the footprints overlaying the simulated maximum wind velocity vectors (indicated by blue arrows) onto the locations of hail damage. The hail damage is represented by vectors pointing orthogonally to the damaged facades (represented by pink arrows). The comparison between the facades damage, which is related to the trajectory of hail transported by the strong downburst-related outflow, and the simulated maximum

velocity reveals interesting findings. Specifically, the best solution and Cluster 1 solutions exhibit the strongest alignment between the maximum wind velocity vectors and hail damage vectors, particularly in the central part of the city and along the path of the downburst. In contrast, Cluster 2 and Cluster 3 demonstrate a consistent deviation of the maximum velocity, with Cluster 2 deviating northward and Cluster 3 deviating southward, relative to the hail trajectories. This observation suggests that the actual downburst event likely followed a pattern more closely resembling Cluster 1 rather than the other two potential solutions.





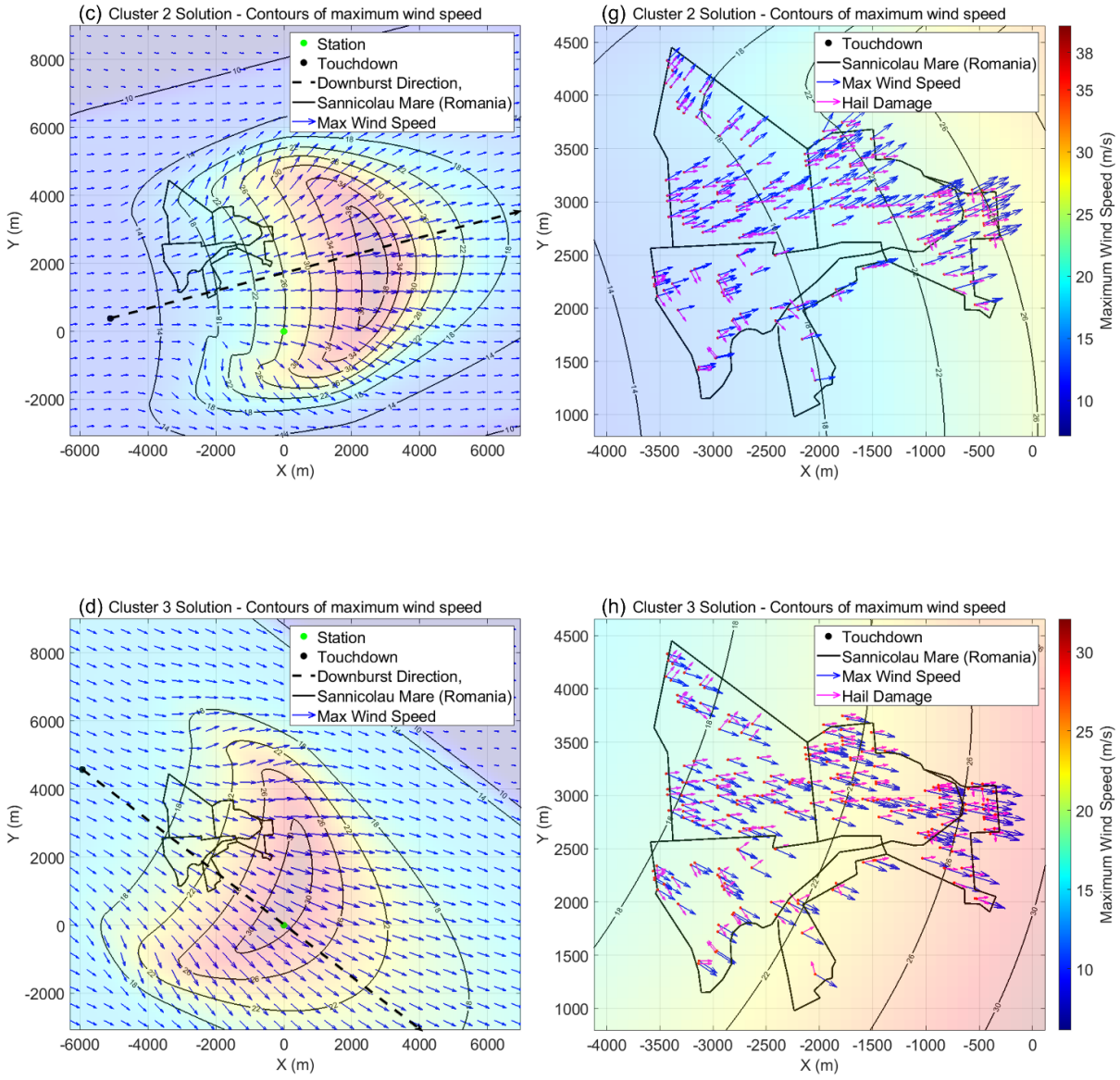


Figure 17. Simulated footprints of the downburst that occurred in Sânnicolau Mare. Panels from (a) to (d) representing the footprints for the best solution, Cluster 1, Cluster 2, and Cluster 3 respectively. Panels form (e) to (h), representing comparison between hail damage and maximum simulated wind speed for the best solution, Cluster 1, Cluster 2, and Cluster 3 respectively.

These observations lead to the conclusion that the optimal (best) solution, which minimizes the objective function *F*, is the most reliable among all possible solutions. In the current study, this has been achieved through a comprehensive approach involving numerous simulations, specifically tailored for cases where only anemometric data is available, despite having access to additional data types like the Radar images. This choice was driven by the higher likelihood and frequency of availability of anemometric data in practical scenarios, thus providing a more universally applicable context for the analytical downburst model. The methodology involved conducting a large number of simulations to thoroughly explore the solution space, given the data-limited scenario. Consequently, this approach enabled the extraction of kinematic and geometric parameters of the downburst outflow wind field exclusively from anemometric data.

599 However, it is important to acknowledge that in scenarios where additional data types such as Radar or Lidar or other
600 sensors are available, the approach to reconstructing the downburst wind field would differ significantly. In such
