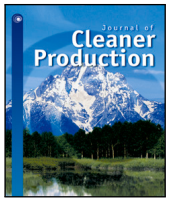
Journal of Cleaner Production 414 (2023) 137612

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro



Sustainable supply chain design under correlated uncertainty in energy and carbon prices

Florencia Lujan Garcia-Castro *, Ruben Ruiz-Femenia *, Raquel Salcedo-Diaz, Jose A. Caballero

Institute of Chemical Process Engineering, University of Alicante, Apartado de Correos 99, Alicante, 03080, Comunidad valenciana, Spain

ARTICLE INFO

Handling Editor: Panos Seferlis

Keywords:
Correlated uncertainty
Energy price
Carbon price
Supply chain
Stochastic programming

ABSTRACT

This paper aims to provide an improvement in the modeling of supply chain designs by incorporating correlated uncertainty among multiple parameters, resulting in a more resilient design. A new methodology to generate forecasts for historically correlated time series, regardless of their underlying probability distributions, is presented and applied to generate scenarios for energy and carbon prices, which historically proved to be correlated. These scenarios are then used in a stochastic computation to obtain a three-echelon supply chain design in Europe maximizing the economic performance. The emissions were monetarized through the incorporation of the European Union cap-and-trade emissions trading system into the model. The social impact of the supply chain network is measured in terms of the direct, indirect and induced jobs it creates, which are proportional to the economic performance. By combining the developed methodology with data mining algorithms, a reduction in the number of required scenarios by more than 90% was achieved. The numerical case study moreover shows that the stochastic design ensures an average reduction of emissions by more than 3 ktons compared to the use of a deterministic approach. In comparison, the computation of a stochastic supply chain design without parameter correlation takes 5 times longer.

1. Introduction

There are several possible sources of uncertainty in a supply chain model. One of them is referred to as structural uncertainty and describes the inability of a model to represent real-world situations, such as disregarding legal regulations. If regulatory uncertainty is not considered when designing a supply chain, companies may for example hesitate to invest in sustainable changes to their production process. Furthermore, the resulting design may lack flexibility in meeting future regulations, leading to unexpected investment costs. Carbon regulations have become an increasingly relevant topic in recent years, and the introduction of an Emissions Trading System (ETS) by governments is a key driver of this trend. Both customers and production companies are encouraged to move towards products with a better carbon footprint, and it has been confirmed that the most efficient approach to reduce emissions is through carbon pricing (World Bank Group, 2019). To link the emission to overall profit, the model incorporates the cap-and-trade system from the European Union. As a result, the total emissions will be indirectly minimized (European-Union, 2016).

Another type of uncertainty is the so-called parameter uncertainty, and it occurs when the values of different parameters, such as product demand on different markets, energy prices or raw material prices, are assumed for the entire time horizon under study and not known beforehand. In this manuscript, uncertainty in energy and carbon prices will be studied. Both have undergone major changes during the last years, and exposed the difficulties of supply chains to handle them. Not accounting for uncertainty in energy and carbon prices may lead to less flexible supply chain designs, and low flexibility may lead to higher emissions and costs than necessary. To incorporate uncertainty into a stochastic numerical model, various methods have been investigated in the literature. One of them is Chance Constraint Optimization (CCO), in which constraints do not have to be satisfied for each scenario, but rather only for a certain level of confidence. CCO can be used to investigate and balance the trade-off between cost and risk and has been employed for example by Mitra et al. (2008), who considered uncertainty in the demand of 34 products of a supply chain network.

Another method is Robust Optimization (RO), in which uncertain parameters are assumed to move in a known range. RO then aims to optimize the objective for the worst-case scenario, while ensuring feasibility for the entire parameter range. Robust optimization was employed by Khorshidvand et al. (2021), who presented a novel modeling approach for closed-loop supply chains with uncertain demand, and by Krishnan et al. (2022) who used it to design a food supply chain under supply uncertainty. In the present manuscript, