POLITECNICO DI TORINO

Master's Degree
in Energy Engineering

Master's Thesis

Surrogate Model for Fluid-Dynamic Simulation and Hydrogen Blending Tracking in Natural Gas Distribution Grids.

Supervisors

Prof. Pierluigi Leone

Marco Cavana, PhD

Candidate

Davide Rubinetto

Contents

Abstract

The injection of hydrogen into natural gas distribution networks represents a viable strategy for decarbonizing the European energy sector. However, the variability of the thermophysical properties of the mixture and the fluctuations in demand require advanced tools to monitor gas quality and network integrity. Traditional fluid dynamics solvers based on differential equations, although accurate, involve high computational costs that limit their applicability for operational control.

This thesis proposes an innovative computational framework applied to the real distribution network of the city of Riccione. In the first phase, a steady-state physical solver was validated in MATLAB, able to simulate multicomponent transport and track hydrogen propagation. The validation with respect to the commercial software SAInt confirmed the accuracy of the model, with an $R^2 > 0.98$ for nodal pressures and composition tracking. The model was employed to conduct a parametric analysis of the hydrogen concentration distribution across the network, varying one parameter at a time. The investigation specifically focused on the displacement of zero-flow points and/or mixing points, where hydrogen concentration shows the highest variations. While this approach enabled a detailed understanding of the underlying physical mechanisms, it did not allow for a generalization of the analysis to a fully multivariate framework.

To overcome computational limitations, the research implemented a surrogate model based on Vector-Valued Kernel Regression (VVRR) techniques. The model was trained on a large dataset generated by Monte Carlo simulations, varying critical parameters such as supply pressure, pipe's roughness, consumptions and hydrogen concentration at the sources, to verify its reliability under different operating conditions. For each set of inputs, the physical solver calculated the corresponding state of the network, creating the training and test datasets. The model was designed to accurately predict hydraulic quantities, which were then used to assess gas quality and analyse the nodal concentration of hydrogen in the entire infrastructure. Validation was performed by comparing the predictions with the outputs of the physical solver, quantifying the errors and using graphical tools such as scatter plots.

The process was applied to three main case studies of increasing complexity: a single injection case with 4 variable parameters, a double injection case with 5 variable parameters and an advanced double injection scenario with 30 variable parameters, characterized by partially independent consumption clusters. The surrogate model showed excellent accuracy in predicting fluid dynamic parameters, allowing precise calculation of nodal concentration. Despite increasing complexity, the quality of the hydraulic prediction remained high, with only slight discrepancies in the concentration calculation, attributable to minimal uncertainties in the estimation of flow rates and consumption. The greatest differences occur near the interconnections between adjacent loops, where the flow direction is sometimes predicted incorrectly. However, given the low frequency of these outliers, the results confirm the suitability of the model for operational analysis.

The analysis provides DSOs with a rapid and sufficiently effective predictive tool for managing future multi-component networks, offering new research opportunities in an area that is still relatively unexplored but could play a key role in the energy transition.

Chapter 1

1 Introduction and Energy Context

1.1 Climate Goals and the Strategic Role of Hydrogen

1.1.1 Climate Context and European Goals

The European Union has taken the lead in the global energy transition, setting ambitious targets to transform Europe into the first climate-neutral continent by 2050. This commitment has been formalised through the European Green Deal [1], a package of strategic initiatives that aims to decouple economic growth from resource use. To support this long-term vision, the “Fit for 55” legislative package has been introduced, setting a mandatory intermediate of reducing net greenhouse gas emissions by at least 55% by 2030 compared to 1990 levels.

In this scenario of deep transformation, the role of natural gas is set to change radically. Currently, natural gas still represents a fundamental component of the European energy mix, covering approximately 24% of total primary energy supply [2].

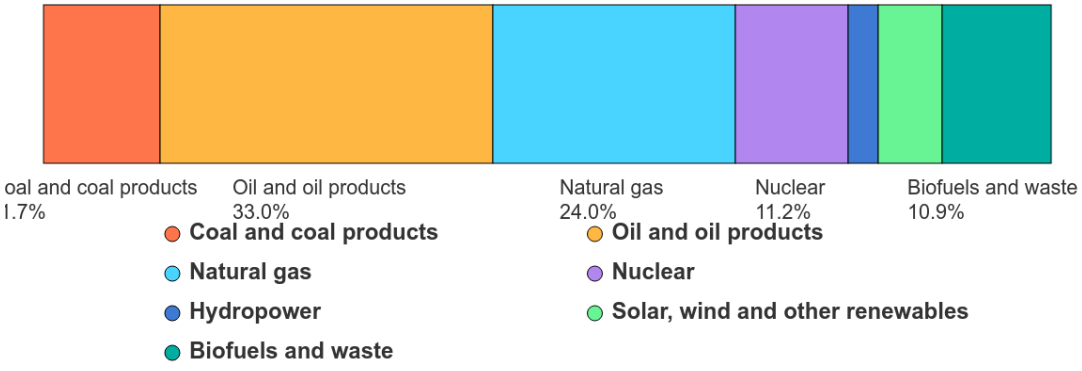


Figure 1.1. Total energy supply by source, Europe, 2023 [2]

However, in order to align with the expected decarbonisation targets, the share of fossil gas will have to decline gradually over the coming decades. This space will gradually be filled by renewable sources and low-carbon energy carriers, including green hydrogen, which are essential for sectors that are difficult to electrify.

The necessity of this transition has been further accelerated by recent geopolitical instability. The turbulence in the global energy market following Russia's invasion of Ukraine and the war in Iran have dramatically highlighted the energy security risks associated with dependence on fossil gas and oil imports, particularly for Europe.

1.1.2 Challenges in Direct Electrification and RES Variability

Although direct electrification is the primary road to decarbonisation, it still faces significant technical and structural obstacles. One of the main challenges concerns grid reliability and resilience.

The massive integration of non-programmable renewable energy sources (RES), which are naturally intermittent, requires major infrastructure upgrades to manage power variability and ensure system stability.

Furthermore, the expansion of electrochemical storage is constrained by the limited availability of critical minerals such as lithium and cobalt, which are essential resources for battery production but subject to geopolitical and supply chain constraints.

These factors highlight how a strategy based only on electrification is not sufficient to cover the entire energy demand in a safe and sustainable manner [3].

1.1.3 Hydrogen as a Strategic Energy Carrier

In this scenario, hydrogen emerges as a versatile and strategic energy carrier, able to overcome the limitations of direct electrification. One of its main advantages is its ability to decarbonise "hard-to-abate" sectors, i.e. those industrial and logistical sectors that require high energy density that are difficult to electrify, such as steelmaking, long-haul heavy transport and maritime transport [3].

Furthermore, hydrogen can play an important role in seasonal energy storage. While batteries are ideal for short-term balancing, hydrogen allows energy to be stored for long periods, transferring it from one season to another. This seasonal storage capacity is essential to compensate for the intermittent nature of renewable sources, ensuring energy security even during periods of low wind or solar production [3].

1.2 Hydrogen Blending in Natural Gas Networks

1.2.1 The Concept of Blending and Grid Repurposing

Blending involves physically injecting hydrogen into the existing natural gas infrastructure, creating a binary mixture often referred to as H2NG. This strategy represents a fundamental step in the transitional phase of decarbonisation, as it reduces the carbon intensity of the gas distributed and, at the same time, stimulates hydrogen production by offering an immediate market outlet without waiting for the development of a dedicated network.

Closely related to blending is the concept of grid repurposing, i.e. the redevelopment and reuse of existing gas transport and distribution infrastructure. Europe has a capillary and well-developed network of gas pipelines; exploiting these assets for the transport of hydrogen offers substantial economic advantages over the construction of new dedicated hydrogen pipelines.

Repurposing makes it possible to drastically reduce capital expenditure (CAPEX) costs and shorten project implementation times, avoiding the bureaucratic and territorial complexities associated with laying new pipelines.

However, the transition from pure methane to hydrogen mixtures is not without technical challenges. The feasibility of repurposing depends strictly on verifying the compatibility of the existing infrastructure: it is necessary to carefully assess the interaction of hydrogen with pipe materials and the adequacy of operating pressures to avoid structural integrity risks and ensure transport safety [4].

1.2.2 Comparative Analysis of Thermophysical Properties of H_2 and NG

The injection of hydrogen modifies the operating conditions of the network due to its marked chemical and physical differences compared to methane. In particular, the reduced density and different volumetric energy content of hydrogen directly impacts transport capacity, pressure management and safety requirements.

The following table compares the fundamental thermophysical properties of the two fluids, highlighting the critical parameters for infrastructure simulation and operation.

Property	Natural Gas (CH_4)	Hydrogen (H_2)	Network Impact
Molar Mass [g/mol]	16.043	2.016	H_2 is 8 times lighter, increasing flow velocity at equal pressure.
Density [kg/m^3]	0.68	0.085	H_2 has 8 times lower density, reducing energy transported per unit volume.
LHV (mass) [MJ/kg]	50	120	H_2 has a much higher energy content per unit mass (approximately 2.5x).
LHV (volume) [MJ/Nm^3]	35.8	10.8	H_2 carries about 1/3 of the energy per unit volume compared to methane.
Compressibility Factor (Z)	0.998	1.0006	H_2 is less compressible and it behaves more like an ideal gas.
Speed of Sound [m/s]	446	1308	H_2 has 3 times higher sound velocity, increasing instability risks.
Flammability Limit [% Vol]	5.0 - 15.0	4.0 – 77.0	The explosive range of H_2 is much wider, increasing safety risks.
Diffusivity Coefficient [cm^2/s]	0.16	0.76	H_2 diffuses approximately 4 times faster, increasing the risk of permeation and leaks.
Laminar Flame Speed [m/s]	0.4	3	Increase of operational risks for domestic and industrial burners.

Table 1.1. Comparison of thermophysical and safety properties between hydrogen (H_2) and methane (CH_4) under standard conditions (15°C and 1 atm) [5–7].

1.2.3 Technical Challenges and Operational Risks of Hydrogen Integration

The introduction of hydrogen into natural gas networks radically changes the thermophysical properties of the mixture and the combustion dynamics, affecting the operational management of the network, the integrity of the infrastructure and the safety of end uses.

From a thermodynamic point of view, hydrogen is characterised by a low volumetric energy density that changes the energy profile of the gas transported. Analyses conducted on H2NG mixtures indicate that the addition of hydrogen reduces both the Higher Heating Value (HHV) and the Lower Heating Value (LHV) of the mixture; for example, a 30% concentration of hydrogen can result in a 20-25% reduction in heating on a volumetric (or molar) basis value [6].

Consequently, to guarantee the same energy supply to users, an increase in storage volume of between 30% and 55% is required, as well as an increase in flow rates, which risks pushing the network beyond standard operating limits (20-25 m/s), triggering erosion and instability.

A critical parameter for gas exchangeability is the Wobbe Index (WI), which tends to decrease with hydrogen injections due to the reduced energy per unit volume of the mixture. However, the most significant impact on end-user safety concerns flame behaviour: hydrogen has a significantly higher laminar flame speed (approximately 3 m/s) than methane (0.4 m/s). This dramatic increase in reactivity poses specific operational risks for domestic and industrial burners, such as flashbacks, requiring careful re-qualification of appliances [4].

In parallel with combustion challenges, hydrogen injection requires rigorous verification of material compatibility. For steel pipes, the main critical issue is hydrogen embrittlement: the small atomic size allows hydrogen to diffuse into the metal's crystal lattice, degrading its ductility and fracture toughness, accelerating crack propagation and increasing susceptibility to fatigue, especially in the presence of pressure fluctuations. In polyethylene distribution networks, on the other hand, the challenges relate to higher permeation rates and the potential for accelerated ageing of the material. Finally, hydrogen compression is thermodynamically more costly, requiring greater compression work and the use of multi-stage systems with intercooling to manage the high discharge temperatures [7,8].

Operating networks at high pressures (above 50 bar) increases the risks to structural integrity. Under these conditions, the fracture toughness of steel can be reduced by up to 30% and permeation through the walls increases significantly, raising the likelihood of failure [8].

To mitigate these risks and enable sustainable transport of H2NG mixtures, integrated corrective strategies must be implemented. In addition to selecting resistant materials (such as austenitic stainless steels) and adopting distributed compression to reduce local stresses, the use of gaseous inhibitors to limit hydrogen adsorption on metal surfaces is being considered [6].

These measures must be accompanied by an adjustment of the operating pressure and flow limits to keep the system within acceptable safety ranges.

1.2.4 Regulatory Limits and Standardization

The injection and blending of hydrogen into existing natural gas networks requires transmission and distribution system operators to obtain specific regulatory authorisations. Currently, the global regulatory landscape is characterised by

considerable heterogeneity and a lack of a specific approach to hydrogen, which in many cases is still simply considered among dangerous gases.

In Europe, the regulatory situation reflects this fragmentation, with blending limits varying dramatically from country to country. Although the European Commission, through its 2020 European Hydrogen Backbone (EHB) initiative [9], recognises blending as a useful transitional solution for reducing emissions in the short term, there is still no harmonised framework among member states. This lack of coordination is critical, especially for the cross-border transport of hydrogen blends between countries with different regulations.

Most European countries do not have an explicitly stated legal limit, which often forces operators to adhere to accepted safety standards that limit hydrogen to very low percentages, typically well below 10% by volume. In other cases, limits exist based on historical operations or specific network restrictions [4].

Below is a summary of the maximum permitted limits (% by volume) in various European countries [10]:

- Germany: There is no general value, it depends on the suitability of applications and infrastructure.
- France: The limit is currently set at 2%.
- Spain: Up to 5% is permitted for unconventional gases.
- Austria: The limit is 10%, but depending on the specific application a further individual assessment may be necessary.
- Netherlands: the current threshold for H_2 is 0,02 % for TSO High Pressure grids and 0,5 % for TSO regional grids and DSO grids.
- Italy: Current legislation (2022) limits the amount of hydrogen that can be blended with natural gas to 2%.

The development of new standards is hindered by several technical and contractual factors. Regulators need more data on technical performance and safety to define new limits. For example, surveys among European regulators have highlighted concerns regarding end-equipment tolerance and industrial customers' sensitivity to variations in gas composition.

Finally, a crucial aspect for standardisation concerns the traceability of injected hydrogen. To address this, the European CertifHy project launched the first Guarantee of Origin (GO) scheme specifically for hydrogen in Europe. This system labels the product's origin, distinguishing between Green and Low Carbon hydrogen. By allowing end-users to certify their sustainable consumption, the CertifHy GOs provide transparency and create a direct market pull for clean hydrogen [11].

1.3 Literature Review: Machine Learning in Gas Network Operation

The integration of machine learning (ML) techniques into the natural gas sector has expanded rapidly over the past decade [12–18]. The current scientific literature identifies four primary application domains, which can be classified according to the underlying energy-related problem: short-term demand forecasting, state estimation, leak detection and operational optimization.

Despite these advances, the application of ML methodologies to hydrogen blending in natural gas networks remains an emerging research area. This field is still characterized by significant open challenges, including limited availability of representative datasets, increased system complexity due to hydrogen's distinct physical properties and the need for models capable of ensuring both predictive accuracy and operational reliability.

1.3.1 ML Applications in Conventional Natural Gas Networks

In traditional methane networks, the use of ML is now well established in several areas:

- Gas Load Forecasting: Consumption forecasting represents the most mature area in gas load applications, with current literature consistently showing that advanced data-driven approaches outperform traditional statistical models. This class of problems can be framed as a classical time-series modelling task, in which gas consumption, measured, for example, by smart meters, is recorded over several months or years with sampling intervals of a few hours (e.g., every 6 hours). Such sampling frequencies yield multiple observations per day and, consequently, a potentially large number of training samples, ranging from several hundreds to tens of thousands. A key characteristic of gas consumption is its pronounced intra-day variability. Demand typically fluctuates throughout the day, for instance, it may decrease during certain morning or evening periods and increase during peak daytime hours. This daily pattern, combined with the substantial amount of available historical data, has encouraged researchers to develop customized machine learning and deep learning architectures, particularly those based on advanced artificial neural networks (ANNs). For example, [18] proposed a block-oriented ANN approach in which gas consumption is modelled as a function of exogenous meteorological variables, including temperature, wind speed, precipitation and solar radiation. The training dataset consisted of six months of gas consumption data, yielding several hundred samples corresponding to the monitoring period. Similarly, [17] introduced an advanced hybrid deep learning architecture specifically designed for sequential and time-series data. The proposed model aims to achieve higher predictive accuracy compared to classical statistical approaches such as ARIMA. Its effectiveness is demonstrated through a comparative evaluation against multiple alternative forecasting techniques. The dataset used in this study includes natural gas pipeline load measurements collected over nearly three years, comprising nearly 25-thousands observations. The sampling interval is one hour. The data are partitioned into training, validation and testing

subsets, with 80% used for training, 10% for validation and the remaining 10% reserved for out-of-sample testing, following standard practice for this class of applications.

- State Estimation, optimization and multi energy scenario: In the context of hydraulic state estimation and optimization, particularly for pressures and flow rates, research is progressively moving away from computationally intensive physical solvers based on non-convex differential equations. Instead, increasing attention is being devoted to machine learning–based surrogate models, predominantly built upon neural network architectures. Recent studies have proposed ML-driven approaches for a variety of applications, including the estimation of fluid-dynamic variables and the solution of complex optimization problems arising in multi-energy systems. A crucial challenge in estimating the hydraulic state is managing noise in measurement data. In this context, a recent study [12] introduced an innovative data-driven model used to learn the direct mapping between historical network measurements (typically affected by instrumental errors) and the entire fluid dynamic state of the system (nodal pressures and mass flows). Instead of filtering the data in advance and risking losing real information, the model estimates the actual fluid dynamic relationships of the system network, converting them into a simple approximation problem while keeping all the essential data. This approach has shown better stability and accuracy than traditional numerical methods by bypassing the complex direct solution of differential equations. In addition, surrogate models are being integrated into optimization frameworks to reduce computational burden while preserving adequate accuracy. For instance, [16] presents a neural-network-constrained optimization method aimed at supporting the integration of renewable energy sources into energy systems, demonstrating how data-driven models can effectively replace or augment traditional physics-based solvers in large-scale optimization tasks.

1.3.2 Emerging Trends in Hydrogen-Blended Networks

With the introduction of hydrogen, the properties of the fluid (density, viscosity) change, making it necessary to upgrade the analysis instruments. Current literature on blending focuses mainly on two aspects:

- Safety & Leak Detection: They constitute the most extensively investigated research area, mainly due to the critical risks associated with hydrogen's physical properties, such as its wide flammability range, low ignition energy and high diffusivity. These characteristics significantly increase the complexity of monitoring and risk mitigation in hydrogen-blended gas networks. Recent studies propose advanced hybrid methodologies tailored to specific infrastructure contexts. For urban service tunnels, one approach combines computational fluid dynamics simulations with a multi-task learning neural network based on Long Short-Term Memory units. This integrated framework enables the simultaneous prediction of both the location and the intensity of a leak. The neural network architecture is specifically designed for the analysis of sequential and time-series data, allowing it to capture the dynamic evolution of gas dispersion over time [15]. For long-distance transmission pipelines,

alternative solutions integrate real-time transient flow models with artificial intelligence, based leak detection systems. In these cases, the detection module combines convolutional neural networks with neuro-fuzzy inference systems, improving pattern recognition performance while maintaining a degree of interpretability and robustness under uncertain operating conditions [14].

- Blending Optimization: Hydrogen injection optimization is primarily addressed through mixed-integer nonlinear programming models, which are used for the design and operational planning of hydrogen-enriched natural gas networks. These optimization frameworks aim to identify cost-effective infrastructure configurations while ensuring compliance with technical and regulatory constraints. When applied to real-world case studies, such as the Ontario gas distribution network [13], these models enable the determination of the optimal system topology, including the placement of power-to-gas facilities and hydrogen mixing stations. The objective is typically to minimize overall investment and operational costs while satisfying hydrogen concentration limits imposed for safety and compatibility reasons, which generally range between 5% and 15% by volume.

1.3.3 Research Gaps and Contribution

Despite the progress outlined above, the literature reveals a significant gap in gas quality monitoring, a topic expected to play a central role in future machine learning applications for hydrogen-blended networks. Most existing studies rely on detailed hydraulic simulators, which are computationally intensive and lack surrogate models capable of linking hydraulic variables directly to gas composition parameters. Moreover, no clear agreements has yet emerged regarding the most suitable machine learning approach for this task.

To address this gap, this work proposes a Kernel-Based Vector-Valued Regularized Regression framework as an alternative to deep neural networks, which typically require large training datasets. The proposed method operates with fewer samples, ensures rapid training and guaranteed convergence through convex optimization and enables the construction of a robust and efficient surrogate model. The resulting model supports gas quality tracking in distribution networks at negligible computational cost compared to physics-based solvers, providing a concrete foundation for further research and methodological advancements.

1.4 Limitations of Physics-Based Modelling for Complex Networks

The main limitation of physics-based modelling is the high computational cost required to solve the systems of non-linear differential equations that govern gas flow. Although classical numerical solvers guarantee a high degree of physical fidelity by respecting conservation laws, their application on complex networks quickly becomes impractical in terms of execution time.

This critical point was highlighted by Yin X et al. (2022) [19], in their study on surrogate models for rapid flow control. The authors point out that traditional numerical methods for solving fluid dynamics equations are intrinsically 'computationally intensive and time-consuming'. This structural slowness makes these models unsuitable or even inapplicable for scenarios that require real-time responses or that require a very high number of iterations, such as the analysis of thousands of operating scenarios in capillary distribution networks.

In the context of DSO (Distribution System Operators) networks, characterised by a high number of nodes and branches, the exclusive use of physics-based models creates a computational bottleneck that prevents dynamic and reactive network management, driving research towards alternative solutions capable of drastically reducing calculation times while maintaining acceptable accuracy.

This computational challenge is compounded by the growing complexity of gas quality management, a critical issue formally recognized by the UNI EN 16726 standard [20]. The document highlights how networks receive gas from multiple entry points and heterogeneous sources. The increase in biomethane and hydrogen injections, which have a significantly lower density than methane, will introduce further local variations. These factors can cause sudden fluctuations in the Wobbe Index and other properties of the distributed gas.

The standard emphasises that a high rate of variation in these parameters can compromise the operation of critical end applications, such as gas engines, turbines and industrial processes. Almost instantaneous changes in gas quality can occur for several reasons, including the presence of different mixtures along the main flow direction or the occurrence of bidirectional flows of gas with different qualities. In the latter case, "zero flow points" can be created along the pipeline, where a rapid transition between two different gas qualities occurs. The position of this point varies continuously depending on consumption. Similarly, if gas from two different sources is subsequently injected at the same inlet point, there is almost no mixing within the pipeline. As a result, the maximum variation in gas quality occurs at the exact moment when the switch from one source to another takes place.

Currently, as highlighted by the norm, the prediction, monitoring and control of this rate of change do not represent an established technical standard for every application and largely escape the direct control of network operators. This observation further reinforces the need to develop rapid and efficient predictive models capable of overcoming the computational limitations of traditional solvers in order to map and manage sudden changes in quality in increasingly complex operating scenarios.

1.5 Proposed Methodology: Surrogate Modelling for Network Analysis

Considering the computational limitations and operational challenges discussed in the previous sections, the main objective of this thesis is the development and validation of a surrogate model based on Machine Learning (ML) techniques. This approach aims to

overcome the rigidity of traditional physical simulators by offering a predictive tool capable of assessing the impact of hydrogen blending in distribution networks with an accuracy comparable to that of fluid dynamics codes, but with drastically reduced computation times.

In this context, the surrogate model acts as a functional approximation of the "high-fidelity" physical model. Instead of iteratively solving the complex system of non-linear differential equations for each new operating scenario, the ML model learns the mathematical relationship between the inputs (e.g. network topology, demand profiles, etc.) and the outputs of interest (e.g. node pressures, flow rates, fluid velocities). Once trained, the surrogate can make near-instantaneous predictions, enabling multi-scenario analysis and real-time optimisations that would otherwise be impractical.

The proposed methodology consists of three fundamental phases:

- Dataset generation: Use of a stationary fluid dynamics calculation code (Physics-Based) to simulate a large number of operating scenarios, varying critical parameters such as user load and injected hydrogen concentration. The results of these simulations will constitute the training database.
- ML model development: Design and training of a machine learning algorithm able to capture the non-linearities present in the fluid dynamics equations and to make predictions about the output.
- Validation on Case Studies: Application of the surrogate model to specific case studies representative of real distribution networks (DSOs). The effectiveness of the methodology will be verified by comparing the predictions of the surrogate model with the results of the physical model.

Through this approach, the thesis aims to demonstrate how the integration of physical modelling and artificial intelligence can provide network operators with an effective tool for planning the transition to hydrogen, managing the uncertainty and complexity of existing infrastructure.

Chapter 2

2 Physics-Based Modelling and Steady-State Simulation

2.1 Topological Characterization of the Network

Gas infrastructure, like the electricity grid and many other civil infrastructures, has an intrinsic network structure, consisting of a set of physically interconnected elements. In this context, graph theory is a powerful mathematical tool for the analysis, simulation and optimization of such systems. In fact, every physical network can be described as a directed graph. In mathematical terms, a graph is defined as a pair $\mathcal{G} = (\mathcal{V},\mathscr{E})$, where: $\mathcal{V}$ is a set of elements called nodes; $\mathscr{E}$ is a set of pairs of nodes called directed branches.

Transferring these definitions to the physical representation of the network, a directed branch represents any element of the infrastructure that connects an inlet node to an outlet node in a defined flow direction (as a pipeline). Therefore, nodes represent the points of interconnection between two or more adjacent branches, except for branches that originate (or terminate) in a peripheral node. To define the problem, it is necessary to collect the topological and physical data required to model the elements of $\mathcal{G}$. Specifically:

- Nodes: They are identified by geographical coordinates (latitude and longitude) and by their altitude (z), a fundamental parameter for hydraulic analysis.
- Pipelines: Each branch is characterized by its geometric and physical properties, in particular length (L), internal diameter (D) and absolute roughness (ε).

This set of data forms the base for the topological representation of the network, which must be translated into algebraic form to enable numerical simulation. This step is performed by calculating the Incidence Matrix A. This matrix mathematically defines the connectivity between nodes and pipelines, acting as a link between the network structure and the mathematical model that governs gas flow behaviour.

The incidence matrix A is obtained as:

$$A = A^{+} + A^{-} \tag{2.1}$$

The structure of incidence matrix A can be described by distinguishing between inflows and outflows:

- A^{+}: Identifies the inflow connections. It is defined as a binary matrix in which each element $a_{i,j}^{+}$ has the following values:

$$a_{i,j}^{+} = \begin{cases} 1 & if\ arc\ j\ \ terminates\ at\ node\ i, \\ 0 & otherwise. \end{cases} \tag{2.2}$$

- A^{-}: Identify outflow connections. Similarly, each element $a_{i,j}^{-}$ is defined as:

$$a_{i,j}^{-} = \begin{cases} -1 & if\ arc\ j\ \ originates\ from\ node\ i, \\ 0 & otherwise. \end{cases} \tag{2.3}$$

The resulting matrix A (given by the sum of the input and output components) has dimensions $n \times b$, where:

- *n:* total number of nodes in the network.
- *b:* total number of pipes in the network.

This representation allows for an accurate description of the topological connections of the network, making the calculation of physical quantities easier (such as mass conservation at the nodes).

2.2 Fluid Properties and Equations of State

For a simulation of network behaviour that is as close to reality as possible, the fluid cannot be considered as an ideal.

The model implemented considers the gas as a real fluid, whose thermodynamic properties vary according to pressure p, temperature T and chemical composition.

The standard reference conditions (STD Conditions) are defined in the model as:

- Standard pressure: p_{std} = 101325 bar.
- Standard temperature: T_{std} = 288.15 K (15°C).

The gas density ρ at each point in the network is calculated using the equation of state for real gases:

$$\rho = \frac{p}{Z(p,T,y) \cdot R_{mix} \cdot T} \tag{2.4}$$

Where:

- Z is the compressibility factor, calculated using advanced equations of state (the study refers to the GERG-2008 [21]) as function of the molar composition y.
- R_{mix} (or R_b) is the specific gas constant (J/kg K), derived from the universal gas constant UGC and the molar mass of the mixture M_{mix}.

$$R_{mix} = \frac{UGC}{M_{mix}}\ , \qquad M_{mix} = \sum_i y_i \cdot M_i \tag{2.5}$$

Where:

- UGC: Universal Gas Constant, equal to $8.314\ \mathrm{J/(mol \cdot K)}$.
- M_{mix}: Mean molecular weight of the gas [kg/mol].
- y_i: Molar fraction of component i.
- M_i:Molecular weight of component i [kg/mol].

The algorithm uses an iterative update of these properties: the average pressure p_{avg} along each pipe is estimated using the approximation formula for parabolic profiles:

$$p_{avg} = \frac{2}{3} \cdot \left(p_{in} + p_{out} - \frac{p_{in} \cdot p_{out}}{p_{in} + p_{out}} \right) \tag{2.6}$$

This value is used to recalculate dynamic viscosity, friction factor λ and compressibility factor Z.

For subsequent calculations, for determining density and for the accumulation terms in the continuity equation, it is necessary to define the parameter cc2n (or cc2b). This term represents the square of the isothermal sound velocity and it corrects the relationship between pressure and density based on the actual properties of the gas.

The relationship used is:

$$cc2n = Z(p, T, y) \cdot R_b \cdot T \tag{2.7}$$

As a result, the gas density is calculated in the model as:

$$\rho = \frac{p}{cc2n} \tag{2.8}$$

2.3 Governing Equations and Numerical Resolution

The behaviour of gas within the network is governed by two fundamental equations, which describe respectively the dynamics of the fluid along the branches and the conservation of mass at the nodes:

- Momentum Equation: It describes the balance of forces along a pipe, considering the effects of viscous friction and changes in potential energy due to differences in altitude (gravitational terms).
- Continuity Equation: It ensures mass conservation at every node in the network.

The combination of these equations, with the appropriate boundary conditions, allows the determination of the unknown state variables of the system:

- p: Nodal pressures (p_k).
- G: Mass flow rates along the pipelines (G_k).

However, the non-linear nature of these equations makes a direct analytical solution impossible. It is necessary to adopt an iterative numerical approach: starting from an initial guess, the solution is progressively refined until convergence is achieved, i.e. when the difference between two successive iterations falls below a predefined tolerance ($\approx 10^{-6}$ - 10^{-4}).

2.3.1 Momentum Equation

To model the behaviour of gas inside pipes, the solver implements the momentum conservation equation derived from fluid mechanics for compressible flows.

In general form, the equation balances the forces of pressure, inertia, friction and gravity. However, for steady-state simulation, the time terms are neglected (or set to zero), simplifying the starting equation to the balance between pressure gradient, friction losses and potential energy variation.

The original differential equation considered is:

$$\frac{dp}{dx} + \frac{\lambda G|G|}{2DA^2\rho} + \rho g sin(\alpha) = 0 \tag{2.9}$$

Where:

- G: Mass flow rate.
- λ: Friction factor (Darcy friction factor).
- ρ: Gas density.
- α: Pipe angle of inclination.

Compared to models for incompressible fluids that use a linear gravitational term, in the case of compressible fluids it is necessary to implement an analytical formulation that considers the compressibility of the gas along the height difference.

The relationship between the pressures at inlet (in) and outlet (out) nodes and the viscous pressure drop is expressed as:

$$p_{in} - p_{out} \cdot e^{\frac{s}{2}} = \Delta p_{friction} \tag{2.10}$$

The exponential term $e^{\frac{s}{2}}$ adjusts the downstream node pressure to account for the weight of the gas column. The parameter s is calculated as:

$$s = \frac{2g \cdot \Delta H}{cc2n} \tag{2.11}$$

Where:

- $\Delta H = z_{out} - z_{in}$ is the elevation difference.
- cc2n is the square of the isothermal sound velocity, defined as $Z \cdot R_b \cdot T$.

The friction losses are modelled according to the Darcy-Weisbach equation. In the solver, the fluid dynamic resistance Rf is calculated at each iteration k using the following formulation:

$$R_f = \frac{16 \cdot \lambda \cdot cc2n \cdot \Delta x}{\pi^2 \cdot D^5 \cdot (p_{in} + p_{out})} \tag{2.12}$$

To solve the system on the entire network scale, the pressure gradient is built using the ADP (Pressure Gradient Matrix). In the solver, this matrix is constructed by combining the network topology with exponential term of altitude:

$$ADP = (A^+ - A_s^-)^T \tag{2.13}$$

Where:

- A^+: Positive incident matrix (1 at the input nodes, 0 otherwise).
- A_s^-: Exponentially weighted negative incident matrix, defined as:

$$A_s^- = A^- \cdot e^{\frac{s}{2}} \tag{2.14}$$

Consequently, the momentum equation for the entire network for each iteration k in linearized matrix form, becomes:

$$ADP \cdot p_k = R_f \cdot |G_k| \cdot G_k \tag{2.15}$$

Although the general formulation of the momentum conservation equation includes an inertial term (R_i), which is necessary to describe fluid dynamic transients and resistance to temporal variations in flow rate, this contribution is zero in steady-state analyses. Under steady-state conditions, the state variables of the system are assumed to be constant over time. Consequently, the equation (2.15) is simplified to a static equilibrium balance between pressure forces, viscous friction forces and gravitational forces, neglecting the dynamic effects related to the inertia of the fluid mass.

2.3.2 Friction Factor

The calculation of the friction factor λ is a fundamental component in the physics-based modelling of natural gas transport networks. In the developed solver, the traditional approach based on separate formulas for the various regimes (Hagen-Poiseuille for laminar flow and Colebrook-White [22] for turbulent flow) has been replaced by a single explicit correlation.

This approach allows the calculation of λ by means of a single continuous equation valid for the entire Reynolds number domain, removing numerical discontinuities in the transition zone that often cause instability in iterative algorithms.

In order to calculate the friction factor, the model determines the flow regime in each pipe by evaluating the Reynolds number (Re). This dimensionless parameter represents the ratio between the inertial forces and the viscous forces within the fluid [23]. The relationship used in the code for calculating Re is expressed as:

$$Re = \frac{G \cdot D}{A \cdot \mu} \tag{2.16}$$

Where:

- G: Mass flow rate of gas [kg/s].
- A: Cross-sectional area of the pipe [m^2].
- D: Internal diameter of the pipe [m].
- μ: Dynamic viscosity of gas [Pa s]. The model assumes a fixed value of $\mu = 1.0 \cdot 10^{-5}\ Pa \cdot s$.

This equation allows the classification of flow into three different regimes, defined in the code by the following numerical thresholds:

- Laminar flow: Re < 2300.
- Transition regime: $2300 \leq Re \leq 4000$.
- Turbulent flow: Re > 4000.

The implemented function uses a mathematical structure based on two coefficients, a and b, which regulate the transition between different physical laws as a function of Re (laminar vs turbulent); this equation is called Cheng's correlation [24].

The coefficients are calculated as follows:

- Laminar flow coefficient (a):

$$a = \frac{1}{1 + \left(\frac{Re}{2720}\right)^9} \tag{2.17}$$

  This term converges to 1 for low Re (laminar domain) and drops rapidly to 0 for Re > 2720.

- Transition coefficient (b):

$$b = \frac{1}{1 + \left(\frac{Re}{160 \cdot D/\varepsilon}\right)^2} \tag{2.18}$$

  Where ε is the absolute roughness of the pipe [m]. This term controls the influence of relative roughness in the transition to turbulent flow.

The final expression for the friction factor λ combines these terms into a single inverse power law:

$$\lambda = \left[\left(\frac{Re}{64}\right)^{a} \cdot \left(1.8 \cdot log_{10}\frac{Re}{6.8}\right)^{2(1-a)b} \cdot \left(2.0 \cdot log_{10}\frac{3.7 \cdot D}{\varepsilon}\right)^{2(1-a)(1-b)} \right]^{-1} \quad (2.19)$$

This formulation has significant numerical advantages:

- For $Re \to 0$ (Laminar flow): Since $a \to 1$, The equation is automatically reduced to $\lambda = 64/Re$ (Hagen-Poiseuille's law).
- For $Re \to \infty$ (Turbulent flow): Since $a \to 0$, the equation converges to the Colebrook-White correlation for rough pipes.
- Transition zone: Instead of defining a rigid interval where manual interpolation is forced (e.g. between 2300 and 4000), this formula controls the flow transition in a natural way. The coefficients a and b act as variable 'weights': as the Reynolds number increases, the equation gradually turns off the laminar contribution and turns on the turbulent one. This creates a perfectly smooth curve, without sudden jumps, avoiding issues when the flow is exactly between the two regimes.

The calculated friction factor λ is then used to determine the fluid dynamic resistance Rf in the Momentum Equation (Eq. 2.12), affecting the calculation of pressure drops along the network.

2.3.3 Continuity Equation (Mass Conservation)

The continuity equation ensures the conservation of mass within the network. It requires that for each control volume (node), the sum of incoming and outgoing flows is balanced by any accumulation of mass over time. The general formulation for a generic node i is expressed by the following equation:

$$\sum_{in} G_{in} - \sum_{out} G_{out} + L_{ext} = \frac{\partial m}{\partial t} \quad (2.20)$$

Where:

- $G_{in/out}$: Mass flow rates entering and exiting the pipes connected to the node.
- L_{ext}: Mass flow exchanged with the outside (source or user).
- $\frac{\partial m}{\partial t}$: Time variation of the mass of gas stored in the node (accumulation term).

Using the incidence matrix A defined in section 2.1, which topologically maps the connections between arcs and nodes, the equation can be rewritten in compact matrix form:

$$A \cdot G + L = \frac{dm}{dt} \quad (2.21)$$

Where G is the vector of flow rates in the branches and L is the vector of flows at the nodes.

To take into account the compressibility of the gas, the accumulation term must be expressed as a function of the state variable, i.e. the nodal pressure p. The mass m in a node of volume V is given by $m = \rho \cdot V$

Since density ρ depends on pressure according to the equation of state ($\rho = p/cc2n$), the change in mass can be written as:

$$\frac{\partial m}{\partial t} = \frac{V}{cc2n}\frac{\partial \rho}{\partial t} \tag{2.22}$$

By using a discrete approach in time with step Δt, the time derivative of pressure is approximated as:

$$\frac{\partial p}{\partial t} \approx \frac{p_k - p_n}{\Delta t} \tag{2.23}$$

Where p_k is the pressure at the current iteration of the module solving the linearized pressure equation loop and p_n is the pressure at the previous timestep. In the numerical model, this term is handled by introducing the Accumulation Matrix Φ, defined as a diagonal matrix whose elements are:

$$\Phi = \mathrm{diag}\left(\frac{V}{cc2n \cdot \Delta t}\right) \tag{2.24}$$

Consequently, the linearized accumulation term becomes:

$$\frac{dm}{dt} \approx \Phi \cdot p_k - \Phi \cdot \mathrm{p_n} \tag{2.25}$$

By combining the topological equation (2.21) with the linearized accumulation term (2.25) and introducing the identity matrix II to manage external flows, we obtain the final form of the Continuity Equation used in the solver.:

$$\Phi \cdot \mathrm{p_k} + A \cdot G_k + \mathrm{II} \cdot \mathrm{L_k} = \Phi \cdot \mathrm{p_n} \tag{2.26}$$

This equation represents the complete mass balance implemented in the code.

It is important to observe that, in the specific case of pure steady-state simulations, the time-related terms are negligible. In this configuration, the matrix Φ is zero and the equation is reduced to a static flow balance:

$$A \cdot G_k + \mathrm{II} \cdot \mathrm{L_k} = 0 \tag{2.27}$$

However, the general structure (2.26) is kept in the code to allow the solver to be generalized in a transient form.

2.4 Boundary Conditions

In order to find a single solution to the system of differential equations described in the previous sections, it is necessary to define an appropriate set of boundary conditions. These conditions constrain the degrees of freedom of the system, reflecting the real operating conditions of the network.

In the model developed, there are two types of boundary conditions:

- Set pressure nodes. The pressure value for the generic i-th node is set to a known p_{set}. This condition must be assigned to at least one node in the network (typically the supply nodes) to act as a reference for the entire hydraulic system.

$$p_i = p_{set} \tag{2.28}$$

- Set flow nodes. The value of the mass flow exchanged with the outside L is known. This condition is applied to most nodes and represents the delivery points to customers (users).

$$L_i = G_{ext} \tag{2.29}$$

In the solver, these conditions are not handled separately but are integrated directly into the global matrix system through specific constraint equations.

The model uses three auxiliary matrix structures to define these constraints:

- OOnn: Coefficient matrix for pressure variables.
- III: Matrix for external flow variables.
- TN_L: Known Terms Vector.

The generic constraint equation for the i-th node can be written as a linear equation:

$$\mathrm{OOnn}_{i,i} \cdot p_i + \mathrm{III}_{\mathrm{i,i}} \cdot L_i = TN_{L_i} \tag{2.30}$$

The configuration of the matrices determines the type of condition applied:

- For a set pressure node, the diagonal element of OOnn is set to 1 and the corresponding element of III is cleared. The known term TN_{L_i} assumes the value of the set-point pressure.

$$1 \cdot p_i + 0 \cdot L_i = p_{set} \tag{2.31}$$

- For a set flow node, the diagonal element of OOnn is set to 0 and the element of III is kept at 1. The known term TN_{L_i} assumes the value of the imposed withdrawal.

$$0 \cdot p_i + 1 \cdot L_i = G_{cons} \tag{2.32}$$

This formulation allows to change the type of control on each node (from pressure control to flow control) simply by updating the binary matrices, without modifying the structure of the main solver.

2.5 Global System Assembly and Solution

To solve the fluid dynamics problem for the entire network, the local equations derived in the previous sections (Momentum and Continuity) are assembled with the boundary conditions into a single global linear system. This approach allows all unknown variables (pressures and flow rates) to be solved simultaneously.

2.5.1 Matrix Structure

The resulting linear system, solved at each iteration k of the algorithm, takes the following block matrix structure:

$$\begin{bmatrix} \Phi & A & II \\ ADP & -R_k & 0 \\ OOnn & OOnb & III \end{bmatrix} \cdot \begin{bmatrix} p_k \\ G_k \\ L_k \end{bmatrix} = \begin{bmatrix} TN_p \\ TN_M \\ TN_L \end{bmatrix} \tag{2.33}$$

The sub-matrices that make up the system are defined as follows:

- Φ: Diagonal matrix containing the accumulation terms at the nodes (related to volume and compressibility), as defined in Eq. 2.24.
- A: Topological incidence matrix mapping branch capacities on nodal balances.
- ADP: Generalized pressure gradient matrix, including exponential correction factors for altitude (Eq. 2.13).
- R_k: Diagonal matrix of the linearized total resistance. In the code, this term is calculated as:

$$R_k = diag(2 \cdot R_f \cdot |G_{k-1}| + R_i) \tag{2.34}$$

  The factor 2 derives from the linearisation of the quadratic term (G^2) using Taylor series, while R_i is the inertial resistance (zero in steady state).
- II, III: Identity matrices used to map external flows in the mass balance and boundary conditions.
- OOnn, OOnb: Matrices used for imposing boundary conditions on pressures and flows.

The vector of unknowns X_k contains:

- p_k: Nodal pressure at the current iteration.
- G_k: Mass flow rate in branches.
- L_k: Flows exchanged with the outside (sources or loads).

The vector of known terms is composed of:

- TN_p Known terms of the continuity equation (related to accumulation in the previous iteration, $\Phi \cdot p_n$).
- TN_M: Known terms of the momentum equation, which balance the linearization of resistance.
- TN_L: Numerical values of boundary conditions (set-point pressures p_{set} or supply flow ratesG_{ext}).

2.5.2 Boundary Conditions Implementation

The boundary conditions are implemented by modifying the rows of matrices OOnn and III and vector TN_L.

- Set pressure nodes, i.e. for supply nodes where the pressure is set at p_{set}:
  - The OOnn matrix assumes a value of 1 on the corresponding diagonal (eye)
  - The III matrix is reset to zero to decouple the imposed flow (which becomes a calculated unknown).
  - The known term TN_L is set equal to p_{set}.
- Set flow nodes, i.e. for consumption where the flow G_{ext} is known:
  - The OOnn matrix is reset to zero.
  - The III matrix assumes a value of 1 on the diagonal.
  - The known term TN_L is set equal to G_{ext} (positive for inputs, negative for consumption, zero for transit nodes).

2.5.3 Iterative Solution Algorithm

Since the system depends non-linearly on unknown variables (resistance R_k depends on G and gas properties ρ, Z depend on p), the solution is obtained through an iterative cycle (Inner Loop). Furthermore, for networks that transport variable gas mixtures, there is an outer cycle (Outer Loop) that updates the chemical composition at the nodes.

The algorithm implemented for fluid dynamics resolution follows these steps:

- Initialization: Assignment of initial guess values for pressures and flow rates.
- Property Update: Calculation of physical properties (friction factor λ, compressibility factor Z, density ρ) and updating of matrices R_k and ADP.
- Matrix Assembly & Solving: Construction of the global system and resolution by matrix inversion (operator \ in MATLAB):

$$X_k = matrix_k^{-1} \cdot TN_k \tag{2.35}$$

- Under-relaxation: The variables are updated using under-relaxation factors to stabilize the solution.:

$$p_{new} = \alpha_P \cdot p_{old} + (1 - \alpha_P) \cdot p_{calc} \tag{2.36}$$

The convergence control implemented in the code is based on the Relative Percentage Error. At each iteration k, the error is calculated for each individual node (for pressure) and for each pipe (for flow rate). The percentage error for pressure ε_P and flow rate ε_G is defined as:

$$\varepsilon^{(k)} = \frac{\left|X^{(k)} - X^{(k-1)}\right|}{|X^{(k-1)}| + \delta} \cdot 100 \tag{2.37}$$

Where:

- $X^{(k)}$ is the value of the variable (pressure or flow rate) at the current iteration.
- $X^{(k-1)}$ is the value at the previous iteration.
- $\delta = 10^{-12}$ is a numerical safety factor added to the denominator. This term is essential for avoiding division by zero errors (NaN) in network sections where the flow rate tends towards zero.

The stop condition verifies that the maximum error across the entire network is less than the set tolerance:

$$max(max(\varepsilon_P), max(\varepsilon_G) < toll \tag{2.38}$$

In the model, tolerance is set to $toll = 10^{-6}$ (i.e. a relative error of 0.0001%).

Furthermore, the algorithm includes a consistency check: if in any iteration the unknown variables assume complex values (if ~ isreal), the residual is forced to a high value of 10^6 to indicate non-convergence and prevent non-physical results.

2.6 Gas Quality Tracking and Mixing Model

The case studies analysed in this paper do not focus on the transport of natural gas alone but examine scenarios in which methane (CH_4) and hydrogen (H_2) mixture are transported through the infrastructure. The introduction of hydrogen significantly changes the thermophysical properties of the mixture: fundamental parameters such as the compressibility factor Z, density ρ and sound velocity c^2 depend on the local concentration of H_2, which can vary from node to node depending on the injection

points and mixing flows. Consequently, the fluid dynamics model cannot operate with a fixed composition but must be coupled with a Gas Quality Tracking Model. From a computational point of view, this coupling is achieved through a double iterative cycle architecture:

- An Outer Loop, which solves the chemical species balance and updates the nodal composition.
- An Inner Loop, described in sections 2.3–2.5, which solves the pressure and flow equations using the physical properties updated by the outer loop.

The algorithm solves the problem of chemical species propagation based on the Perfect Mixing Assumption: it is assumed that the flows entering a node mix instantly and homogeneously before continuing towards the downstream pipes.

2.6.1 Dynamic Topology Update

Unlike pressure, which propagates everywhere, the quality of gas changes only when it physically moves: its composition travels only when carried by the flow rate. However, in a meshed network, the direction of flow in each pipe is not known and may be reversed during the iterations of the fluid dynamics solver. To correctly manage mass transport, the code implements a dynamic update of the topology at each iteration of the composition cycle. A "correct" Incidence Matrix (in the model Ai) is defined, which reflects the actual physical direction of the gas:

- If the flow rate G_k on a branch is negative (flow opposite to the initial topological convention), the input and output nodes are logically inverted.:

$$if\ G_k < 0 \rightarrow Inlet_{new} = Outlet_{old}\ and\ Outlet_{new} = Inlet_{old} \tag{2.39}$$

- The matrices A^+ e A^- are recalculated to ensure that A^+ always identifies the node from which the flow physically exits and enters the pipeline and A^- the node where the flow arrives.

This operation ensures that mass transport calculations are always consistent with the flow velocity ($G_{kfi} = |G_{kf}|$).

2.6.2 Species Mass Balance (Mixing Equation)

The composition is calculated by solving the mass balance for each individual chemical species 'c' (e.g. methane, hydrogen) at each node i.

Since mass is conserved, the balance is formulated in terms of mass fractions ($x_{i,c}$):

$$x_{i,c}^{(k+1)} = \frac{\sum_{j \in pipes_{in}} \dot{m}_{j \rightarrow i} \cdot x_{j,c} + \dot{m}_{ext,i} \cdot x_{ext,c}}{\sum_{m \in pipes_{out}} \dot{m}_{i \rightarrow m} + \dot{m}_{load,i}} \tag{2.40}$$

Where:

- $x_{i,c}^{(k+1)}$ is the mass fraction of species c at the considered node i (at the current iteration).
- $\dot{m}_{j\to i}$ is the mass flow rate entering node i from an upstream node j.
- $x_{j,c}$ is the mass fraction of species c in the flow coming from the upstream node j.
- $\dot{m}_{ext,i}$ is the mass flow rate of external injections entering node i.
- $x_{ext,c}$ is the mass fraction of species c in the external injection.
- $\dot{m}_{i\to m}$ is the mass flow rate leaving node i towards a downstream node m.
- $\dot{m}_{load,i}$ is the mass flow rate delivered to a user (load) leaving the network at node i.

In the model, this system is solved in compact matrix form for all species simultaneously:

$$XX_c^{(k+1)} = \frac{B^- \cdot XX_c^{(k)} + I^- \cdot XX_{ext}}{\sum flow} \tag{2.41}$$

Where:

- XX_c: Mass concentration matrix (Nodes x Components).
- B^-: Advection matrix that transports the composition from upstream nodes to downstream nodes, weighted by the flow rates of the pipelines.
- I^- : Diagonal matrix that regulates mass injections from external sources.

2.6.3 Mass-to-Molar Conversion and Convergence

Although the mixing balance is only rigorous for mass fractions, the equations of state (GERG-2008) require the molar composition ($y_{i,c}$) as input.

Therefore, the algorithm includes a conversion step at each iteration:

- Mass to Molar Conversion:

$$y_{i,c} = \frac{x_{i,c}/M_c}{\sum_k (x_{i,k}/M_k)} \tag{2.42}$$

  In the model, this is handled vectorially using the molar masses of the individual species to calculate the average molar mass of the mixture at the node.

- Convergence Check: The composition cycle ends when the variation in concentrations between two successive iterations is less than the tolerance ($toll_C$):

$$\max\left(\frac{XX_c^{(k+1)} - XX_c^{(k)}}{XX_c^{(k)} + \delta}\right) < \varepsilon \tag{2.43}$$

In this case as well, a check is performed on the real nature of the numbers (isreal) to ensure the stability of the chemical solution.

At the end of the process, the new molar composition YY_c is used to update the physical properties of the gas (density, viscosity, sound velocity) for the next step of the hydraulic calculation, closing the coupling between the fluid dynamics and chemistry of the problem.

Chapter 3

3 Case Study – The Riccione Gas Network

This chapter is dedicated to the modelling and hydraulic analysis of the gas distribution network of the Municipality of Riccione (INRETE). In addition to the topological characterization of the infrastructure, numerical simulations will be shown for nominal cases, both in the current natural gas configuration and in a transition scenario with hydrogen blending.

These nominal scenarios will serve as a reference for subsequent analyses, in which the network parameters will be significantly varied through the training of a surrogate ML model.

3.1 Geographical Context and Demographics

Located on the southern Romagna coast, in the province of Rimini, the city of Riccione covers an area of approximately 17.5 km² [25], characterized by a predominantly flat morphology along the coast that tends to become slightly hilly as it moves inland. This topography is important for the distribution network, as it minimizes problems related to significant differences in altitude, allowing for relatively uniform pressure regulation.

From a demographic point of view, the town has a resident population of approximately 34,600 inhabitants [25]. Although Riccione is a well-known tourist resort subject to strong summer fluctuations (which mainly impact the water and electricity systems), for the purposes of gas network modelling, the main variables are the resident population and winter seasonality.

3.2 Network Topology and Infrastructure

The infrastructure under analysis is configured as a local distribution network (managed by a Distribution System Operator - DSO), which acts as an interface between the national high-pressure transmission network and end users. In accordance with Italian regulatory classification (D.M. of April 16, 2008 [26]), the network analysed falls within Type IV. This category includes pipelines with a maximum operating pressure in the range $1.5\ bar \leq p_{op} \leq 5\ bar$.

It is therefore a medium-pressure network, the backbone of urban distribution, which feeds the subsequent reductions for domestic users or directly supplies large-scale users.

The system is powered through interconnections with the regional transmission network (TSO - Transmission System Operator). Gas transfer and the necessary pressure reduction are guaranteed by two Regulation and Measurement Cabins (called ReMi

stations or City Gate Stations), which constitute the source boundary conditions of the hydraulic model.

The two stations are characterized by different pressure set points:

- Primary Station: Set to a nominal pressure of 4.5 bar.
- Secondary Station: Set to a nominal pressure of 4.1 bar.

From a numerical modelling point of view, the network topology was represented by a connected graph composed of the following elements:

- 241 Nodes: Including demand nodes, transit nodes and 2 source nodes (ReMi).
- 249 Pipes: Connection elements characterized by different lengths and diameters. The pipes are made of steel with a (nominal) roughness of $1.4 \times 10^{-4}\, m$.

Figure 3.1. Topological scheme of the Riccione gas distribution network. The schematic highlights the two injection points from the TSO (Primary in the southeast and Secondary in the west) acting as boundary conditions for the system.

Figure 3.1 shows the topological representation of the system, in which the green markers identify the two injection points (ReMi stations) that act as source nodes, while the red markers identify the final reduction stations. In particular, the source in the southeast is the primary station, while the one in the west is the secondary station.

3.3 Reference Scenario: Pure NG Configuration

This section presents the results of steady-state simulation, feeding the network exclusively with natural gas. The boundary conditions imposed on the source nodes (ReMi) were set at 4.5 bar (Main Station) and 4.1 bar (Secondary Station) respectively. The nominal data (node pressures, consumption, pipe lengths, diameters, etc…) used for the following analyses were provided by the distributor INRETE.

3.3.1 Nodal Pressure Analysis

The spatial distribution of operating pressures along the network is shown graphically in Figure 3.2. The colour map allows us to appreciate the trend of the pressure gradient and the extent of the distributed pressure losses that are generated as we move away from the source nodes.

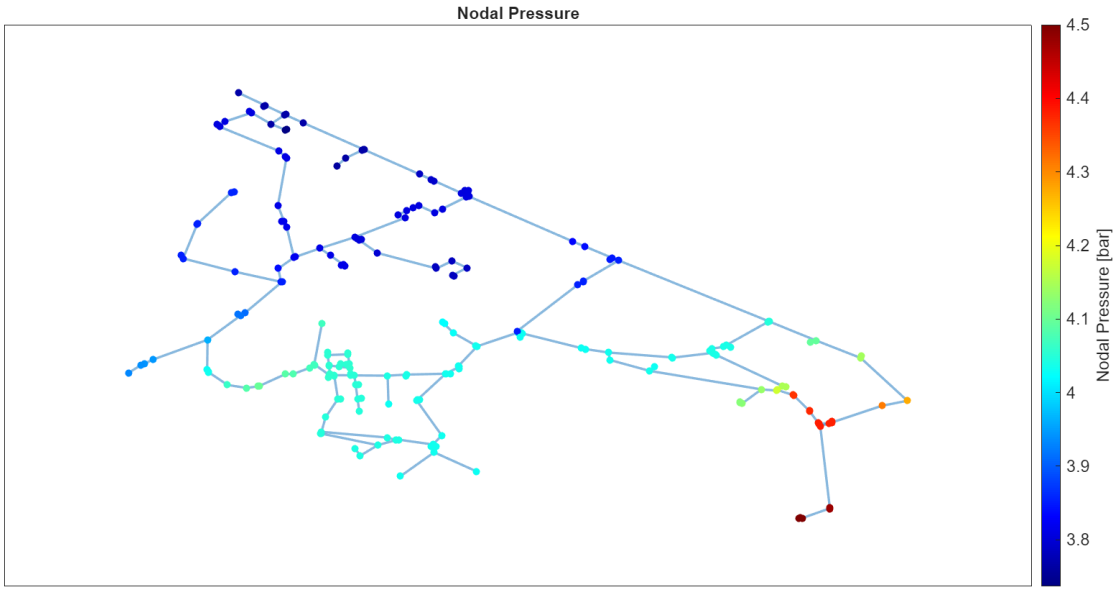


Figure 3.2. Spatial visualization of operating pressures in the Riccione network under nominal conditions in case of 100% NG

The graph clearly shows a zonal distribution of pressures, consistent with the topology of the network and the location of the supply stations:

- High Pressure Zone (Southeast): The area shown in red/orange corresponds to the area around the primary ReMi. Here, the pressure remains close to the injection value of 4.5 bar, ensuring sufficient hydraulic power to sustain the flow towards the northern areas.
- Medium Pressure Zone (Southwest): At the secondary ReMi, represented by green-cyan tones, values of around 4.1 bar are observed. It is interesting to note that this second injection point locally supports the pressure, avoiding excessive depressions in the western part of the network, but without dominating the overall gradient imposed by the primary station.

- The areas shown in dark blue identify the peripheral branches where minimum pressure values are measured, as they are located at the maximum topological distance from the supply points.

From the quantitative analysis of the results, the minimum pressure value recorded is 3.74 bar. This value is well above the lower operating limit for Type IV networks (p_{min}<1.5 bar). The safety margin observed ($\Delta P \approx 2.2\ bar$ from the limit) confirms that, in the nominal natural gas scenario, the network operates under conditions of broad stability, with no critical issues that could compromise supply to users, even in the terminal areas of the system.

3.3.2 Analysis of Pipeline Mass Flowrates and Velocities

To complete the characterization of the reference scenario, the dynamic parameters along the network pipes were analysed. While pressure provides information on the energy status of the nodes, the analysis of flow rates and velocities are essential for identifying any bottlenecks in transport and verifying the operating margins for future blending.

Figure 3.3 e 3.4 illustrate the mass flow rate distribution and the velocity field resulting from the simulation in the nominal scenario (100% NG).

The map of mass flow rates, illustrated in Figure 3.3, clearly shows the structure of the distribution network. The highest flows (represented in shades ranging from yellow to red) are concentrated in the pipelines immediately downstream of the Primary ReMi (bottom right). It is possible to visualize the primary gas path, which starts from the main injection point, branches out toward the city centre and then dissipates into the capillary branches (shown in dark blue), where the flow rate is drastically reduced to serve the end users. This configuration confirms that the main pipes are correctly sized to handle the current mass load without saturation.

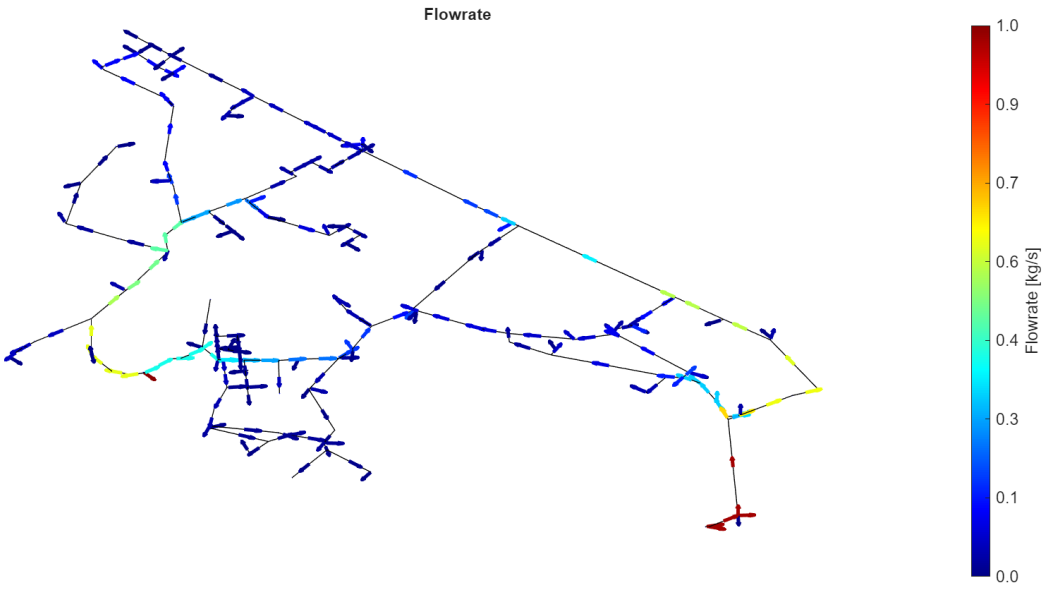


Figure 3.3. Color-coded map of gas flow rates in the pipelines under nominal conditions.

The distribution of velocities (Figure 3.4) is the most critical parameter in the transition to hydrogen. This analysis of the natural gas-only scenario reveals a situation of general stability:

- Most of the network operates at very low speeds (blue zone), below 2-3 m/s.
- Peak values are recorded exclusively in a few sections adjacent to the injection nodes and in some junctions of the central network.

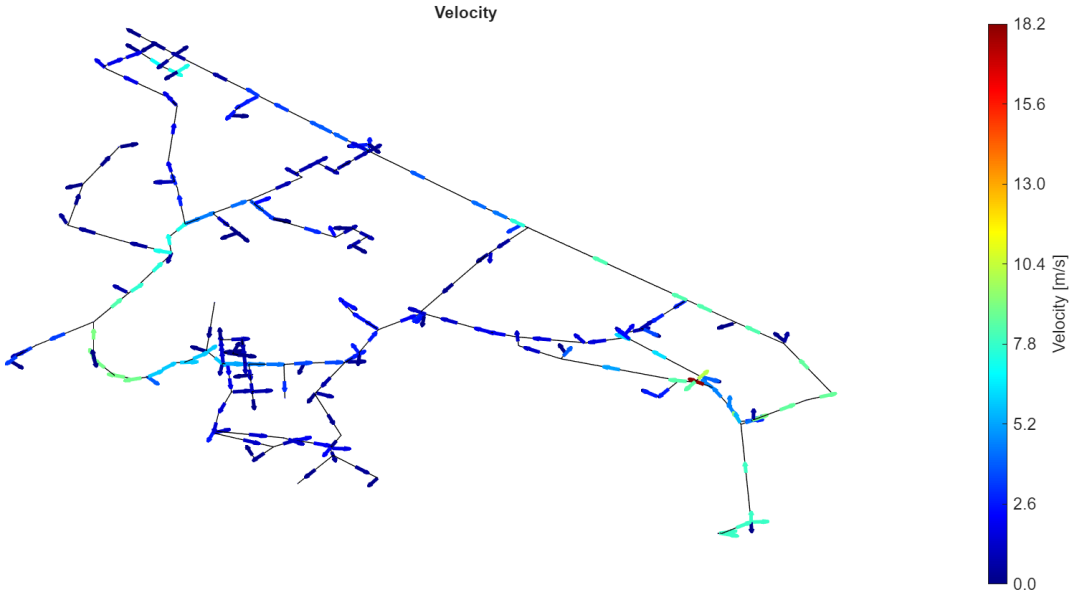


Figure 3.4. Gas velocity distribution across the network pipelines.

The maximum speed value is 18.23 m/s, slightly below the technical safety threshold, generally set at 20-25 m/s for steel/PE distribution networks (UNI 9165 [27]) in order to avoid erosion, excessive vibrations and noise pollution.

The peak value measured is a critical parameter in view of the introduction of hydrogen blending into the network. Blending with hydrogen, which has a significantly lower density than natural gas, will lead to an increase in volumetric velocities under the same energy or pressure conditions. Consequently, monitoring this parameter becomes a priority in order to prevent further increases in velocity from exceeding acceptable limits.

3.4 Simulation of Hydrogen Blending Scenario in the Riccione Network

The introduction of hydrogen into the natural gas transport network causes significant changes in the thermophysical properties of the fluid mixture, with direct impacts on the fluid dynamics of the pipeline. Since hydrogen has a significantly lower density than methane and a lower volumetric calorific value, maintaining the same energy output requires an increase in volumetric flow rates and, consequently, in flow velocities.

The following section presents the results of a scenario in which an H2NG mixture is injected with the aim of evaluating the impact of blending on the distribution of pressure, velocity and gas composition throughout the network.

3.4.1 Double-Point Hydrogen Injection with Differentiated Blending

This simulated scenario represents a more complex operating configuration, as both network supply sources deliver a mixture of H2NG, but with different concentrations:

- Source 1 (Primary ReMI): Mixture with $y_{H_2} = 10\%$ e $y_{CH_4} = 90\%$.
- Source 2 (Secondary ReMI): Mixture with $y_{H_2} = 5\%$ e $y_{CH_4} = 95\%$.

This asymmetry in composition allows us to analyse the behaviour of the network in the presence of concentration gradients and to evaluate the hydraulic stress induced by the injection of a percentage of hydrogen.

As shown in Figure 3.5, the entire network is filled with hydrogen, but its distribution is not uniform. The areas of influence of the two sources are clearly visible: the branches supplied by the primary ReMI have a molar fraction close to 0.10 (dark red), while the areas served by the secondary ReMI have values close to 0.05 (dark blue). In the area of hydraulic convergence of the two flows, a mixture with an intermediate concentration (yellow) is created, depending on the instantaneous flow rates.

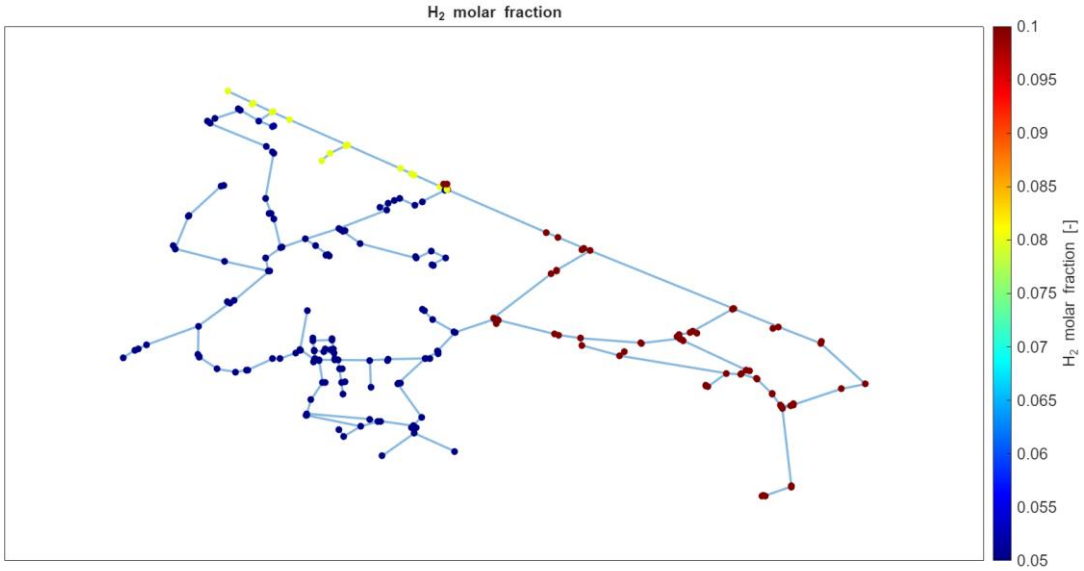


Figure 3.5. H2 molar fraction distribution. The network exhibits a non-uniform blend composition resulting from the interaction between the primary source (10% H2) and the secondary source (5% H2).

Blending has a greater impact on the fluid's range of motion. In particular, flow velocities increase compared to the scenario with pure natural gas. As shown in Figure 3.6, peak values of around 19.5 are reached in the most stressed sections. Although this value is still within the technical limits of operability, it represents a considerable increase that reduces operational safety margins and increases mechanical stress on the pipelines.

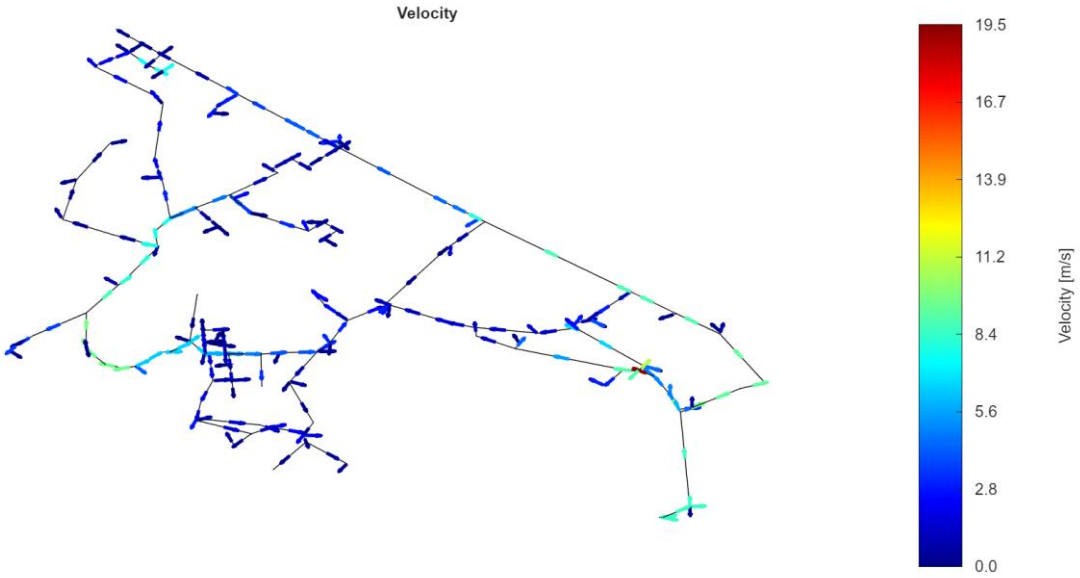


Figure 3.6. Fluid velocity profile. Higher hydrogen content leads to a further increase in flow velocities, reaching peak values around 20 m/s in critical branches.

The increase in speed necessarily results in an increase in distributed pressure losses. Comparing Figure 3.7 with the previous scenario, a slight decrease in node pressures can be observed. The pressure at the terminal nodes (shown in dark blue) reaches the minimum values recorded in the simulations, approximately 3.6 bar (a value not appreciable from the graph). Although the pressure drop remains well within safe operating limits, the trend confirms that higher blending percentages may require an increase in the inlet pressure at the ReMI to ensure minimum service levels at the most disadvantaged delivery points.

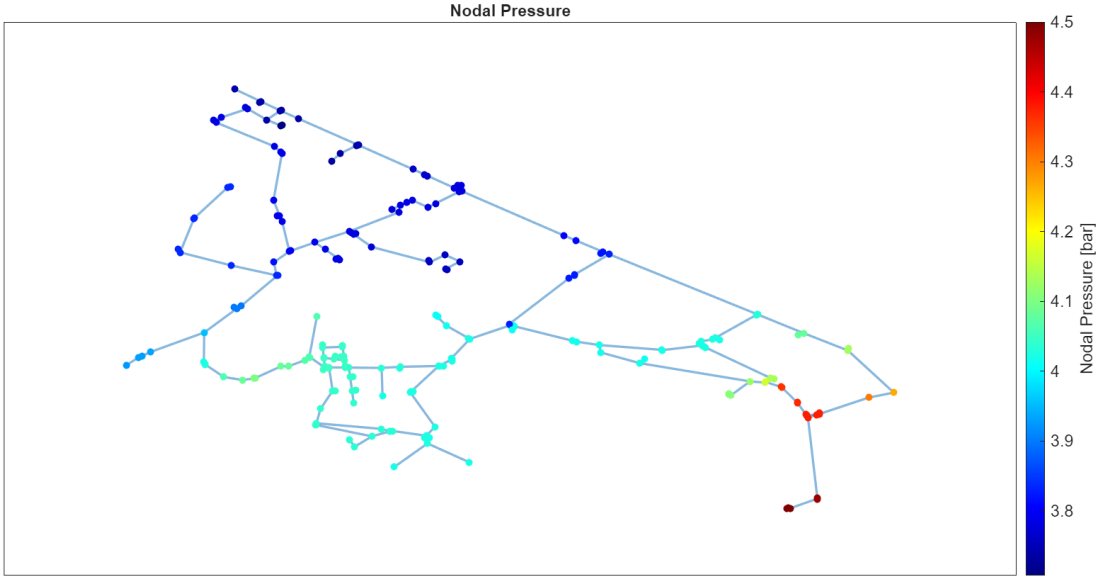


Figure 3.7. Nodal pressure distribution (2nd Scenario). The generalized increase in velocity results in higher pressure drops compared to the single-injection scenario, leading to lower available pressures at the peripheral nodes.

3.5 Validation of the Fluid Dynamics Model

Before going ahead with generating the massive dataset needed to train the surrogate model (which will be discussed in later chapters), it was essential to check the accuracy of the numerical solver developed in this research. The validation phase aims to ensure that the pressure and velocity profiles and the propagation of hydrogen concentration in the network are consistent with those calculated by established industry standards.

3.5.1 Reference Software: SAInt

The comparison was carried out using SAInt (Scenario Analysis Interface for Energy Systems) [28], commercial software developed by encoord. SAInt is a benchmark in the planning and modelling of integrated energy systems and was chosen as a standard for its ability to handle complex energy transition scenarios.

The main features of the software relevant to this validation include:

- Multi-Energy Modelling: SAInt simulates interactions between gas transport networks, electricity networks and thermal networks (multi-vector systems), providing an integrated view of energy flows.
- Hydrogen Support: The software is explicitly designed to analyse blending scenarios, allowing the simulation of gas mixture transport and the integration of renewable sources.
- State and Motion Equations: Unlike other commercial software that also solves the complete energy balance (including nodal temperature), SAInt adopts an isothermal approach regarding temperature. To calculate the friction factor, it uses the Hofer equation [29], a simplified variant of the Darcy-Weisbach and Colebrook-White equations [22].

3.5.2 Analysis of Results and Validation

Validation was carried out by comparing the results obtained on the 241 nodes and 249 pipelines of the test network. The comparison focused on three fundamental state variables: nodal pressure (p), mass flowrate in the pipes (G) and hydrogen molar fraction in the nodes (y_{H2}). To evaluate the accuracy of the proposed solver compared to the commercial software SAInt, an error analysis was conducted by calculating the Mean Absolute Error (MEA), the Maximum Relative Error (MRE) and the coefficient of determination (R^2).

These variables are calculated in this way:

- MEA: $$MEA = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.1}$$
- MRE: $$MEA = \max\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right) \tag{3.2}$$

- R^2:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{3.3}$$

Where:

- y_i: Values calculated by SAInt
- $\hat{y}_i$: Values calculated by the solver
- $\bar{y}_i$: Mean of the values calculated by SAInt
- n: Total number of values (nodes or pipes)

Figure 3.8 shows the parity plots for the quantities examined. In these graphs, the x-axis represents the value calculated by SAInt, while the y-axis shows the value returned by the solver developed. The distribution of the points shows a remarkable alignment with the bisector of the first quadrant, demonstrating the high overall correspondence between the solver and the commercial software.

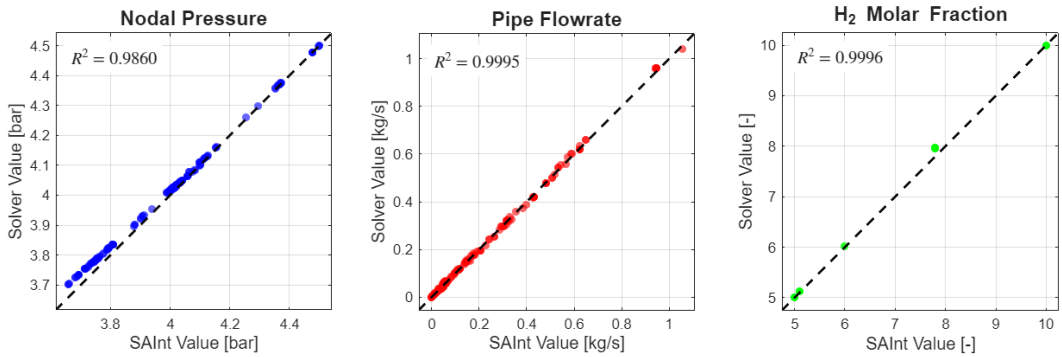


Figure 3.8. Parity plots comparing the developed solver against SAInt. From left to right: Nodal Pressure, Pipe Flowrate and Hydrogen Molar Fraction. The dashed line represents perfect agreement.

In quantitative terms, the error metrics calculated across the entire computational domain are summarized in the Table 3.1.

Variable	MAE	MRE [%]	R^2
Nodal Pressure	0.0195 bar	1.19	0.9860
Hydrogen Fraction	0.0127	2.20	0.9996
Pipe Flowrate	0.0030 kg/s	36.39	0.9995

Table 3.1. Statistical performance metrics of the proposed solver compared to SAInt.

The following findings were obtained from the data analysis:

- Nodal pressure: The model shows excellent accuracy with a coefficient of determination R^2 = 0.9860. The mean absolute error is 0.0195 bar (approximately 20 mbar), a value that is almost negligible compared to the operating pressures of the network (3.7 - 4.5 bar). Also, the maximum relative error recorded (1.19%) confirmed the absence of critical local deviations.
- Nodal hydrogen concentration: The validation of the hydrogen concentration at the nodes is excellent (R^2 = 0.9996). The almost perfect coincidence of the concentration values validates the composition tracking algorithm and the correct resolution of the mass balances at the mixing points.
- Pipe flow rate: Regarding flow rates, an R^2 close to unity (0.9995) and a low mean absolute error (0.0030 kg/s) are observed. The high value of the maximum

relative error (36.39%) is located exclusively in branches of the network characterized by flow rates close to zero ($\approx 10^{-4}$ Kg/s). In these cases, the denominator of the relative error formula tends to zero, numerically amplifying differences that, in absolute terms (MAE), do not have a significant impact on the physical functioning of the network.

It is important to note that the fluid dynamics solver and the commercial software SAInt use different formulas to calculate the friction factor (Cheng's equation and Hofer's equation, respectively). This difference, which is not insignificant, influences the validation results.

The validation therefore confirms that the solver developed is able to accurately reproduce the physical behaviour of the network, making it a reliable tool for generating the dataset needed to train the surrogate model described in the next chapter.

3.6 Parametric Analysis of Hydrogen Mixing Zones

At the conclusion of the fluid dynamics analysis and before the development of the surrogate model described in the next chapter, a parametric sensitivity analysis was conducted to physically understand how variations in operating conditions impact hydrogen propagation in the network.

The main goal of this phase was to monitor the behaviour of the mixing points, i.e., those nodes in the network where the flows from the two sources (ReMI 1 and ReMI 2) converge and where the resulting hydrogen concentration is determined by the local hydraulic equilibrium.

To simplify the interpretation of the results, the critical nodes where mixing occurs most frequently have been grouped into three macro-areas, called Zone 1 (orange box), Zone 2 (green box) and Zone 3 (red box), as shown in Figure 3.9. These zones represent the topological points where the interaction between the flows from the two sources occurs.

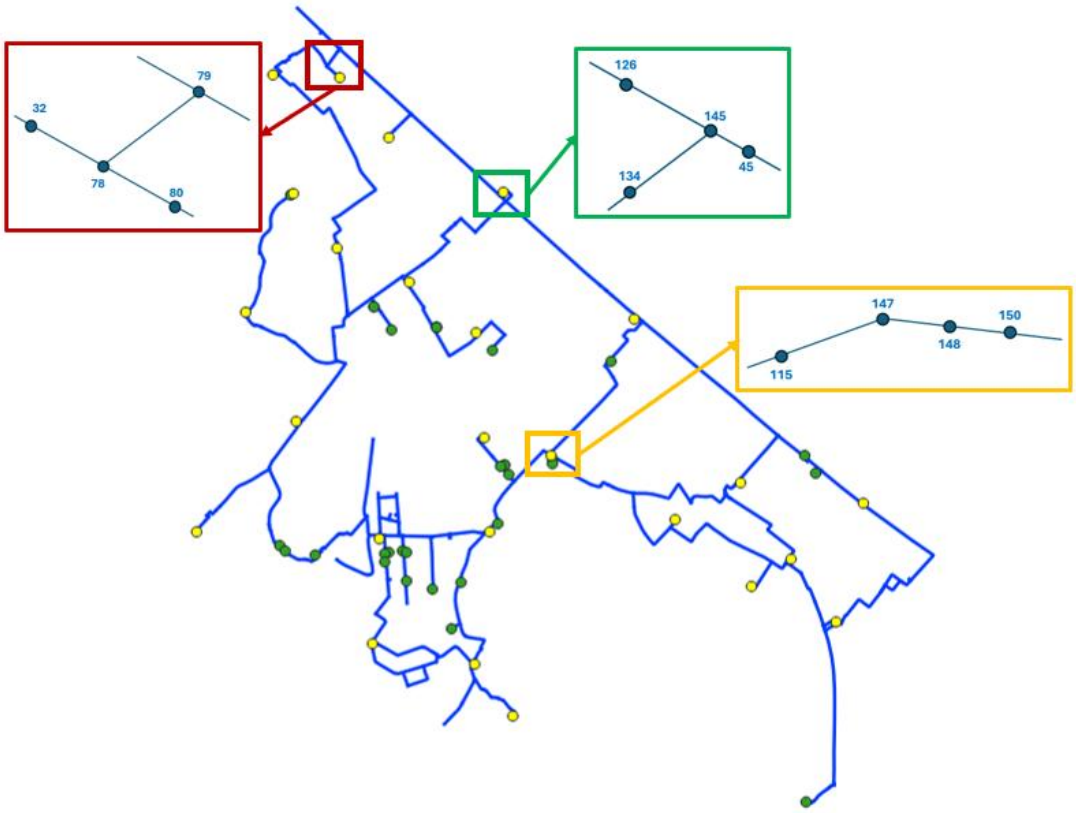


Figure 3.9. Identification of the three critical mixing zones: Zone 1 (orange box), Zone 2 (green box), Zone 3 (red box) within the network topology where hydrogen concentration variability is analysed.

To capture the full spectrum of influencing variables, two distinct numerical approaches were developed:

- One factor at a time analysis: Applied to hydraulic parameters (consumption, supply pressures and pipe roughness). This methodology involves perturbing a single variable in a defined percentage range (such as ±20% of the nominal value), while keeping the others constant. This approach allows to evaluate the impact of each parameter on the position of the mixing front. To highlight the interaction between flows of different compositions, the simulation was set up with an asymmetric injection scenario, in which the primary source injects a mixture of 10% H2 and the secondary source injects 5%.

- Two-dimensional surface analysis: Applied specifically to injected hydrogen concentrations. The composition in the downstream nodes depends on the simultaneous interaction of the two sources; therefore, a two-dimensional grid simulation was implemented, simultaneously varying the molar fractions at the two ReMI to evaluate the complete response of the system.

The following sections present the results obtained for each parameter analysed.

3.6.1 Sensitivity to Global Network Load Variations (G_{cons})

This analysis investigates the impact of gas demand on hydrogen blending dynamics. The global consumption parameter, G_{cons}, varied within the range ±20% relative to nominal conditions.

Zone 1: Zone 1 represents the most hydraulically sensitive area. While nodes 115 and 150 remain stable under the influence of their respective sources, node 148 shows critical “step” behaviour (Figure 3.10). Under nominal conditions, it is supplied by the source at 10%, but once consumption exceeds +10%, the nodal hydrogen concentration quickly drops to 5%, indicating a shift in the zero flow point from node 147 to node 148. This effect introduces sudden variations in gas quality that are difficult to manage [20].

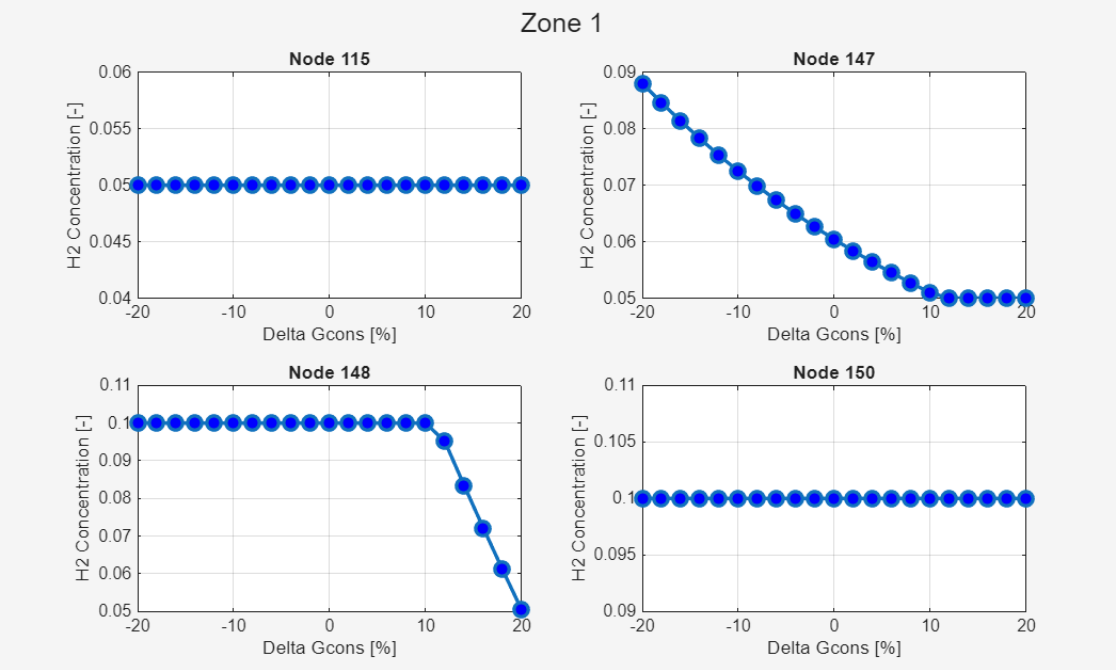


Figure 3.10. Sensitivity analysis for Zone 1. Node 148 exhibits a sharp non-linear transition when the loads increase exceeds +10%.

Zone 2: In Zone 2 (Figure 3.11), the response is linear. The mixing node is 145 and undergoes progressive dilution as the load increases: the influence of the primary source gradually gives way to that of the secondary source, causing a drop in concentration (from ~10% to ~6.5%), without the discontinuities observed in Zone 1. Node 126 follows the same behaviour as it is adjacent to node 145.

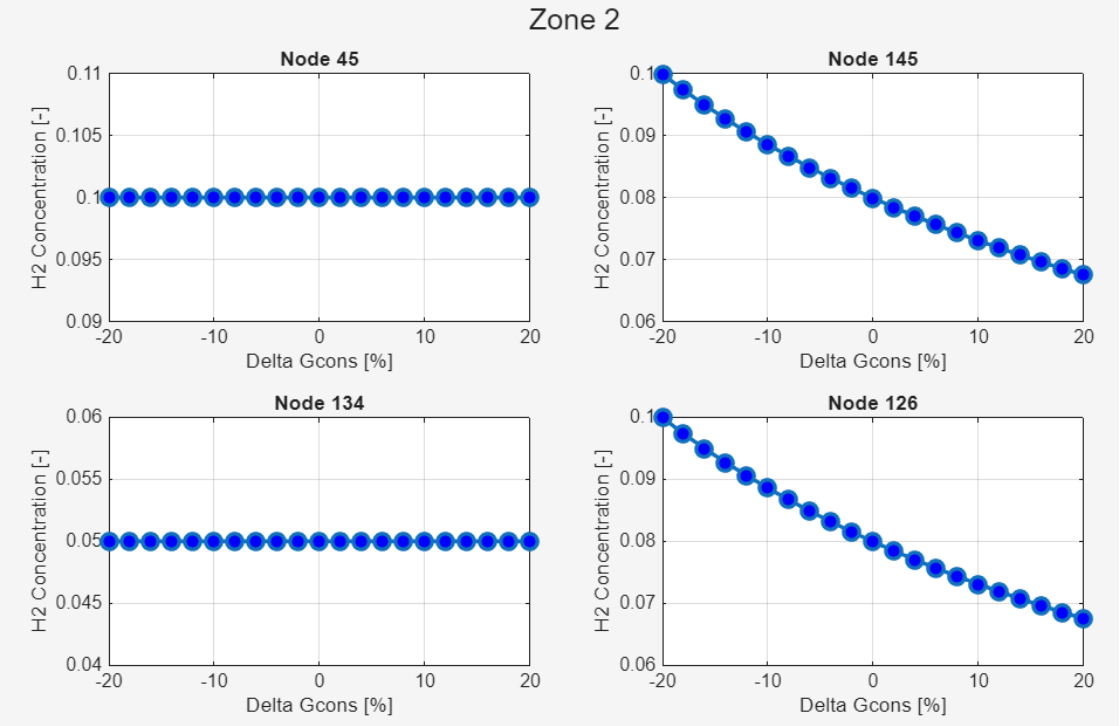


Figure 3.11. Sensitivity analysis for Zone 2. The hydrogen concentration decreases almost linearly as the global demand increases.

Zone 3: In Zone 3 (Figure 3.12), the gradients are greatly reduced: nodes 79, 78 and 80 show minimal variations in concentration, highlighting excellent stability that makes this area less sensitive to fluctuations in demand.

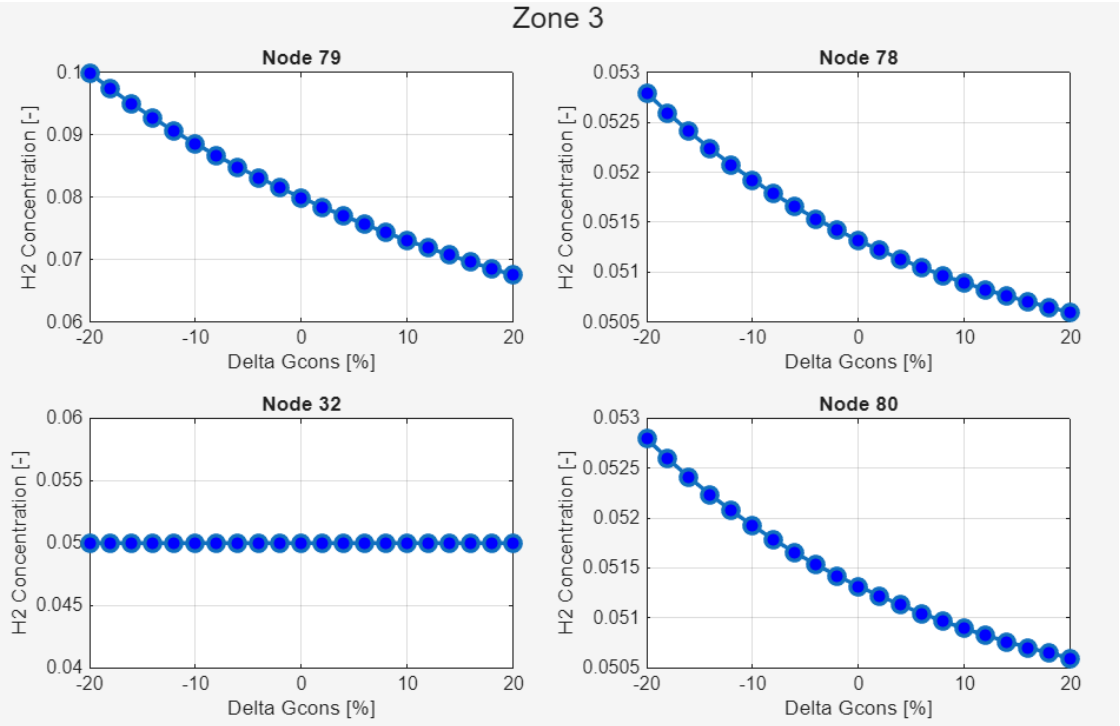


Figure 3.12. Sensitivity analysis for Zone 3. The mixing response is smoother compared to other zones.

3.6.2 Sensitivity to Supply Pressure Variations (P_{set})

The second analysis examines the influence of the supply pressure at the outlet of the ReMI cabins. The P_{set} parameter varied by 7% from the nominal value. The analysis of the results shows a consistent but opposite response of the system compared to the case of consumption variation.

Zone 1: In Zone 1 (Figure 3.13), a different behaviour is observed compared to the load variation scenario. Node 148, previously identified as critical (and a mixing node), shows perfect stability at a value of 10% H2. Node 147, on the other hand, shows a positive linear trend: as the set-point pressure increases, the hydrogen concentration increases (from approximately 5.4% to 6.7%). This suggests that higher overall pressure enhances the "push" of the Primary Source and that, by varying the pressure of the sources, the mixing point remains fixed.

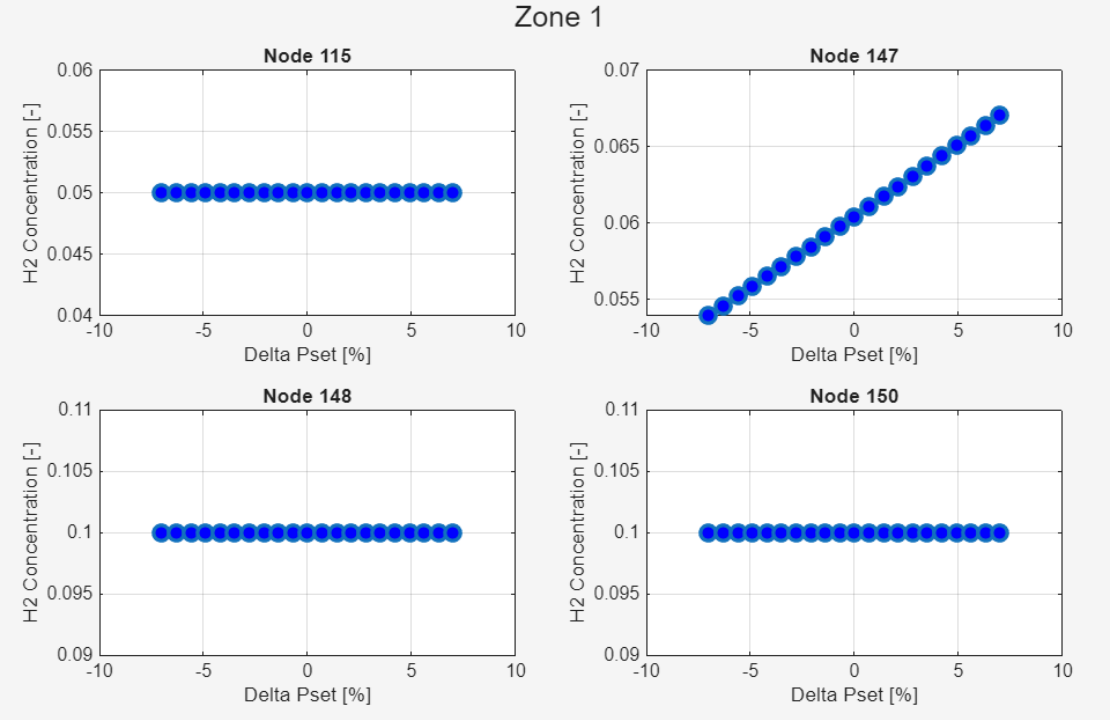


Figure 3.13. Sensitivity analysis for Zone 1 (P_{set}). Unlike the load scenario, Node 148 remains stable. Node 147 shows a linear increase in H2 concentration as pressure rises.

Zone 2: In Zone 2 (Figure 3.14) the response is also linear and positive. The mixing point remains constant at node 145 and there is a constant increase in hydrogen concentration as the network pressure increases. This confirms that the increase in P_{set} extends the range of the primary source (10%) and reduces the dilution effect caused by the secondary source.

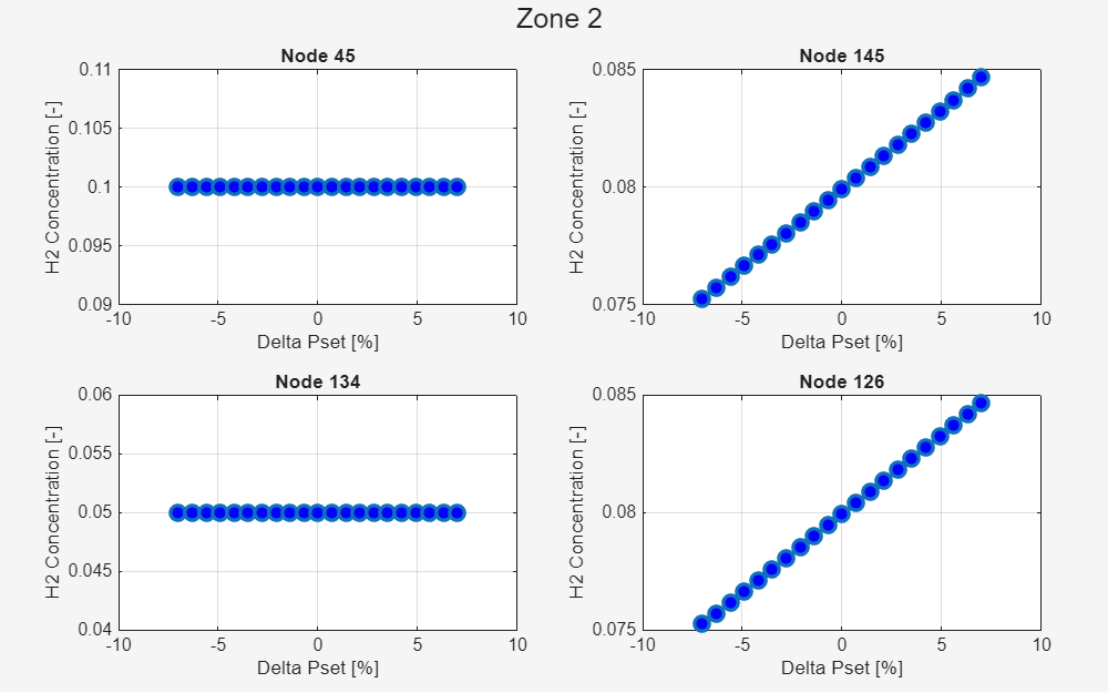


Figure 3.14. Sensitivity analysis for Zone 2 (P_{set}). The correlation indicates that higher system pressures extend the influence of the primary source (10% H2).

Zone 3: Zone 3 (Figure 3.15) confirms more moderate variations, as in the previous case. However, the trend remains consistent with the other zones: nodes 79, 78 and 80 all show a positive slope, although the increase in absolute value is modest.

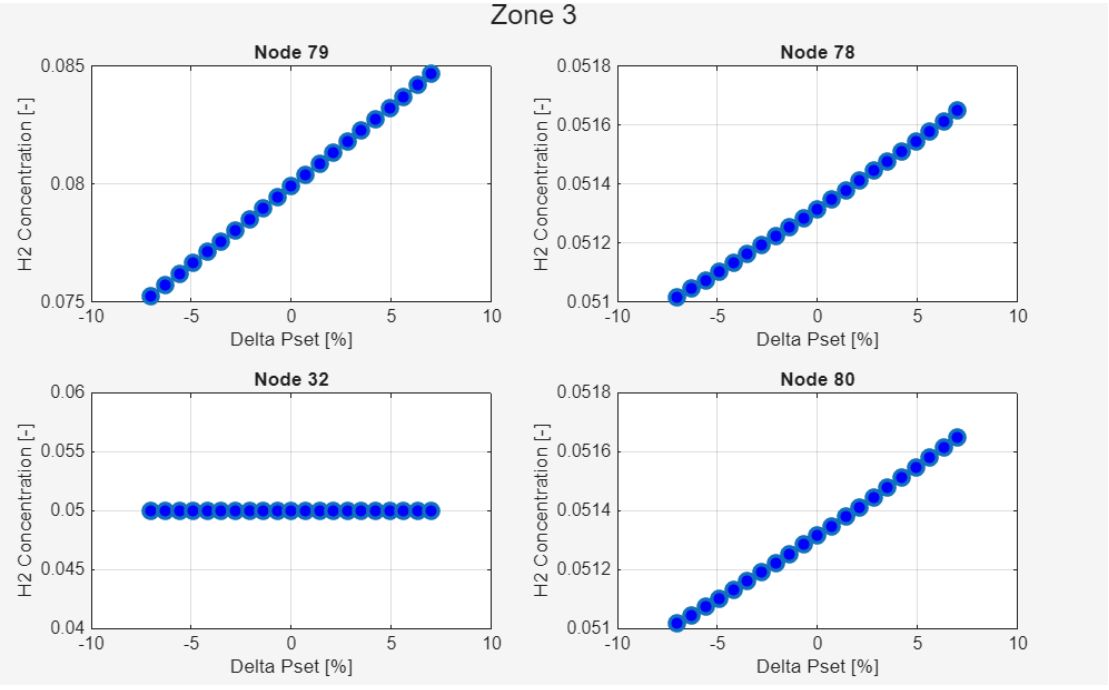


Figure 3.15. Sensitivity analysis for Zone 3 (P_{set}). A slight but positive trend confirms the hydraulic expansion of the primary blend.

3.6.3 Sensitivity to Pipe Roughness Variability (ε)

The latest analysis of hydraulic parameters focuses on the impact of the absolute roughness of the pipes. This parameter is often subject to uncertainty during modelling, depending on the age and condition of the pipes. In this analysis, roughness was uniformly disturbed across the entire network in the range ±10%.

Zone 1: In Zone 1 (Figure 3.16), nodes 115, 148 and 150 show complete stability. It is interesting to note that node 148, which is critical for load variations, remains fixed at 10% concentration in this case, suggesting that friction variations in this range are not sufficient to trigger the flow inversion observed in Section 3.6.1. Node 147, on the other hand, shows a slight negative linear correlation. As roughness increases, the concentration decreases slightly (from approximately 0.062 to 0.059).

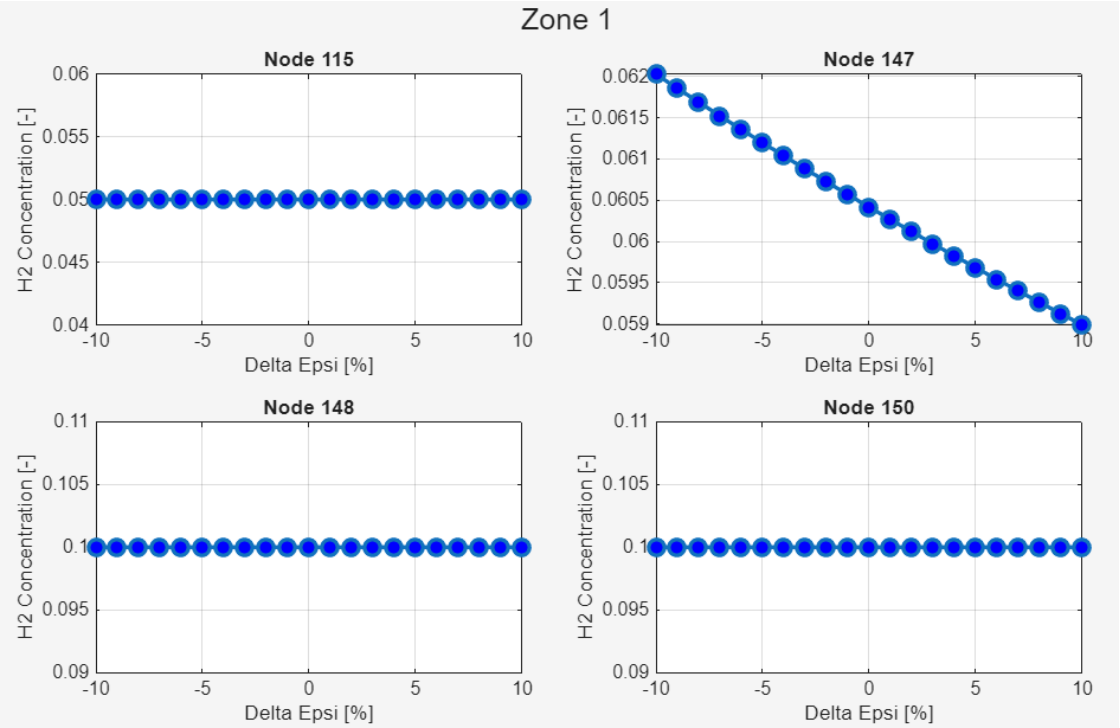


Figure 3.16. Sensitivity analysis for Zone 1 (ε). Increasing pipe roughness leads to a linear decrease in H_2 concentration at node 147, while node 148 remains stable.

Zone 2: Zone 2 (Figure 3.17) confirms the negative trend at the mixing node. Node 145 remains the only mixing point and shows a linear and constant decrease in the molar fraction of hydrogen as the friction coefficient increases.

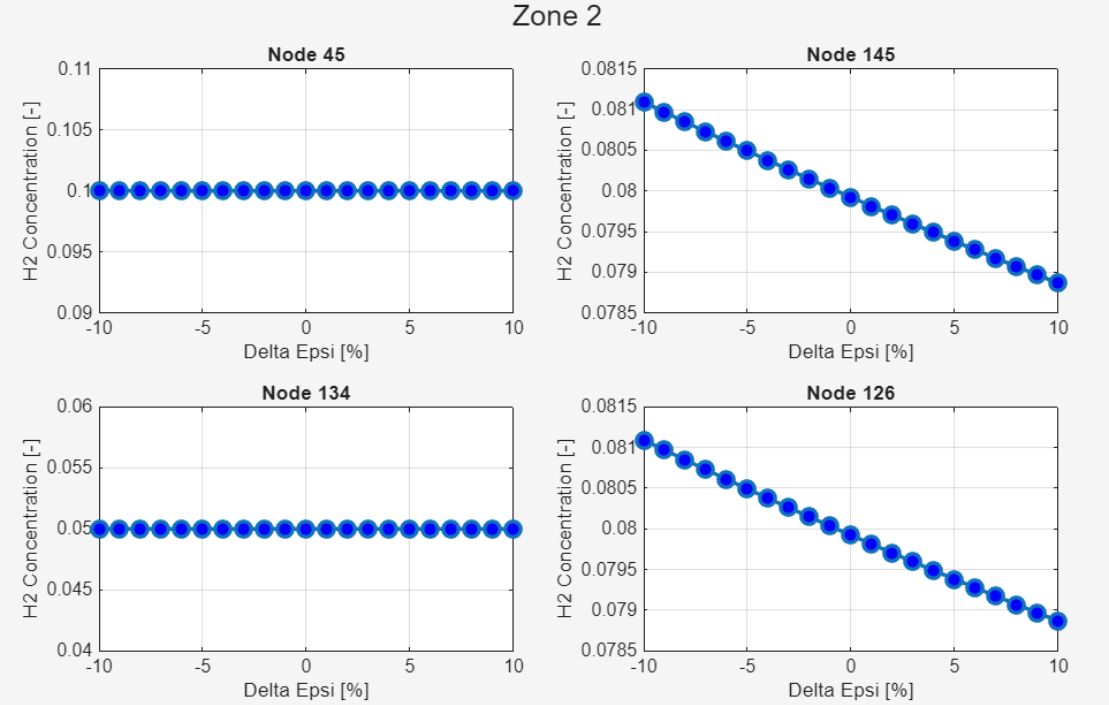


Figure 3.17. Sensitivity analysis for Zone 2 (ε). A negative trend is observed: higher friction losses lead to lower local H2 concentrations.

Zone 3: Zone 3 (Figure 3.18) also follows the pattern described above. Nodes 79, 78 and 80 all show a negative slope. The changes in absolute value are small and the response is free of discontinuity.

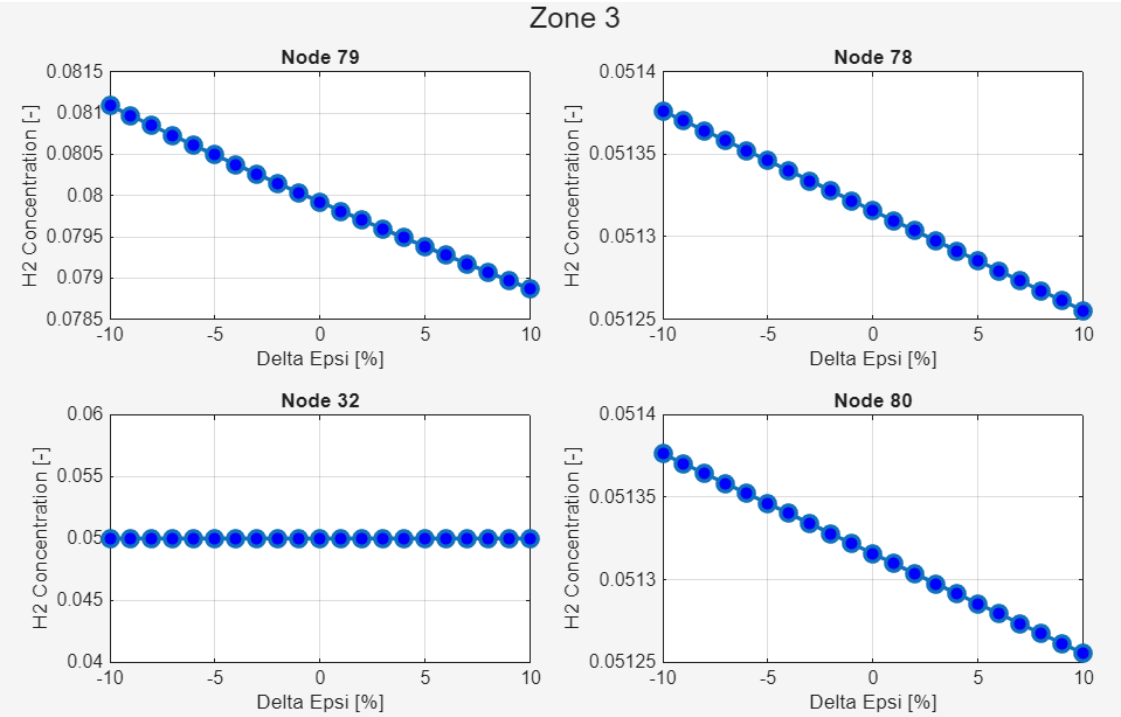


Figure 3.18. Sensitivity analysis for Zone 3 (ε). By increasing the pipes roughness, a gentle, linear reduction in hydrogen content occurs.

3.6.4 Sensitivity to Hydrogen Blending Ratios at Injection Points

To conclude the parametric analysis, the impact of variations in the quality of the injected gas was examined. Differently from the previous sections, in this scenario the hydraulic state of the network is maintained constant (consumption, nominal pressure set points and pipe roughness). The independent variables are the molar fractions of hydrogen at the two sources, which were varied simultaneously in the range ±95% with respect to the nominal value of 10%.

The results are represented by surface graphs, where the slope of the plane indicates the degree of dependence of the node on one source or the other.

Zona 1: In Figure 3.19, nodes 115, 148 and 150 have surfaces inclined exclusively along one axis. Node 115 depends only on the 2nd Source, while nodes 148 and 150 depend only on the 1st Source. This confirms that, hydraulically, these nodes are supplied only by their respective sources. Node 147, on the other hand, shows a diagonally inclined surface. The resulting concentration is a linear combination of both inputs, confirming that the node represents the zero flow point where the two flows converge.

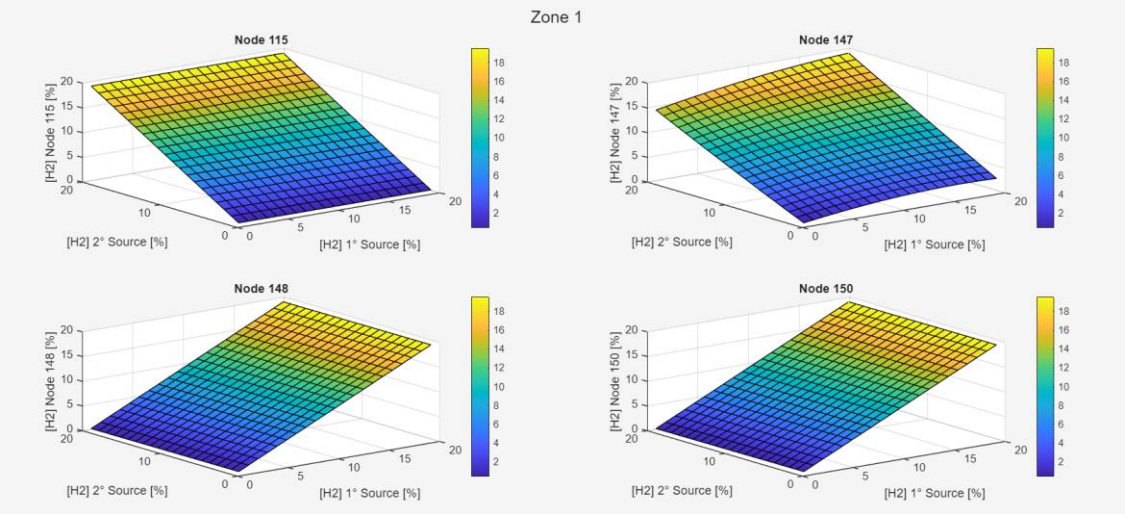


Figure 3.19. Surface response for Zone 1. Node 115 and 150 show dependence on a single source, while Node 147 exhibits a diagonal slope, indicating active mixing.

Zone 2: In Zone 2 (Figure 3.20), nodes 45 and 134 show an exclusive dependence on one of the two sources, as evidenced by the zero slope along a single axis. Node 145 (and therefore also 126) has a diagonal surface, so there is a significant contribution from both sources.