601 situations, the availability of more diverse data allows for a more constrained and targeted reconstruction process. By
602 integrating specific parameters from these additional data sources, like storm speed and direction and ABL wind speed
603 and direction, the solution space can be narrowed down more effectively, potentially reducing the number of simulations
604 needed and enhancing the precision of the model.

605 **6 Conclusions**

606 This study focuses on the analysis of solutions obtained by combining an analytical model (Xhelaj et al., 2020) with a
607 global metaheuristic optimization algorithm for the reconstruction of the wind field generated during the Sânnicolau Mare
608 downburst event in Romania on June 25, 2021. The analytical model and optimization algorithm are coupled using the
609 Teaching Learning Optimization Algorithm (TLBO) to estimate the kinematic parameters of the downburst outflow. The
610 procedure for this coupling and parameter estimation is described in detail in the study by Xhelaj et al. (2022). Therefore,
611 the objective was to analyse the differences among the solutions provided by the optimization algorithm and to assess
612 their physical validity as alternatives to the optimal solution. In the presence of multiple physically sounding solutions, it
613 has been demonstrated that additional data describing the downburst thunderstorm event is necessary to determine which
614 solution best represents reality. To support the analysis a comprehensive damage survey was conducted in collaboration
615 with the University of Genoa (Italy) and the University of Bucharest (Romania) to assess the extent and location of hail
616 damage on buildings in the affected area. This survey, along with the wind speed and direction signals recorded during
617 the downburst event by a telecommunication tower located approximately 1 km from the city, significantly enhances the
618 information available for the reconstruction and simulation of the downburst using the optimization procedure. The
619 analysis of the solutions generated by the optimization algorithm involves multivariate data analysis (MDA) techniques,
620 specifically agglomerative hierarchical clustering coupled with the K-Means algorithm (AHK-MC) and principal
621 component analysis (PCA). The AHK-MC is used for classifying the solutions into different clusters based on their
622 features, while PCA is employed to determine the importance of the variables in the analytical model for the downburst
623 event reconstruction.

624 The application of AHK-MC resulted in the identification of three main clusters, each with distinct characteristics, among
625 the 1024 solutions.

626 - Solutions belonging to Cluster 1 are characterized by a slow storm motion, small touch down distance from the
627 city of Sânnicolau Mare and by long duration of the downburst event. The best overall solution belongs to Cluster
628 1.
629 - Solutions belonging to Cluster 2 are characterized by a moderate storm motion and moderate distance of the
630 touch-down from the town of Sânnicolau Mare. These solutions are also characterized by high duration of the
631 intensification period of the downburst event.
632 - Solutions belonging to Cluster 3 are characterized by a high storm motion and high distance of the touch-down
633 from Sânnicolau Mare. They are also characterized by low duration of the downburst event and low values of
634 the maximum radial velocity.

635 The result of the MDA allows also to establish at least for the case under consideration that the set of variables
636 $\{\boldsymbol{V_t}, \boldsymbol{X_{C0}}, \boldsymbol{T_{max}}, \boldsymbol{V_{rmax}}, \boldsymbol{F}, \boldsymbol{Y_{C0}}\}$ which are ordered from the strongest to the weakest are the more important for the
637 reconstruction/simulation of the downburst event. The remaining variables $\{\boldsymbol{T_{end}}, \boldsymbol{\rho}, \boldsymbol{R}\}$ have a lower contribution. It is

important to observe the partitioning in strongest variables and weakest ones does not represent a general case, since the partition depends on the downburst case under study.

Finally, the comparison between the facades damage, which are related to the trajectory of hails transported by the strong downburst-related outflow and the simulated maximum velocity shows that the best solution and Cluster 1 solutions seem to have a "good" overlapping between maximum wind velocity vectors and hail damage vectors. Considering the solutions of Cluster 2 and 3, it seems that the match between maximum wind velocity vectors gradually decreases, with the worst case represented by Cluster 3 solutions. These observations allow to conclude that the optimal solution, that is, the one that minimizes the objective function F, is the best with respect to the other three cluster solutions, also from the point of view of the damage analysis. As a result, for the specific case being examined, relying on the best overall solution provided by the optimization algorithm appears to yield promising results for reconstructing the downburst wind field. Obviously, an analysis of this type, conducted on several downburst events, will be able to better confirm this statement.

Author contributions

This paper is based on the Ph.D. thesis of Andi Xhelaj, under the guidance of Prof. Giovanni Solari and Prof. Massimiliano Burlando. Andi Xhelaj played a crucial role in conceptualizing the study, developing the methodology, organizing the data, conducting data analysis, and preparing the manuscripts and figures. The study was supervised by Prof. Massimiliano Burlando, who provided guidance and conducted internal review process.

Declaration of competing interest

The authors affirm that they have no known financial conflicts of interest or personal relationships that could have influenced the findings presented in this paper.

Acknowledgments

The authors would like to acknowledge the valuable contributions of I. Calotescu, X. Li, M.T. Mengistu, and M.P. Repetto for providing the time histories of the recorded data and the hail damage map from the survey of the Thunderstorm event in Sânnicolau Mare, Romania, on 25 June 2021. The monitoring system and damage survey were carried out as part of a research collaboration between the University of Genoa (UniGe) and the Technical University of Civil Engineering Bucharest (UTCB), funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 741273) for the Project THUNDERR - Detection, simulation, modeling, and loading of thunderstorm outflows to design wind-safer and cost-efficient structures, supported by an Advanced Grant (AdG) 2016. As the corresponding author is not a native English speaker, he utilized the OpenAI's GPT-4 model as an editing tool during the creation of this paper to review and amend grammatical and spelling mistakes and to ensure linguistic consistency and coherence throughout this paper.

References

1. Abd-Elaal, E.S., Mills, J.E., Ma, X., 2013. A coupled parametric-CFD study for determining ages of downbursts through investigation of different field parameters. J. Wind Eng. Ind. Aerod. 123, 30–42.
2. Abdi, H., and Williams, L., J.: Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics 2(4):433 – 459, 2010.
3. Amato, F., Guignard, F., Robert, S. Kanevski, M.,: A novel framework for spatio-temporal prediction of environmental data using deep learning. Sci Rep 10, 22243, doi:10.1038/s41598-020-79148-7, 2020.

4. Bjerknes, J., Solberg, H.: Life cycle of cyclones and polar front theory of atmospheric circulation. Geophysiks Publikationer 3, 3-18, 1922.
5. Bogensperger, A., Fabel, Y. A practical approach to cluster validation in the energy sector. Energy Inform 4 (Suppl 3), 18 (2021). https://doi.org/10.1186/s42162-021-00177-1.
6. Burlando, M., Antonelli M., Ratto, C., F.: Mesoscale wind climate analysis: identification of anemological regions and wind regimes. Int. J. Climatol. 28(5), 629–641. DOI: 10.1002/joc.1561 – ISSN 1097-0088, 2008.
7. Burlando, M.: The synoptic-scale surface wind climate regimes of the Mediterranean Sea according to the cluster analysis of ERA-40 wind fields. Theor. Appl. Climatol. 96(1), 69-83. DOI: 10.1007/s00704-008-0033-5 – ISSN 0177-798X – eISSN 1434-4483, 2009.
8. Burlando, M., Romanic, D., Solari, G., Hangan, H., Zhang, S.: Field data analysis and weather scenario of a downburst event in Livorno, Italy on 1 October 2012. Mon. Weather Rev. 145, 3507-3527, 2017.
9. Calotescu, I., Bîtcă D., Repetto, M., P.: Full-scale behaviour of a telecommunication lattice tower under wind loading. Lightweight Structures in Civil Engineering, XXVII LSCE Łódź, 2nd – 3rd of December 2021, Szafran J. and Kaminski M, Eds. p. 15-18, 2021.
10. Calotescu, I., Repetto, M., P.: Wind and structural monitoring system for a Telecommunication lattice tower. 14th Americas Conference on Wind Engineering, May 17-19, 2022, Lubbok, TX, 2022.
11. Calotescu, I., Li, X., Mengistu, M.T., Repetto, M., P.:Post-event Survey and Damage Analysis of An Intense Thunderstorm in Sannicolau Mare, Romania. 14th Americas Conference on Wind Engineering, May 17-19, 2022, Lubbok, TX, 2022.
12. Calotescu, I., Li, X., Mengistu, M.T., Repetto, M., P.: Thunderstorm impact on the built environment: A Full-scale measurement and post-event survey case study. J Wind Eng. Ind. Aerodyn, 2024.
13. Chay, M.T., Albermani, F., Wilson, B., 2006. Numerical and analytical simulation of downburst wind loads. Eng. Struct. 28, 240–254.
14. Davenport, A., G.: The application of statistical concepts to the wind loading of structures. P. I. Civ. Eng. 19, 449-472., 1961.
15. Fujita, T., T.: Manual of downburst identification for project Nimrod. Satellite and Mesometeorology Research Paper 156, Dept. of Geophysical Sciences, University of Chicago, 104 pp, 1978.
16. Fujita, T., T.:Downburst: Microburst and Macroburst. Univ. Chic. Press II, p. 122, 1985.
17. Glauert, M.B., 1956. The wall jet. J. Fluid Mech. 1, 625–643.
18. Hartigan, J., A.: Clustering Algorithms. Wiley, New York, 1975.
19. Hartigan, J., A. and Wong, M., A.: A K-means clustering algorithm. Applied Statistics 28, 100–108, 1979.
20. Hjelmfelt, M.R., 1988. Structure and life cycle of microburst outflows observed in Colorado. J. Appl. Meteorol. 27, 900–927 (August).
21. Hjelmfelt, M.R.: Microburst and Macroburst: windstorms and blowdowns. In: Johnson, E.A., Miyanishi, K., (Eds.), Plant Disturbance Echology. Academic Press, Amsterdam, pp. 59-102, 2007.
22. Holmes, J.D., Oliver, S.E., 2000. An empirical model of a downburst. Eng. Struct. 22, 1167–1172.
23. Husson, F., Lê, S., & Pagès, J.: Exploratory Multivariate Analysis by Example Using R (2nd edition). CRC Press, 2017.
24. Ivan, M., 1986. A ring-vortex downburst model for flight simulations. J. Aircraft 23 (3), 232–236.

25. Jiang, Y., Cooley, D., Wehner, M., F.: Principal Component Analysis for Extremes and Application to U.S. Precipitation. Journal of Climate, Volume 33, 15, 2020.
26. Kassambara, A.: Practical Guide to Principal Component Methods in R (Multivariate Analysis Book 2) (1st edition). STHDA. ASIN: B0754LHRMV, 2017.
27. Kaufman, L. and Rousseuw, P.: Finding Groups in Data. An Introduction to Cluster Analysis, Wiley & Sons, New York, 1990.
28. Le, T.H., Caracoglia, L., 2017. Computer-based model for the transient dynamics of tall building during digitally simulated Andrews AFB thunderstorm. Comput. Struct. 193, 44–72.
29. Letchford, C.W., Mans, C., Chay, M.T.: Thunderstorms – their importance in wind engineering (a case for the next generation wind tunnel). J. Wind Eng. Ind. Aerodyn., 90, 1415-1433, 2002.
30. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, eds L. M. Le Cam & J. Neyman, 1, pp. 281–297, 1967.
31. Markowski, P., & Richardson, Y.: Mesoscale Meteorology in Midlatitudes. Wiley-Blackwell, 2010. ISBN: 978-0470742136.
32. McCarthy, J., Wilson, J.W., Fujita, T.T., The Joint Airport Weather Studies Project, 1982.Bulletin of the American Meteorological Society 63, 15–22.
33. Oseguera, R.M., Bowles, R.L., 1988. A simple analytic 3-dimensional downburst model based on boundary layer stagnation flow. NASA Tech. Memo. 100632.
34. Parker, M. D., Johnson, R. H., 2004: Structures and dynamics of quasi-2D mesoscale convective systems. J. Atmos. Sci., 61, 545–567.
35. Proctor, F.H., 1987a. The terminal area simulation system - Part I: theoretical formulation. NASA Contractor Report 4046.
36. Proctor, F.H., 1987b. The terminal area simulation system - Part II: verification cases. NASA Contractor Report 4047.
37. Rao, S., J., and Sengupta, A.: Topics in Circular Statistics. World Scientific, 2001.
38. Rao, R.V., Savsani, V.J. and Vakharia, D.P.: Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems, Comput. Aided Des., 43(3), pp. 303–315, 2011.
39. Solari, G.: Wind science and engineering. Springer, Switzerland, 2019.
40. Solari, G.: Emerging issues and new frameworks for wind loading on structures in mixed climates. Wind Struct. 19, 295-320, 2014.
41. Solari, G., Burlando, M., De Gaetano, P., Repetto, M., P.: Characteristics of thunderstorms relevant to the wind loading of structures. Wind Struct. 20, 763-791, 2015.
42. Solari, G., Burlando, M., Repetto, M., P.: Detection, simulation, modelling and loading of thunderstorm outflows to design wind-safer and cost-efficient structures. J. Wind Eng. Ind. Aerodyn., 200, 2020.
43. Vicroy, D.D., 1991. A simple, analytical, axisymmetric microbust model for downdraft estimation. NASA Technical Memorandum 104053.
44. Vicroy, D.D., 1992. Assessment of microburst models for downdraft estimation. J. Aircraft 29, 1043–1048.
45. Ward, J., H., Jr.: Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association, Volume 58, Issue 301, 236-244, 1963.

46. Weisman, M. L.: Bow echoes: A tribute to T.T. Fujita. Bulletin of the American Meteorological Society, 82(1), 97-116, 2001.
47. Xhelaj, A., Burlando, M., Solari, G.: A general-purpose analytical model for reconstructing the thunderstorm outflows of travelling downbursts immersed in ABL flows. J Wind Eng. Ind. Aerodyn. 207 104373, 2020.
48. Xhelaj, A., Burlando, M.: Application of metaheuristic optimization algorithms to evaluate the geometric and kinematic parameters of downbursts. Advances in Engineering Software. Volume 173, November 2022, 103203., 2022.