European Commission

JRC SCIENCE FOR POLICY REPORT

Security of Gas Supply with the ProGasNet Simulator: an Uncertainty & Sensitivity Analysis Exercise

Mara T. A., Kopustinskas V., Rosati R.,
Stakelyte G., Praks P.

2017

Joint Research Centre

EUR 28953 EN

Contact information
Name: Daniel Albrecht
Address: European Commission, Joint Research Centre, TP 361, via E. Fermi 2749, 21027 Ispra (VA), Italy
Email: daniel.albrecht@ec.europa.eu
Tel.: +39 0332 78 9237

JRC Science Hub
https://ec.europa.eu/jrc

JRC109751

EUR 28953 EN

PDF ISBN 978-92-79-77155-2 ISSN 1831-9424 doi:10.2760/818537

Luxembourg: Publications Office of the European Union, 2017

How to cite this report: Mara T. A., Kopustinskas V., Rosati R., Stakelytė G., Praks P., *Security of gas supply with the ProGasNet simulator: an uncertainty and sensitivity analysis exercise*, EUR 28953 EN, Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-77155-2, doi:10.2760/818537, JRC109751

Security of Gas Supply with the ProGasNet simulator: an Uncertainty & Sensitivity Analysis Exercise

Abstract
Security of gas supply is a crucial stake for the European countries. The ProGasNet simulator has been developed by the European Commission to study the vulnerability of the gas network to different threats. With the simulator the European gas network (or a portion of the network) can be modelled. But several uncertainties are present in the models. Therefore, it is important to analyse the impact of these uncertainties on the model-based inferences. This report presents an uncertainty and sensitivity exercise of the security of gas supply model of an anonymised EU gas transmission network of several members. The uncertainty analysis shows that, for some scenarios of gas supply disruption, due to the uncertainties in the input parameters it cannot be clearly concluded whether the network is reliable (i.e. can supply the demand of the countries) or not. The sensitivity analysis points out the uncertain inputs mostly responsible for this lack of precision. It is found that priority should be given to the better assessment of the peak demand of the studied countries if one wants to get reliable results. Additionally, the study also highlights areas where the studied infrastructure can be improved.

Contents

Foreword

This deliverable was carried out under the umbrella of the European Commission Competence Centre on Modelling.

The Competence Centre on Modelling promotes a transparent, coherent and responsible use of modelling to underpin the evidence base for EU policies. It leverages the modelling capacity and competences across the Commission and beyond. Starting with the Commission-wide modelling inventory MIDAS, it supports a proper documentation, use, and reuse of models. It further helps identifying common approaches to quality and transparency of model use, and establishes a Community of Practice on Modelling.

Within the Competence Centre on Modelling, the SAMO[1] team has the mission to carry out uncertainty and sensitivity analyses of EC workhorse models, to conduct research in this field, to provide tools, training and ad hoc scientific support to model users in order to enhance the robustness of model-based evidences in the European Commission.

For more information on the Competence Centre please visit https://ec.europa.eu/jrc/en/modelling.

[1] Acronym for Sensitivity Analysis of Model Output

Acknowledgements

The authors are thankful for the support, discussions and ideas for this work to Stefano Tarantola, Ricardo Bolado-Lavin from Energy security, distribution and markets unit of the Joint Research Centre, Directorate C. The authors also thank Daniel Albrecht and Paul Smits of the Modelling, Indicators and Impact Evaluation Unit, Directorate I for their careful proofreading of the report.

Authors

Mara T. A., Kopustinskas V., Rosati R., Stakelytė G., Praks P.

Executive summary

The report presents an uncertainty and sensitivity analysis exercise of the security of gas supply model implemented in the probabilistic gas network simulator ProGasNet. The study aims to identify and rank the model parameters that most significantly affect the security of supply. The sensitivity analysis results are important for model improvements, better understanding of the simulation process and especially for identification of areas where to target further research.

Policy context

The ProGasNet simulator has been developed with the primary purpose to quantify the security of gas supply situation in probabilistic metrics. The simulator can be used to perform risk assessment of the gas transmission network as required by the EC Reg 994/2010. In addition, the simulator can evaluate infrastructure development plans and projects of common interest. ProGasNet was used in vulnerability assessment which could be of interest to the critical infrastructure protection policy-makers.

Key conclusions

The study showed the potential and usefulness of sensitivity study applied to the ProGasNet gas network model. It has not only identified the most important model parameters for which more attention should be paid during the estimation process of the input parameter values, but also provided useful insights into the simulation process by confirming, e.g., the heterogeneity of the network from a sensitivity analysis perspective.

Main findings

The model was run for four different scenarios representing different disruption situations. The study confirms the results already observed in other studies that some disruption scenarios affect only part of the network (e.g. specific countries) while other parts of the network are not affected. This clearly indicates heterogeneity of the network and the need for further infrastructure development. The most important parameters for each country are identified, and peak demand value is the main parameter for the three countries. Therefore better assessment of the peak demand is crucial to guarantee the reliability of the results provided by ProGasNet.

Related and future JRC work

The JRC currently develops two other models for the gas transmission network analysis: GEMFLOW mass-balance model and EUGas physical model. The models have different granularity, geographical coverage and outputs, but they all have many common input data needs. They can be used to estimate the values of some of the ProGasNet input parameters. The sensitivity analysis has highlighted important directions toward which to investigate in order to enhance the reliability of the ProGasNet model responses.

This work was carried out under the umbrella of the European Commission Competence Centre on Modelling. The Competence Centre promotes a transparent, coherent and responsible use of modelling to underpin the evidence base for EU policies. The Competence Centre, which was launched in 2017, is hosted by JRC and draws on the expertise of modellers and policy officers across the Commission.

Quick guide

The report is structured in three main parts:

— Description of the ProGasNet simulator;

— Description of the methodology to perform sensitivity and uncertainty analysis;

— Analysis of the results and discussion;

1 Introduction

The Competence Centre on Modelling, hosted by the Modelling, Indicators and Impact Evaluation unit of the Joint Research Centre has been created to *promote a responsible, coherent and transparent use of modelling to underpin the evidence base for EU policies*. The Sensitivity Analysis group (SAMO), part of this Competence Centre, promotes the responsible use of models by acknowledging and accounting for uncertainties in model inputs. For this purpose, the SAMO group proposes training, ad hoc support and statistical tools to model users to carry out this kind of analysis.

The present report describes one of the first applications of uncertainty and sensitivity analysis of model responses conducted by the SAMO group. It is a joint work between Unit I.1 and Unit C.5 of the Joint Research Centre, an ad hoc support to the study of the vulnerability of a gas natural network between three European countries.

Natural gas networks can be viewed as complex technical systems, which are exposed to various threats, for example technical failures, natural disasters and human/political uncertainties. As a consequence of these threats, a subset of network components might fail. In order to simulate security of gas supply due to component failures/attacks, the probabilistic gas network simulator ProGasNet (Probabilistic Gas Network Simulator) has been developed.

ProGasNet is able to model, in a single computer model, capacity and reliability constraints of a natural gas network. The physical model is based on graph theory (maximum flow algorithm), whereas component failures are simulated by the Monte-Carlo method.

The ProGasNet simulator has been developed with the primary purpose to quantify the security of gas supply situation in probabilistic metrics. The simulator can be used to perform risk assessment of the gas transmission network as required by the EC Regulation 994/2010. In addition, the simulator can evaluate infrastructure development plans and the proposed projects of common interest. ProGasNet was used in vulnerability assessment which could be of interest to critical infrastructure protection policy makers.

Despite of the insightful information that can be obtained regarding the security of gas supply within European regions with this software, still ProGasNet remains a strong approximation of the reality of gas supply in Europe (simplified assumptions, quasi steady-state modelling, lack of knowledge about some parameter values, etc). Therefore, it is important to assess the impact of epistemic uncertainties in gas network models in order to point out those sources of uncertainty that have an impact on the model responses of interest. Identifying important sources of (epistemic) uncertainty allows to guide further investigation for possible improvements of gas network models.

Therefore, an uncertainty and sensitivity analysis of a gas network model has been carried out. The selected network was an EU gas transmission network of several member states. The model is based on realistic network topology and data, but due to sensitivity of information, the network is anonymised. The study aims to identify and rank the model parameters that most significantly affect the security of supply. Additionally, the analysis has also provided fruitful information about the studied gas transmission network model by highlighting for instance its weaknesses.

2 The ProGasNet simulator

The ProGasNet simulator is the JRC in-house developed software tool currently in use at the Energy Security, Systems and Systems and Market Unit of the Directorate of Energy, Transport and Climate. ProGasNet is used for experimental simulation-based security of supply analyses of selected European gas transmission networks. Usually, one million Monte-Carlo simulations are automatically solved within one hour on a single-core processor computer. The software tool can make use of a multi-core computer, as multiple simulations (Monte Carlo runs) can be evaluated independently.

In order to concurrently model both reliability and capacity constraints of gas transmission networks, ProGasNet uses a stochastic network representation, where each node and edge of the flow network can randomly fail, according to a given probabilistic model. The component failures are sampled by the Monte-Carlo simulations. In each simulation, a maximum-flow optimization problem with a user defined priority of supply pattern is solved.

In ProGasNet, a stochastic network flow model with a priority supply pattern is assumed and used for probabilistic reliability assessment of gas networks. The ProGasNet algorithm is robust as well as flexible and also able to deal with various priority-supply strategies. The default version of the algorithm uses a priority supply pattern based on geographical distance from the source node: Nodes geographically closer to the gas source are served first.

The ProGasNet software tool is suitable for evaluation of new energy infrastructure, either real or virtual (Praks, Kopustinskas, & Masera, 2015), bottleneck analysis (Kopustinskas & Praks, 2015), time-dependent gas storage analysis and component importance ranking. In order to detect and analyse weak parts of gas networks, we also test approaches for reusing ProGasNet results for vulnerability and resilience analyses (Praks, Kopustinskas, & Masera, 2017).

2.1 The ProGasNet computational algorithm

The ProGasNet estimates consequences by applying a maximum flow (MF) algorithm. This algorithm has already been tested on gas transmission networks of several EU countries. The mathematical description of the MF problem is a standard problem in graph theory (Deo, 2008).

Let us consider a network having m nodes and n arcs which carries natural gas. We associate with each arc (i, j) a non-negative integer flow with an upper bound of u_{ij}. We shall assume throughout the development that the u-values (arc capacities) are finite integers. In such a network, we wish to find the maximum amount of flow from the source node 1 to the sink node m.