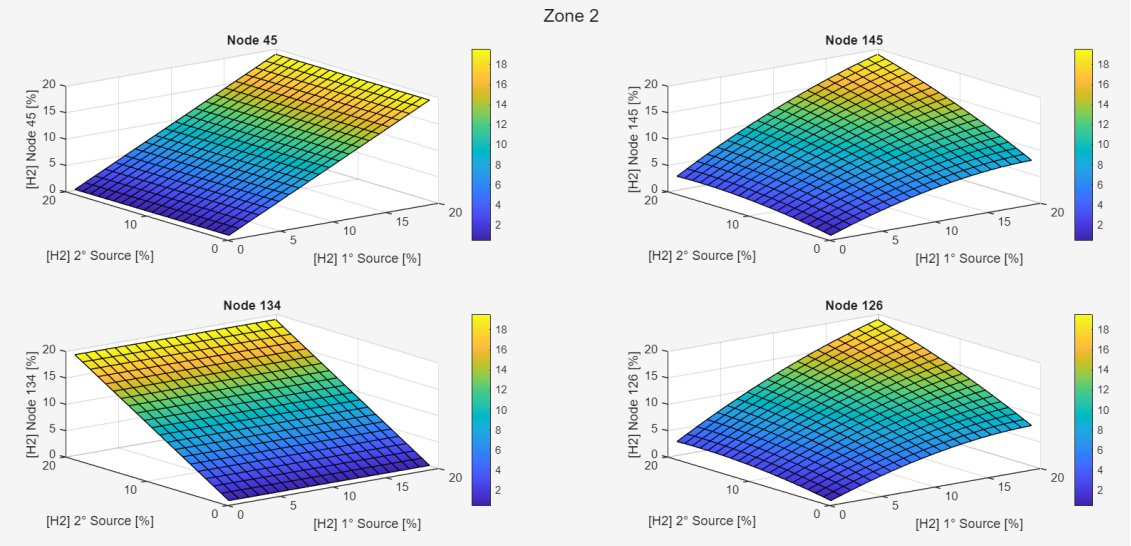


Figure 3.20. Surface response for Zone 2. Nodes 45 and 134 are dominated by a single source, whereas Nodes 145 and 126 show a mixed composition.

Zone 3: In Zone 3 (Figure 3.21), the hydrogen concentration at node 32 depends only on the secondary source; at node 79, however, a mixture dependent on both sources arrives (the same mixture as at node 145 in Zone 2). Node 78 remains a stable mixing point, showing a slight dependence on the primary source as well, imposing the same composition on node 80.

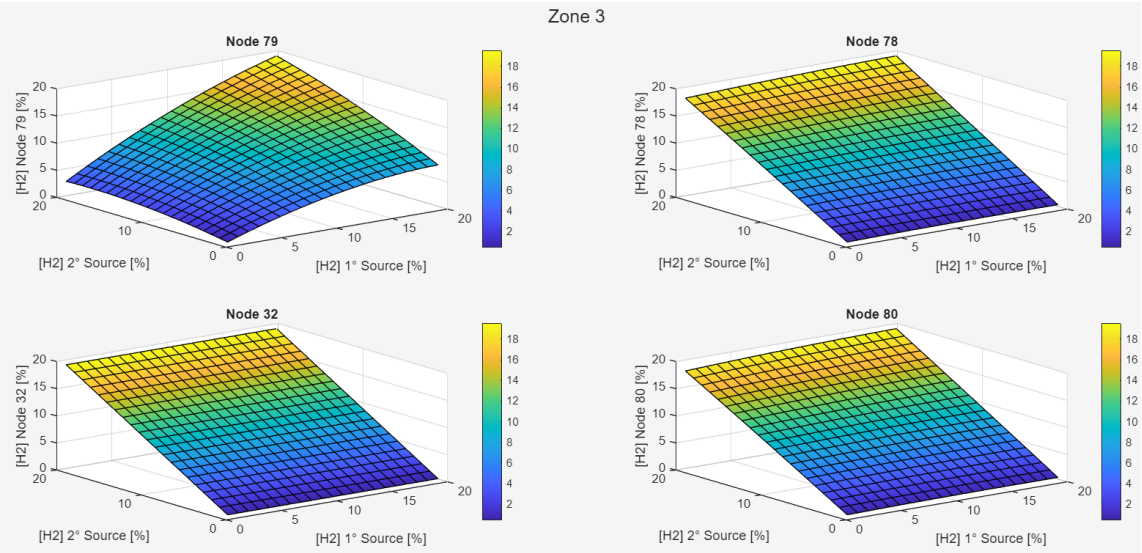


Figure 3.21. Surface response for Zone 3. Node 32 is dominated by a single source; Node 79 shows the same mixed composition from Node 145 (Zone 2). Node 78 is a stable mixing point, with a clear dependence of both sources.

The parametric analysis showed that changes in gas demand have the greatest impact on hydrogen blending dynamics, generating rapid and non-linear changes in the concentration of some nodes in zone 1 due to the displacement of the zero flow point. This information will support the choice of ranges for variation in fluid dynamics parameters.

Chapter 4

4 Surrogate Modelling and Case Study Analysis

The purpose of this chapter is to develop a data-driven predictive model able to estimate the behaviour of the gas network in hydrogen blending scenarios. Since the physical state of the system is defined simultaneously by multiple variables (node pressures, flow rates, velocities and H_2 concentrations), the problem is vectorial and multidimensional.

The selection of the Machine Learning algorithm was driven by the need to balance accuracy and computational cost. Although neural networks (non-convex optimization problems) are the standard for complex nonlinear problems, their training is difficult as it requires massive datasets, making them less practical when the data comes from computationally expensive engineering simulations.

On the other hand, Kernel-based methods offer significant advantages for this type of application:

- Data efficiency: They guarantee high predictive performance even with limited training datasets.
- Mathematical robustness: They are based on convex optimization, which avoids the local minimum typical of neural networks.
- Computational lightness: They do not require specialized hardware.
- Extension for multi-output [30]

Based on these considerations and supported by recent literature [30], an approach based on Compressed Vector Kernel Ridge Regression was adopted. This methodology allows the capture of spatial correlations between network nodes, providing a fast and reliable surrogate model for subsequent statistical analysis.

4.1 Surrogate Model Development and Implementation

The computational core of this study consists of a model that combines the physics-based fluid dynamics solver (described and validated in Chapter 3) with machine learning algorithms for vector regression. The entire architecture was implemented in MATLAB and structured around a main script that sequentially executes the initialization, data generation, training and accuracy assessment phases of the model.

The primary objective of the implementation is to overcome computational limitations intrinsic to iterative numerical solvers. While the physical model requires the solution of nonlinear systems of equations (mass and momentum balances) at each load variation, the surrogate model is designed to provide an almost instantaneous evaluation through direct input-output mapping.

The code aims to build an approximation function able to predict the complete state of the network in response to a vector of uncertain operating conditions that vary with each simulation. For each operating condition, the surrogate model returns the pressures at the nodes, the flow rates and the flow velocities in the pipes. Afterwards, using the flow rates, the nodal gas exchanges and the hydrogen concentrations at the sources, nodal balances are performed at each node to calculate the nodal concentration throughout the network.

The workflow implemented in the code follows a sequential logic divided into five macro-phases, detailed in the following sections:

- Network Initialization: Loading of the network topology and definition of operating conditions by imposing ranges of variation for input parameters.
- Data Collection: Running Monte Carlo simulations using the physics-based solver to build the training and test dataset.
- Model Training: Training the Kernel regression model to learn nonlinear relationships between inputs and outputs.
- Quality Mixing Module: With the fluid dynamic variables predicted by the input model, a weighted average of the gas composition converging in each node is calculated.
- Surrogate Model Accuracy Assessment: Validation of performance on data not used for training and analysis of prediction errors.

Below is a flowchart (Figure 4.1) showing the structure of the model just described.

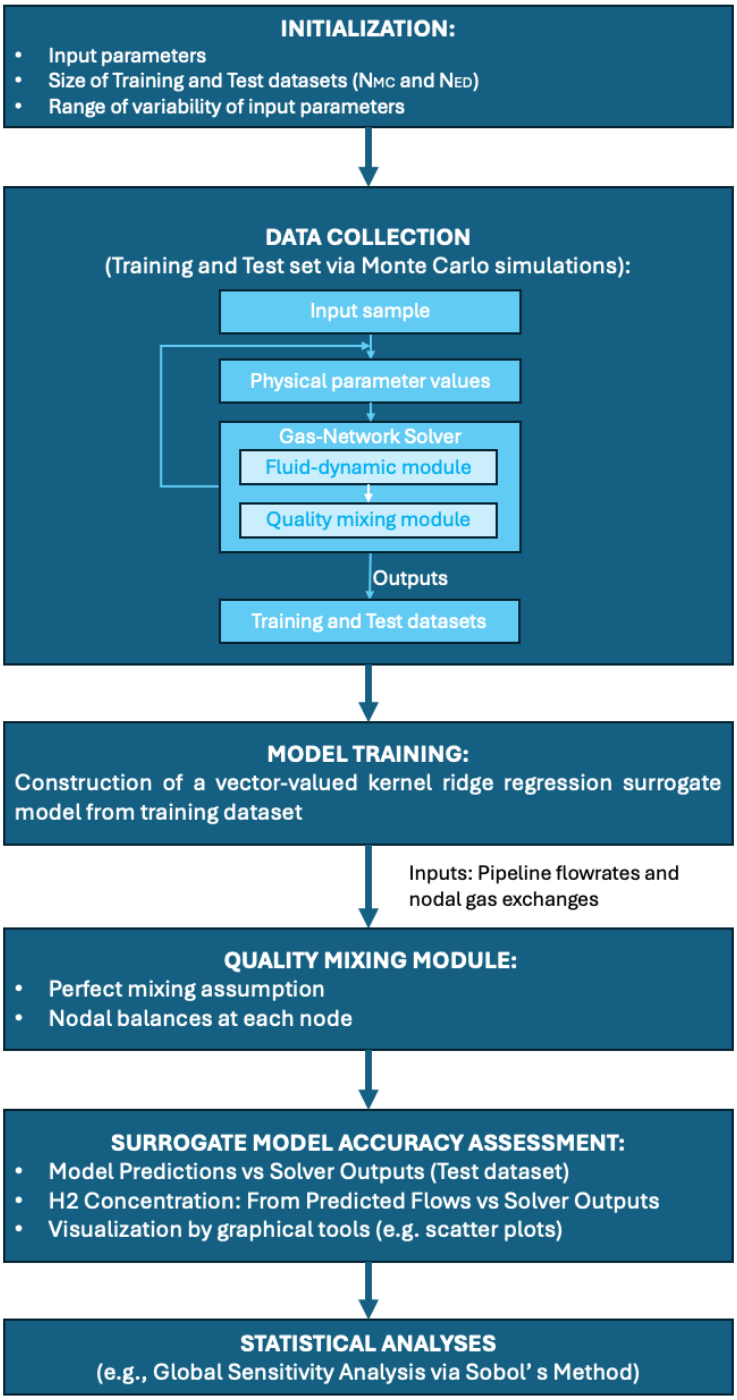


Figure 4.1. Flowchart summarizing the structure of the model divided into all its macro phases.

4.1.1 Network Initialization and Definition of Parametric Variability

The construction of the surrogate model begins with a rigorous definition of the dataset size and training strategy. Two fundamental parameters control the data generation process:

- N_{MC} (Monte Carlo Samples): Represents the total number of simulations performed using the physical solver. This forms the global knowledge base available for study.
- N_{ED} (Experimental Design Samples): Represents the subset of samples used specifically for training the machine learning model.

The difference between these two values ($N_{TEST} = N_{MC} - N_{ED}$) constitutes the Validation Set. It is essential to point out that the data contained in this set are never "seen" by the algorithm during the learning phase. They are used exclusively to verify the predictive capacity of the surrogate on new scenarios.

In this thesis, we chose to use a deliberately low number of training samples (N_{ED}) compared to the complexity of the problem. This strategy aims to demonstrate the efficiency of the Kernel method and its ability to generalize correctly and predict complex nonlinear behaviours even when faced with sparse training data.

To test the model under different operating conditions, the code uses a structure that defines the range of variability for each input variable. As detailed in the Case Studies section (Section 4.3), different scenarios will be analysed by varying both the number of active parameters (from 4 to 30) and their perturbation intensity.

The physical and topological initialization of the network is managed by a specific function that constructs the graph by importing essential geometric parameters, such as node connectivity, lengths L, diameters D and absolute roughness ε directly from external datasets.

At the same time, the vectors of nominal values for nodal consumption (G_{cons}) and set-point pressures (P_{set}) are loaded. Furthermore, a crucial aspect of this routine concerns the definition of the chemical boundary conditions necessary to simulate blending. The function receives the molar fractions of hydrogen at the sources as input and dynamically updates the composition matrix, ensuring that the mass balance closes at unity ($y_{CH4} + y_{H2} = 1$).

4.1.2 Monte Carlo Simulation and Dataset Collection

Once the variability ranges have been defined, the next step involves generating the dataset needed to train the model. Given the high computational cost of fluid dynamics simulations, it is important to use a sampling technique that maximizes the information collected with as few simulations as possible. For this purpose, the code replaces simple random sampling with Latin Hypercube Sampling (LHS), implemented using the MATLAB lhsdesign function.

This statistical stratification technique ensures that each dimension of the parameter space is sampled with uniform probability density, avoiding sample overlap and leaving fewer unexplored areas. The generated vector, $x_{LHS} \in [0,1]^N$, is then normalized to the interval [-1, 1] using the linear transformation:

$$x_{norm} = 2 \cdot x_{LHS} - 1 \tag{4.1}$$

This vector x_{norm} is the input for the Monte Carlo simulation cycle.

In the iterative cycle, a dedicated function translates the vector x_{norm} into real physical quantities to be used as input for the solver. The perturbation occurs as follows:

$$P_{sim} = P_{nom} \cdot (1 + x_i \cdot \delta_i) \quad (4.2)$$

where δ_i is the maximum amplitude of the variation defined during initialization (see section 4.1.1).

The modified physical variables belong to different categories.

Firstly, the conditions at the injection nodes (ReMI) are changed, modifying both the set-point pressures (P_{set}) to simulate fluctuations in the transport network and the quality of the injected gas (Y_{H2}), ensuring chemical consistency by recalculating the methane fraction to close the mass balance.

In parallel, the model considers the variability of user demand, varying the consumption at nodes (G_{cons}) uniformly or partially differentiated by geographic cluster, in order to reproduce the variability of daily or seasonal consumption profiles.

Finally, to include uncertainties related to the condition of the infrastructure, variability is introduced in the absolute roughness of the pipes, a factor that directly influences pressure losses as materials age.

For each iteration, the physical solver calculates the network system. The code includes a convergence check (if ~pos), where if the simulation converges, the network state is extracted and saved in the Y_{MC} output matrix. The stored variables include nodal pressures (P), flow rates in the pipes (Q), velocities (v) and nodal hydrogen concentrations (Y_{H2}). If the simulation diverges, the sample is discarded to avoid introducing numerical noise or non-physical values into the training dataset.

At the end of the process, the complete dataset consists of the pair of matrices X_{MC} and Y_{MC}, which are ready for the training phase.

4.1.3 Machine Learning Model Training

After completing the dataset generation using Monte Carlo simulations, the next step involves the actual training of the surrogate model. The code first implements a random data partitioning procedure to ensure rigorous validation. A random permutation of the indices (randperm) is generated from the total matrix of convergent simulations. This shuffling ensures that there is no bias in the data selection. The dataset is then divided into two subsets:

- Training Set (X_{ED}, Y_{ED}): Consists of N_{ED} samples, used by the algorithm to learn input-output relationships.
- Test Set (X_{TEST}, Y_{TEST}): Consists of the remaining samples (N_{MC} - N_{ED}) and is used exclusively for performance verification.

Once trained, the models are used to generate predictions on the Test set, whose results (Y_{pred}) will be analysed in the next section.

4.1.4 Quality Mixing Module

The last component of the computational framework is the Quality Mixing module. This module operates downstream of the surrogate model training, using the predicted hydraulic variables to calculate the distribution of hydrogen concentration throughout the entire network.

The fluid dynamic variables needed by the surrogate model to solve the problem of hydrogen mixing and quality monitoring are:

$$G_{ext} \rightarrow vector\ of\ all\ the\ \dot{m}_{ext_i}\ (for\ all\ the\ nodes)$$

$$FR \rightarrow vector\ of\ all\ the\ \dot{m}_j\ (for\ all\ the\ pipes)$$

It is essential to specify the nature of the mass variables involved in the calculation: while nodal consumption (G_{ext}) represents data input to the system (subject to the variability described in Section 4.1.2), the surrogate model has been trained to predict the net injection flows at the sources (where the pressure is set) necessary to meet this instantaneous demand. Using the reconstructed motion field and mass balances at the injection nodes, the module determines the propagation of the mixture, ensuring the strict conservation of chemical species.

The algorithm receives as input the flow rates in the pipelines and the gas exchanges at the nodes generated by the ML models described in Section 4.1.3. It also acquires the boundary conditions relating to the gas composition at the injection points, which are updated for each scenario according to the specific parameters of the test set.

The fundamental principle on which the module is based is the resolution of the mass conservation equation for each node. Physically, the composition in a generic node is determined by the weighted average of the mass flow rates of the compositions of all the flows converging in that node (both from upstream pipes and from exchanges with the outside).

The algorithm calculates the mixing by solving the nodal balance through a weighted average. Under a perfect mixing assumption, the following formula is used:

$$\left[w_{(c)}\right]_{i^*} = \frac{-\sum_j a^-_{i^*,j}\,\dot{m}_j \cdot \left[w_{(c)}\right]_j - \dot{m}^{(-)}{}_{ext_{i^*}} \cdot \left[w_{(c)}\right]_{ext_{i^*}}}{\sum_j a^+_{i^*,j}\,\dot{m}_j + \dot{m}^{(+)}{}_{ext_{i^*}}} \tag{4.3}$$

where:

$$a^+_{i^*,j} = \begin{cases} +1, & \text{pipe } j \text{ is outgoing from junction node } i^* \\ 0, & \text{pipe } j \text{ is incoming to junction node } i^* \\ 0, & \text{pipe } j \text{ has no connections with junction node } i^* \end{cases}$$

$\dot{m}^{(+)}{}_{ext_{i^*}}$ is the withdrawn gas flow from junction node i^*;

$$a^-_{i^*,j} = \begin{cases} 0, & \text{pipe } j \text{ is outgoing from junction node } i^* \\ -1, & \text{pipe } j \text{ is incoming to junction node } i^* \\ 0, & \text{pipe } j \text{ has no connections with junction node } i^* \end{cases}$$

$\dot{m}^{(-)}{}_{ext_{i^*}}$ is the injected gas flow in junction node i^*;

and:

- $[w_{(c)}]_{i^*}$ is the mass fraction of the component c at the junction node i^*; it is the unknown of the equation, resulting from the perfect mixing of the incoming fluxes.
- $[w_{(c)}]_j$ is the mass fraction of the component c at all the adjoining nodes that are connected to the junction node i^* through the j^{th} pipe.
- $[w_{(c)}]_{ext_{i^*}}$ is the mass fraction of the component c within the mass flux that is injected from outside in the network.

The system of equations is solved iteratively. Starting from a first guess condition (typically pure methane), the compositions are propagated through the network until the maximum relative error between two consecutive iterations falls below a tolerance threshold ($toll_C = 1 \cdot 10^{-8}$).

4.1.5 Surrogate Model Accuracy Assessment and Visualization of Results

The final stage of the computational workflow is dedicated to validating the model's performance and graphically displaying errors. This post-processing routine is essential to ensure that the surrogate has effectively learned the physics of the problem and is able to generalize on unseen data, avoiding overfitting (i.e., the critical condition in which the algorithm adapts excessively to the training data, learning even the statistical noise and losing predictive power on new operating scenarios).

The model loads the prediction matrices (Y_{pred}) generated by the ML model and compares them with the reference matrices of the Test Set (Y_{TEST}) calculated by the physical solver. The main tool used is the Parity Plot (or comparison Scatter Plot). For each physical quantity of interest (pressure, mass flow rate, velocity and external flow rate), the code generates a graph in which the x-axis represents the reference value (output of the physical solver) and the y-axis represents the prediction value (output of the surrogate model).

In this representation, a perfect prediction corresponds to a point that is exactly on the bisector of the first quadrant (the straight-line y = x). The dispersion of points around this diagonal line provides an immediate visual measure of the residual error: points far from the diagonal indicate systematic errors in the model.

4.2 Definition of Case Studies

To assess the robustness and accuracy of the surrogate model, the computational framework was tested on three main scenarios of increasing complexity. The aim was to analyse the network's response, starting from a base case and progressing to a highly

realistic simulation, gradually increasing the number of parameters varied during the Monte Carlo simulation for the creation of the training and test datasets.

4.2.1 Classification of Scenarios

The three case studies differ in terms of injection topology and the degree of freedom left to the input variables:

- Case Study 1 (Low Complexity - 4 Parameters): One of the two sources injects a mixture of hydrogen and natural gas, while the other injects pure natural gas. In this base scenario, the stochastic perturbation acts on four global variables: the sources set pressure, the global consumption of the network (varying uniformly), the absolute roughness of the pipes and the percentage of hydrogen injection at the main source.
- Case Study 2 (Medium Complexity - 5 Parameters): Injects an H2NG mixture (dual blending) into the second source as well. Compared to the previous case, a fifth variable parameter is added: the concentration of hydrogen injected by the secondary source, making the mixtures fed into the two nodes independent of each other.
- Case Study 3 (High Complexity - 30 Parameters): This represents the most realistic and challenging scenario for the Machine Learning algorithm. In addition to the fluid dynamics and gas quality parameters seen above, the set-point pressures are varied independently and the gas demand is no longer varied uniformly. Consumption has been divided into 25 independent clusters. The variation in the load of each cluster is given by the superposition of two components: a percentage variation common to all clusters (which simulates seasonal fluctuation) and an independent variation for each cluster (which models the variability linked to the different consumption habits of local users).

4.2.2 Nominal Data and Operating Range Selection

To ensure the industrial validity of the study, the ‘nominal’ consumption profiles used as a starting point for the simulations were extrapolated from real data. Specifically, it used the 2023 annual dataset for the city of Riccione, provided by the Distribution System Operator (DSO) INRETE.

Since the fluid dynamics of the network were analysed under steady-state conditions, it was necessary to identify a reference operating point. It was decided to stress the system by analysing it under the most demanding operating conditions: the nominal operating point therefore corresponds to the maximum consumption point recorded throughout the entire year.

The amplitude of the disturbance (the range of consumption variation to be explored using the Monte Carlo method) was defined by analysing the actual load trend provided by the DSO. A time window of three critical winter days (from 19/01/2023 to 21/01/2023) was isolated, during which the annual peak demand occurs. As can be seen from Figure 4.2, Riccione shows an average value of around 1.2 kg/s in that period. Compared to this average, withdrawals fluctuate, falling to a minimum value of around

0.6 kg/s and reaching maximum peaks of around 1.8 kg/s. This trend shows a load fluctuation of approximately ±50% compared to the daily average value.

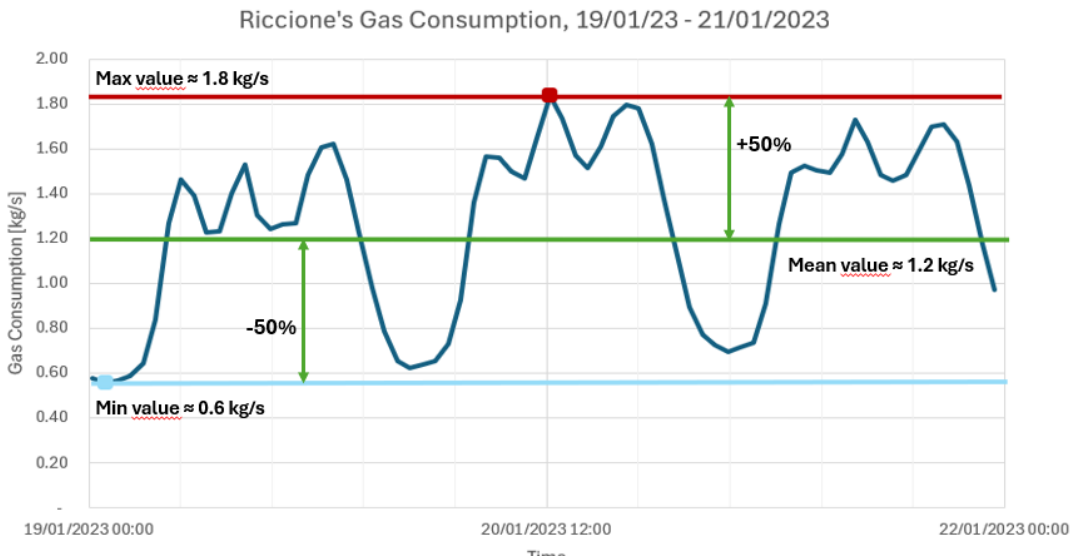


Figure 4.2. Gas consumption profile of the Riccione distribution network during three peak winter days (January 19-21, 2023), highlighting the ±50% fluctuation around the mean value.

Based on this observation, it was decided to adopt a consumption variation range of ±50% for the generation of the dataset. However, in order to test the model under stress conditions, this percentage variation was not applied to the average value, but rather to the absolute maximum consumption value recorded during the year. This methodological approach ensures that the surrogate model is trained on scenarios that simulate severe winter conditions and withdrawals beyond standard operating conditions.

4.2.3 Input Parameters and Perturbation Ranges

To clarify the experimental setup, the input parameters selected as stochastic variables during the Monte Carlo simulations are summarized below, along with their percentage variation ranges for each case study:

1) 4 input parameters with one hydrogen injection:
    - Par. 1) Entry pressures at both ReMi stations (with a variation $\Delta P = \pm 7\%$).
    - Par. 2) H_2 blending at the primary ReMi station ($10\% \pm \Delta H = \pm 95\%$).
    - Par. 3) Pipeline roughness ($\Delta\varepsilon = \pm 10\%$).
    - Par. 4) Exit flow to Riccione users ($\Delta G_{ext} = \pm 50\%$).
2) 5 input parameters with two hydrogen injections:
    - Par. 1) Entry pressures at both ReMi stations (with a variation $\Delta P = \pm 7\%$).
    - Par. 2) H_2 blending at the primary ReMi station ($10\% \pm \Delta H = \pm 95\%$).
    - Par. 3) H_2 blending at the secondary ReMi station ($10\% \pm \Delta H = \pm 95\%$).
    - Par. 4) Pipeline roughness ($\Delta\varepsilon = \pm 10\%$).

- Par. 5) Exit flow to Riccione users ($\Delta G_{ext} = \pm 50\%$).

3) 30 input parameters test case with two hydrogen injections:
    - Par. 1) Entry pressure at ReMi station 1 (with a variation $\Delta P = \pm 7\%$).
    - Par. 2) Entry pressure at ReMi station 2 (with a variation $\Delta P = \pm 7\%$).
    - Par. 3) H_2 blending at the primary ReMi station ($10\% \pm \Delta H = \pm 95\%$).
    - Par. 4) H_2 blending at the secondary ReMi station ($10\% \pm \Delta H = \pm 95\%$).
    - Par. 5) Pipeline roughness ($\Delta \varepsilon = \pm 10\%$).
    - Par. 6-30) Exit flow to 25 Riccione clusters of users (common variation = ±35% + independent variation = ±15%, maintaining an overall maximum range of $\Delta G_{ext} = \pm 50\%$).

4.3 Machine Learning Outcomes

The Machine Learning methodology described above was applied to generate a surrogate model capable of predicting the main fluid dynamic variables of the network. Specifically, the framework produced predictions for each of the following variables:

- P – nodal pressures.
- FR - pipeline mass flow rates.
- G_{ext} - mass flows exchanged with the external environment (representing actual injections and consumption)).
- V - gas velocity in the pipes.

To evaluate and visually illustrate the performance and accuracy of the surrogate model, scatter plots were produced for each of the physical variables mentioned. Within these graphs, the x-axis shows the “real” values calculated by the traditional physical solver and belonging to the test dataset, while the y-axis shows the corresponding predictions generated by the machine learning model. In this representation, a perfect alignment of the points along the quadrant bisector indicates a zero prediction error.

In addition to the fluid dynamics analysis, the model's predictions for exchanged flow rates and flow rates in the pipes were used as input for calculating nodal hydrogen concentrations. To validate the robustness of this hybrid approach and verify the accuracy of gas quality tracking, an additional scatter plot specifically dedicated to hydrogen mixtures was produced.

The results of this analysis are presented and discussed in the following paragraphs, organized to analyse the three case studies of increasing complexity defined above.

4.3.1 4 Input Parameters with One Hydrogen Injection

For the analysis of this first operating scenario, characterised by four variable input parameters and a single hydrogen injection, the model was trained and validated on a total dataset of 500 Monte Carlo simulations ($N_{MC} = 500$). Specifically, the training

phase required the use of only 200 samples (N_{ED} = 200), showing the remarkable efficiency of the algorithm in learning network dynamics with a limited amount of data.

Figure 4.3 illustrates the behaviour of surrogate models for fluid dynamic variables. The alignment of points along the bisector is almost perfect for all four variables analysed: pressure (P), external flow rate (G_{ext}) flow rate in the branches (FR) and velocity (V). This result certifies the remarkable ability of Machine Learning to replicate the complex non-linear physical equations that govern gas flow. It should be noted that, within the graph relating to G_{ext}, the negative values represent the injection flow rates at the network inlet, calculated at the supply points (ReMi cabins).

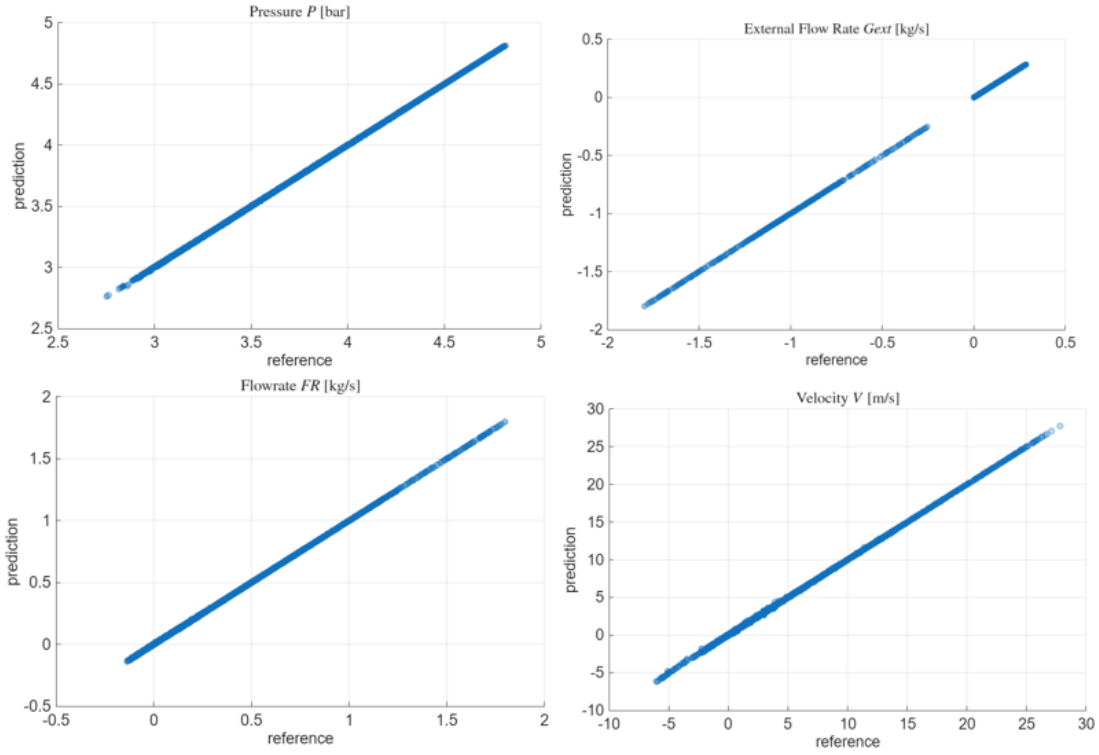


Figure 4.3. Scatter plots of predicted versus reference values for nodal pressures (P), external gas exchanges (G_{ext}), pipeline mass flow rates (FR) and gas velocities (V) in Case Study 1. The surrogate model demonstrates nearly perfect accuracy across the parameter variation range (N_{MC} = 500, N_{ED} = 200).

In the next step, the values predicted by the surrogate model for branch flow rates (FR) and node exchanges (G_{ext}) were used as inputs in the mixing formula for species mass balance. This step provided the prediction of the node hydrogen concentration (Y_{H2}).

The result of this gas quality tracking is shown in Figure 4.4. The scatter plot shows a good overall alignment between the predicted values and the reference values calculated by the physical solver. Although there are a slight dispersion and the presence of some outliers compared to the absolute accuracy obtained for the fluid dynamic variables, the overall accuracy remains widely acceptable, confirming the validity of this approach for the specific network configuration analysed.

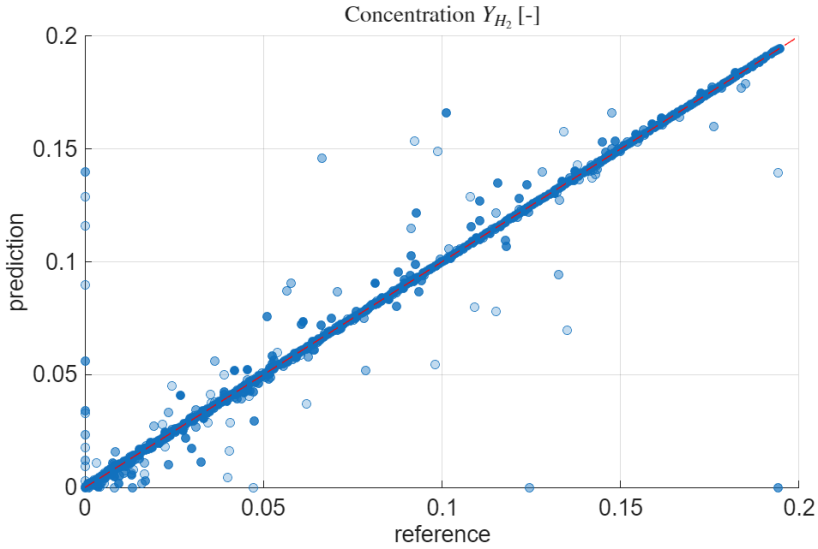


Figure 4.4. Scatter plot of predicted versus reference values for nodal hydrogen concentrations (YH2) in Case Study 1. The nodal concentrations are deterministically calculated using the ML-predicted fluid dynamic variables (FR and G_{ext}).

4.3.2 5 Input Parameters with Two Hydrogen Injections

This second scenario is a direct extension of the previous case, in which hydrogen-containing mixtures are fed into the network from both entry points (ReMi cabins) and their respective molar fractions can be adjusted completely independently. It is important to note that, despite the slight increase in complexity introduced by the fifth stochastic parameter, the size of the overall dataset and the number of samples used for training were kept identical to the base case, i.e. 500 Monte Carlo simulations (N_{MC} = 500) e 200 training samples (N_{ED} = 200).

Figure 4.5 shows the scatter plots for purely fluid dynamic variables. As can be seen from the graphs, the alignment of the points along the bisector remains excellent. This demonstrates that the surrogate model is able to handle the addition of an extra input parameter, accurately representing the complex fluid dynamics of the network without requiring an expansion of the training database.

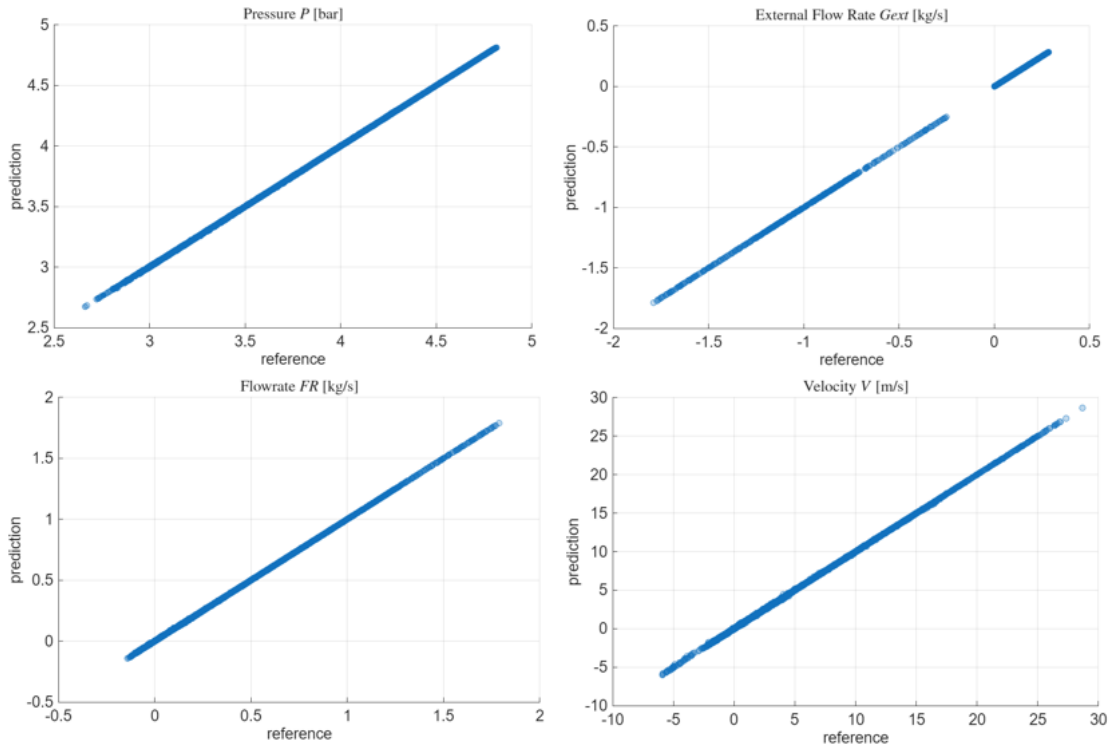


Figure 4.5. Scatter plots of predicted versus reference values for nodal pressures (P), external gas exchanges (G_{ext}), pipeline mass flow rates (FR) and gas velocities (V) in Case Study 2. Despite the addition of a fifth input parameter, the surrogate model maintains high predictive accuracy using the same dataset size as the baseline scenario (N_{MC} = 500, N_{ED} = 200).

Subsequently, fluid dynamics predictions were used to calculate the nodal concentrations of hydrogen resulting from the double injection. Figure 4.6 shows the scatter plot for the Y_{H2} variable. Even in this dual blending scenario, where the interaction between flows from two different sources makes calculating gas quality more complex, the model maintains remarkable predictive robustness.

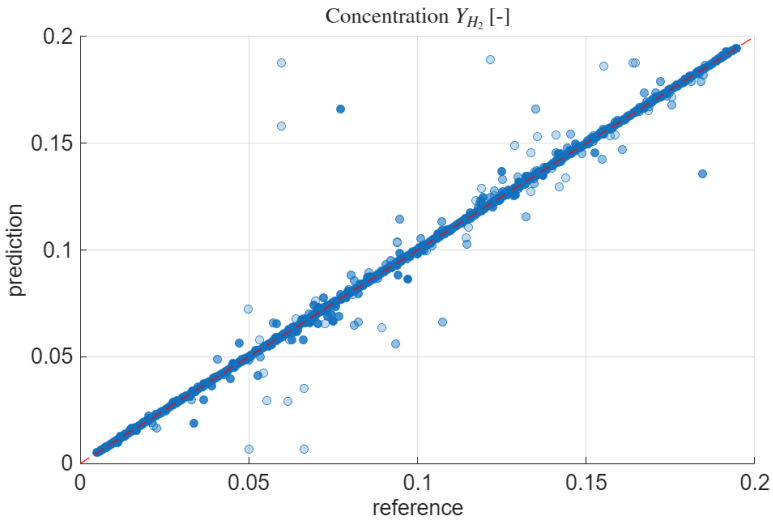


Figure 4.6. Scatter plot of predicted versus reference values for nodal hydrogen concentrations (Y_{H2}) in Case Study 2. The plot demonstrates the model's ability to accurately track gas quality in a dual blending scenario, achieving good results without increasing the computational cost of the training phase.

4.3.3 30 Parameters with Independent Cluster Variations

In this third case study, the test is further expanded by allowing independent variation of set-point pressures and partially independent variation of consumption divided into 25 clusters. This means testing the gas infrastructure over an extremely wide and general operating domain.

This high variability and the high dimensionality of the problem (30 parameters) explain the need to increase the size of the dataset compared to previous cases. In an initial exploratory phase, the model was trained keeping N_{ED} = 200 (with N_{MC} = 1000); however, as expected for a 30-dimensional problem, the predictive accuracy was not optimal. As can be seen in Figure 4.7, the bisectors remain evident but have become significantly thicker due to the increased complexity of the problem. Therefore, the size of the training sample was more than doubled, bringing the training samples to N_{ED} = 500.

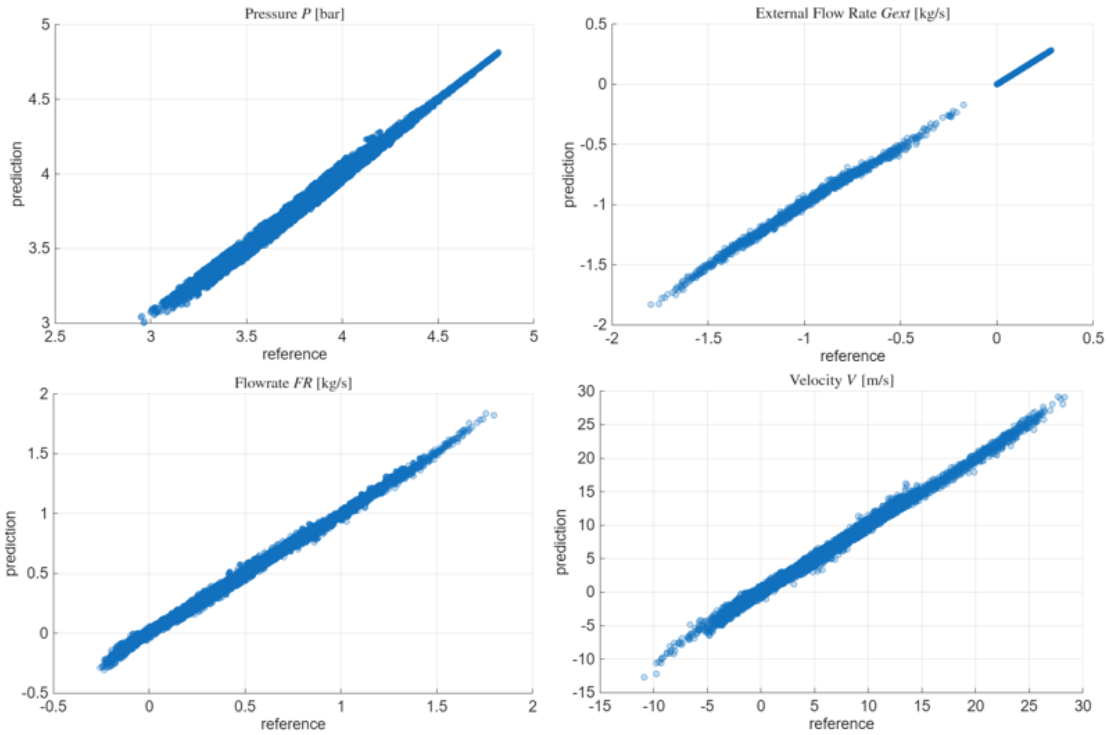


Figure 4.7. Scatter plots of predicted versus reference values for the fluid-dynamic variables and hydrogen concentration in Case Study 3, using a preliminary training set of N_{ED} = 200. The high dimensionality of the problem leads to a degradation in prediction accuracy.

With the dataset extended to N_{ED} = 500, the model's performance improves. The scatter plots in Figure 4.8 show that, on the fluid dynamic side (pressures, external exchanges, flow rates in the branches and velocities), the predictions are once again sufficiently aligned with the reference values of the physical model, but producing thicker bisecting scatter plots than in the two previous cases.

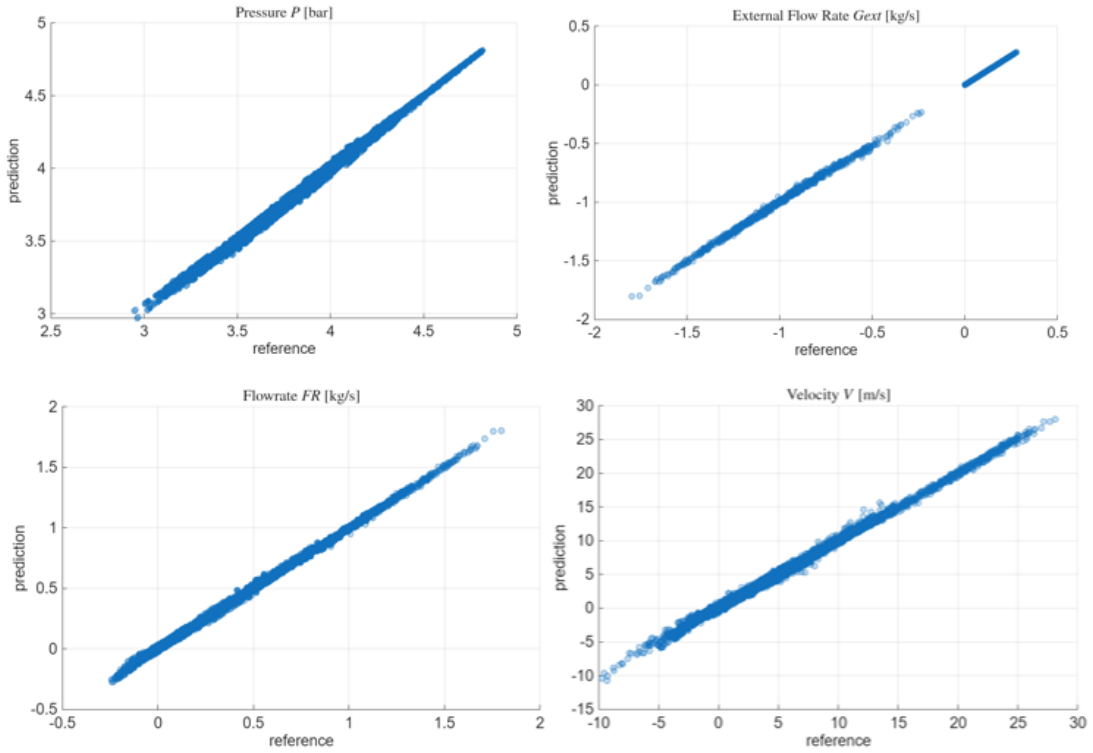


Figure 4.8. Scatter plots of predicted versus reference values for nodal pressures (P), external gas exchanges (G_{ext}), pipeline mass flow rates (FR) and gas velocities (V) in Case Study 3 (N_{MC} = 1000, N_{ED} = 500). The predictions are highly accurate despite the 30-parameter variability.

As regards the scatter plot of nodal hydrogen concentrations (Y_{H2}), shown in Figure 4.9, most samples lie close to the bisector, indicating good overall agreement. However, a limited number of points show more marked deviations. This dispersion is due to the effect of the independent fluctuations in the 25 consumption clusters and source set point pressures which continuously shift the zero flow points and mixing zones. In these areas, the formula for calculating the nodal hydrogen concentration amplifies the micro-uncertainties present in the flow rate predictions, generating deviations in the local concentration.

Further attempts to narrow this prediction band would provide limited practical benefit, especially considering the primary objective of keeping the training dataset within modest dimensions (in the order of a few hundred samples).

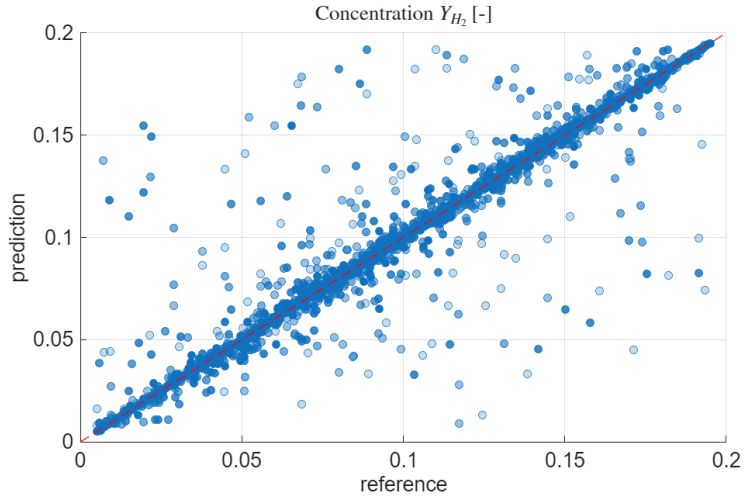


Figure 4.9. Scatter plot of predicted versus reference values for nodal hydrogen concentrations (Y_{H2}) in Case Study 3 (N_{MC} = 1000, N_{ED} = 500). While most points align with the bisector, some dispersion is visible due to the shifting of zero flow points caused by independent cluster variations.

To verify the impact of these discrepancies in gas quality, the Root Mean Square Error (RMSE) was calculated for the hydrogen concentration in each node i, defined as:

$$RMSE_i = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\hat{y}_{i,k} - y_{i,k}\right)^2} \tag{4.4}$$

Where n is the number of validation samples, $\hat{y}_{i,k}$ is the predicted hydrogen concentration value in the i-th node for the k-th test sample and $y_{i,k}$ is the corresponding value obtained by the physical solver.

The left panel of Figure 4.10 shows the spatial distribution of the RMSE for each node in the network. The predictive error is not distributed uniformly, but is localised in specific areas which, from a topological perspective, correspond exactly to the connections between the network loops (the mixing zones). However, the magnitude of the RMSE remains small when compared to the operating range of concentrations (which varies from 0.5% to 19.5%), remaining below 1.2% in the worst nodes. The panel on the right illustrates the distribution of the RMSE across the nodes, highlighting that approximately 20% of the network (just over 40 nodes out of 241) has an RMSE greater than 0.6%.

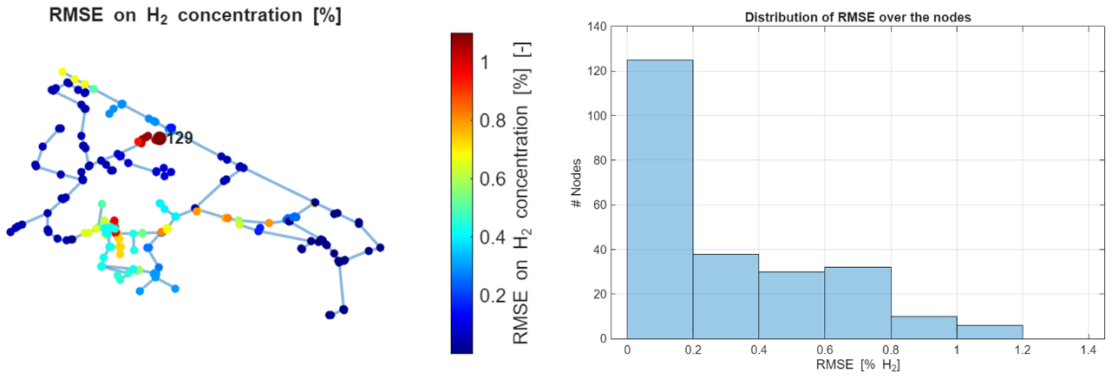


Figure 4.10. Spatial distribution of the RMSE on H_2 concentration across the network (left) and the corresponding histogram (right). The maximum errors are topographically localized at loop connections, with most nodes exhibiting an RMSE well below 0.6%

For a more in-depth analysis, the node with the maximum absolute RMSE value (Node 129) was examined. The absolute error between prediction and reference was evaluated for all validation samples. The normalised distribution in Figure 4.11 (left) shows that the frequency of the highest errors (around 10-12% deviation) is in the order of 10^{-3}, making them practically negligible and comparable to rare outliers. Finally, Figure 4.11 (right) shows the overlap of the Probability Density Function (PDF) of the concentrations calculated in Node 129 by the physical model compared to those predicted by the surrogate model. The excellent fit of the two distributions confirms that the surrogate model has excellent overall predictive power, demonstrating that the few points far from the diagonal in the previous scatter plot do not affect the statistical validity of the framework, which is sufficiently accurate even for this complex operational scenario.

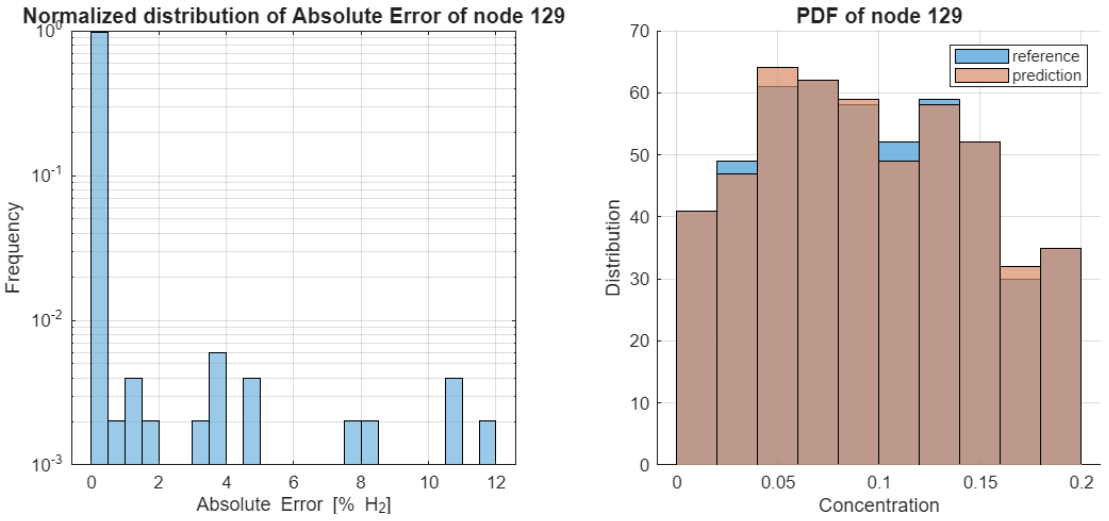


Figure 4.11. Normalized distribution of Absolute Error (left) and comparison of the Probability Density Function (right) for Node 129, which exhibited the highest RMSE. The PDF overlap demonstrates the high statistical reliability of the surrogate model.

4.4 Surrogate Model Applicability

Once the surrogate model has been trained and its accuracy validated, it can be effectively used to replace the slower physical model in performing rapid variational

analyses. This approach allows for instantaneous exploration of a large domain of operating conditions, assessing the sensitivity of the gas infrastructure to parameter variations and mapping the statistical behaviour of the entire network topology. It is important to point out that a specific surrogate model was developed for each application case, due to variations in the type and number of parameters. The case studies are presented in increasing order of complexity and generalisation.

4.4.1 Case Study 1: 4 Parameters with Single H2 Injection

This scenario represents a distribution network in which one of the two entry points (node 240) is characterised by the injection of a mixture of natural gas and hydrogen, while the second (node 241) supplies only pure natural gas. Since the surrogate model was trained on a uniform distribution of hydrogen composition (with an average value of 10% and a variation of ±95%), it is possible to investigate an extremely wide mixing domain. As for the fluid dynamic boundary conditions, the pressure set points at the two ReMi cabins vary simultaneously by ±7%. Since this variation is governed by a single parameter (rigid variation), the ratio between the two inlet pressures remains constant in all configurations; consequently, the cabin with the higher set point will always impose the macroscopic direction of the main flow. Finally, the consumption of the entire network varies rigidly by ±50%: this assumption simulates homogeneous user behaviour.

4.4.1.1 Statistical Mapping of Fluid-Dynamics and Gas Quality

To visualise the impact of these variations, the surrogate model was used to predict the network response on a set of 500 parameter configurations randomly generated from these variations.

Figure 4.12 shows the topology of the Riccione network mapped with the mean values and standard deviations of the node pressures. The standard deviation reflects the variability associated with parameter uncertainty and is a fundamental statistical metric for quantifying the dispersion of data around their mean value. The nodes highlighted in red in the standard deviation graph correspond to the peripheral areas of the network where the impact of the combined variations in inlet pressure and withdrawals is most significant, causing the greatest pressure fluctuations.

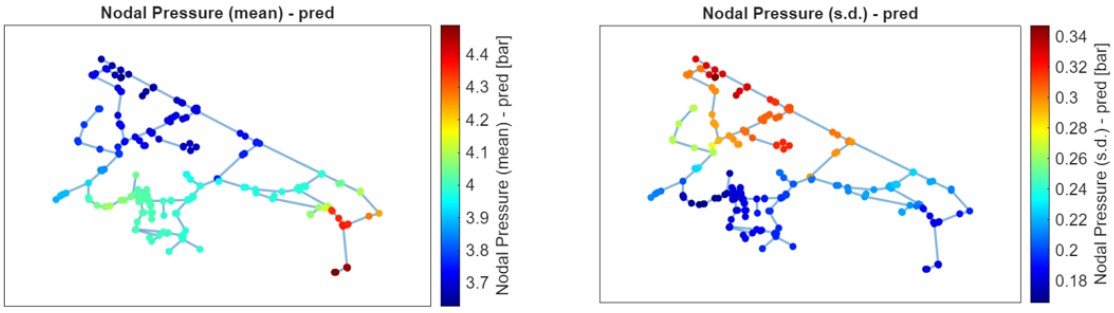


Figure 4.12. Spatial distribution of the mean values (left) and standard deviations (right) of nodal pressures, predicted by the surrogate model for Case Study 1.

Similar considerations can be made by analysing Figure 4.13, which illustrates the maps relating to mass flow rates in the branches and gas velocities.

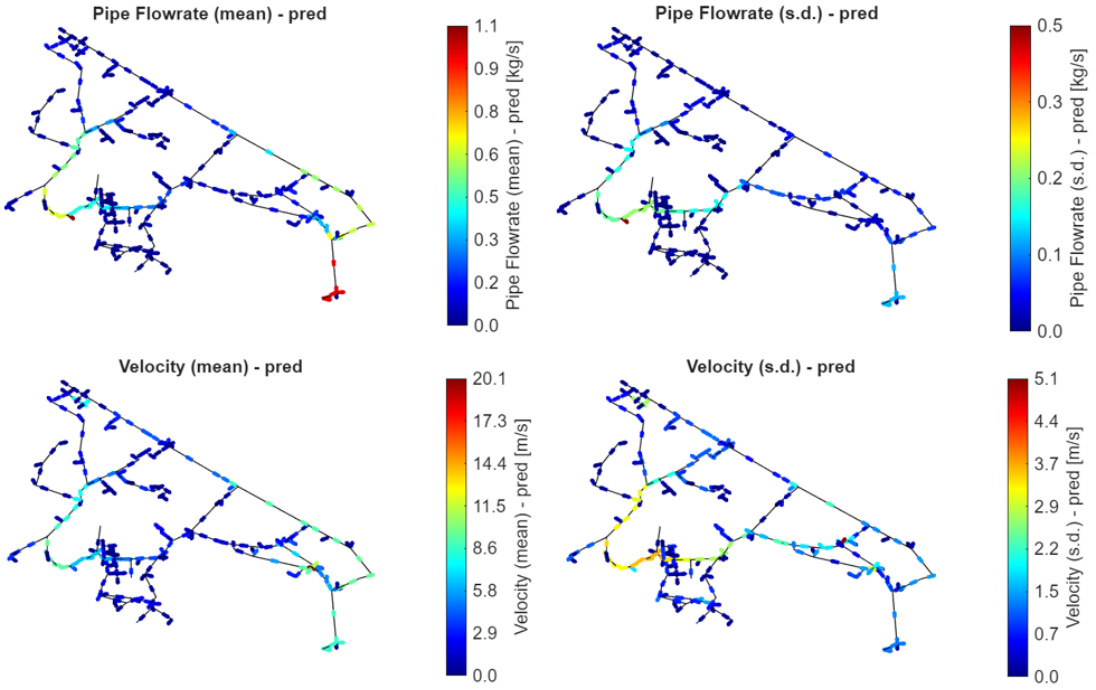


Figure 4.13. Spatial distribution of mean values and standard deviations for pipe flow rates (top) and gas velocities (bottom) across the network topology.

The analysis becomes interesting by observing the propagation of hydrogen within the infrastructure. Figure 4.14 shows the average nodal concentration of H_2 and the relative standard deviation.

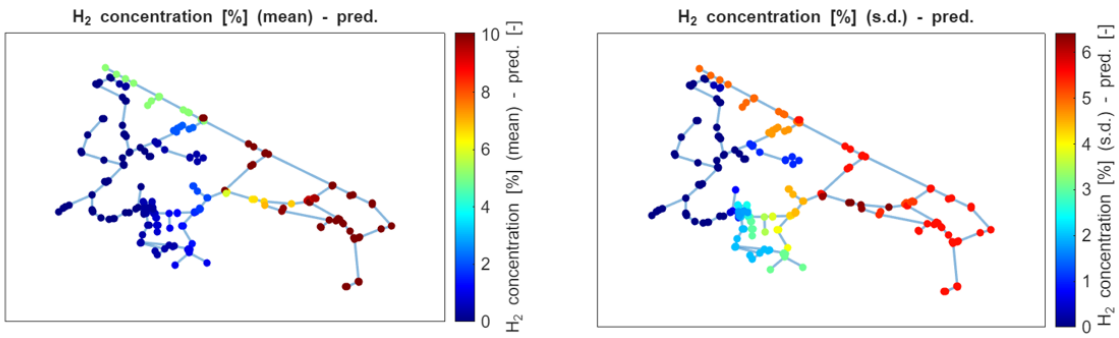


Figure 4.14. Spatial distribution of the mean values and standard deviations of nodal hydrogen concentration (Y_{H2}).

From the map of average values (left), it is clear that across the entire range of pressure and consumption variations, the network is divided into three distinct macro-areas in terms of gas quality, plus a fourth area that can be defined as a 'transition' area. By combining this information with the standard deviation map (right), it is possible to deduce not only the amplitude of the variation, but also the topological rigidity of these areas:

- Blue Area (0% average H_2): The nodes in this portion of the network show an average concentration of zero and a standard deviation of zero across the entire range of variation. This means that this area is powered exclusively by the pure natural gas source (node 241) and behaves, in terms of concentration, exactly like the source node itself.

- Dark Red Area (10% average H_2): This portion exhibits an average concentration of 10% and a standard deviation of approximately 5.5%. Since these values replicate the stochastic behaviour of the blending injection source (node 240) identically, it can be seen that the entire group of nodes belonging to this area is dominated only by the mixed flow, without undergoing dilution.
- Green Area (~5% average hydrogen): Located mainly in the northern part of the network, this is an area where physical mixing between the two flows always occurs. The resulting concentration (which averages around 5%) depends on the hydrogen fractions injected at the sources and the overall fluid dynamics of the system.
- Transition Area (Yellow/Orange/Blue): Unlike the previous areas, this group of nodes exhibits highly heterogeneous behaviour. These nodes do not have rigid mixing behaviour but represent a moving boundary between the two adjacent concentration areas. Depending on the specific fluid dynamic conditions of the network at a given moment, they can assume different states: receiving mixed gas (such as node 240), pure gas (such as node 241) or acting as an effective point of collision between the flows.

A quantitative visualisation of this partition is provided by the bar charts in Figure 4.15. The information scattered throughout the network topology is reordered in descending order of average hydrogen concentration to facilitate visualisation of the behaviour of each individual node. For graphical reasons it would be difficult to indicate all the node ID labels in the network However, the nodes labelled in the x-axis can be located in the specific zones on the network map following the same colour code.

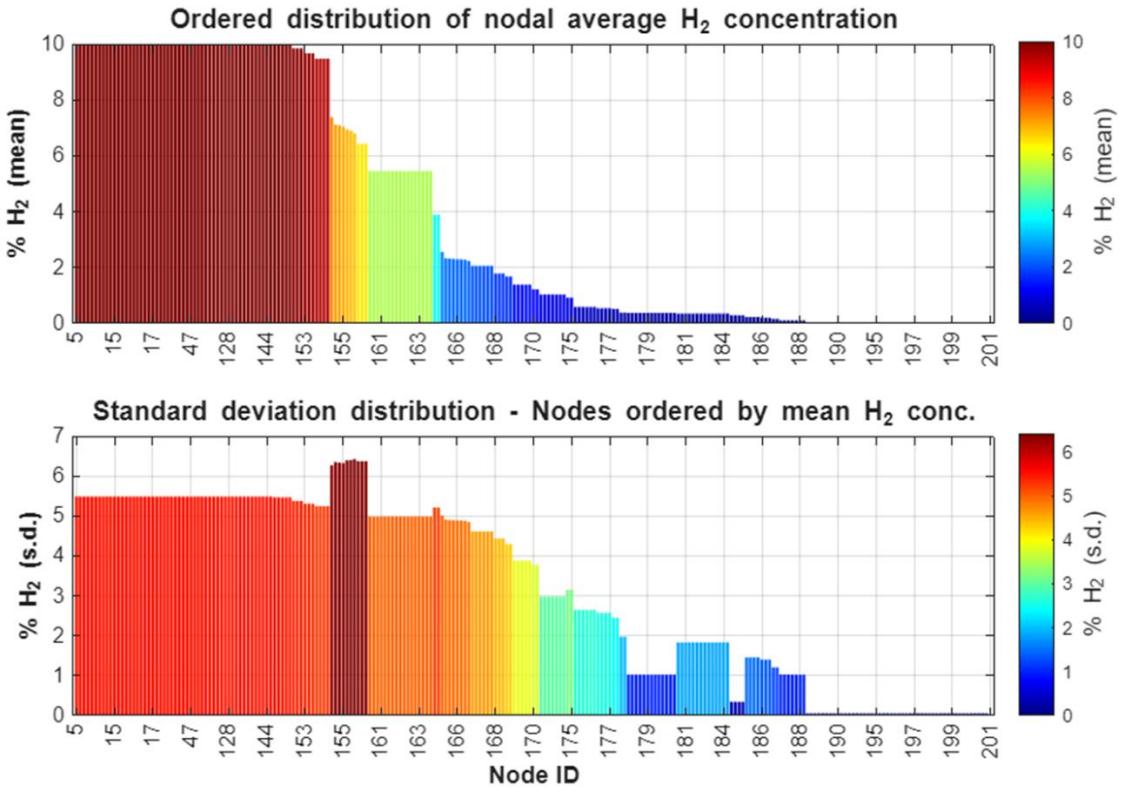


Figure 4.15. Ordered distribution of nodal average H_2 concentration (top) and the corresponding standard deviation distribution (bottom) for the network nodes.

The upper graph visually confirms the existence of the areas described above: the large block of nodes fuelled by pure mixture (dark red bars), the constant mixture area (green bars), the pure natural gas area (dark blue bars on the right) and the transition 'staircase' (yellow, orange and light blue bars). The lower graph of standard deviations further supports the analysis quantitatively: there is a specific peak in standard deviation that even exceeds the variability of the injection source. This peak mathematically identifies some of the nodes in the transition area, demonstrating how the collision of gas fronts generates very high statistical dispersion, confirming the mobile and unstable nature of this operating boundary.

4.4.1.2 Nodal Frequency Analysis and Mixing Probabilities

A more detailed analysis (node by node) can be conducted by examining the probability distribution that a specific node will have a given concentration with respect to one of the sources. In this context, it is particularly interesting to evaluate the behaviour of the nodes in relation to the ReMi cabin where the hydrogen mixture is injected (node 240).

By exploiting the computational speed of the surrogate model, a dataset of 1000 nodal concentration vectors was generated, derived from the same number of stochastic combinations of the four input parameters. Based on these results, the Probability Density Function (PDF) for the network nodes was calculated. To illustrate the mixing dynamics, the analysis focused on a specific sequence of contiguous nodes highlighted in Figure 4.16. This branch was selected because it crosses a transition area identified in the previous paragraph: starting from the nodes furthest from the hydrogen source (and closest to the pure natural gas inlet), the path gradually moves towards the nodes closest to the blending injection.

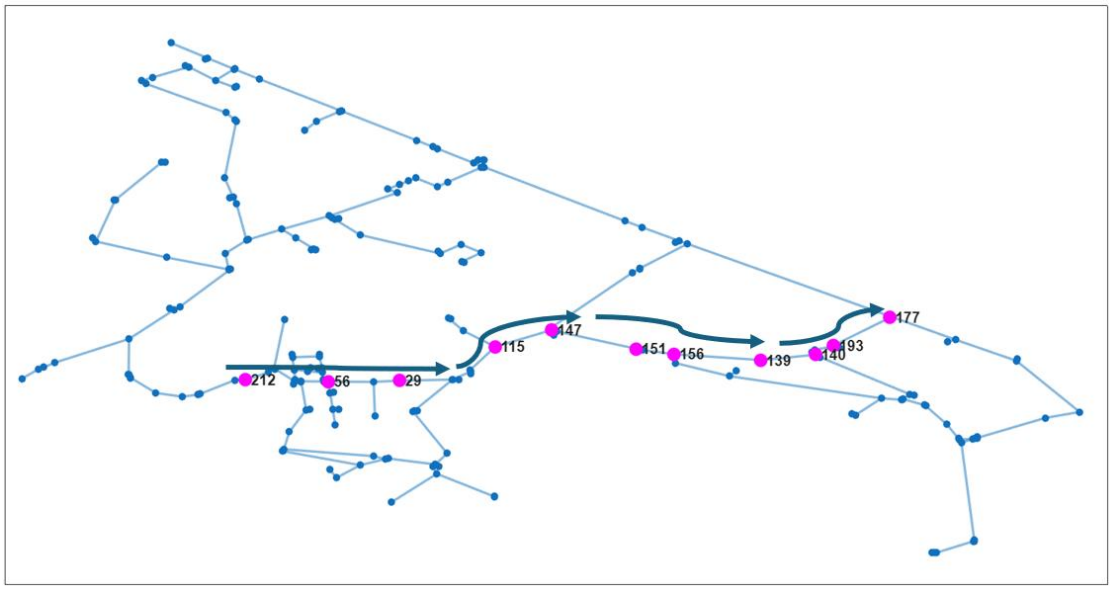
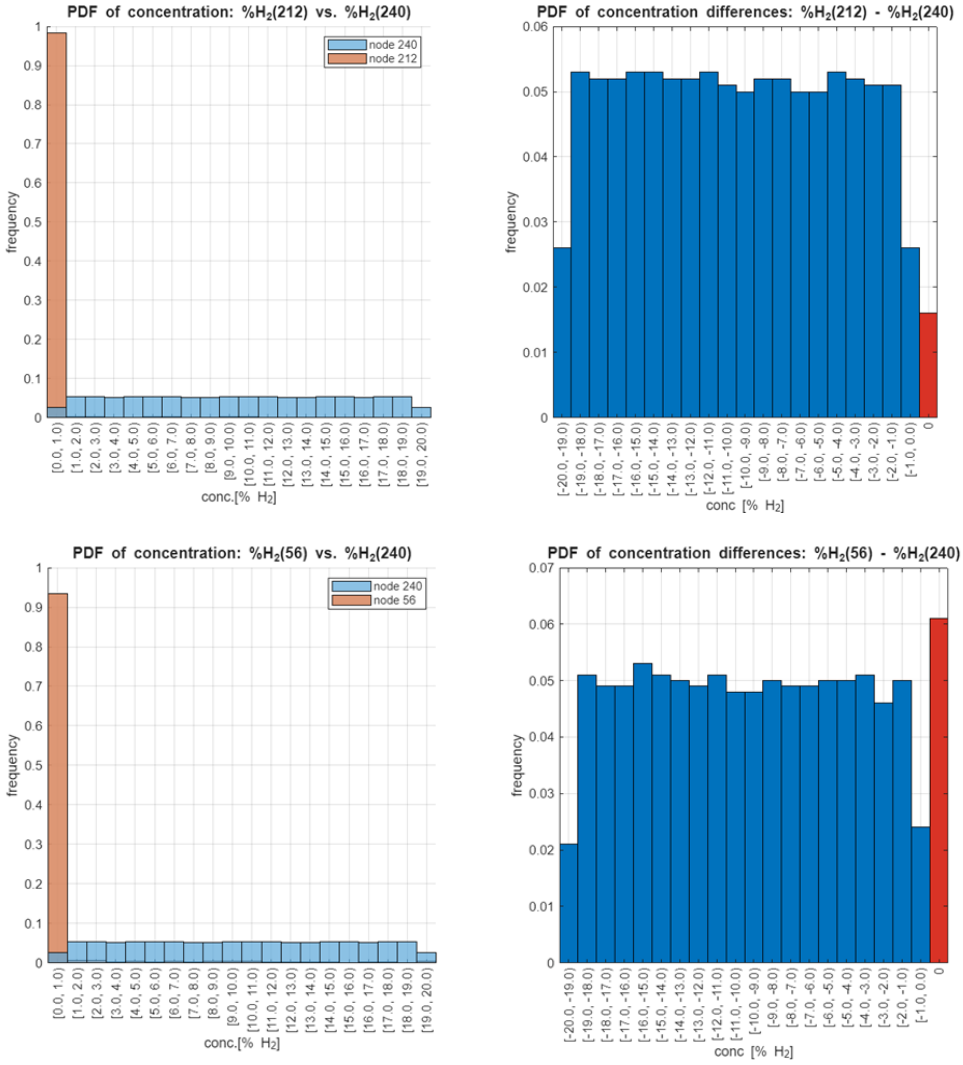


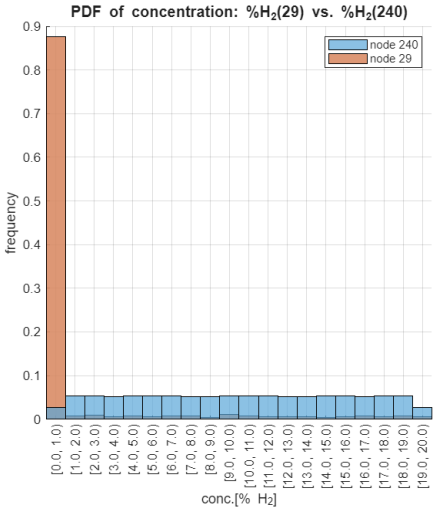
Figure 4.16. Topological layout highlighting the selected branch of nodes (from node 212 to 177) across the transitional area, analysed for hydrogen concentration frequencies.

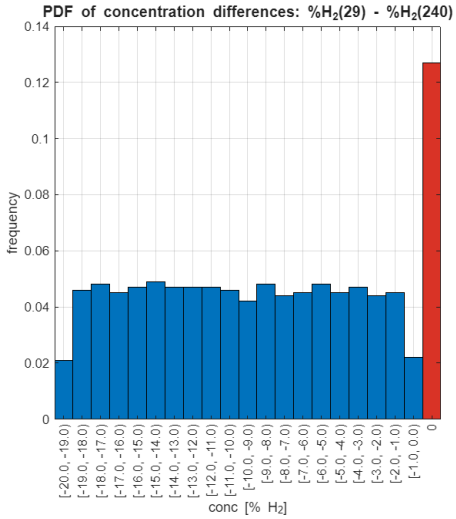
For each of the selected nodes, two side-by-side graphs were produced, shown vertically in succession in Figure 4.17.

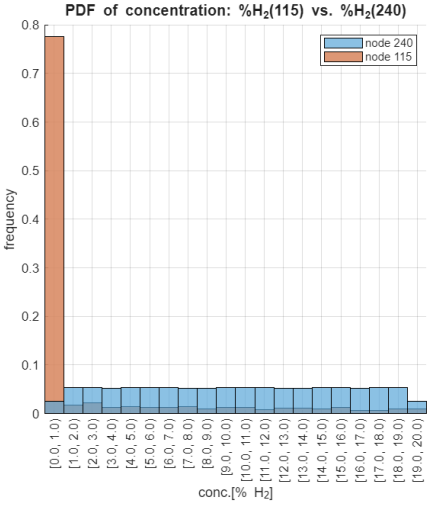
The left panel of each row compares the PDF of the absolute hydrogen concentration of the node under consideration (orange bars) superimposed on the PDF of source node 240 (blue bars). The frequency is sampled in 1% bins, in the range from 0% to 20%. It is essential to remember that the concentration at source 240 is itself a uniform stochastic variable in the domain analysed.

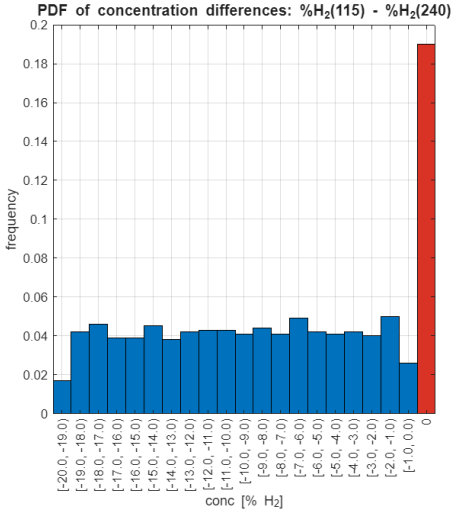
The right-hand panel shows the PDF of the concentration difference between the analysed node and the source node ($Y_{H2,node}$ - $Y_{H2,240}$). In this graph, a special bin (highlighted in red) counts the occurrences where this difference is zero. This parameter is crucial: when the difference is zero, it means that in that specific operating configuration, the node is not undergoing any mixing, but is behaving exactly like source node 240, receiving the pure mixture.

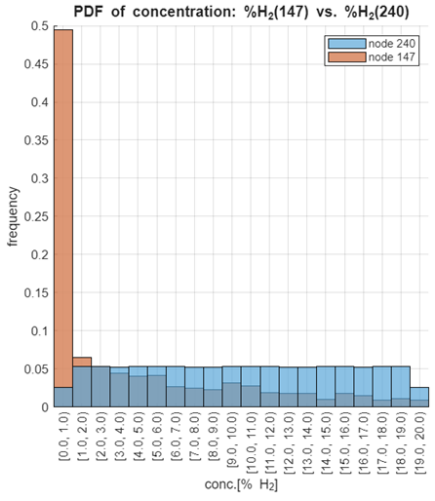


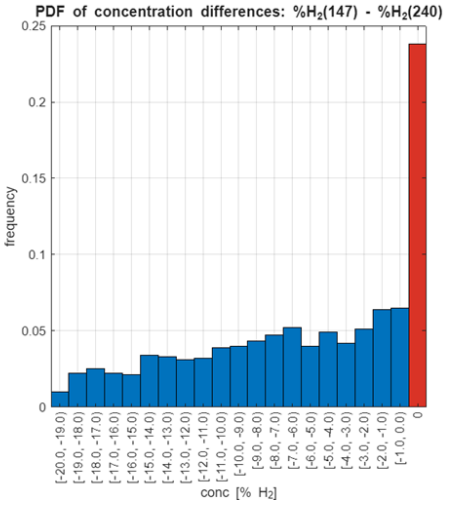

PDF of concentration: %H2(29) vs. %H2(240)
node 240
node 29
frequency
conc.[% H2]

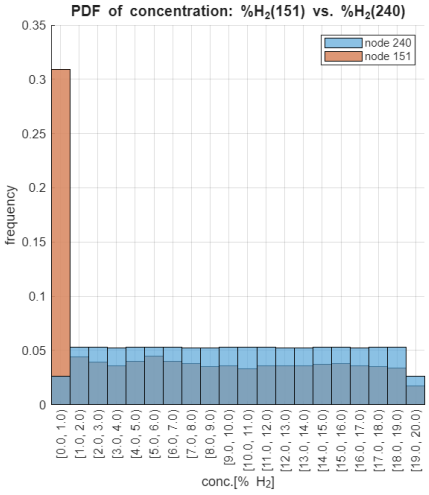

PDF of concentration differences: %H2(29) - %H2(240)
frequency
conc [% H2]

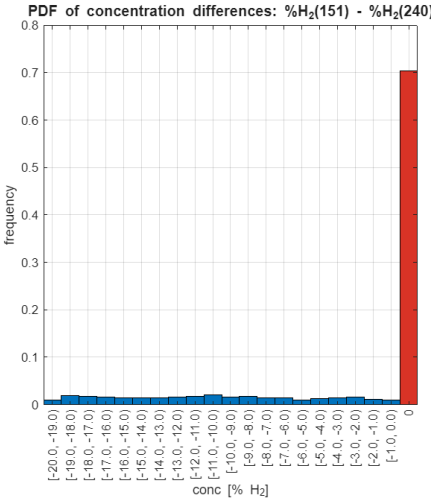

PDF of concentration: %H2(115) vs. %H2(240)
node 240
node 115
frequency
conc.[% H2]

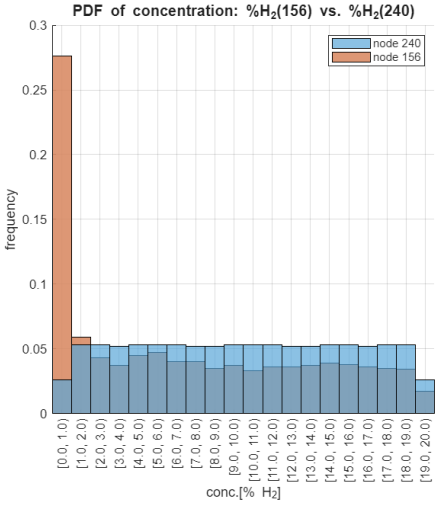

PDF of concentration differences: %H2(115) - %H2(240)
frequency
conc [% H2]

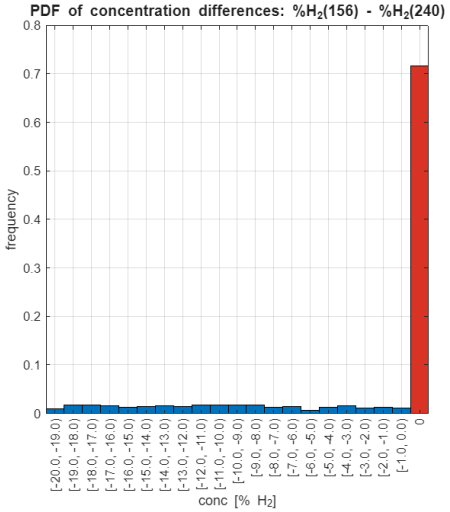

PDF of concentration: %H2(147) vs. %H2(240)
node 240
node 147
frequency
conc.[% H2]

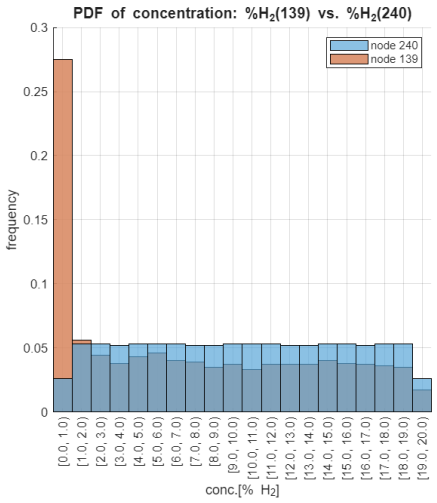

PDF of concentration differences: %H2(147) - %H2(240)
frequency
conc [% H2]

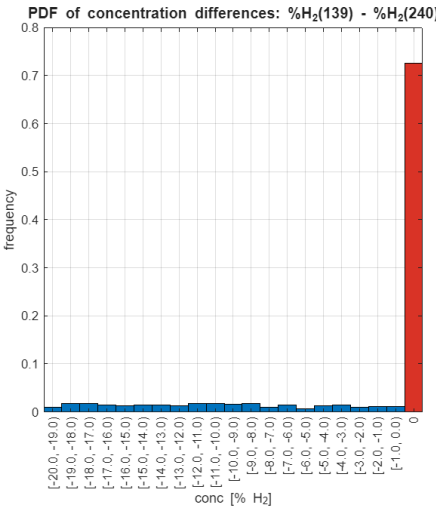

PDF of concentration: %H2(151) vs. %H2(240)
node 240
node 151
frequency
conc.[% H2]


PDF of concentration differences: %H2(151) - %H2(240)
frequency
conc [% H2]


PDF of concentration: %H2(156) vs. %H2(240)
node 240
node 156
frequency
conc.[% H2]


PDF of concentration differences: %H2(156) - %H2(240)
frequency
conc [% H2]


PDF of concentration: %H2(139) vs. %H2(240)
node 240
node 139
frequency
conc.[% H2]


PDF of concentration differences: %H2(139) - %H2(240)
frequency
conc [% H2]

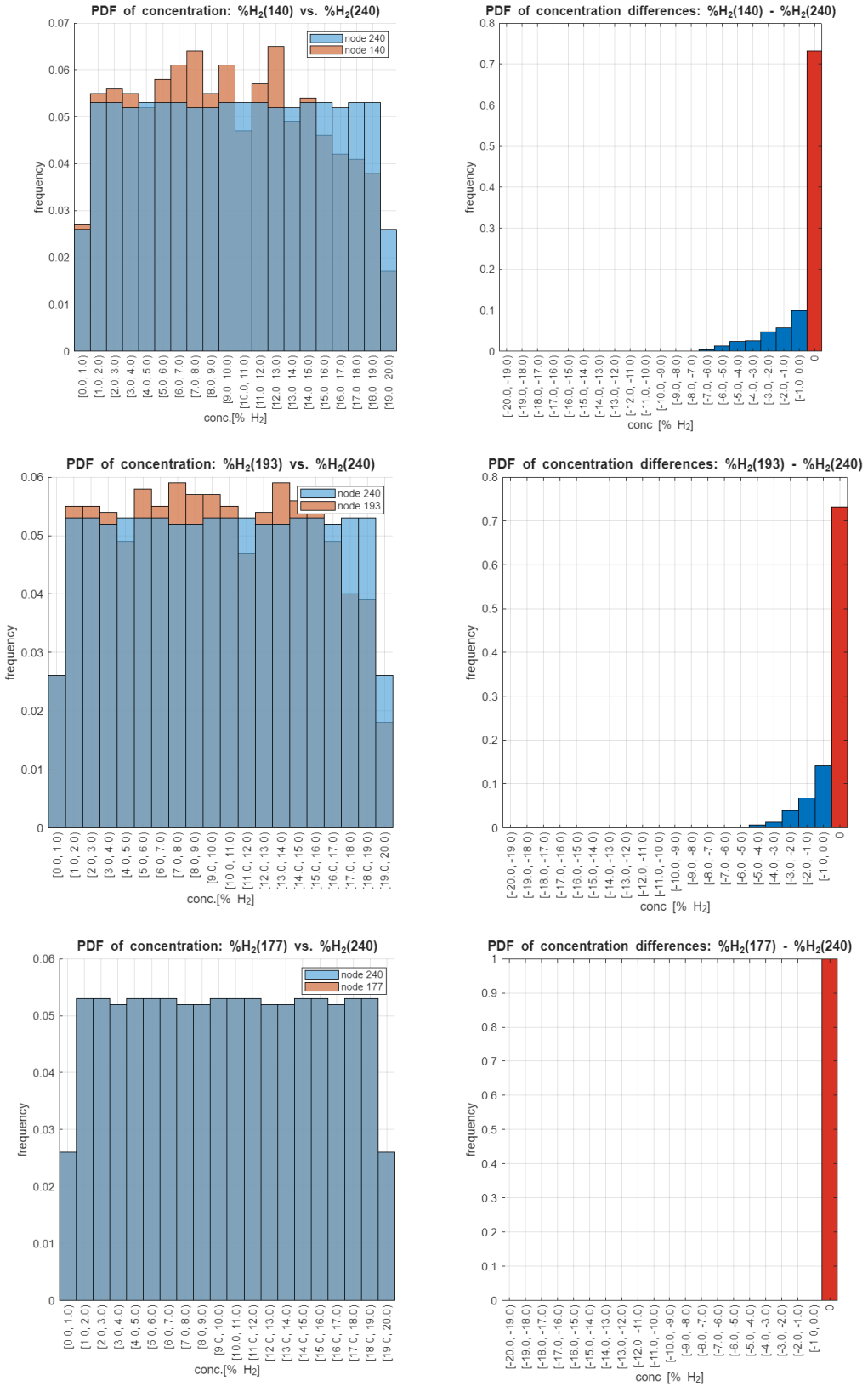


Figure 4.17. Probability Density Functions of absolute H_2 concentration (left panels) and concentration differences with respect to the source node 240 (right panels) for the selected branch of nodes (from 212 to 177). The vertical sequence illustrates the spatial transition of the gas mixing front, showing how the probability of receiving pure blended gas (red bars) increases moving towards the hydrogen source.

Looking at the graphs for the nodes farthest from the hydrogen source (nodes 212, 56 and 29), the absolute PDF shows a very high peak in the first bin [0, 1). This indicates that for most parameter combinations, these nodes behave like the blue area (pure natural gas) and the red bar of zero difference is consequently smaller.

However, proceeding along the branch, we enter the transition zone. Already from node 147, the probability of having concentrations close to zero drops quickly. From the difference graph, we can see that the red bar begins to grow progressively and steadily. While for node 115 the probability of receiving a pure mixture from node 240 is only about 19%, passing through node 147 this rises to almost 30%, then surges and exceeds 70% in nodes 151 and 156. In these nodes, mixing occurs more frequently depending on the instantaneous fluid dynamics, oscillating between receiving diluted flows and receiving pure flows.

Continuing the analysis towards the nodes closest to the hydrogen source, the shape of the absolute PDF of nodes 193 and 177 begins to adapt and increasingly closely follow the uniform distribution of source node 240. In node 177, the probability that the concentration difference is zero reaches 100%, indicating that this node is permanently dominated by injection with blending, regardless of load variations.

4.4.1.3 Global Sensitivity Analysis via Sobol' Indices

Once the accuracy of the surrogate models has been validated, they can be used to perform a global sensitivity analysis aimed to quantify the impact of each input parameter on output variability. This assessment is particularly effective for identifying potential critical areas within the network, i.e. those areas where certain parameters have a predominant influence, leading to significant variations in the variables of interest.

Specifically, for this study, the Sobol' method, a variance-based sensitivity analysis approach, was applied. This method divides the total variance of the output into fractions that can be attributed to each input parameter considered separately.

Total-effect Sobol' indices are calculated according to the following formulation:

$$S_{T_i} = \frac{\mathrm{E}_{X_{\sim i}}\left(\mathrm{Var}_{X_i}(Y|X_{\sim i})\right)}{\mathrm{Var}(Y)} \tag{4.5}$$

These indices measure the contribution of the i-th input parameter x_i to the variance of the output, including all its interactions with the other input parameters. In the above formula, *Var(Y)* represents the total variance of the output y, while $Y|X_{\sim i}$ indicates the output y calculated by keeping all inputs constant except x_i. Consequently, the term in the numerator $E_{X_{\sim i}}\left(Var_{X_i}(Y|X_{\sim i})\right)$ represents the expected value of the variances with respect to x_i, calculated by setting the other inputs to a constant (but random) value.

In practice, the statistical quantities described above cannot be calculated analytically but are instead estimated using quasi-Monte Carlo methods. A common approach to improving the efficiency of estimators is to generate an appropriate low-discrepancy

sequence, such as the Sobol' sequence. The implemented algorithmic procedure consists of the following steps:

- Generation of a sampling matrix of size $N\ x\ 2p$, where p is the number of input parameters. This matrix is constructed using a Sobol' sequence with N points in a 2p-dimensional hyperspace, following the same statistical distribution as the input variables.
- Division of the matrix into two distinct sets of samples, named A and B, composed respectively of the first and last p columns of the original matrix.
- Construction of p new matrices A_B^i, each of size $N\ x\ p$, with $i\ =\ 1, \dots\ , p$. These matrices are composed in such a way that the i-th column comes from matrix B, while the remaining columns are taken from matrix A.
- Estimate of the index numerator S_{T_i} using the following approximation:

$$E_{\boldsymbol{X}_{\sim i}}\left(\mathrm{Var}_{X_i}(Y|\boldsymbol{X}_{\sim i})\right) \approx \frac{1}{2N}\sum_{j=1}^{N}\left(f(\boldsymbol{A})_j - f\left(\boldsymbol{A}_B^i\right)_j\right)^2$$

  where $f(A)_j$ e $f(A_B^i)_j$ represent the model evaluated for the j-th sample of set A and matrix A_B^i, respectively.

The length N of the sequence defines the accuracy of the estimator; typical values used are in the order of $N \sim 10^3\ or\ 10^4$. To complete this analysis, a total of $N(p+1)$ model evaluations are required: N evaluations for dataset A and N evaluations for each of the p datasets A_B^i.

This volume of calculation is prohibitive for the physical solver, but using the surrogate model, global sensitivity analysis can be completed at negligible computational cost.

The global sensitivity analysis is discussed below, applying the methodology based on Sobol' indices presented above. The outputs on which this assessment focuses are the nodal hydrogen concentrations relating to the same sequence of nodes analysed in the previous paragraph (from node 212 to node 177), which cross the transition area of the network.

The aim is to assess which of the following stochastic input parameters has the greatest impact on the variability of local concentration at these specific points:

- P: set point of nodal pressures at the inlet to the ReMi cabins.
- E: absolute roughness of pipes.
- G_{ext}: mass flows exchanged with the external environment.
- Y_1: hydrogen concentration in the mixture injected by the primary source (node 240).

The results of the Sobol' index calculation for the first half of the branch (from node 212 to node 151) are shown in Figure 4.18. In the case of nodes furthest from the hydrogen source and facing the pure natural gas area (such as nodes 212, 56, 29, 115 and 147), the parameter that contributes most significantly to concentration variability is clearly the overall network consumption (G_{ext}). The influence of the concentration at

the source (Y_1) is present but secondary. This behaviour is physically consistent: in these border nodes, the fluctuation in gas quality is determined almost exclusively by the macroscopic displacement of the mixing "front", a phenomenon governed primarily by the volume of withdrawals by users in this scenario.

However, proceeding to the right along the transition branch (nodes 147 and 151), the impact of parameter Y_1 starts to increase progressively and steadily.

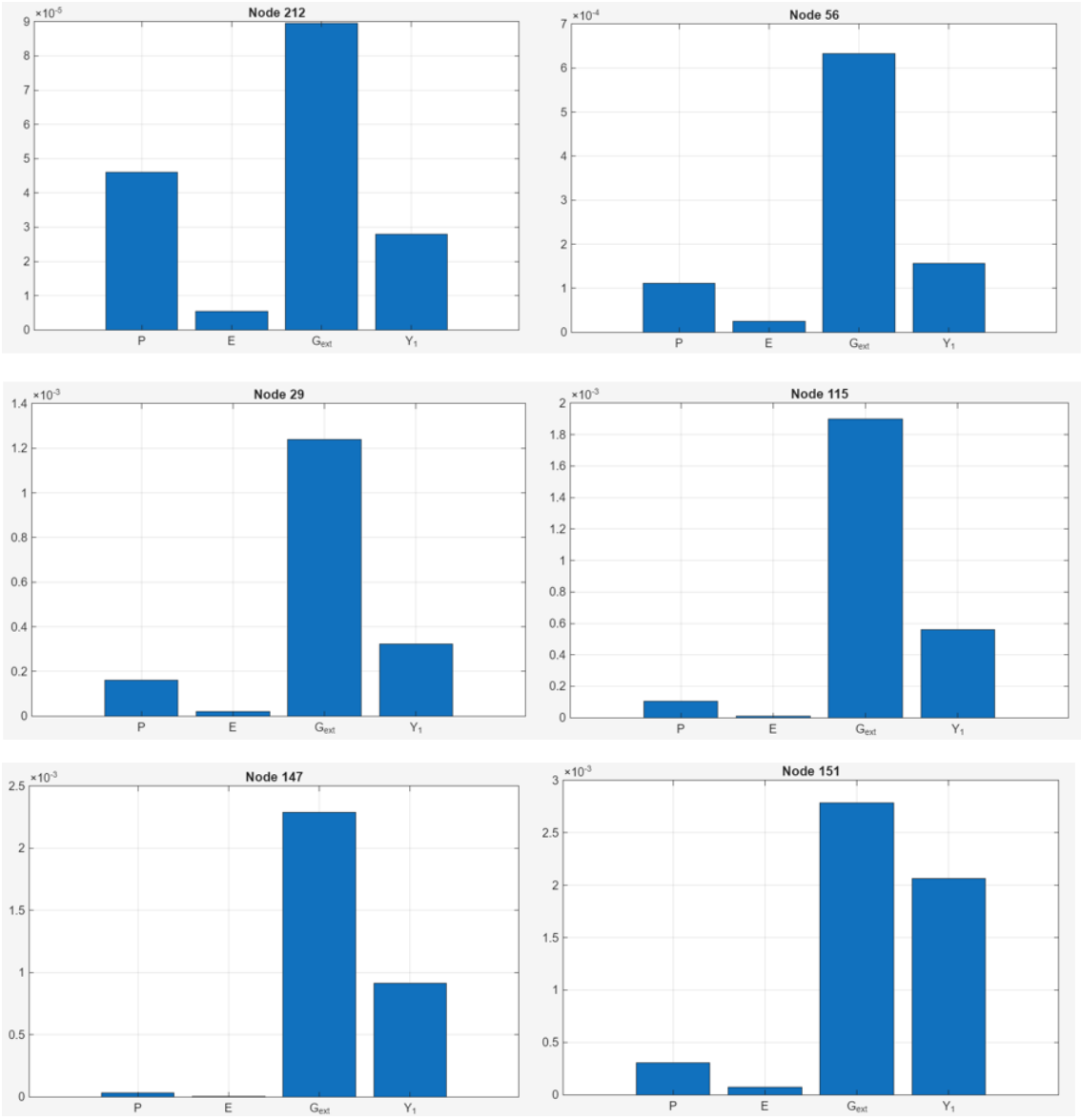


Figure 4.18.Total-effect Sobol' indices for the hydrogen concentration in the first segment of the transitional branch, from node 212 to node 151.

The change in pattern is seen in the nodes closest to the hydrogen injection (nodes 140, 193 and 177), the results of which are shown in Figure 4.19. As it approaches the area dominated by blending, the uncertainty in the local concentration stops depending on fluid dynamics and begins to be strongly influenced by the variability of the source itself. The influence of consumption drops and the concentration at the source becomes the dominant parameter. At node 177, the Sobol index for Y_1 is the only one to show a significant contribution, making the other parameters completely negligible.

This result confirms and mathematically quantifies the conclusions derived from the frequency analysis: since node 177 receives the pure mixture from the source in all operating configurations, its variance depends exclusively on how much hydrogen is injected upstream, making it completely insensitive to fluctuations in network loads.

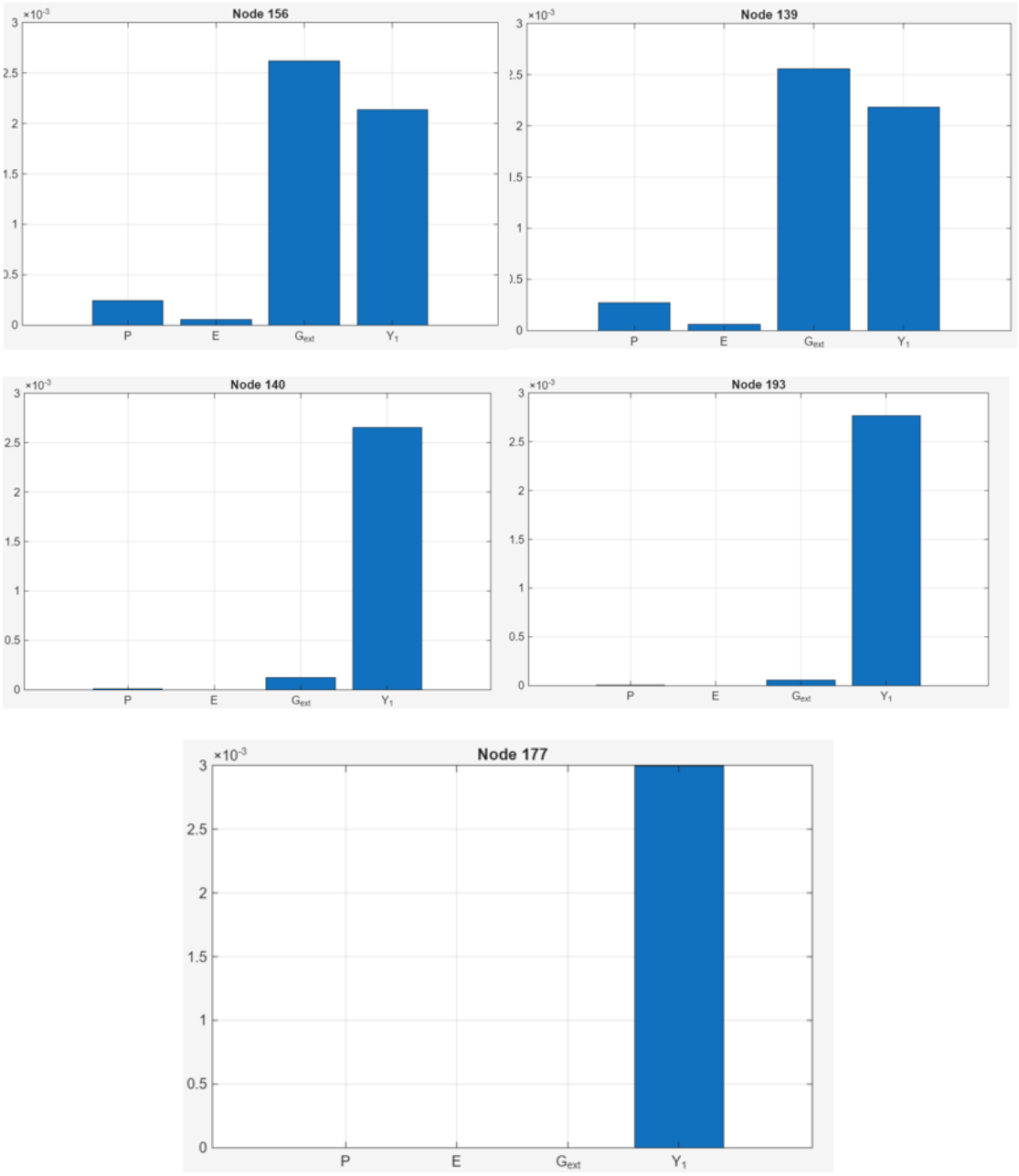


Figure 4.19. Total-effect Sobol' indices for the hydrogen concentration in the second segment of the transitional branch, from node 156 to node 177.

In addition, it is useful to highlight how the other areas of the network would behave if subjected to the same analysis. If Sobol's indices were calculated for nodes deeply immersed in the “Blue Area” (where the concentration is constantly equal to 0%), the result would return virtually zero variances for all parameters, as there would be no fluctuation. On the contrary, the entire ‘Red Area’ would behave exactly like node 177, showing total dependence on parameter Y_1 alone.

Finally, Figure 4.20 illustrates the results of Sobol's analysis for node 75, chosen as a representative point of the “Green Area” (the area north of the network where mixing always occurs). For this area of constant mixing, the overall consumption of the network (G_{ext}) is once again the parameter with the greatest impact, followed significantly by the concentration of the blending source (Y_1).

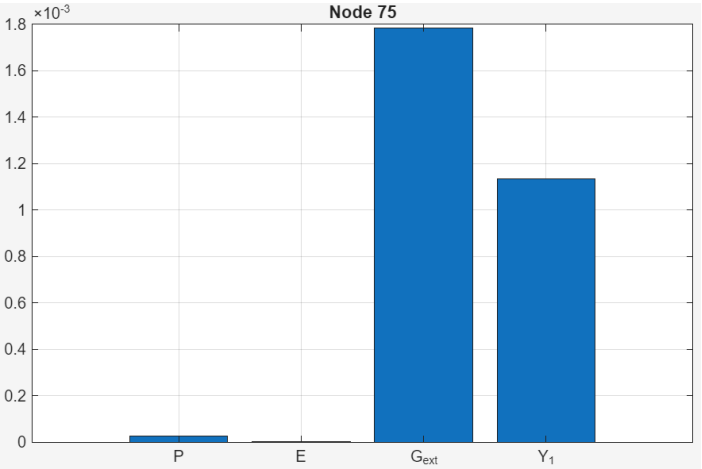


Figure 4.20. Total-effect Sobol' indices for the hydrogen concentration in node 75, chosen as representative of the "Green Area" (constant mixing zone).

4.4.1.4 Discussion of Case Study 1

This first test case analysed the response of the distribution network subject to hydrogen blending injection from one of its two entry points. The exploration of a large operating domain, characterised by simultaneous variations in pressure, overall consumption, pipe roughness and hydrogen concentration at the source, made it possible to trace the spatial and statistical propagation of gas quality.

The results extracted from the surrogate model show that the behaviour of the network is relatively stable, leading to the formation of three well-defined areas of concentration (Blue Area for pure natural gas, Red Area for pure blending and Green Area for constant mixing).

However, the real critical point from a fluid dynamics perspective is the moving boundaries that separate these zones, i.e. the area defined as the “transition zone”. The quantitative results obtained from the analysis of probability density functions and Sobol indices take on fundamental physical significance when interpreted in relation to the issues highlighted by standard UNI EN 16726. The guideline emphasises that, in networks supplied by heterogeneous sources, the collision between opposing flows of gas with different qualities generates so-called zero flow points within the pipes.

Frequency analysis has demonstrated exactly this phenomenon: the probability that boundary nodes will experience true intermediate mixing is low. On the contrary, these nodes tend to replicate the concentration of one source or the other. This confirms that the actual mixing of gas along the sections of pipeline where the fronts collide is almost absent: therefore, the transition from one gas quality to another occurs in a sudden way.

Furthermore, the topological position of these zero flow points is not static. As demonstrated by the global sensitivity analysis, in transition nodes, concentration

variability is predominantly driven by consumption fluctuations (G_{ext}). The continuous shift of this front as loads vary means that users located near the boundary are exposed to almost instantaneous changes in local hydrogen concentration.

4.4.2 Case Study 2: 5 Parameters with Double H2 Injection

This scenario represents the direct extension of the previous case to an operating configuration in which both gas entry points (the ReMi cabins at nodes 240 and 241) are characterised by the injection of mixtures containing hydrogen. From a modelling point of view, a fifth input parameter has been added: the hydrogen concentration at the second entry node (node 241). The stochastic variability of the blending at this node was assumed to be identical to that at node 240; therefore, both sources are subject to a uniform variation in hydrogen composition within the same range explored previously. It is essential to emphasise that the two injections vary completely independently of each other.

The surrogate model trained on this five-dimensional space can cover an even larger operating domain, allowing the instantaneous simulation of the impact of any combination of hydrogen fractions introduced by the two sources.

4.4.2.1 Statistical Mapping of Fluid-Dynamics and Gas Quality

This section presents a summary of the statistical mapping. Since the methodological approach is identical to that discussed at length for Case Study 1 (paragraph 4.4.1.1), the analysis will focus on highlighting the differences introduced by the double injection, omitting the already known results.

For purely fluid dynamic quantities, Figure 4.21 shows the spatial distributions of the mean values and standard deviations for nodal pressures, while Figure 4.22 illustrates the same metrics for mass flow rates in the branches and gas velocities. The maps confirm that the overall hydraulic behaviour of the network remains consistent with the previous scenario, maintaining the same areas of higher uncertainty at the borders of the infrastructure.

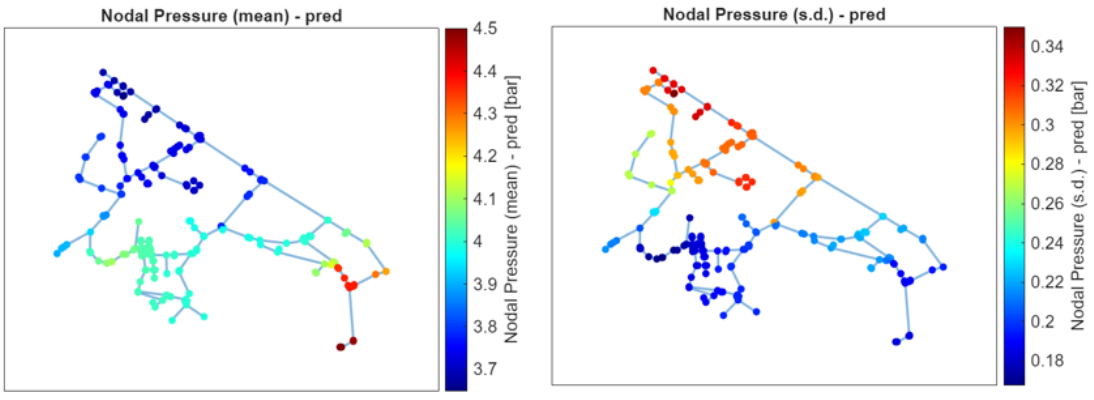


Figure 4.21. Spatial distribution of the mean values (left) and standard deviations (right) of nodal pressures for Case Study 2.

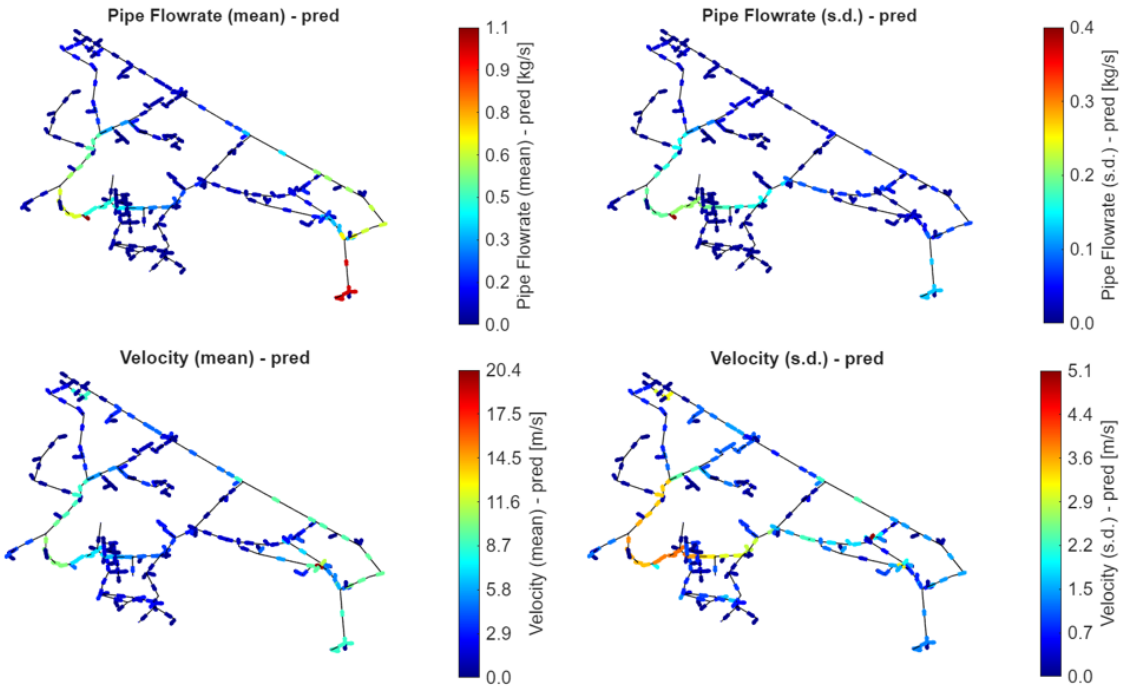


Figure 4.22. Spatial distribution of mean values and standard deviations for pipe flow rates (top) and gas velocities (bottom) across the network topology.

The most relevant aspect emerges from the analysis of hydrogen propagation. Figure 4.23 shows the maps relating to the average nodal concentration and its standard deviation.

Compared to the single injection case, the average hydrogen concentration is now drastically more uniform across the entire network. This phenomenon is the natural statistical consequence of the fact that both sources oscillate uniformly around the same value (approximately 10%): consequently, any mixture generated by their collision will tend to maintain a similar overall average. Despite this apparent uniformity, the graph of average values (on the left) still allows us to distinguish the presence of a persistent mixing area (highlighted in blue and light blue), where the average values deviate slightly (around 9.5%) due to the specific dilution dynamics and fluid dynamic asymmetries of the network.

However, the actual division of the network can be observed by analysing the standard deviation map (graph on the right). The areas coloured in dark red have a high standard deviation (approximately 5.5%), identical to that of the source locations. These “groups of nodes” behave exactly like their reference source, receiving the gas before any mixing takes place. In contrast, the blue and light blue areas (which show a decrease in standard deviation to 4.8%) represent the true mixing zones.

The nodes located at the boundaries between these areas exhibit complex behaviour, acting as a transition between taking on the variance profile of a specific source or that of the constant mixing zones. In the following paragraphs, exploiting the potential of the surrogate model, a more targeted Monte Carlo analysis will be conducted to explore in detail the tracking potential in this complex configuration.

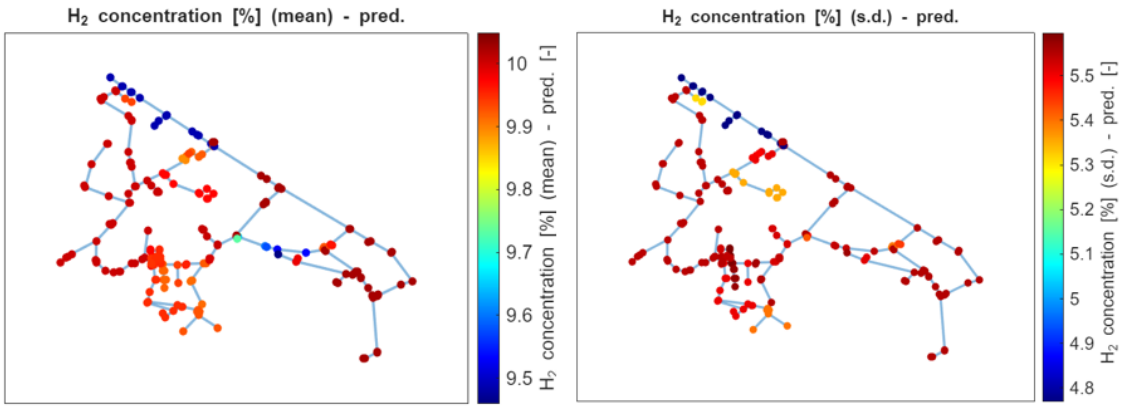


Figure 4.23. Spatial distribution of the mean values (left) and standard deviations (right) of nodal hydrogen concentration (Y_{H2}) in Case Study 2. The dual independent blending leads to a highly uniform mean concentration, while the standard deviation map reveals the true mixing zones (blue areas with reduced variance).

4.4.2.2 Asymmetric Dual Blending Scenario

To further test the surrogate model's ability and simulate more challenging operating conditions, an operational sub-scenario was defined, characterised by asymmetric hydrogen input between the two sources. Specifically, the range of variation of the five stochastic parameters was set as follows:

- Par. 1 (P): Set-point pressures at both ReMi cabins with common variation of ±7%.
- Par. 2 (Y_1): H_2 blending at the primary cabin (node 240) set at 5% with a relative variation of ±10%.
- Par. 3 (Y_2): H2 blending at the primary cabin (node 240) set at 10% with a relative variation of ±50%.
- Par. 4 (E): Pipe roughness with a variation of ±10%
- Par. 5 (G_{ext}): Outflow to users (consumption) with a variation of ±50%.

Using the surrogate model, a Monte Carlo simulation was performed with 1000 samples extracted from a uniform distribution on this specific domain. Figure 4.24 shows the spatial map of the average hydrogen concentration (left) and its standard deviation (right) resulting from this analysis.

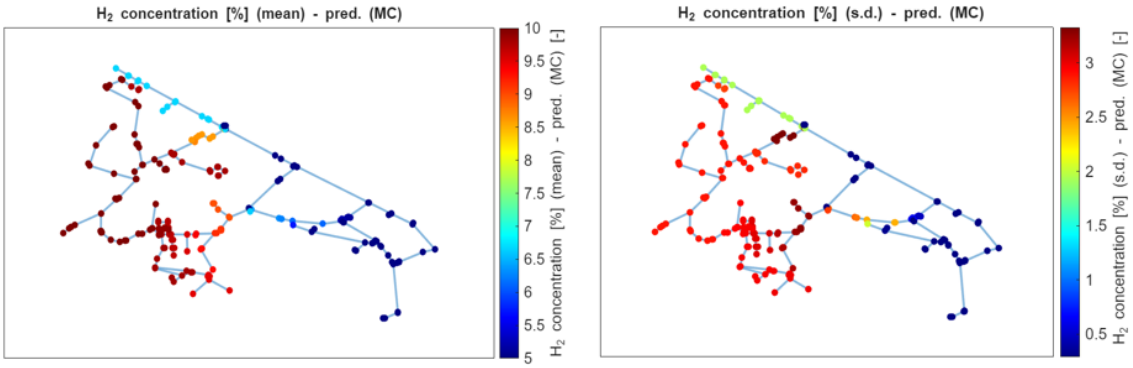


Figure 4.24. Spatial distribution of the mean the standard deviation of nodal hydrogen concentration in the asymmetric dual blending scenario. The network clearly splits into a 10% region (dark red) and a 5% region (dark blue).

Unlike the isotropic case analysed previously, the asymmetry of the sources forces the network to divide into clearly distinct areas of quality, strictly visualised by the bar chart in Figure 4.25. There is a block of nodes with an average concentration of 10% (dark red bars) and a second block with an average concentration of 5% (dark blue bars). It is particularly interesting to observe the behaviour of the standard deviation in the lower graph: the nodes in the 10% area are affected by the high uncertainty of source 241 (with a standard deviation close to 3%), while the nodes in the 5% area show negligible uncertainty, consistent with the low variability imposed on source 240.

Between these two stable domains, a small group of nodes emerges with average concentrations slightly below 10% or above 5% (orange, yellow and blue bars).

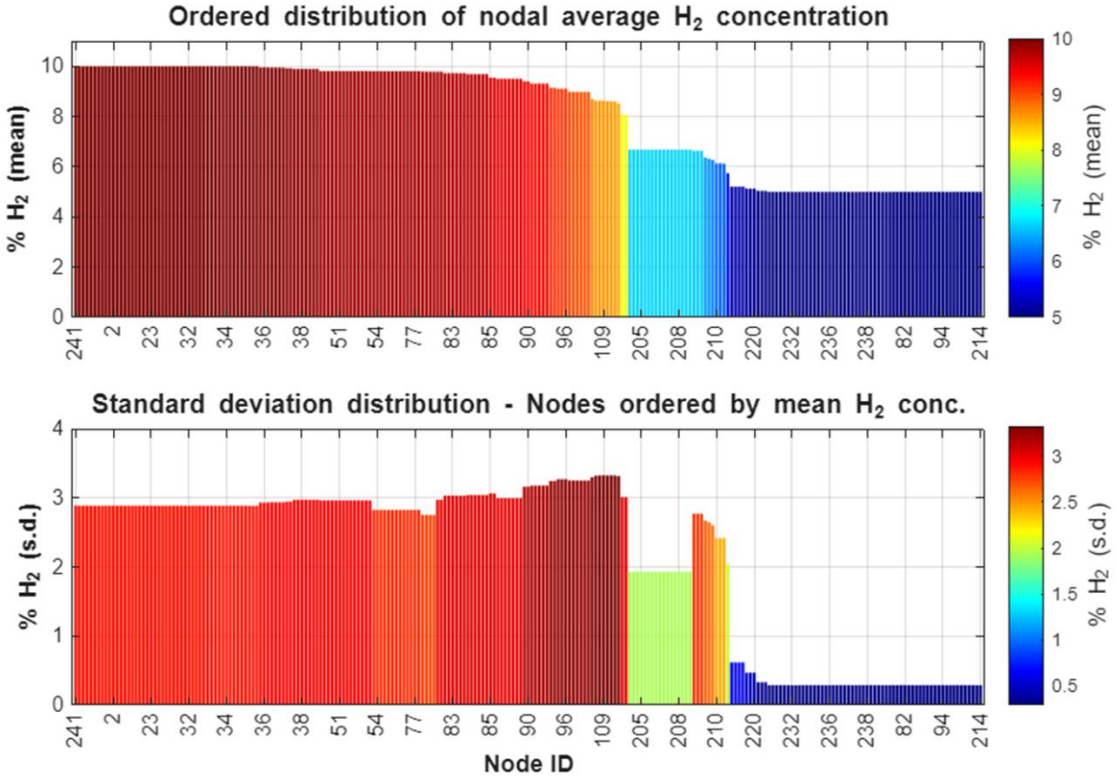


Figure 4.25. Ordered distribution of nodal average H_2 concentration (top) and corresponding standard deviation (bottom) for the asymmetric scenario. The standard deviation perfectly reflects the different variation ranges imposed on the two sources.

Some of these transition nodes have been isolated and highlighted topologically in the map in Figure 4.26. Their location confirms that they represent the moving boundaries between the different quality areas. In addition to the southern branch already identified in Case Study 1 (nodes 212 to 177), this analysis will consider an additional group of critical nodes in the central-northern area of the network (nodes 134, 218 and 68), as they have the highest standard deviation value.

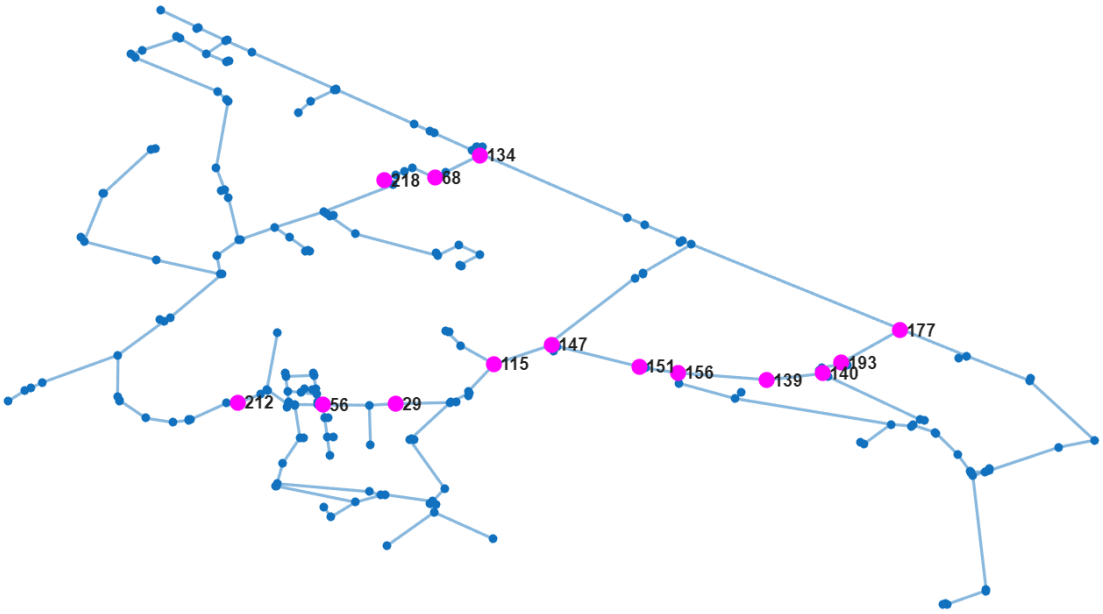


Figure 4.26. Topological layout highlighting the critical nodes belonging to the moving boundaries in the asymmetric scenario. The set includes the previously identified southern branch and new critical nodes in the north-central area.

The behaviour of these boundary nodes is complex. The results indicate that the average concentration at these points differs only marginally from that of the adjacent areas; therefore, from a strictly operational point of view, a network operator could decide to assimilate them to the nearest macro-area in order to simplify gas quality control.

However, in order to fully understand the dynamics governing these fluctuations, it is necessary to investigate which parameters dominate local variance. For the sake of brevity, the detailed analysis of nodal frequencies (PDF) is not repeated here, as it would lead to methodological conclusions similar to those already discussed at length in section 4.4.1.2. Instead, the following section presents Sobol's Global Sensitivity Analysis dedicated to these specific nodes, to quantify the cross-impact of the different parameters.

4.4.2.3 Global Sensitivity Analysis via Sobol' Indices

To understand which parameters control the fluctuation in gas quality at the boundaries of macro-areas, a Sobol' Global Sensitivity Analysis was performed. The investigation focuses on the two branches of critical nodes (from node 212 to node 177 and from 134 to 218) identified in the previous paragraph. In this five-parameter scenario, the objective is to quantify the combined influence of: nodal pressures (P), pipes roughness (E), global consumption (G_{ext}), hydrogen concentration at the primary source (Y_1) and concentration at the secondary source (Y_2). It is important to specify that this analysis was carried out under the symmetrical scenario (section 4.4.2.1), in which the sources have the same average value and the same range of variation in hydrogen concentration.

In contrast to Case Study 1, where the dynamics of the moving boundary were dominated only by consumption due to the presence of a single blending injection, the double injection and the independence of the two sources create a more complex statistical competition.

The results relating to the western portion of the first branch (nodes 212, 56, 29 and 115) are shown in Figure 4.27. Being topologically closer to the area dominated by the secondary ReMi substation, these nodes show a strong dependence on parameter Y_2. The local variance is driven by the high uncertainty associated with this input (±95%), while consumption (G_{ext}) begins to have only a marginal influence as one moves towards the centre of the branch (nodes 29 and 115). The impact of the primary source Y_1 in this area is slight.

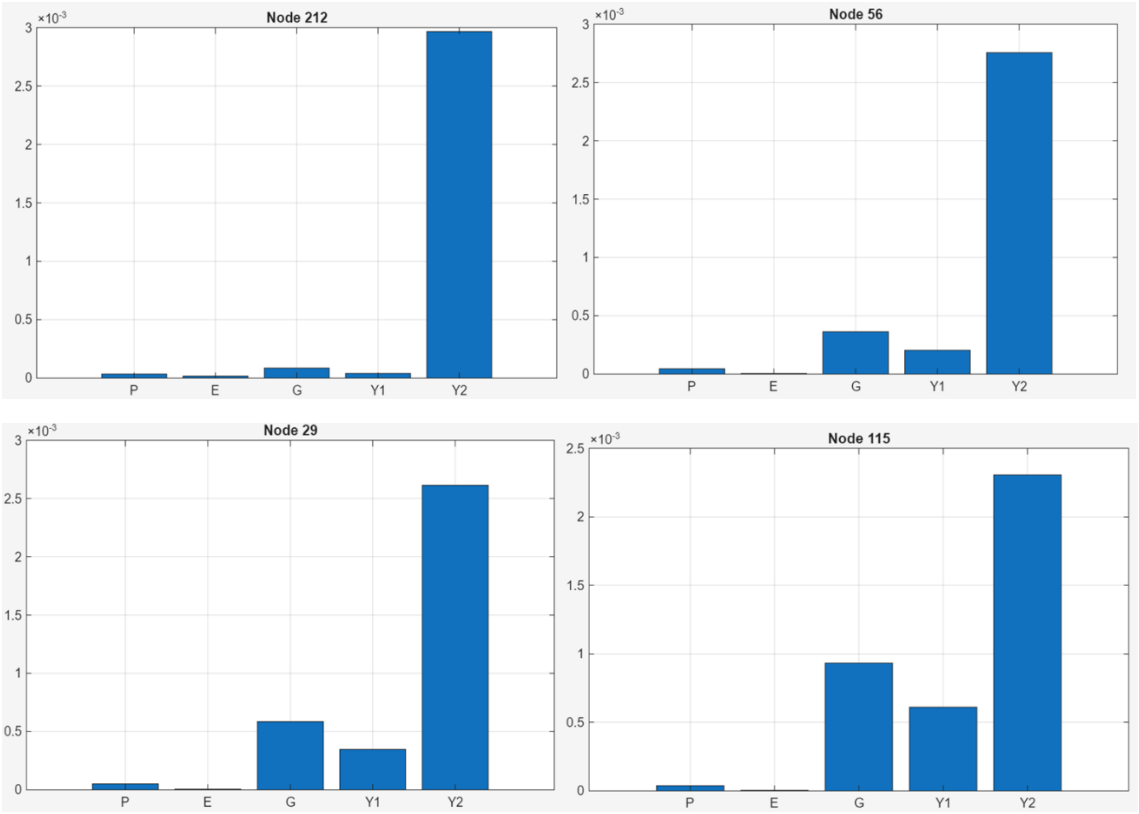


Figure 4.27. Total-effect Sobol' indices for nodes 212, 56, 29 and 115. The variance in this western segment of the transitional branch is heavily dominated by the secondary source concentration (Y_2).

Proceeding along the branch, the analysis reaches the centre of the moving boundary, located exactly at nodes 147 and 151 (Figure 4.28). In these two nodes, there is an almost perfect balance between the forces involved: the impact of the two sources and the displacement effect induced by consumption reach comparable values. In particular, in node 147, the influence of Y_2 is still slightly predominant compared to G_{ext} and Y_1, while in node 151 the situation is reversed, with Y_1 exceeding consumption and the influence of Y_2 decreasing. This inversion identifies the exact collision line between the two gas fronts.

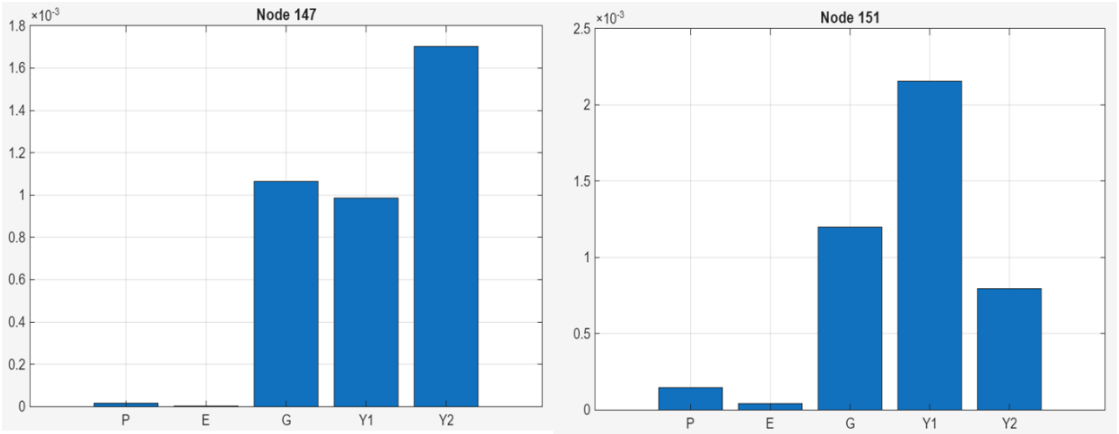
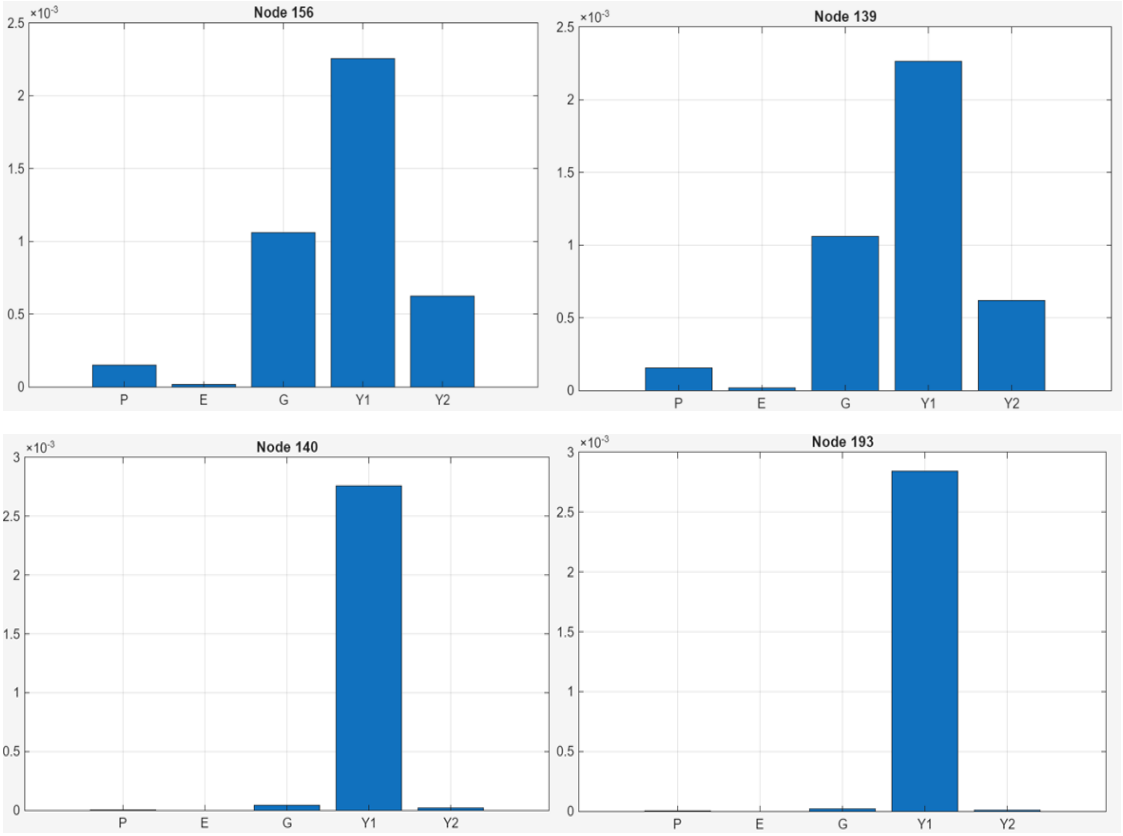


Figure 4.28. Total-effect Sobol' indices for nodes 147 and 151, identifying the core of the moving boundary. A perfect intersection of influences is visible, with the dominant parameter shifting from Y_2 (node 147) to Y_1 (node 151).

Finally, Figure 4.29 shows Sobol's indices for the eastern portion of the branch (nodes 156, 139, 140, 193 and 177). By entering this area, the influence of the secondary source quickly diminishes and consumption (G_{ext}) loses its ability to govern variability. Local concentration becomes controlled by the primary source (Y_1).

The trend continues in node 177, whose graph shows absolute control (100%) of parameter Y_1. Consistent with the findings in Case Study 1, although this node is geographically close to the boundary, it behaves fluid-dynamically as a rigid extension of the upstream injection source, resulting in complete isolation from the fluctuations introduced by the other ReMi cabin and other parameters.



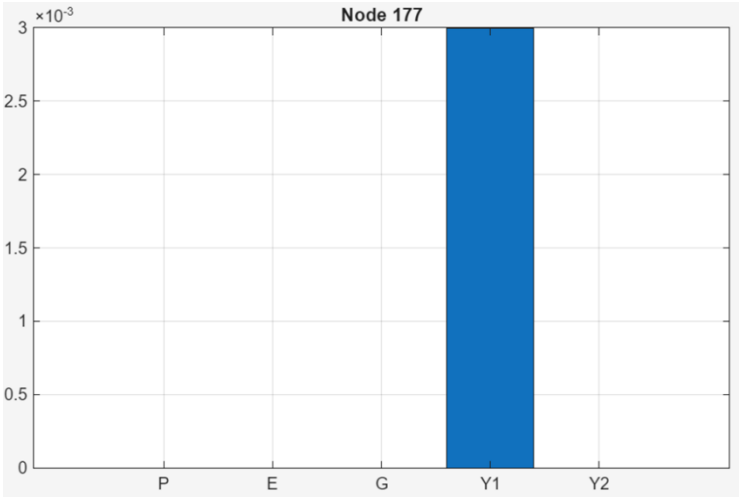


Figure 4.29. Total-effect Sobol' indices for the eastern segment of the branch (nodes 156, 139, 140, 193 and 177). Moving towards the primary ReMi station, the influence of Y_1 rises exponentially, reaching 100% dominance at node 177.

To complete the analysis, Figure 4.30 shows the Sobol indices for nodes 134, 68 and 218, which follow each other topologically along the critical central-northern area. Unlike the clear inversions of influence observed along the southern branch, this region shows considerable spatial continuity, characterised by a complex three-way interaction that remains stable along the entire segment. In all three nodes, the variance in concentration is governed primarily by the secondary source, followed closely by global consumption and, to a smaller but constant measure, by the primary source. This high dependency on the secondary source is because these nodes are topologically closer to the secondary ReMi.

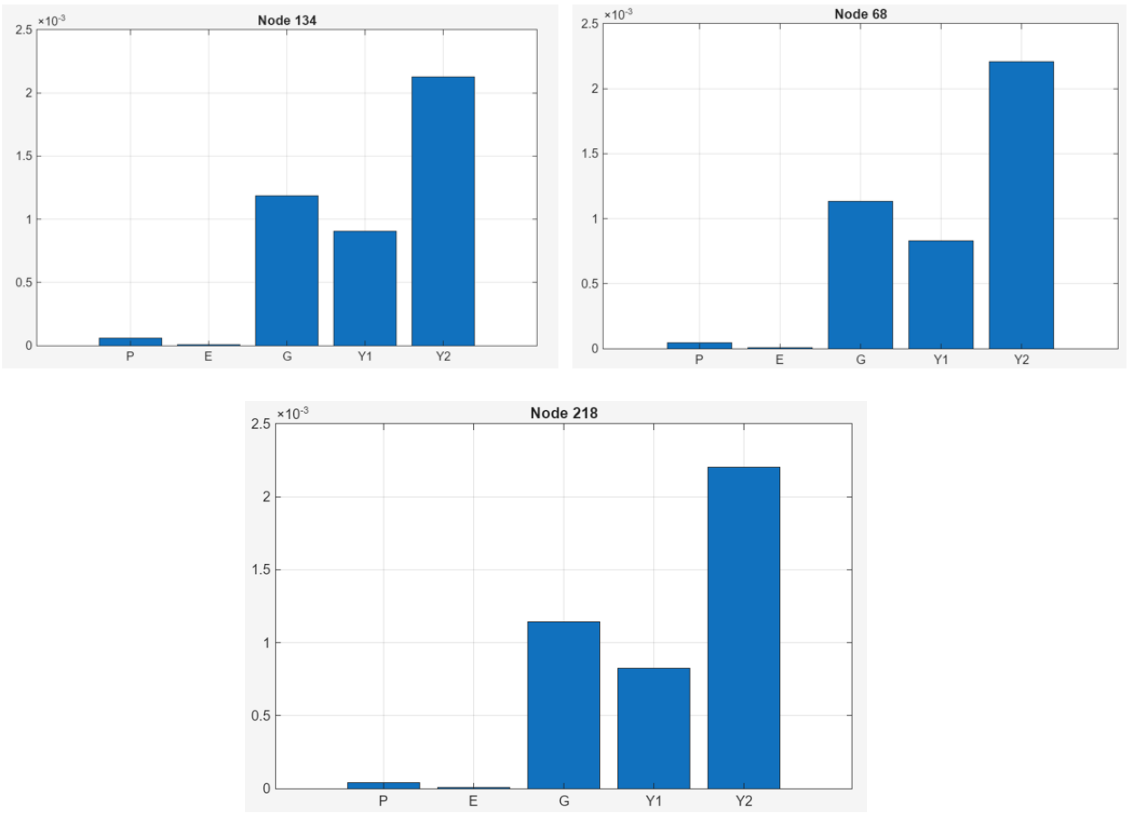


Figure 4.30. Total-effect Sobol' indices for nodes 218, 68 and 134, representative of the north-central moving boundary. Unlike the southern branch, this sequence shows a spatially consistent three-way interaction heavily weighted towards the secondary source (Y_2) due to its larger variance range.

4.4.2.4 Discussion of Case Study 2

In this second test case, parametric analysis using surrogate models was extended to an operating configuration in which both network entry points have independent variability in gas quality (hydrogen injection). This scenario is of fundamental importance for infrastructures that involve the integration of multiple injection points, such as green hydrogen production hubs distributed across the territory.

The presence of dual blending generates a slightly more complex statistical scenario than single injection. As shown by the standard deviation maps, the collision between two independent stochastic flows increases the presence and topological spread of zero flow points.

The introduction of the asymmetric scenario and the subsequent application of Sobol's Global Sensitivity Analysis made it possible to quantify the dynamics of these boundaries. It has been shown that along the transition branches, the position of the mixing points is strongly governed by the variability of the individual competing sources.

In summary, the use of the surrogate model in this second case study demonstrated its flexibility in capturing not only the macroscopic propagation of gas, but also the interactions between multiple inputs, preparing the field for the analysis of networks characterised by extreme parametric heterogeneity.

4.4.3 Case Study 3: 30 Parameters with Independent Cluster Variations

This third and final test phase represents a significant generalisation of the predictive framework. The same distribution infrastructure is simulated on a very high-dimensional variation domain, consisting of 30 independent stochastic parameters. Compared to the previous scenarios, the increase in complexity is driven by two key factors:

- Independent set-point pressures (2 parameters): The set-points at the two ReMi cabins do not vary synchronously and rigidly but fluctuate independently from each other. This greatly expands the operating domain, allowing the model to explore conditions in which the fluid dynamic roles of the entry points are completely reversed, i.e. the cabin that usually acts as the primary inlet can become secondary, altering the direction of the flows.
- Clustered consumption (25 parameters): The overall network load is no longer varied as a single group. Consumption has been divided into 25 distinct geographical clusters, corresponding to the final reduction stations that supply downstream users. To maintain a high degree of physical realism without reducing the generality of the test, the variability of these clusters has been split into two contributions: a fluctuation common to all nodes (±35%, representative of daily or seasonal macro-trends shared by the entire urban area) overlapped by a totally independent local variation for each cluster (±15%). The sum of these effects ensures that the maximum variation limit of ±50% already explored in previous cases is maintained but introduces complex spatial asymmetries in withdrawals.

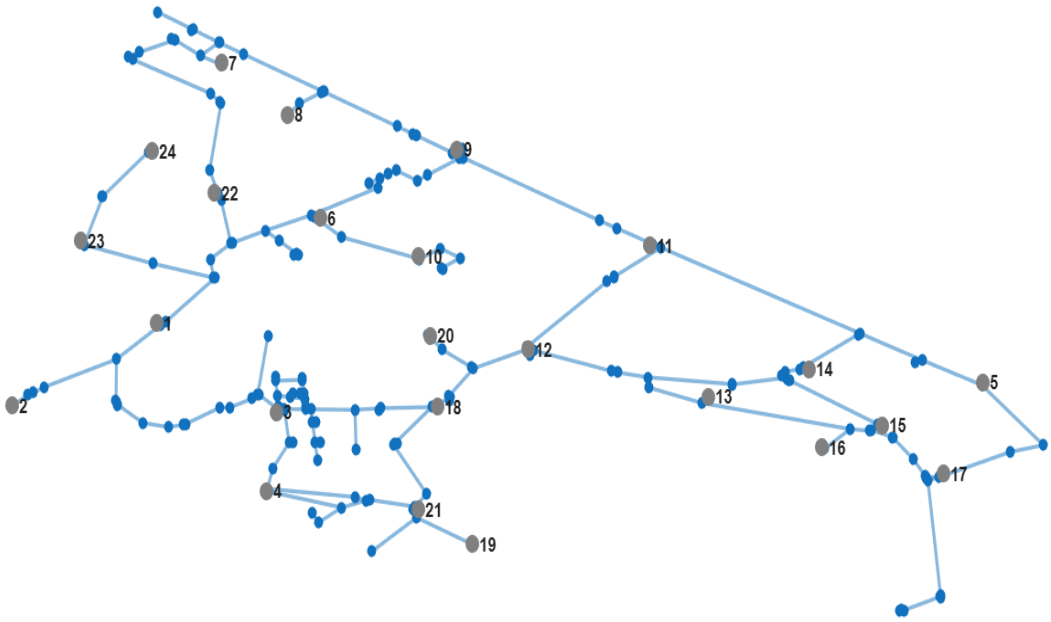


Figure 4.31. Topological distribution of the 25 independent gas consumption clusters (Final Reduction Stations) across the network infrastructure.

The 30-dimensional space is completed by variations in pipe roughness (E) and the two hydrogen concentrations at the sources (Y_1 and Y_2), which are kept independent as in Case Study 2.

4.4.3.1 Statistical Mapping of Fluid-Dynamics and Gas Quality

Using the 30-parameter surrogate model, a Monte Carlo simulation was performed on a dataset of 1000 operating configurations in order to extract the statistical mapping of the network.

Figures 4.32 and 4.33 show the spatial distributions of the mean values and standard deviations for nodal pressures, mass flow rates and gas velocities in the branches, respectively.

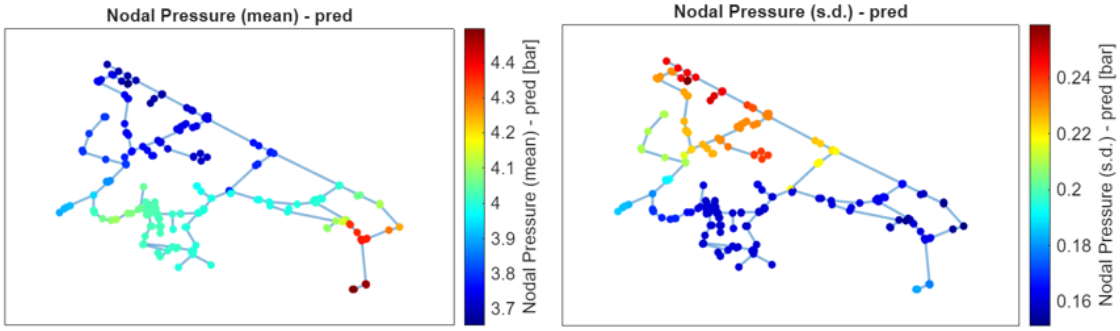


Figure 4.32. Spatial distribution of the mean values (left) and standard deviations (right) of nodal pressures for the 30-parameter scenario. The independent ReMi set points induce a wider and more complex variance profile across the network.

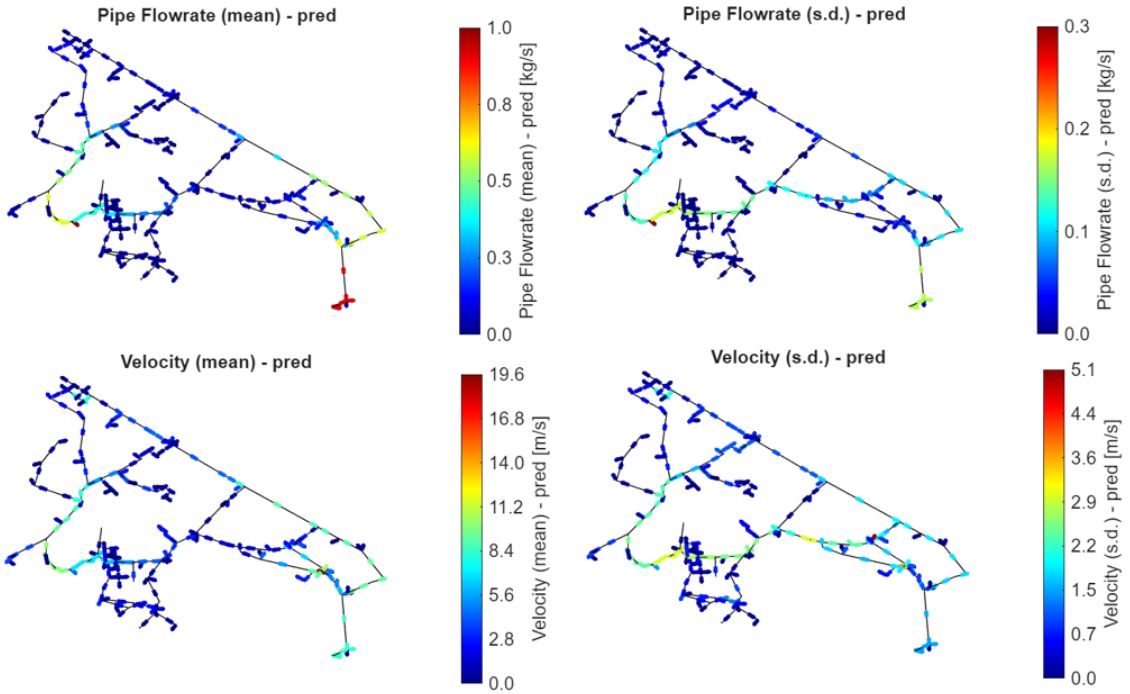


Figure 4.33. Spatial distribution of mean values and standard deviations for pipe flow rates (top) and gas velocities (bottom).

The impact of this complex hydraulics is directly reflected in the tracking of mixture quality. Figure 4.34 shows the topological distribution of the average hydrogen concentration (left) and the corresponding standard deviation (right).

Similar to what was observed in the isotropic scenario of Case 2, the global average concentration tends towards the nominal value (around 10%), around which both sources oscillate uniformly. However, visual analysis of the standard deviation map reveals a profound change in the behaviour of the network. While in the 5-parameter case the deep mixing zones (blue areas with low variance) and pure gas zones (dark red areas with high variance) were separated by relatively small and well-defined boundaries, in this 30-dimensional scenario the nodes that deviate from the statistical behaviour of the sources are extremely numerous and spread across almost the entire infrastructure.

The decoupling of consumption clusters means that the areas where flows collide (zero flow points) no longer move along a single predictable axis but disperse chaotically into multiple secondary branches. This demonstrates that the addition of 25 degrees of spatial freedom has subdivided the macro-areas of quality, generating extremely complex and extensive mixing behaviour.

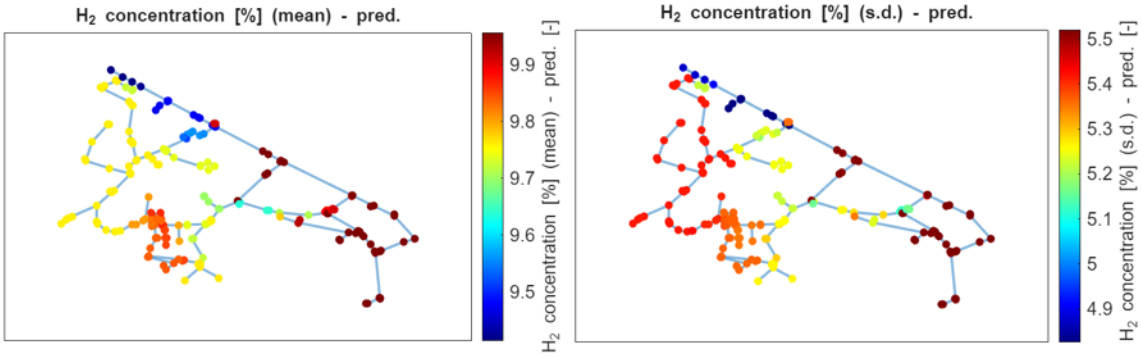


Figure 4.34. Spatial distribution of the mean values (left) and standard deviations (right) of nodal hydrogen concentration in Case Study 3. The widespread presence of mixed variance (yellow/orange nodes) indicates that the 30-parameter variation fragments the moving boundaries, creating a highly diffuse mixing behaviour.

In the following paragraph, following the same logic used in previous cases, the surrogate model will be applied to a specific asymmetric sub-scenario to investigate in detail the position and amplitude of these new fragmented boundaries.

4.4.3.2 Statistical Mapping of Fluid-Dynamics and Gas Quality

To fully explore the potential of the 30-dimensional surrogate model, an operational sub-scenario was defined characterised by asymmetric hydrogen injections, similar to that tested for Case 2, but immersed in a significantly more complex fluid dynamic variation domain. Specifically, the stochastic parameters were set as follows:

- Par. 1-2 (P_1, P_2): Set-point pressures at the two ReMi cabins, independently variable by ±7%.
- Par. 3 (Y_1): H_2 blending at the primary cabin (node 240) set at 5% with a relative variation of ±10%.

- Par. 4 (Y_2): H2 blending at the primary cabin (node 240) set at 10% with a relative variation of ±50%.
- Par. 5 (E): Pipe roughness with variation of ±10%.
- Par. 6-30 ($G_{ext,i}$): Output flows to users, divided into 25 independent clusters, each with a total variation of ±50% (combination of common and local fluctuation).

Using the surrogate model, a Monte Carlo simulation of 1000 samples with uniform distribution was performed with the input set just defined. Figure 4.35 shows the spatial maps of the average concentration and relative standard deviation for this specific test.

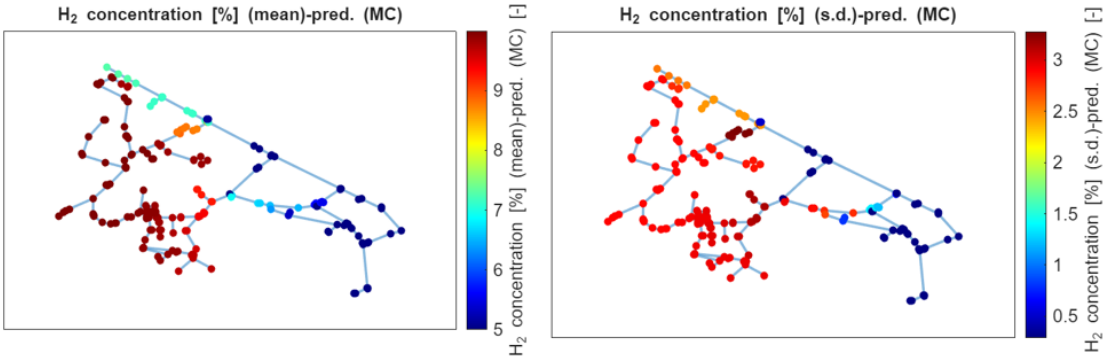


Figure 4.35. Spatial distribution of the mean values (left) and standard deviations (right) of nodal hydrogen concentration in the 30-parameter asymmetric sub-scenario.

To understand the fragmentation of quality areas, it is essential to analyse the orderly distribution of nodal concentrations shown in Figure 4.36. Unlike cases with reduced dimensionality, where the gap between the concentration of one source and another was more clearly, here the transition between concentration levels is much more gradual (as can be seen from the central "staircase" of yellow, green and blue nodes) . This behaviour of the mixing front is physically expected: the total independence of the two pressure set points and the partial independence of the 25 stations (consumption clusters) generate an extremely dynamic fluid dynamic field, in which the zero-flow points move continuously and over greater distances. As a result, it is more complex to clearly define the boundaries of areas with homogeneous quality.

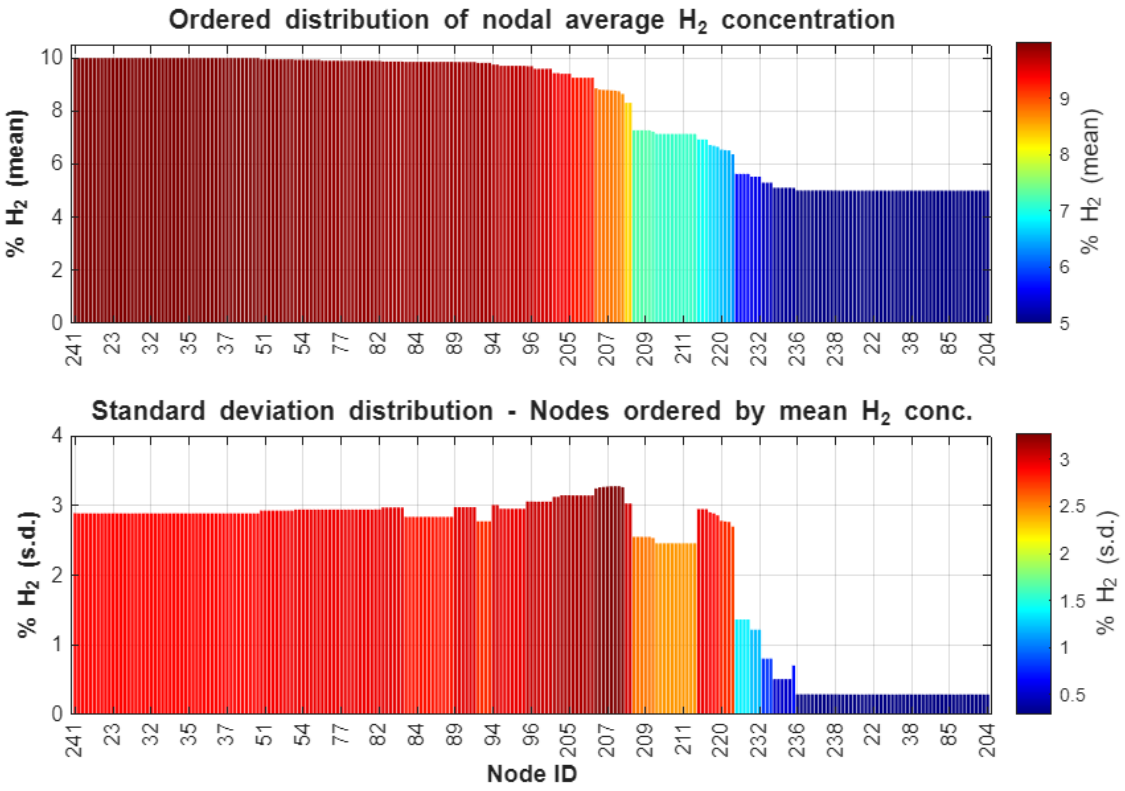


Figure 4.36. Ordered distribution of nodal average H_2 concentration (top) and corresponding standard deviation (bottom) for the 30-parameter asymmetric scenario. The transition between concentration domains is noticeably more gradual due to the uncorrelated fluctuations of consumptions and inlet pressures.

However, despite the extreme variability induced by the 30 parameters, the intermediate mixing effect remains confined to a specific number of nodes. Looking at the standard deviation graph in Figure 4.36 (below), we can see that the large areas stably fed by a fixed mixture (the 10% block with high variance on the left and the 5% block with low variance on the right) are still well defined and numerous.

The nodes that make up this large and blurred moving boundary have been located on the network map in Figure 4.37. They follow the critical branches already identified in Case 2 (the southern branch from node 212 to 177 and the central-northern area with nodes 218, 68 and 134), demonstrating that the infrastructure is still able to contain and direct the collision of flows.

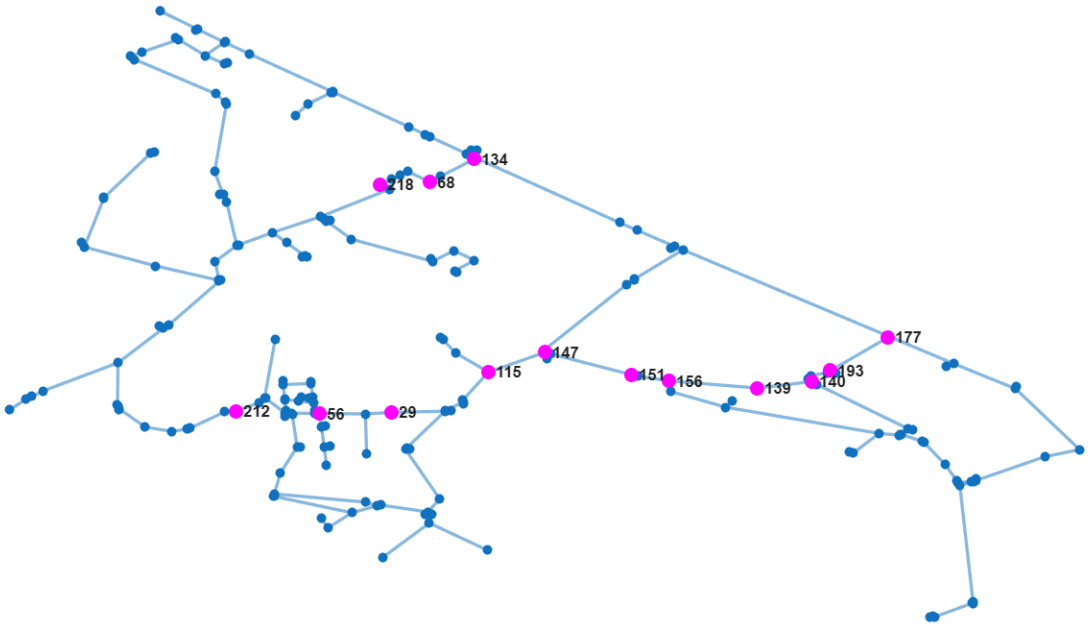


Figure 4.37. Topological layout highlighting the critical nodes belonging to the moving boundaries in the 30-parameter asymmetric scenario.

In the concluding paragraph dedicated to this case study, the complex interaction governing this transition zone will be analysed mathematically using Sobol's Global Sensitivity Analysis, allowing the effective impact of the variable parameters given as input to be weighed.

4.4.3.3 Global Sensitivity Analysis via Sobol' Indices

To explain the complex interaction generated by the 30 stochastic parameters, Global Sensitivity Analysis was applied to the critical nodes previously identified. Due to the high dimensionality, the location of the moving boundaries is less clear than in simpler scenarios, appearing as a very blurred transition band. The investigation focuses on the nodes of the southern branch (from 212 to 177) and the central-northern branch (from 134 to 218). It is important to specify that this analysis was carried out under the symmetrical scenario (section 4.4.3.1), in which the sources have the same average value and the same range of variation in hydrogen concentration.

In contrast to previous cases, the overall effect of consumption (G_{COMM}) has been separated from the fluctuations of the individual clusters (G_{ext1} - G_{ext25}) and, especially, the injection pressures (P_1 and P_2) have been made completely independent.

The analysis of the western portion of the transition branch (nodes 212, 56, 29 and 115) is shown in Figure 4.38. In this region, the local concentration remains clearly dominated by the concentration set at the secondary entry node (Y_2), due to its proximity. However, moving towards the centre of the network (node 115), the common background fluctuation of consumption (G_{COMM}) and the primary source parameter (Y_1) begin to increase. The impact of individual local consumption clusters is mathematically negligible, probably due to the smaller range of variation.

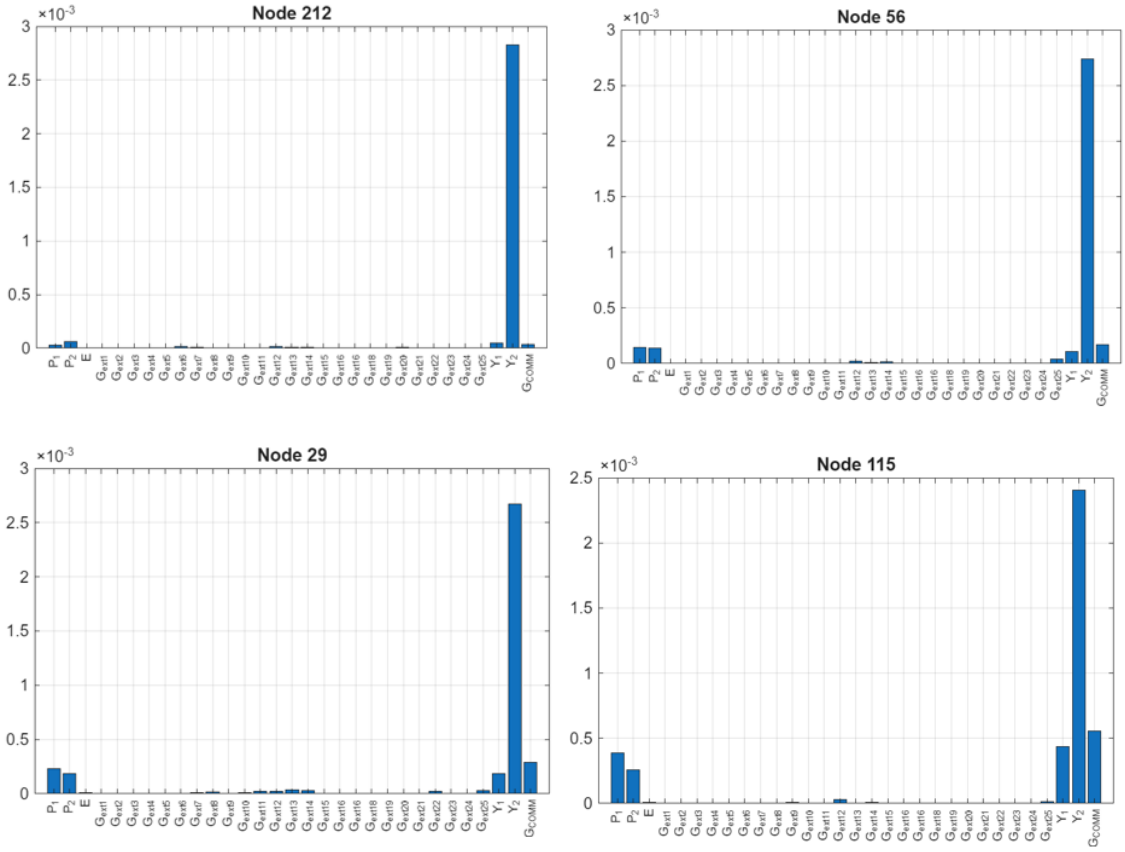


Figure 4.38. Total-effect Sobol' indices for nodes 212, 56, 29 and 115 in the 30-parameter scenario. The western segment is strongly governed by the secondary source concentration (Y_2).

In the centre of the moving boundary, at nodes 147 and 151 (Figure 4.39), there is a change in behaviour. At node 147, Y_2 is still the prevailing parameter, but there is a marked increase in the influence of set-point pressures (P_1 and P_2).

Node 151 shows even more complex behaviour: the domain changes in favour of the primary source (Y_1), but parameters P_1 and P_2 are the determining forces for local variability, even exceeding the common fluctuation in consumption (G_{COMM}). This result is because, having decoupled the pressures at the ReMi cabins, the model simulates the reversal of hydraulic thrust between the two inlets. The competition between these two pressures physically shifts the zero flow point.

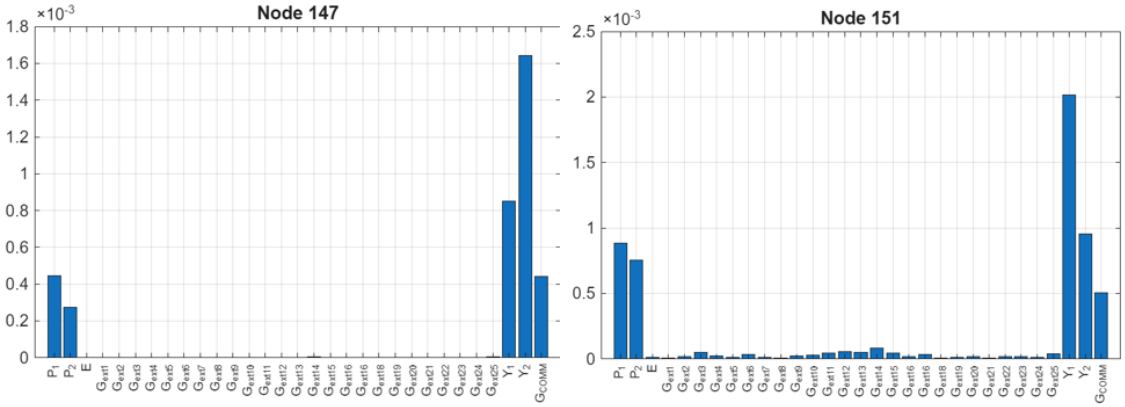


Figure 4.39. Total-effect Sobol' indices for the core transitional nodes 147 and 151. The uncoupling of the inlet pressures (P_1, P_2) radically alters the sensitivity profile, highlighting their crucial role in shifting the zero-flow point.

Continuing eastward (nodes 156, 139, 140, 193 and 177 in Figure 4.40), the transition is complete. The influence of Y_2, pressures and consumption rapidly decline. At node 177, the Sobol' index certifies almost total control by the primary source concentration (Y_1), confirming that the nodes topologically closest to the ReMi 240 cabin almost always behave like the source, without being influenced by the variability of the other parameters.

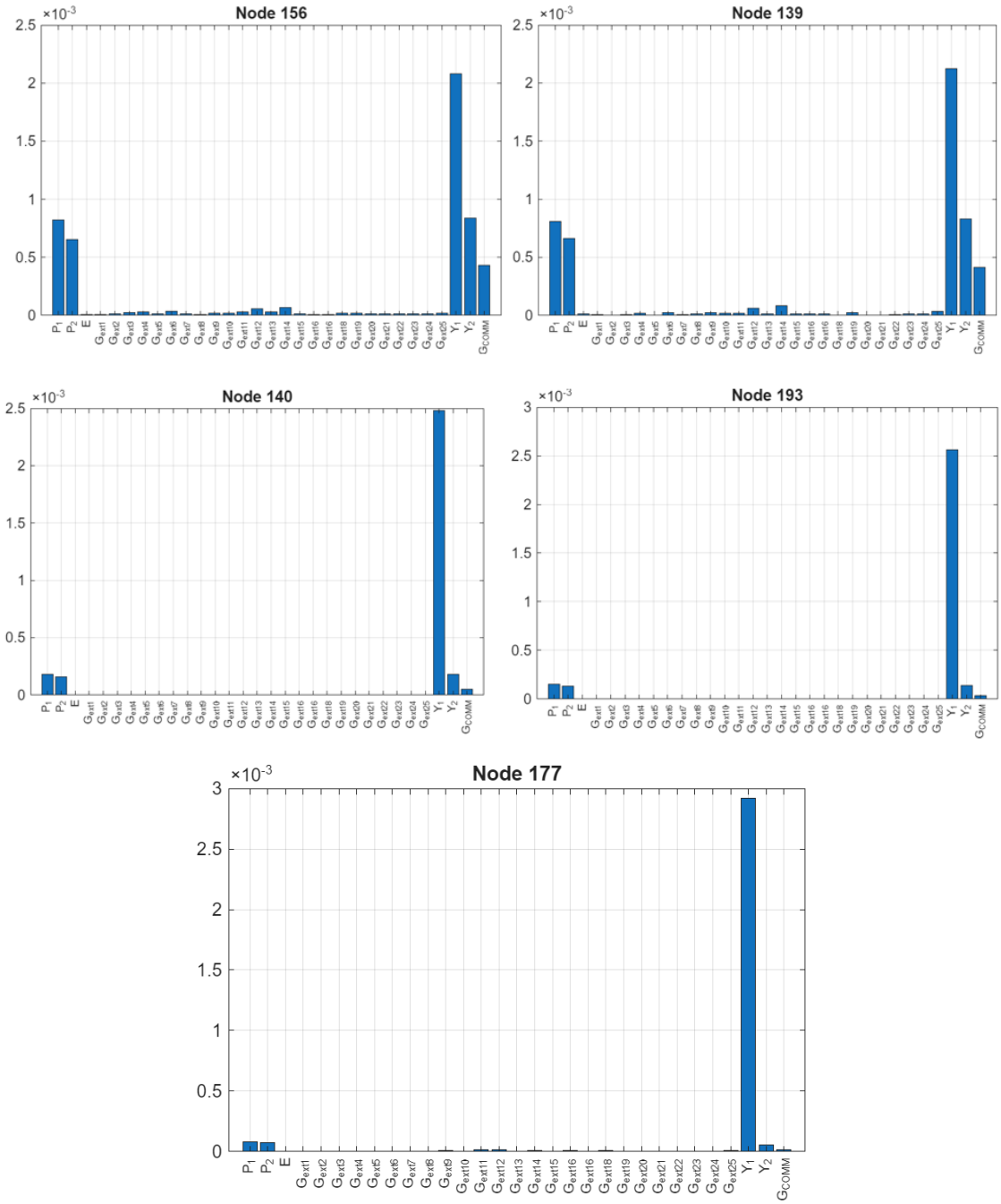


Figure 4.40. Total-effect Sobol' indices for the eastern segment (nodes 156, 139, 140, 193 and 177). The sensitivity returns strictly to the primary source concentration (Y_1).

To complete the investigation, the analysis was focused on the critical central-northern area; this specific region of the network shows the highest standard deviation for hydrogen concentration.

Figure 4.41 shows the Sobol' indices for nodes 134, 68 and 218, which follow each other topologically along this mobile boundary with high variance.

In this area, the parameter relating to the secondary source (Y_2) dominates in all three nodes due to its proximity, but is counterbalanced by a block of competing quantities: the primary concentration (Y_1), the set of common consumption (G_{COMM}) and the injection pressures (P_1 and P_2). It is also of great physical interest to note that, precisely in these border nodes, the contributions (albeit minor) of some specific local consumption clusters (e.g., for 6, 9 and 11) begin to emerge.

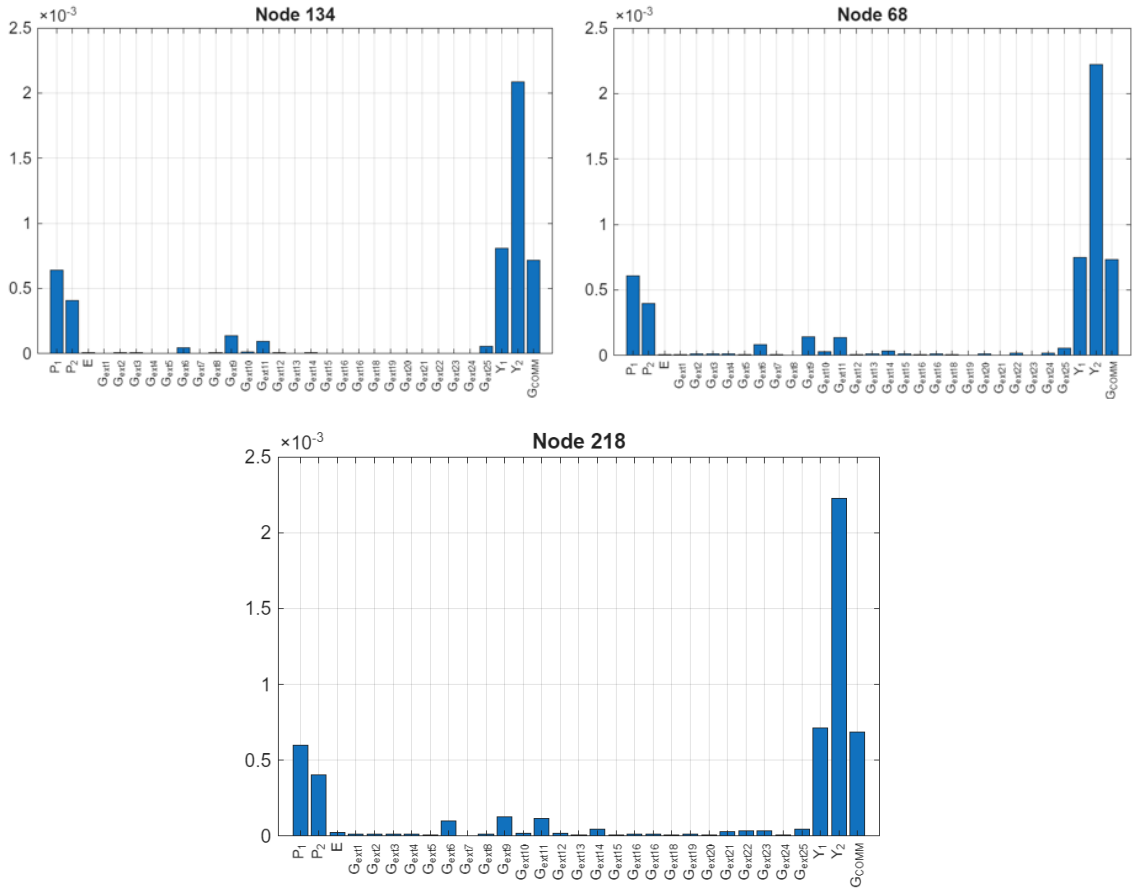


Figure 4.41. Total-effect Sobol' indices for nodes 218, 68 and 134, forming the north-central moving boundary. This sequence, which exhibits the highest standard deviation in the network, shows a consistent interaction governed heavily by Y_2, but with significant impact from pressures, global consumption and specific local clusters.

Shifting the analysis to the northern area, characterised by constant mixing of flows, Figure 4.42 shows the Sobol indices for node 75. In this peripheral area, the variance in concentration is governed primarily by the two injections, with the primary source (Y_1) prevailing over the secondary source (Y_2).

In this area, there is a balanced sensitivity to both pressure set points. Since the flows converge continuously without repelling each other, the mixture is sensitive to the overall hydraulic balance between the two ReMi cabins. Finally, there is a slight contribution from global consumes (G_{COMM}) and specific local clusters (such as the station 7), confirming that the gas quality in these areas of permanent mixing is the result of the overall balance of the infrastructure.

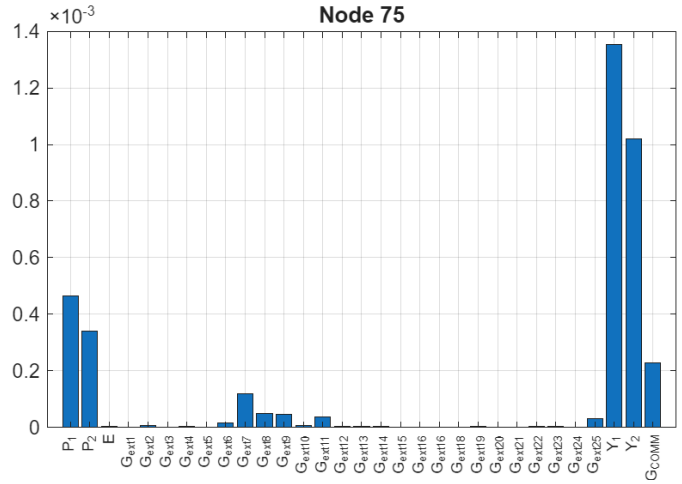


Figure 4.42. Total-effect Sobol' indices for node 75, representative of the stable mixing area in the North. The profile highlights a dominant dependence on both source concentrations and a remarkably balanced sensitivity to both inlet pressures.

4.4.3.4 Discussion of Case Study 3

Compared to the previous scenarios, the results of the third case study show much more complex network behaviour. The spatial distribution of hydrogen is fragmented and the transition zones (moving boundaries) extend over larger portions of the infrastructure. This effect is due to the increase in stochastic parameters from 5 to 30. To make the model more realistic, gas consumption no longer varies uniformly across the network but has been divided into 25 independent clusters, representing the final reduction stations. Each cluster has been assigned a common base variation (±35%) to which an independent local variation (±15%) is added. In addition, the set-point pressures at the two ReMi substations have been decoupled, varying completely independently.

To understand which parameters cause this high fragmentation, the Global Sensitivity Analysis (Sobol' indices) provided fundamental results, showing a change in hierarchies compared to simpler cases. As expected, the concentration of hydrogen at the injection points remains the parameter that most influences the quality of the gas in each node analysed. However, having made the inlet pressures independent, these have become decisive. In the areas where the flows interact and in the mixing areas, the impact of pressures even exceeds that of the overall variation in consumption. On the contrary, fluctuations in individual local clusters have a minor impact, limited to generating modest variations in the surrounding nodes. In this case, it is the pressure difference between the entry points that determines the displacement of the mixing front, much more so than load variations. It should be noted that the lower impact of consumption clusters could be partly due to the reduced range of variability imposed on them during the input phase.

The application of the surrogate model to this scenario made it possible to test its limits under conditions of high uncertainty. As highlighted during the validation phase, in this configuration there is a slight reduction in the predictive accuracy, but this occurs only within the areas where flows interact. The probability of making significant errors on critical nodes remains extremely low. In the vast portions of the network with a stable composition (mainly supplied by a single source), the accuracy of the model remains

very high. This confirms that the surrogate model is a reliable tool for mapping areas of homogeneous quality and identifying transition zones that require specific monitoring.

Finally, comparing all the case studies, the increase in degrees of freedom has not altered the basic physics of the infrastructure. The nodes identified as critical in the 4 and 5-parameter models are also confirmed as such in the 30-parameter case. However, in the third scenario, the variability introduced by independent pressures and fragmented loads causes the critical areas to expand, by shifting the moving boundaries to topologically larger areas of the network.

Conclusions

This thesis addressed a current engineering challenge for gas distribution network operators: safely predicting and managing the introduction of hydrogen blending into infrastructure historically designed for natural gas only. Since hydrogen blending significantly changes the thermophysical properties of the fluid, making its propagation dependent on continuously varying network loads and pressures, the exclusive use of traditional fluid dynamics solvers is computationally unsustainable for large-scale parametric exploration.

To overcome this limitation, the research successfully validated the implementation of a data-driven predictive framework based on a Vector-Valued Kernel Regression surrogate model. Trained on a limited set of physical simulations, the model was able to map the hydraulic response of the entire real network of Riccione (pressures, velocities and mass flow rates), allowing the subsequent gas-tracking module to obtain the local hydrogen concentration without significant loss of accuracy. The most significant methodological achievement of this approach is its extreme scalability: the model managed to maintain its predictive robustness by scaling from basic scenarios with four stochastic variables to a very large domain governed by 30 independent parameters.

The application of this framework through Monte Carlo simulations has not only reduced calculation times but also revealed the intrinsically statistical nature of hydrogen propagation. In all the case studies analysed, the network topology imposed the formation of large macro-areas with stable gas quality, divided by moving boundaries where flows from different sources converge and interact. This thesis demonstrated how the identification of zero flow points is of crucial importance: it is in these transition nodes, characterised by high standard deviation, that the most sudden fluctuations in gas composition occur. The ability to locate these critical areas provides a direct response to the requirements of the UNI EN 16726 standard, allowing the mapping and prevention of unacceptable variations in gas quality that could compromise the safety and efficiency of end-user devices.

The most in-depth physical investigation was achieved using a Global Sensitivity Analysis (Sobol' indices), which revealed hierarchies of influence as the degrees of freedom of the system increased. While in the simplified scenarios the movement of the mixing boundaries was mainly driven by global load variations, the decoupling of parameters in the 30-dimensional model revealed a different hierarchy. Apart from the obvious absolute dominance of the concentrations injected at the ReMi, the balance of the input set-point pressures had a significant impact on the hydrogen concentration of the moving boundary nodes. In the centre of the moving boundaries (zero flow points), the difference in hydraulic thrust was able to overcome the impact of macro-variations in consumption, physically shifting the mixing front much more than the fluctuations of individual local clusters.

In conclusion, the transition from a low-dimensional domain to a complex scenario confirmed that the independent variation of input set-point pressures and the geographical variability of withdrawals divides and spreads critical areas but does not change the main topological guidelines of the network. By integrating artificial intelligence and statistics, Distribution System Operators (DSOs) can now move beyond the deterministic testing of individual scenarios to adopt a probabilistic approach, enabling them to:

- Anticipate how hydrogen blending variability will propagate spatially across the infrastructure.
- Identify large regions with stable quality, suitable for the application of simplified and cost-effective monitoring strategies.
- Detect moving boundaries (zero flow points) that require advanced supervision.
- Evaluate the robustness of hydraulic set points before their operational implementation.

The tool developed in this work is therefore a fundamental methodology for supporting decisions, essential for ensuring the safe and efficient management of gas networks in a future scenario where hydrogen blending will be an established operational practice.

Bibliography

[1] Salim SS, Luxembourg SL, Smekens K, Dalla Longa F, van der Zwaan B. Pathways to climate neutrality: Europe's energy transition under the Green Deal. Renewable and Sustainable Energy Reviews 2026;226:116272. https://doi.org/10.1016/J.RSER.2025.116272.

[2] Europe – Countries & Regions - IEA n.d. https://www.iea.org/regions/europe/natural-gas (accessed December 28, 2025).

[3] Karmaker SC, Chapman AJ, Sen KK, Saha BB. Electrification or Hydrogen? A strategic pathway for Japan's energy transition. Int J Hydrogen Energy 2026;201:152989. https://doi.org/10.1016/J.IJHYDENE.2025.152989.

[4] Erdener BC, Sergi B, Guerra OJ, Lazaro Chueca A, Pambour K, Brancucci C, et al. A review of technical and regulatory limits for hydrogen blending in natural gas pipelines. Int J Hydrogen Energy 2023;48:5595–617. https://doi.org/10.1016/J.IJHYDENE.2022.10.254.

[5] Hydrogen - Thermophysical Properties n.d. https://www.engineeringtoolbox.com/hydrogen-d_1419.html (accessed January 5, 2026).

[6] Kumar L, Sleiti AK, Al-Ammari WA. Thermo-economic analysis of blending hydrogen into natural gas pipeline with gaseous inhibitors for sustainable hydrogen transportation. Chem Eng Sci 2025;311:121619. https://doi.org/10.1016/J.CES.2025.121619.

[7] Zhou L, Liu J, Chen D, Huang C, Fu S, Wang Y, et al. Hydrogen blending into natural gas networks: global practices, infrastructure compatibility, and end-user implications. Appl Energy 2026;402:127057. https://doi.org/10.1016/J.APENERGY.2025.127057.

[8] Zekun Y, Zhanli Y, Hao Y, Yan Z, Kai X. Hydrogen embrittlement in welded joints of high-strength pipeline steels: A review of mechanisms, characterization, and mitigation strategies. International Journal of Pressure Vessels and Piping 2025;218:105615. https://doi.org/10.1016/J.IJPVP.2025.105615.

[9] The European Hydrogen Backbone (EHB) initiative | EHB European Hydrogen Backbone n.d. https://ehb.eu/ (accessed January 2, 2026).

[10] Gas Quality Standardisation and Monitoring | ENTSOG n.d. https://www.entsog.eu/gas-quality-standardisation-and-monitoring (accessed March 22, 2026).

[11] Towards a new Hydrogen market - CertifHy Green Hydrogen Guarantees of Origin are launched - CERTIFHY n.d. https://www.certifhy.eu/sin-categoria/towards-a-new-hydrogen-market-certifhy-green-hydrogen-guarantees-of-origin-are-launched/ (accessed March 22, 2026).

[12] Huang Y, Feng L, Liu Y. A Data-Driven State Estimation Framework for Natural Gas Networks With Measurement Noise. IEEE Access 2023;11:30888–98. https://doi.org/10.1109/ACCESS.2023.3262415.

[13] Ogbe E, Mukherjee U, Fowler M, Almansoori A, Elkamel A. Integrated Design and Operation Optimization of Hydrogen Commingled with Natural Gas in Pipeline Networks. Ind Eng Chem Res 2019;59:1584–95. https://doi.org/10.1021/acs.iecr.9b04550.

[14] Cristello J, Dang Z, Hugo R, Park SS. Artificial intelligence based leak detection in blended hydrogen and natural gas pipelines. Int J Hydrogen Energy 2024;91:744–64. https://doi.org/10.1016/j.ijhydene.2024.10.146.

[15] Zhao J, Bai Y, Li J, Cu W, Zhou W, Zhang Y, et al. A leakage detection method for hydrogen-blended natural gas pipelines in utility tunnels based on multi-task LSTM and CFD simulation. Int J Hydrogen Energy 2025;97:1335–47. https://doi.org/10.1016/j.ijhydene.2024.12.031.

[16] Arrigo A, Dolányi M, Bruninx K, Toubeau J-F. Machine Learning for Improved Gas Network Models in Coordinated Energy Systems 2022.

[17] Zhao M, Guo G, Fan L, Han L, Yu Q, Wang Z. Short-term natural gas load forecasting based on EL-VMD-Transformer-ResLSTM. Scientific Reports 2024 14:1 2024;14:20343-. https://doi.org/10.1038/s41598-024-70384-9.

[18] Sharma V, Cali Ü, Sardana B, Kuzlu M, Banga D, Pipattanasomporn M. Data-driven short-term natural gas demand forecasting with machine learning techniques. J Pet Sci Eng 2021;206:108979. https://doi.org/10.1016/j.petrol.2021.108979.

[19] Yin X, Wen K, Wu Y, Han X, Mukhtar Y, Gong J. A machine learning-based surrogate model for the rapid control of piping flow: Application to a natural gas flowmeter calibration system. J Nat Gas Sci Eng 2022;98:104384. https://doi.org/10.1016/J.JNGSE.2021.104384.

[20] UNI EN 16726:2025 - UNI Ente Italiano di Normazione n.d. https://store.uni.com/uni-en-16726-2025 (accessed February 16, 2026).

[21] Kunz O, Wagner W. The GERG-2008 Wide-Range Equation of State for Natural Gases and Other Mixtures: An Expansion of GERG-2004. J Chem Eng Data 2012;57:3032–91. https://doi.org/10.1021/JE300655B.

[22] COLEBROOK CF, WHITE CM. THE REDUCTION OF CARRYING CAPACITY OF PIPES WITH AGE. Journal of the Institution of Civil Engineers 1937;7:99–118. https://doi.org/10.1680/IJOTI.1937.14682.

[23] Numero di Reynolds - Wikipedia n.d. https://it.wikipedia.org/wiki/Numero_di_Reynolds (accessed February 4, 2026).

[24] Cheng N-S. Formulas for Friction Factor in Transitional Regimes. Journal of Hydraulic Engineering 2008;134:1357–62. https://doi.org/10.1061/(ASCE)0733-9429(2008)134:9(1357).

[25] Riccione - Wikipedia n.d. https://it.wikipedia.org/wiki/Riccione (accessed January 12, 2026).

[26] DM 16 aprile 2008 - Distribuzione gas naturale n.d. https://mauromalizia.it/norme-di-prevenzione-incendi/dm-16-aprile-2008/ (accessed January 12, 2026).

[27] UNI 9165:2020 - UNI Ente Italiano di Normazione n.d. https://store.uni.com/uni-9165-2020 (accessed January 13, 2026).

[28] SAInt - An Integrated Systems Modeling Platform | encoord n.d. https://www.encoord.com/saint (accessed January 21, 2026).

[29] Benner P, Grundel S, Himpe C, Huck C, Streubel T, Tischendorf C. Gas Network Benchmark Models 2018:171–97. https://doi.org/10.1007/11221_2018_5.

[30] Soleimani N, Trinchero R, Canavero FG. Bridging the Gap Between Artificial Neural Networks and Kernel Regressions for Vector-Valued Problems in Microwave Applications. IEEE Trans Microw Theory Tech 2023;71:2319–32. https://doi.org/10.1109/TMTT.2022.3232895.

List of Figures

List of Tables