Let f represent the amount of flow in the network from node 1, called the source, to node m, called the sink. Then the maximal flow problem may be mathematically formulated as an optimization problem with constraints:

Maximize f subject to

$$\sum_{j=1}^{m} F_{ij} - \sum_{k=1}^{m} F_{ki} = \begin{cases} f & if\ i = 1 \\ 0 & if\ i \neq 1 \text{ or } m \\ -f & if\ i = m \end{cases} ,$$

$$F_{ij} \leq u_{ij},\ i, j = 1, \dots, m \ ,$$

$$F_{ij} \geq 0,\ i, j = 1, \dots, m \ ,$$

where the sums and inequalities are taken over existing arcs in the network. Here the symbol F_{ij} represents the flow from node i to node j. This is called the node-arc formulation of the MF problem since the constraint matrix is a node-arc incidence matrix.

The aim of the optimization problem is to maximize the value of f which represents the amount of flow passing from the source node to the sink node. The first constrain represents a conservation of flows: the sum of the flows entering a node must equal the sum of the flows exiting a node, except for the source node and the sink node. The second constrain represents capacity: the flow of an edge is non-negative and cannot exceed its capacity. The optimization problem of the MF can be solved by various approaches, for example by linear programming or Ford–Fulkerson algorithm, which finds directed paths from the source node to the sink node with available capacity on edges in this path. In the algorithm, this path-searching process is repeated until no additional flow can be added to this directed path. The algorithm uses the concept of so called residual networks, which represents the maximum additional flow that can be sent from any node i to any node j using the arcs (i, j) and (j, i). The case of multiple sources and sinks, involving several source nodes s_1, s_2,…, s_k and several sink nodes t_1, t_2,…, t_r, and where the flow from any source can be sent to any sink, it is known as Multiple Sources and Sinks problem, and can be straightforwardly converted into a one-source and one-sink problem: let us introduce a virtual source s (virtual source node) with edges (of unlimited capacity) directed from this virtual source s to all source nodes s_1, s_2,…, s_k. Furthermore, let us introduce a virtual sink t (virtual source node) with edges (also of unlimited capacity) directed from all sink nodes t_1, t_2,…, t_r to the virtual sink t. Then the problem of maximizing the total value of the flow from all sources is then the same as that of maximizing the value of the flow from s to t.

In each Monte-Carlo simulation step, the deterministic maximum flow model with sampled component network failures is solved as described above. The algorithm uses linear algebra operations for the priority supply pattern, in order to reorder network elements according to the distance based approach (Praks, Kopustinskas, & Masera, 2015). In order to improve the readability of the pseudo-code, the algorithm is expressed by the matrix form. The stochastic flow network algorithm with priority supply pattern based on distance from source includes the following steps:

a. $\boldsymbol{P}$ = *initP(*$\boldsymbol{L}$*,* p_f*) {Initializing and defining failure probability matrix of network elements}*
b. *for i=1,2, …* n_{steps} *{Main Monte-Carlo loop}*
c. $\boldsymbol{C_{rnd}}$ = *randpert(*$\boldsymbol{C}$*,* $\boldsymbol{P}$*) {Failed elements have reduced capacity}*
d. $\boldsymbol{L_{rnd}}$ = *clear_failured_elements(*$\boldsymbol{L}$*) {Totally failed elements are not reachable}*
e. $\boldsymbol{d}$ = *distance (*$\boldsymbol{L_{rnd}}$*).*
f. *[*$\boldsymbol{d_{sorted}}$*,* $\boldsymbol{ix}$*] = sort(*$\boldsymbol{d}$*) {The row vector* $\boldsymbol{d}$ *is sorted by the ascending order}*
g. $\boldsymbol{\Pi}$ = *speye(t) {Sparse identity matrix of order t }*
h. $\boldsymbol{\Pi}$ = $\boldsymbol{\Pi}(\boldsymbol{ix},:)$ *{Permutation matrix of elements according to the distance-based approach}*
i. $\boldsymbol{C_{rnd}} = \boldsymbol{\Pi} \times \boldsymbol{C_{rnd}} \times \boldsymbol{\Pi}^T$ *{Distance -based permutation of the capacity matrix* $\boldsymbol{C_{rnd}}$*}*
j. $\boldsymbol{f}$ = *maxflow(1, t,* $\boldsymbol{C_{rnd}}$*)*
k. $\boldsymbol{F_i} = \boldsymbol{\Pi}^T \times \boldsymbol{f} \times \boldsymbol{\Pi}$*; {Inverse transformation of the flow vector* $\boldsymbol{f}$*}*
l. *end*

Let us describe the algorithm. Step a is used to define the failure probability matrix $\boldsymbol{P}$ of the network elements. Then in Step b, the Monte-Carlo simulation starts. Network elements, which are stored in the capacity matrix $\boldsymbol{C}$, are subject to a random failure, according to the failure probability matrix $\boldsymbol{P}$. In Step c the Monte-Carlo sampled capacity matrix is stored in matrix $\boldsymbol{C_{rnd}}$.

The total failure of the network element causes the component inaccessibility, which can be coded in the algorithm by an infinity distance penalty. For this reason, in Step d the length matrix must be updated. In this way, in case some network component suffers a total failure, the affected network element is not accessible from the source node and the corresponding element of the length matrix $\boldsymbol{L_{rnd}}$ has to be updated.

The algorithm can be also used for modelling the partial failure of a multi-state component, which can be expressed by a partial reduction of the component capacity. Contrary to the total failure, in case of a partial failure the affected network component remains accessible, so the update of the length matrix $\boldsymbol{L_{rnd}}$ in Step d is not necessary.

In Step e the distance vector $\boldsymbol{d}$ is computed. The vector $\boldsymbol{d}$ contains the distance of the shortest paths between the virtual source node and all remaining non-virtual nodes. An entry j of vector $\boldsymbol{d}$ represents the distance of the shortest path from the virtual source node 1 to node j. We used the MATLAB tool Bctnet based on Dijkstra's algorithm. Contrary to the classical Dijkstra's algorithm, it is not necessary to compute the full distance matrix, as only the distance from the virtual source node is used in our algorithm.

In Step e, the distance vector $\boldsymbol{d}$ is sorted by the ascending order to vector $\boldsymbol{d_{sorted}}$, in order to identify the priority for the node commodity supply, because the network elements geographically close to the source node (minimum distance according to length matrix $\boldsymbol{L}$) have to be served first. The vector $\boldsymbol{ix}$ contains the indices satisfying $\boldsymbol{d_{sorted}} = \boldsymbol{d}(\boldsymbol{ix})$.

In Step g the matrix $\boldsymbol{\Pi}$ of order t is created, in order to form the permutation matrix. The matrix $\boldsymbol{\Pi}$ has initially ones on the main diagonal and zeros elsewhere (identity matrix). In order to save computer memory, one can exploit the scarcity pattern of the matrix Π.

In Step h a permutation matrix $\boldsymbol{\Pi}$ of a graph isomorphism problem is computed according to the distance from the gas source, in order to transfer the original model to the distance-based approach by a dynamic reordering of the network elements. Columns of the matrix $\boldsymbol{\Pi}$ are permuted according to the indexes produced in Step f.

In Step i, the graph isomorphism task is computed by linear algebra operations.

Then, in Step j the flow vector $\boldsymbol{f}$ of the Maximum flow algorithm is computed. The aim is to maximize the commodity flow from the virtual source node 1 to the virtual sink node t, according to given constraints. In our computer implementation, we used the above-mentioned Ford–Fulkerson algorithm.

To finish the simulation, in Step k the computed flow vector is transformed back to the original problem by the inversion linear algebra operation. As the permutation matrix $\boldsymbol{\Pi}$ is sparse and orthogonal, the linear algebra operations are very fast and stable. The transformed flow vector is stored in the optimal flow matrix $\boldsymbol{F}$ which, once the Monte-Carlo simulations achieved, is ready for further exploration with statistical methods enabling the monitoring of the flow patterns generated.

2.2 The case study: network/data

Figure 1 shows topology of the study case gas transmission network. It is based on a real regional network topology and data, however geographical location is not displayed. The transmission network GIS data are converted to a graph by creating nodes and links (edges). The nodes are:

- Demand nodes (consumers connected to pressure reduction stations of the transmission network;
- Compressor stations;
- Supply nodes (storages, LNG terminals, import points at cross-borders).

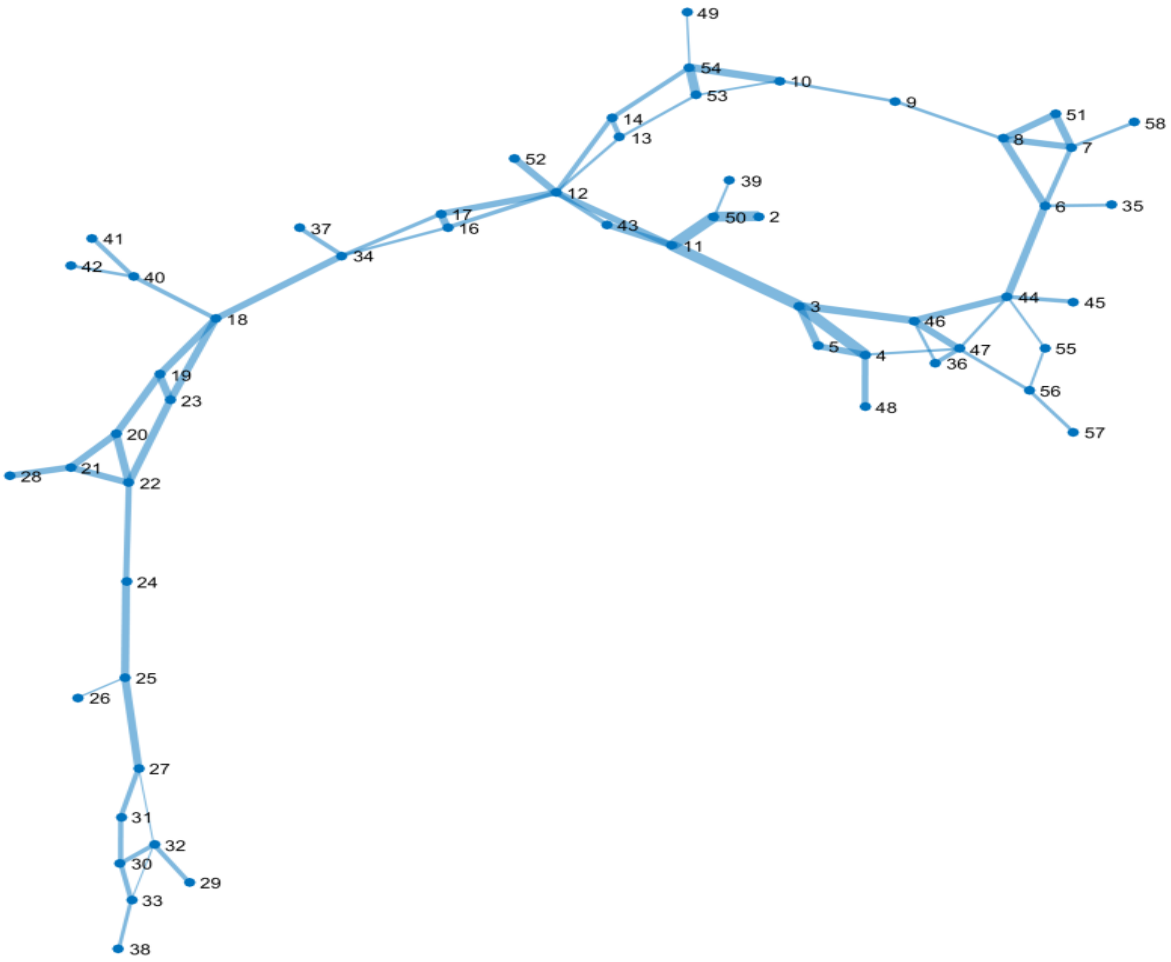
The network links (edges) are typically pipelines. The model explicitly considers two parallel pipelines as two components (double links between nodes).

The basic network data are the same as already reported (Kopustinskas & Praks, 2015). Here we replicate only the most important network data.

The demand nodes are determined by daily demand values, see **Table 1**. These numbers represent peak demand values. But we can use also average winter or summer

consumption values, depending on the purpose of the study. **Table 2** shows maximum capacities and type (pipeline, UGS or LNG) of input supply nodes. In case of underground gas storages (UGS), also the output values of not fully loaded storages can be used. The total maximum supply capacity is 83.6 mcm per day. Note that nodes 2 and 19 are the main gas sources in the network. The total network peak demand is 45.9 mcm/d, so the network in theory has certain degree of spare capacity to compensate supply disruptions. The experience of different analysis already performed shows that depending on where the disruption happens, internal bottlenecks in the network prevent from full usage of this spare capacity (Kopustinskas & Praks, 2015).

Figure 1. Topological layout of the gas network.



Source: JRC developed simplified topology based on data obtained from the countries TSOs.

Table 1. List of non-zero demand nodes in countries no. 0, 1, 2 and 3.

Node	Demand([1])	Country	Node	Demand	Country
id.	(mcm/d)	id.	id.	(mcm/d)	id.
4	0.1	1	36	4.2	1
5	3.2	1	37	1.3	2
6	0.1	1	39	0.3	1

8	0.1	1	41	0.6	2
9	0.1	1	42	0.6	2
10	1	1	43	0.2	1
13	0.5	1	44	0.7	1
17	0.1	1	45	1.3	1
18	8.5	2	47	0.1	1
20	0.6	2	48	1.8	1
25	0.5	3	49	0.2	1
26	0.8	3	51	7	0
27	3	3	52	0.6	1
28	6	0	53	0.1	1
30	0.5	3	55	0.2	1
33	0.5	3	57	0.2	1
34	0.5	2	58	0.3	1
35	0.1	1	60	45.9	SUM([2])

([1]) The gas demand is represented by mcm/d (millions of standardised cubic meters per day).
([2]) Node 60 represents the total demand in the network (SUM).

Source: JRC data elaborated from the countries TSOs data.

Table 2. Maximum possible supply capacity to the network from the source nodes.

Node	Type	Capacity, mcm/day
2	Pipeline	31.2
10	LNG	10.2
11	Pipeline	7
19	UGS	30
29	Pipeline	4
38	Pipeline	1.2

Source: Data from the countries TSOs.

For each network component, failure data must be provided. The following components (nodes) are considered for failures:

- Compressor station (CS) failure: 2.5E-01/yr;
- Underground storage failure: 1.0E-01/yr
- LNG terminal failure: 1.5E-01/yr
- Pipeline failure: 3.5E-05 /km/yr.

The model uses annual failure data (probability of failure per year), however when simulations are performed, one month interval is considered. It is assumed that the same peak consumption in the network is constant during this one month period.

The compressor station node normally is modelled as working or failed, for each state determining the corresponding capacity of the outgoing pipelines. The capacity reduction due to compressor station failure is normally estimated by hydraulic model computations or expert evaluation. As a consequence, due to a CS failure, capacity reduction by 20% of the inlet pipelines and also the outlet pipelines until the next connection node is assumed. This assumption is based on physical flow models, however is not accurate in all cases and also multiple CS failures will have more severe effects on the network operation. In the future, a physical model should be developed in order to estimate the effect of the CS failures more precisely.

2.3 Supply scenarios

In total four different supply scenarios were analysed. The first scenario is the reference scenario during which the system operates under normal conditions without predefined disruptions. The other three scenarios consider some disruption situations.

Scenario 1: All currently available sources. Scenario 1 represents a basic scenario when all sources can be used for supply and the network components can fail randomly according to their reliability parameters.

Scenario 2: Node 2 disruption. In this scenario, supply node 2 (the largest gas source in capacity) is not available. This scenario can test the system for the largest source disruption which can be classified as N-1 situation looking at the network globally.

Scenario 3: Node 19 disruption. Scenario 3 runs the model with disconnected node 19. Thus, the second largest gas source, the underground gas storage, is not available.

Scenario 4: Loss of two largest gas sources (Nodes 2 & 19). In this scenario, the two largest sources, i.e. node 2 and node 19 are not available. The underground gas storage can be unavailable due to technical problems, failures or inability to fill it up during summer period. Scenario 4 simulates a more challenging crisis in which the both sources of the highest capacity are unavailable. This scenario is used to demonstrate vulnerability of the network, when the largest and the second largest gas sources are lost simultaneously. The network can be supplied only with source nodes 10 and 11.

2.4 Results of the ProGasNet simulations

In this section, we illustrate the typical ProGasNet results to assess the security of gas supply. Note that although the network described in **Figure 1** is analysed, the scenarios were defined differently and reflect different disruption situations. The results presented here are only for demonstration purposes and should not be linked to the sensitivity study results presented further in the report.

Table 3 presents ProGasNet results of disruption scenarios at node 60, which represents the total network demand D=45.9 mcm/d. The column "D-mean" shows the expected gas deficit at the node, which is computed as a difference between demand D and the computed mean value of the supply. The results show probabilities of having less than certain predefined percentage (20%, 50%, 80% or 100%) of the required demand or

zero volume of gas (P(Y=0)). The same information is shown graphically by CDF plot in **Figure 2**, but any non-supply percentage level can be seen in the CDF plot.

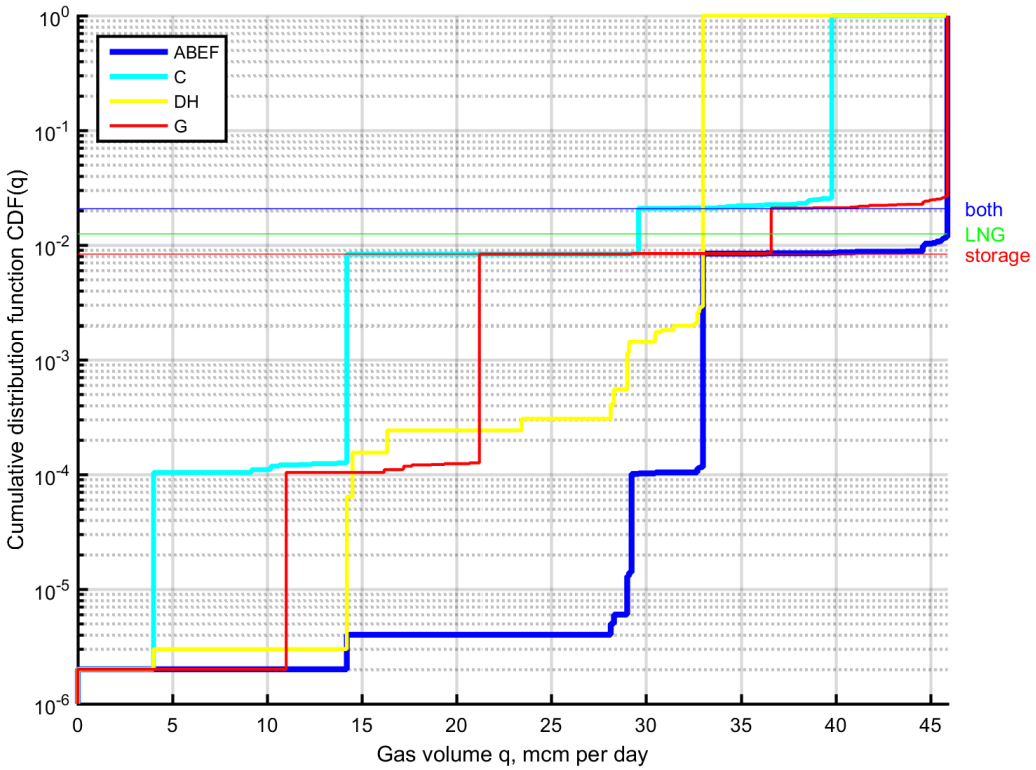
Table 3. The ProGasNet results for the total network supply (demand node 60) during 1 month with peak demand.

Scenario	D-Mean, mcm/d	P(Y=0)	P(Y<0.2D)	P(Y<0.5D)	P(Y<0.8D)	P(Y<D)
DH	12.9	0	1.0E-06	2.4E-04	1	1
C	6.5	0	1.1E-04	8.3E-03	2.2E-02	1
G	0.3	0	0	8.3E-03	2.1E-02	2.7E-02
ABEF	0.1	0	0	2.0E-06	8.5E-03	1.2E-02

Source: JRC calculations

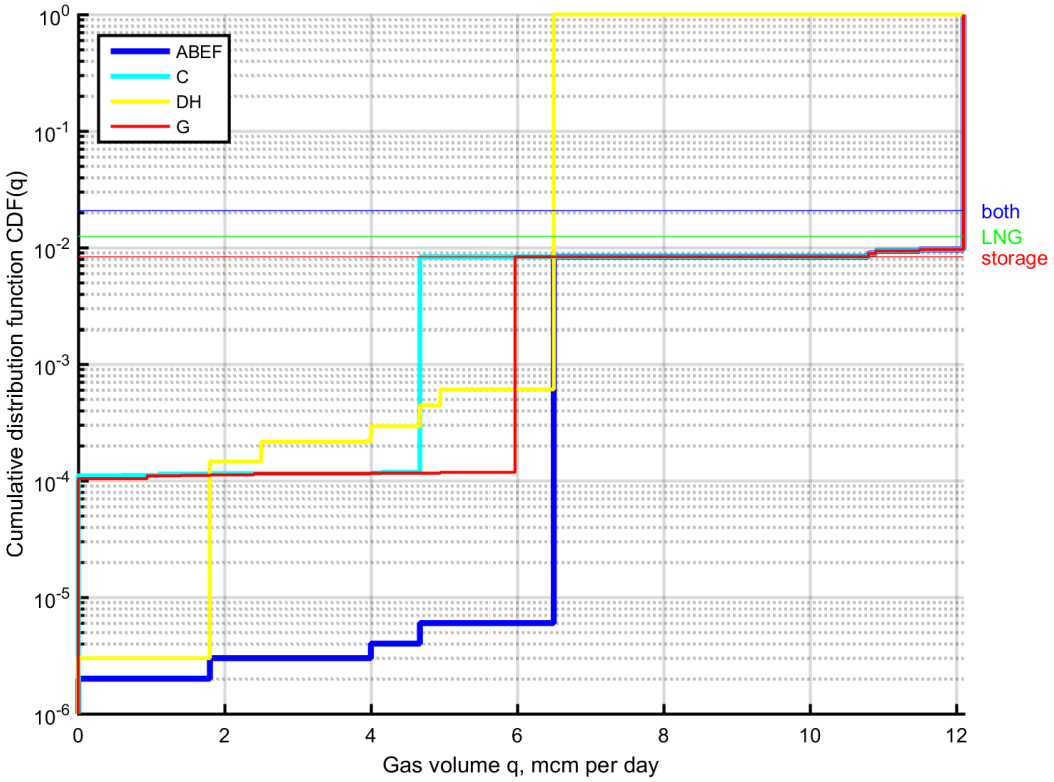
As **Table 3** shows, under scenarios D and H, the mean gas deficit is 12.9 mcm/d and P(Y<0.8D) = P(Y<D) = 1. Under scenario C, the mean gas deficit is 6.5 mcm/d and P(Y<D) = 1, i.e. the required demand cannot be satisfied due to insufficient capacity of available sources or bottlenecks in the network (Kopustinskas & Praks, 2015). The other scenarios have the expected gas deficit close to zero. Moreover, scenarios A, B, E and F have statistically similar results, which are grouped in the same row of the table. Two sample Kolmogorov-Smirnov goodness of fit hypothesis test with p-value=0.005 was used to determine whether two independent random samples are drawn from the same underlying continuous population.

Figure 2. Security of gas supply by the empirical CDF function (total network demand case).



Source: JRC ProGasNet modeling results.

Figure 3. Security of gas supply by the empirical CDF function (one country network demand case).



Source: JRC ProGasNet modeling results.

Figure 2 shows the empirical cumulative distribution function (CDF) of the total network supply for all disruption case studies. It can be interpreted as follows: the monthly probability of having less than specific daily demand (gas volume *q*, on the horizontal axis) is given by the *CDF(q)* value (on the vertical axis). We assume that the daily demand is constant during the month.

Network reliability generally depends on the network topology and also on the reliability of network elements. However, it is possible to identify input parameters of the ProGasNet model that can influence the shape of CDF plots, depending on the disruption scenario. A CDF can have visible "steps". These CDF "steps" are mainly caused by gas facilities failures, which are listed at the y-axis of the CDF plot:

- The expected monthly failure probability of the LNG is, according to the input data of the model, 0.15/12=0.0125, which is shown by the same symbol "LNG".
- The influence of the gas storage is shown by the symbol "storage" by the expected monthly failure probability of the gas storage: 0.10/12=0.00833.
- The expected monthly failure probability that LNG, gas storage or both are failed is expressed by the symbol "both".

Figure 3 shows the empirical cumulative distribution function (CDF) of the partial network demand (e.g. one country). The CDF results can be obtained for any part of the network demand (one city node, one country, a region of countries) depending on the purpose of the study.

We can see that disruption scenarios D and H are the worst cases in terms of security of gas supply. On the other hand, scenarios A, B, E and F represent the best situation. Although these scenarios have approximately the same CDF plots of the gas supply at the global (summary) level, they have not only differences on the redundancy of the gas sources, but also differences at the node level. These aspects can be quantified in ProGasNet as well (Praks, Kopustinskas, & Masera, 2015).

3 Uncertainty & sensitivity analysis methodology

3.1 Probabilistic framework

Uncertainty in model predictions stems from the lack of knowledge about the process of interest (in the present case, about the gas transport), model simplification and parameters' value. For some of the model inputs, it is possible to refine our knowledge about their probable value (that is, decreasing their uncertainty) but such a task is time consuming. Our strategy is i) to assign large but likely uncertainty ranges to the inputs of the gas network model, ii) to check whether this yields large uncertainty in the model predictions of interest and iii) to identify eventually those inputs that are mostly responsible for the predicted uncertainty. In step iii), it is expected that only a few of the uncertain inputs are identified as influential in order to reduce the effort to pay during the subsequent input uncertainty refinement.

In the present study, twenty model inputs of the gas transmission network model described in section 2.2 have been deemed as uncertain. They are gathered in **Table 4**. It can be noted that some are assigned uniform distributions within plausible ranges while some others are assigned normal distribution. It is assumed that the model input values are independent of each other. These uncertainties reflect the experts' belief before further investigation. The prior uncertainties being large reflect the fact that the experts (here the modellers) have a vague knowledge about model inputs uncertainty.

Table 4. Input uncertainty distributions of the gas network model.

Label	Parameter	Baseline value	Distribution (1)	Type (2)	Accuracy (3)
X_1	Capacity of source N2	31.2	U(16,31.2)	A	L
X_2	Capacity of source N19	30.0	U(15,30)	A	L
X_3	Capacity of source N10	10.2	U(5,10.2)	A	L
X_4	Capacity of source N11	7.0	U(3.5,7)	A	M
X_5	Compressor station capacity reduction factor	0.2	$N(0.2,0.05^2)$	E	M
X_6	Peak demand of Country 1	15.5	$N(15.5,0.75^2)$	E	L
X_7	Peak demand of Country 2	12.1	$N(12.1,0.6^2)$	E	L
X_8	Peak demand of Country 3	5.3	$N(5.3,0.4^2)$	E	L
X_9	Failure frequency of LNG	0.15	$N(0.15,0.015^2)$	E	H
X_{10}	Failure frequency of storage facility	0.1	$N(0.1,0.01^2)$	E	M
X_{11}	Failure of compressor station	0.25	$N(0.25,0.025^2)$	E	M
X_{12}	Failure frequency of a	3.5E-05	$N(3.5x10^{-5},(3.5$	E	M

	pipeline		$x10^{-6})^2$)		
X_{13}	Capacity of DN1000 pipeline	30.6	N(30.6,1.5^2)	E	H
X_{14}	Capacity of DN800 pipeline	17.1	N(17.1,0.86^2)	E	H
X_{15}	Capacity of DN700 pipeline	12.1	N(12.1,0.6^2)	E	H
X_{16}	Capacity of DN600 pipeline	8.1	N(8.1,0.40^2)	E	H
X_{17}	Capacity of DN500 pipeline	5.1	N(5.1,0.25^2)	E	H
X_{18}	Capacity of DN400 pipeline	2.8	N(2.8,0.14^2)	E	H
X_{19}	Capacity of DN350 pipeline	2.0	N(2.0,0.1^2)	E	H
X_{20}	Capacity of DN300 pipeline	1.3	N(1.30,0.065^2)	E	H

([1]) U=Uniform distribution, N(μ,σ^2)=Normal distribution of mean μ and variance σ^2.
([2]) E stands for epistemic uncertainty as opposed to A- aleatory uncertainty.
([3]) The modellers belief regarding the assigned prior uncertainty: L=Low, M=Medium and H=High.

Source: JRC estimations.

The independence of the input variables allows writing the input joint probability density function (PDF) as the product of the marginal PDF's, namely,

$$p_x(\boldsymbol{x}) = p_1(x_1)p_2(x_2)\dots p_{20}(x_{20}) \quad (1)$$

with, for instance,

$$p_1(x_1) = U(16,31.2) = \begin{cases} \frac{1}{31.2-16} & \text{if } x_1 \in [16,31.2] \\ 0 & \text{otherwise} \end{cases}$$

$$p_5(x_5) = N(0.2,0.05^2) = \frac{1}{2\pi\times 0.05} e^{-0.5\left(\frac{x_5-0.2}{0.05}\right)^2}.$$

Therefore, the model inputs are treated like random variables as well as the model responses. In the sequel, the model response of interest is denoted Y. Without loss of generality, the latter is assumed scalar and only function of the uncertain inputs $\boldsymbol{X}=(X_1,\dots,X_{20})$, that is

$$y = g(\boldsymbol{x})$$

where the lowercase means the value (e.g. y) taken by the random variable (e.g. Y).

We define the following two mathematical operators:

$$E(Y) = \int_{R^{20}} g(\boldsymbol{x}) p_x(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

$$V(Y) = E\left(\left(Y - E(Y)\right)^2\right)$$

which are respectively the mathematical expectation of Y and its variance. We also define the conditional expectation and conditional variance as follows

$$E(Y|X_i = x_i) = \int_{R^{19}} g(\boldsymbol{x}) p_{x_{-i}}(\boldsymbol{x}_{-i}) \mathrm{d}\boldsymbol{x}_{-i}$$

$$V(Y|X_i = x_i) = E(Y^2|X_i = x_i) - \left(E(Y|X_i = x_i)\right)^2$$

where $\boldsymbol{x}_{-i} = \boldsymbol{x}/x_i$.

3.2 Uncertainty analysis: the Monte Carlo method

Uncertainty analysis aims to characterize the uncertainty on Y knowing the uncertainties on the X-variables. This can be achieved by propagating the input uncertainties through the model response. The uncertainty in Y is fully characterized by its marginal PDF defined as,

$$p_y(Y = y) = \int_{R^{20}} \delta\left(y - g(\boldsymbol{x})\right) p_x(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{2}$$

where δ is the Dirac distribution function. Eq.(2) can be roughly estimated by the following integral,

$$p_y(Y = y) \simeq \frac{1}{\Delta} \int_{R^{20}} \Pi_\Delta\left(y - g(\boldsymbol{x})\right) p_x(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{3}$$

where Π_Δ is the window function defined as,

$$\Pi_\Delta = \begin{cases} 1 \text{ if } g(\boldsymbol{x}) \in [y - \frac{\Delta}{2}, y + \frac{\Delta}{2}] \\ 0 \qquad \text{otherwise} \end{cases}$$

Eq.(3) can be estimated via Monte Carlo simulations in which one randomly draws *N* input set values with respect to the input joint PDF (Eq.(1)), runs the model and evaluates the output vector $\boldsymbol{y}=(y_1,\ldots,y_N)$ and finally, for a given value y and window's width Δ around y, counts how many predicted values in vector $\boldsymbol{y}$ lie within $[y - \frac{\Delta}{2}, y + \frac{\Delta}{2}]$. Denoting N_Δ this number, one can approximate Eq.(3) as follows:

$$\hat{p}_y = \frac{N_\Delta}{\Delta \times \mathrm{N}}$$

Alternatively, Eq.(3) can be evaluated with the smoothing kernel-density approach (Parzen, 1962; Botev, Grotowski, & Kroese, 2010) which is known to provide smoother and more accurate approximations.

3.3 Quantitative sensitivity analysis

3.3.1 The variance-based sensitivity indices

The theory of variance-based sensitivity analysis was elaborated by Ilya M. Sobol' (Sobol, 1993). Sobol' proved that if $g(\boldsymbol{x})$ is square-integrable, then it can be cast onto orthogonal functions of increasing dimensionality as follows,

$$g(\boldsymbol{x}) = g_0 + \sum_{i_1=1}^{n} g_{i_1}(x_{i_1}) + \sum_{i_2>i_1}^{n} g_{i_1 i_2}(x_{i_1}, x_{i_2}) + \sum_{i_3>i_2}^{n} g_{i_1 i_2 i_3}(x_{i_1}, x_{i_2}, x_{i_3}) + \cdots + g_{1\ldots n}(x_1, \ldots, x_n) \quad (4)$$

with $E(g_{i_1 \ldots i_s} \times g_{j_1 \ldots j_t}) = 0$ if $i_1 \ldots i_s \neq j_1 \ldots j_t$ and $E(g_{i_1 \ldots i_s} \times g_{i_1 \ldots i_s}) = D_{i_1 \ldots i_s}$.

Then, it is straightforward to prove that the total variance of y=$g(\boldsymbol{x})$ is,

$$V(f(\boldsymbol{x})) = D_y = \sum_{i_1=1}^{n} D_{i_1} + \sum_{i_2>i_1}^{n} D_{i_1 i_2} + \sum_{i_3>i_2}^{n} D_{i_1 i_2 i_3} + \cdots + D_{1\ldots n} \quad (5)$$

where $D_{i_1 \ldots i_s}$ is a positive partial variance.

This latter equation says that the total variance of the model response is explained by the partial contributions of each input variable. Each variable can contribute solely (e.g., D_{i_1} is the sole contribution of X_{i_1}) or by interaction with the other variables (e.g. $D_{i_1 i_2}$ is the partial contribution due to the interaction between X_{i_1} and X_{i_2}). Dividing Eq.(5) by the total variance yields the so-called Sobol' indices:

$$S_{i_1} = \frac{V(E(g(\boldsymbol{x})|x_{i_1}))}{V(g(\boldsymbol{x}))} = \frac{D_{i_1}}{D_y} \quad (6)$$

$$ST_{i_1} = \frac{E(V(g(\boldsymbol{x})|\boldsymbol{x}_{-i_1}))}{V(g(\boldsymbol{x}))} = \frac{\sum_{\boldsymbol{u} \ni i_1} D_{\boldsymbol{u}}}{D_y} \quad (7)$$

which are normalized between [0,1]. The former is called the first-order sensitivity index while the latter is called the total-order sensitivity index of X_{i_1} and accounts for possible interactions involving this input variable. A model input is claimed irrelevant for Y if its total-order effect is close to zero.

There are several computational methods to assess the Sobol' indices. They can be classified as either sampling-based Monte Carlo approaches (Saltelli A. , 2002; Sobol, 1993), spectral approaches (Saltelli, Tarantola, & Chan, 1999; Cukier, Fortuin, Petschek, & Schaibly, 1973; Sudret, 2008) or metamodeling-based approaches (Oakley & O'Hagan, 2004; Buzzard & Xiu, 2011). In the present work, the polynomial chaos expansion spectral method is employed to assess the Sobol' indices when the model response of interest takes continuous values.

3.3.2 Polynomial chaos expansions

The idea of polynomial chaos expansion (PCE) is to approximate each term in Eq.(4) by multi-dimensional orthonormal polynomials. The rate of convergence of such an expansion of course depends on the regularity properties of $g(\boldsymbol{x})$. By exploiting the Parseval-Plancherel relationship, one obtains a variance decomposition such as Eq.(5). Therefore, we get (Wiener, 1938),

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in N^n} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \quad (8)$$

where $\boldsymbol{\alpha} = \alpha_1 \dots \alpha_n$, with $\alpha_i \in N$, is a multi-index indicating whether $\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ depends on x_i ($\alpha_i > 0$) or not ($\alpha_i = 0$), $a_{\boldsymbol{\alpha}}$ is the polynomial coefficient associated with $\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ which is the so-called multivariate orthonormal polynomial chaos that is written

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \psi_{\alpha_1}(x_1) \times \dots \times \psi_{\alpha_n}(x_n)$$

$\psi_{\alpha_i}(x_i)$ being the α_i-th degree univariate polynomial basis element ($\psi_0 = 1$).

The expression of the univariate polynomial basis elements depends on the PDF assigned to the input variables. If $x_i \sim U(-1,1)$, $\psi_{\alpha_i}(x_i)$ is the Legendre polynomial of degree α_i, while if $x_i \sim N(0,1)$, $\psi_{\alpha_i}(x_i)$ is the Hermite polynomial of degree α_i. One can rely on the Wiener-Askey scheme to choose the appropriate polynomial family (Xiu & Karniadakis, 2002).

Once a PCE expansion such as Eq.(8) is obtained, it is straightforward to prove that the total variance of $g(\boldsymbol{x})$ is,

$$V(g(\boldsymbol{x})) = D_y = \sum_{\boldsymbol{\alpha} \in N^n} a_{\boldsymbol{\alpha}}^2 - a_{0\dots0}^2$$

by exploiting the orthonormality property of the polynomial basis elements, that is,

$$E\left(\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \times \psi_{\boldsymbol{\beta}}(\boldsymbol{x})\right) = \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}$$

where $\delta_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ is the symbol of Kronecker.

Therefore, it is possible to estimate the Sobol' indices from the PCE coefficients as follows,

$$S_{i_1} = \frac{V(E(g(\boldsymbol{x})|x_{i_1}))}{V(g(\boldsymbol{x}))} = \frac{\sum_{\alpha_{i_1} \in N} a_{0\dots0\alpha_{i_1}0\dots0}^2}{\sum_{\boldsymbol{\alpha} \in N^n} a_{\boldsymbol{\alpha}}^2 - a_{0\dots0}^2}$$

$$ST_{i_1} = \frac{E(V(g(\boldsymbol{x})|\boldsymbol{x}_{-i_1}))}{V(g(\boldsymbol{x}))} = \frac{\sum_{\boldsymbol{\alpha} \in N^n : \alpha_{i_1} > 0} a_{\boldsymbol{\alpha}}^2}{\sum_{\boldsymbol{\alpha} \in N^n} a_{\boldsymbol{\alpha}}^2 - a_{0\dots0}^2}$$

Hence, the issue with the PCE approach for variance-based sensitivity analysis is to assess the PCE coefficients. In this work this is achieved with the Bayesian sparse PCE developed in (Shao, Younes, Fahs, & Mara, 2017). With this approach, the variance decomposition is obtained from one single Monte Carlo sample of size *N*. This is very computationally cheap compared with other classical approaches (Saltelli A. , 2002; Saltelli, Tarantola, & Chan, 1999). The cost of the analysis is a criterion to keep in mind with a long-time run model like the ProGasNet.

3.4 Qualitative sensitivity analysis

3.4.1 The Monte Carlo filtering

When the model response is not smooth enough, although the variance decomposition in Eq.(5) is still possible, polynomial approximations such as the PCE might fail at providing accurate results. This is particularly true when the model response takes discrete values (say, Y is Boolean). In that case, one can rely on qualitative sensitivity analysis methods. Such methods allow to identify the irrelevant inputs of the model but do not allow to rank the input variables by order of importance contrarily to the variance-based methods.

Monte Carlo filtering is a qualitative method and it is particularly suited when the model response is Boolean.

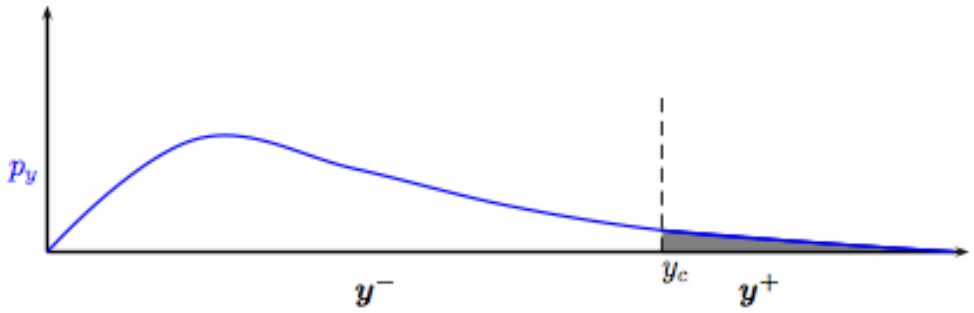
Monte Carlo filtering was introduced in the early 80's to address the following sensitivity analysis issue: *Identifying the model inputs which are mainly responsible for producing model responses in a specific region of the output space* (Hornberger & Spear, 1981). This is illustrated in **Figure 4**. This issue can be addressed with the variance-based sensitivity indices. But, because of the discrete form of the model response (in/out of the region of interest), the Sobol' indices cannot be assessed with a cheap method like PCE or emulator-based approach. Indeed, to our best knowledge, there is no such reliable method to cope with, say, Boolean-type output. In that case, it is more convenient to use the qualitative Monte Carlo filtering (MCF) approach.

In MCF, first the uncertainty on the model inputs is propagated into the model response by a Monte Carlo approach. This implies 1) to generate a random sample of $\boldsymbol{X}$ of size Nxn of the input values, 2) for each combination $(x_{j1},\ldots,x_{jn})$ of the input values, to evaluate the model response y_j, j=1,…,N. This provides the response vector $\boldsymbol{y}=(y_1,\ldots,y_N)$. Then, the input sample $\boldsymbol{X}$ is split into two subsamples, namely $\boldsymbol{X}^+$ and $\boldsymbol{X}^-$. Note that each column of the subsamples contains the realizations (i.e. random values) of each input variable. The subsample $\boldsymbol{X}^+$ provides the response vector $\boldsymbol{y}^+$ that takes values within the region of interest (i.e, $y>y_c$) while $\boldsymbol{X}^-$ is the complementary subsample.

The relevant inputs are identified by column-wisely comparing $\boldsymbol{X}^+$ with $\boldsymbol{X}^-$. Only those inputs that present significant differences in their filtered (split) subsamples are deemed influential. The differences between the pairwise subsamples are measured thanks to the Kolmogorov-Smirnov test (KS-test). The latter measures how likely two subsamples stem from the same probability law. An example of MCF result is given in **Figure 5**. On the left-side of **Figure 5**, the filtered subsamples of X_1 (elements of the first column of $\boldsymbol{X}^+$ and $\boldsymbol{X}^-$) have significant different empirical CDF. This difference means that the filtered values of X_1 gathered in $\boldsymbol{X}^+$ and $\boldsymbol{X}^-$ are not identically distributed. This indicates that to produce model responses in the region of interest (i.e. $y>y_c$), X_1 should be sampled in a specific way. On the right-side of **Figure 5**, the difference between the subsamples is not significant according to the KS-test. Therefore, it is unlikely that X_2 is responsible for producing y-values in the critical region $\boldsymbol{y}^+$.

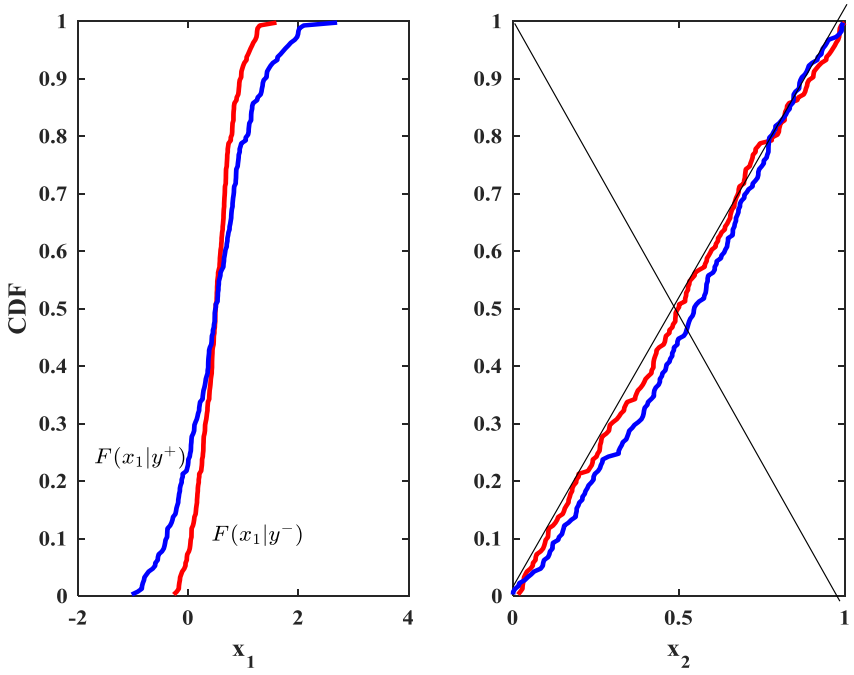
It must be mentioned here that the MCF approach, as described in the previous paragraph, might fail at identifying some relevant inputs. This is because although the KS-test indicates how likely two subsamples stems from the same marginal PDF, it does not indicate how likely they have the same joint PDF. This information is not provided by the (two-sample) KS-test. Therefore, it is also recommended to analyse the Pearson correlations matrices of the subsamples. For instance, if it is found that the correlation coefficient between X_2^+ and X_1^+ (resp. $X_2^-$and X_1^-) is high then it indicates that X_2 is responsible for producing y-values in the critical region $\boldsymbol{y}^+$ mostly by interaction with X_1.

Figure 4. The Monte Carlo Filtering approach.



Source: JRC example.

Figure 5. An example of MCF result. For X_1 (on the left-side), the conditional cdf's are significantly different while for X_2 they are not. It can be concluded that X_1 is likely responsible for producing model responses in the region of interest and not X_2.



Source: JRC example.

4 Uncertainty & sensitivity analysis: results and discussion

4.1 Model output selection

Given the computational time required to run ProGasNet, a sample of size 512 was considered. The input sample was generated according to the probability density function of each input variable (see **Table 4**). The low-discrepancy LPτ sequences of (Sobol', Turchaninov, Levitan, & Shukman, 1992) were employed. It required four days of calculation with ProGasNet to propagate the input uncertainty into the model responses for the four different scenarios under analysis. We recall that the four scenarios are:

- Scenario 1: Reference scenario when all the gas sources are available
- Scenario 2: Simulate a crisis during which pipeline source of node 2 (X_1), the most important source in terms of capacity, is unavailable
- Scenario 3: Source in node 19 (X_2) is unavailable
- Scenario 4: It is a critical case where the both Sources 2 & 19 are unavailable

It is possible to analyse the impact of the model input uncertainty onto different model responses. For this purpose, while propagating the input uncertainty with a Monte Carlo sample, one has to save the different model responses of interest after each model run. In the present work, we have considered 20 different model responses, namely: $\overline{S}$ the mean volume of gas supply, $P(S = 0)$ the probability of none gas supply, $P(S < 0.2D)$ the probability of supplying less than 20% of the demand, $P(S < 0.5D)$ the probability of supplying less than 50% of the demand, $P(S < 0.8D)$ the probability of supplying less than 80% of the demand and $P(S < D)$ the probability of supplying less than the demand. Doing so for each country provided a set of 3x6=18 different model responses.

4.2 Analysis method selection

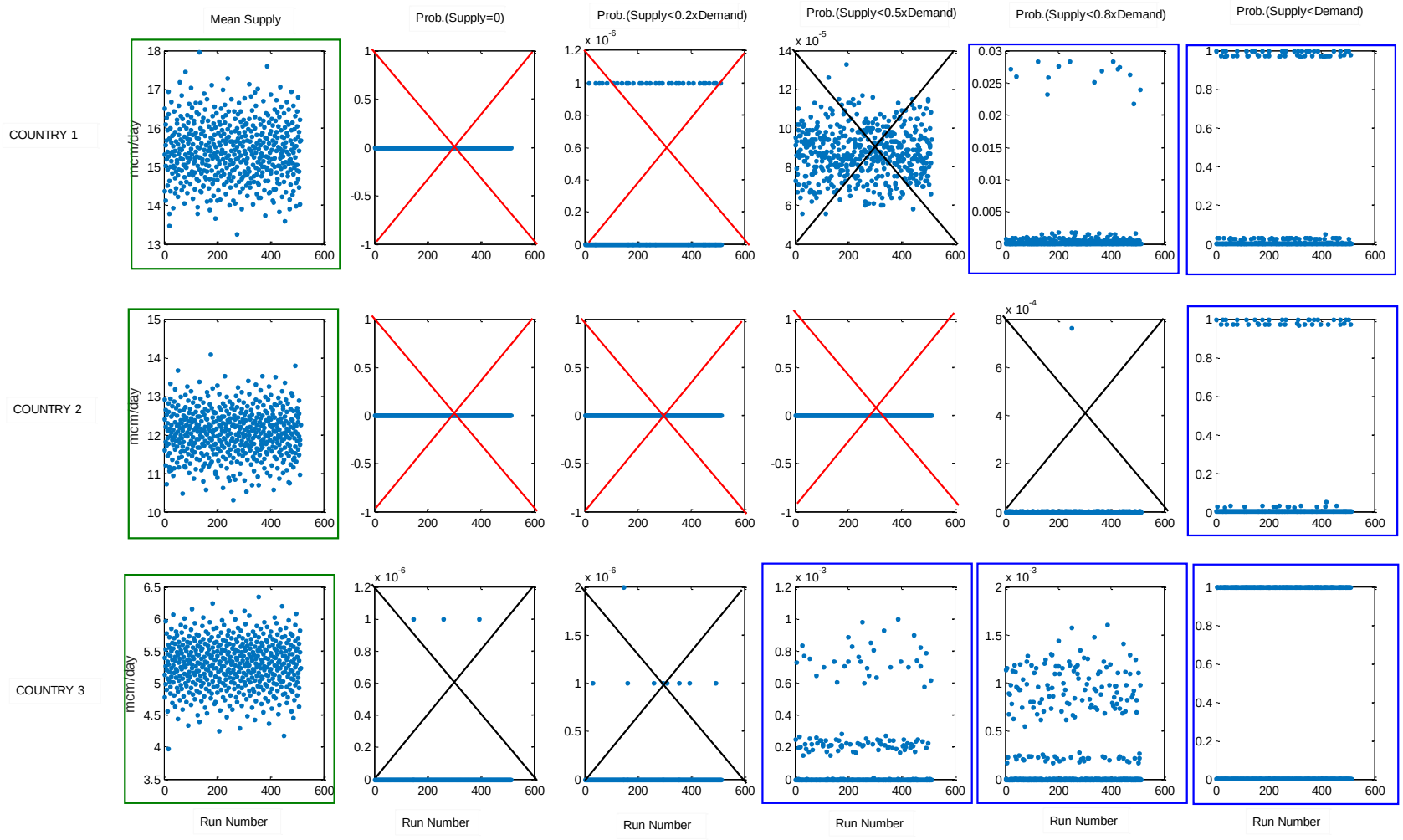
The different model responses are more or less sensitive to the uncertainty in the model inputs. **Figure 6** shows the Monte Carlo simulation results obtained for Scenario 2. In **Figure 6** outputs with a cross either do not change significantly (black cross) or do not change at all (red cross) and are not analysed. Outputs with a green frame are analysed with the PCE approach. The blue-framed outputs are analysed with MCF. It can be noted that some output variables do not vary at all (likewise $P(S = 0)$ for country 1 & 2), some do not vary significantly (values less than 10^{-3}), some others only take discrete values (e.g. $P(S < D)$) while the averages volume of gas supply $\overline{S}$ vary continuously. Consequently, only those model responses that are significantly impacted by the input uncertainty are analysed in the sequel.

The model responses that take continuous values are analysed with the polynomial chaos expansion. The output that take discrete values are analysed with MCF.

4.3 Results for Country 1

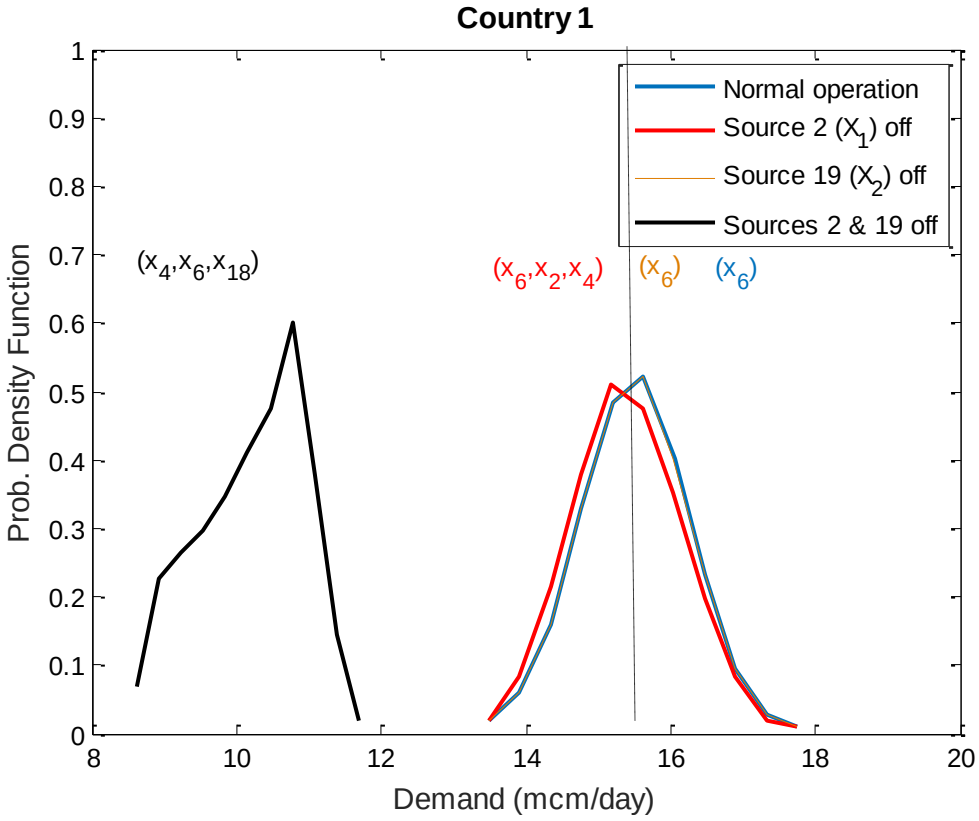
In this section, we analyse the predictive uncertainty on the mean volume of gas $\overline{S}$ supplied by the network to the different countries. The estimated probability density functions of the mean volume of gas supply (in millions of cube meters/day) for the four different scenarios are depicted in **Figure 7**. We note that the PDF's are the same for scenario 1 (normal operation) and scenario 3 (Source 19 off). This means that the system is completely resilient to the failure of this important source regarding the gas supply in Country 1. When the main source of gas is off (scenario 2), the network remains quite resilient. However, as far as scenario 4 is concerned, the system completely fails at satisfying the gas demand (vertical dashed-line) although some quantity of gas is supplied. Notably, in the first three scenarios, according to the model, the ability of the network to supply sufficient gas volume depends on the true value of some of the uncertain input variables. Indeed, **Figure 7** indicates that the system might fail at satisfying the gas demand under certain uncertain conditions, linked to capacity of sources which is subject to strong aleatory uncertainty.

Figure 6. Monte Carlo predictions of the 18 model responses of interest in Scenario 2.



Source: JRC calculations and analysis.

Figure 7. Predicted uncertainty of mean gas supply for country 1 with respect to the different disruption scenarios.



Source: JRC calculations and analysis.

To guess which are the critical uncertain inputs responsible for the variability of this model response we have carried out a global sensitivity analysis for each scenario. Given that this model response takes continuous values, the PCE method is employed (see section 3.3.2). The first-order (S_i) and total-order (ST_i) Sobol' indices are gathered in **Table 5**. We recall that S_i represents the amount of variance of the predicted mean gas supply explained by the input variable while the difference (ST_i-S_i) represents the amount due to the interaction of the variable with the other ones. We note that only scenario 2 is subject to interactions.

The results indicate that the accuracy of the model to predict the mean gas supply to Country 1 heavily depends on the knowledge of the peak demand in that country. This is particularly crucial for scenarios 1 & 3. When source N2 is off, then the model response becomes more complicated involving sources N10 and N11 (see Section 2.2 for their features). When both sources N2 and N19 are off (scenario 4), then the capacity of source N11 prevails for the mean gas volume supplied to country 1, the country peak demand (X_6) becoming less relevant. In summary, to predict accurately the mean gas volume supplied to Country 1 it is crucial to know accurately the value of the following inputs: X_6, X_4, X_2 (by order of importance) and in a less extent X_{18}.

Table 5. Relevant inputs for the predicted mean gas supply for country 1.

Scenario	**Relevant inputs**	S_i	ST_i
1 (Normal)	X_6 = Peak demand of Country 1	100%	100%
2 (X_1 off)	X_6 = Peak demand of Country 1 X_2 = Capacity of Source N19 X_4 = Capacity of Source N11	80% 7% 1%	85% 16% 9%
3 (X_2 off)	X_6 = Peak demand of Country 1	100%	100%
4 (X_1, X_2 off)	X_4 = Capacity of Source N11 X_6 = Peak demand of Country 1 X_{18} = Capacity of DN400	88% 9% 3%	88% 9% 3%

Source: JRC calculations and analysis.

4.4 Results for Country 2

The estimated PDF's of the mean volume of gas supply for the four different scenarios are depicted in **Figure 8**. We note that when source N19 is off (scenarios 3 & 4), the predicted mean value of gas supply is much less than the demand of the country. This indicates that this source is very important for this country. However, the system is resilient to the failure of Source N2 (recall that it is the highest capacity source). This is due to the location of the country with respect to the sources and potential bottlenecks in certain connections. Indeed, bottlenecks can represent a non-negligible storage of gas.

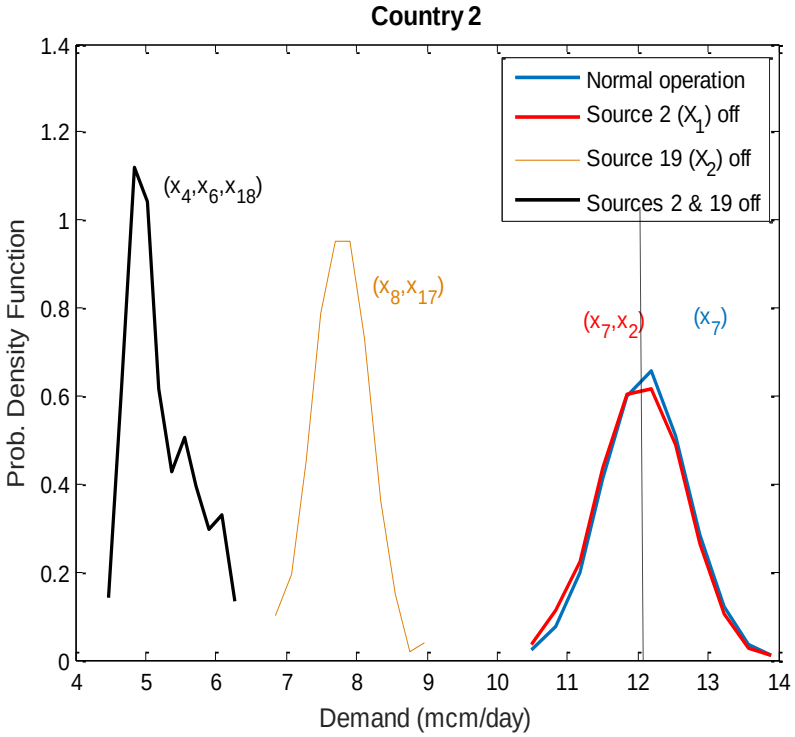
The most important input variable is the peak demand of the country 2 (see **Table 6**) when the system operates normally and when Source N2 is off. Surprisingly, we note that when Source N19 fails, the capacity of the system to provide gas to Country 2 also depends on the peak demand of Country 3 and the peak demand of country 1 (when N2 & N19 off). This result is also a consequence of the priority algorithm implemented in the ProGasNet (See Section 2). Indeed, the latter supplies the closer demand nodes first before serving the other ones. Hence, if a country is far from the main sources of gas, it might be more subject to failure of gas supply.

Table 6. Relevant inputs for the predicted mean gas supply for country 2.

Scenario	Relevant inputs	S_i	ST_i
1 (Normal)	X_7 = Peak demand of Country 2	100%	100%
2 (X_1 off)	X_7 = Peak demand of Country 2 X_2 = Capacity of Source N19	90% 5%	90% 5%
3 (X_2 off)	X_8 = Peak demand of Country 3 X_{17} = Capacity of DN500	60% 40%	60% 40%
4 (X_1, X_2 off)	X_4 = Capacity of Source N11 X_6 = Peak demand of Country 1 X_{18} = Capacity of DN400	70% 20% 10%	70% 20% 10%

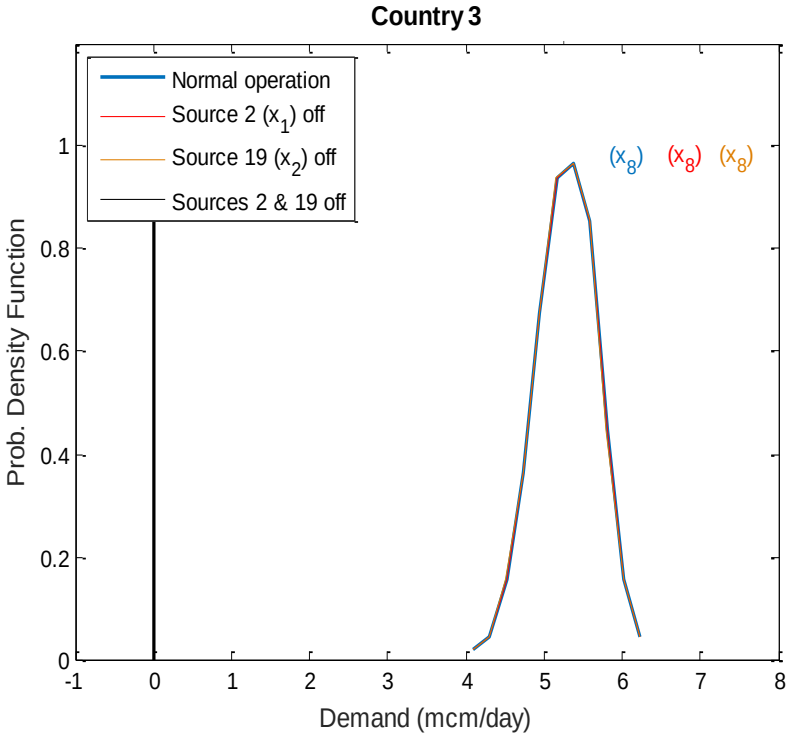
Source: JRC calculations and analysis.

Figure 8. Predicted uncertainty of mean gas supply for country 2 with respect to the different scenarios.



Source: JRC calculations and analysis.

Figure 9. Predicted uncertainty of mean gas supply for country 3 w.r.t. the different scenarios. Note that the results of the first three scenarios overlap.



Source: JRC calculations and analysis.

4.5 Results for Country 3

For Country 3, the system behaviour is simpler. The system is resilient to the failure of one of the main sources but it is incapable to provide gas to Country 3 when both main sources collapse (see **Figure 9**). This is an important result, revealing the country vulnerability to crisis. Crisis simulated by scenario 4 is unlikely to happen. The prediction of volume of gas supply heavily depends on the knowledge of the value of the gas demand in the country.

4.6 Further analysis with MCF

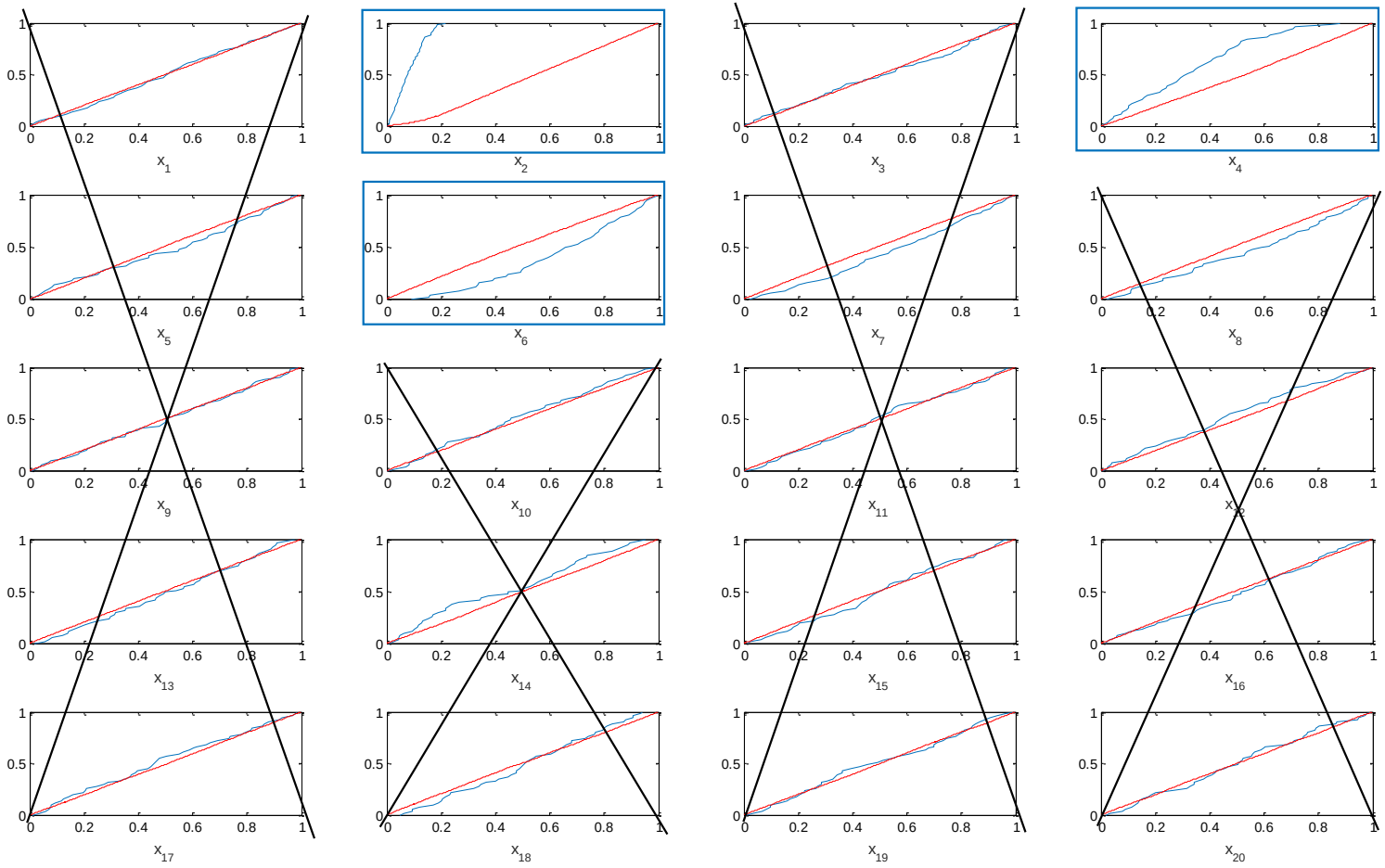
If the analyses of the model responses that take discrete values did not really provide further information, they confirmed those already obtained with the PCE in the previous sections though. The same important inputs have been identified for each country, for instance, (X_2,X_4,X_6) for Country 1. For the sake of completeness, we present here the results obtained for Scenario 2 when considering the probability that the gas supply in Country 1 is less than the demand (i.e. $P(S < D)$, see **Figure 6** first row and last column plot).

The Monte Carlo Filtering was performed by splitting the input sample in two subsamples: the first subsample containing the input sets that provided $P(S < D) \leq 0.5$ and the second one the input sets such that $P(S < D) > 0.5$. The empirical cumulative distribution functions of the different input variables are depicted in **Figure 10**. The red curves in **Figure 10** are the CDFs of the original sample (uniform distribution), the blue curves are the CDFs of the filtered input sample for which $P(S < D) \leq 0.5$. Only the plots with a frame have passed the Kolmogorov-Smirnov test, the remainders have been rejected. The Kolmogorov-Smirnov test concluded that only three variables out of twenty were significantly affected by the filtering. This indicates that the model response is sensitive to these three inputs, namely, (X_2,X_4,X_6). As aforementioned, these three inputs were already identified as being important for the average gas supply.

4.7 Further analysis with other methods

It is worth mentioning that further analyses have been carried with other sensitivity measures and techniques which include: (i) the Pearson correlation coefficient, (ii) Spearman rank correlation coefficient, (iii) contribution to the sample mean and (iv) contribution to the sample variance (Tarantola, Kopustinskas, Bolado-Lavin, Kaliatka, Uspuras, & Vaisnoras, 2012). Pearson correlation coefficient measures the linear relationship between an input and the output. Spearman rank correlation coefficient captures possible nonlinear relationship between the two variables. The last two methods are qualitative graphical methods (Bolado-Lavin, Castaings, & Tarantola, 2009). All these methods are recommended for continuous model responses and as a consequence have been applied only to the mean of gas supply to each country. The results are not shown but they confirm those obtained in Section 4.3, 4.4 and 4.5, albeit some differences in the ranking of the inputs by order of importance.

Figure 10. Monte Carlo filtering for scenario 2 and the output $P(S < D)$ for country 1.



Source: JRC calculations and analysis.

5 Conclusions

The report presents an uncertainty and sensitivity exercise of the security of gas supply model implemented in the probabilistic gas network simulator ProGasNet. The selected network was an EU gas transmission network of several member states. It is based on realistic network topology and data, but due to sensitivity of information, the network is anonymised.

The study aims to identify and rank the model parameters that most significantly affect the security of supply. The model was run for four different scenarios representing different disruption situations. It is found that the peak demand in each country is the key parameter whose precise estimation is very important for accurate model results.

The study confirms, the results already observed in other studies, that some disruption scenarios affect only parts of the network (e.g. specific countries) while other parts of the network are not affected. Even more, due to existing bottlenecks in the system, supply disruptions in one part of the network cannot be restored from the other part even if there is enough gas available. This clearly indicates heterogeneity of the network and the need for further infrastructure development.

Looking at each country, for Country 1, the peak demand is the most important parameter, followed by capacity of sources 19 and 11 and capacity of DN400 pipeline. This is important information for further development of the model and indicates where to target further research efforts.

For Country 2, again the peak demand is the most important parameter, followed by capacity of source 19, DN500 and DN400 pipelines. Interestingly, for some scenarios, peak demands of Countries 1 and 3 are identified as important. This can be well explained by the fact that in certain disruption scenarios, lower demand in neighbouring countries allows better supply to the other countries.

For Country 3, the supply is secured in case of scenarios 2 and 3, but in the case of unlikely scenario 4, the security of supply is threatened.

The study showed the potential and usefulness of sensitivity study applied to the ProGasNet gas network model. It has not only identified the most important parameters of the model for which more attention should be paid during the estimation process, but also provided useful insights into the simulation process by confirming, for instance, the heterogeneity of the network from the sensitivity analysis perspective.

References

Bolado-Lavin, R., Castaings, W., & Tarantola, S. (2009). Contribution to the sample mean for graphical and numerical sensitivity analysis. *Reliability Engineering and System Safety*, 1041-1049.

Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via diffusion. *Annals of Statistics*, 2916-2957.

Buzzard, G. T., & Xiu, D. (2011). Variance-based global sensitivity analysis via sparse-grid interpolation and cubature. *Communications in Computational Physics*, 542--567.

Cukier, R. I., Fortuin, C. M., Petschek, A. G., & Schaibly, J. H. (1973). Sutdy of the sensitivity of coupled reaction systems to uncertainties inrate coefficients. *Hournal of Chemical Physics*, 3873-3878.

Deo, N. (2008). *Graph theory with applications to engineering with computer science.* Prentice Hall.

Hornberger, G. M., & Spear, R. C. (1981). An approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, 7-18.

Kopustinskas, V., & Praks, P. (2015). Bottleneck analysis of the gas transmission network using ProGasNet simulator. *Safety and Reliability of Complex Engineered Systems - Proceedings of the 25th European Safety and Reliability Conference, ESREL 2015* (pp. 1259-1266). Taylor and Francis.

Oakley, J. E., & O'Hagan, A. (2004). Probabilistic sensitivity analysis models: a Bayesian framework. *Journal of Royal Statistical Society, B*, 751-769.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 1065-1076.

Praks, P., Kopustinskas, V., & Masera, M. (2015). *Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure. Reliability Engineering and System Safety Vol. 144*, 254-264.

Praks, P., Kopustinskas, V., & Masera, M. (2017). Monte-Carlo-based reliability and vulnerability assessment of a natural gas transmission system due to random network component failures. *Sustainable and Resilient Infrastructure Vol. 2 , Iss. 3*, 97-107.

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 280-297.

Saltelli, A., Tarantola, S., & Chan, K. (1999). A quantitative model independent method for global sensitivity analysis of model output. *Technometrics*, 39-56.

Shao, Q., Younes, A., Fahs, M., & Mara, T. A. (2017). Bayesian sparse polynomial chaos expansion for global sensitivity analysis. *Computer Methods in Applied Mechanics and Engineering*, 474-496.

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling & Computer Experiments*, 407-414.

Sobol', I. M., Turchaninov, V. I., Levitan, L. Y., & Shukman, V. B. (1992). *Quasi-random Sequence Generator (Routine LPTAU51).* Moscow: Keldysh Institue of Applied Mathematics, Russian Academy of Science.

Sudret, B. (2008). Sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & Reliability Safety*, 964-979.

Tarantola, S., Kopustinskas, V., Bolado-Lavin, R., Kaliatka, A., Uspuras, E., & Vaisnoras, M. (2012). Sensitivity analysis using contribution to sample variance plot: Application to a water hammer model. *Reliability Engineering and System Safety*, 62-73.

Wiener, N. (1938). The homogeneous chaos. *American J. of Mathematics*, 897-936.

Xiu, D., & Karniadakis, E. G. (2002). The Wiener-Askey polynomial chaos for stochastic differential equation. *SIAM J. Science of Computing*, 619-644.

List of abbreviations and definitions

CDF	Cumulative distribution function
CS	Compressor Station
LNG	Liquefied natural gas
MCF	Monte Carlo filtering
PCE	Polynomial Chaos Expansion
PDF	probability density function
ProGasNet	Probabilistic Gas Network simulator
UGS	Underground storage

List of figures

List of tables

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

More information on the European Union is available on the internet (http://europa.eu).

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

KJ-NA-28953-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/818537

ISBN 978-92-79-77155-2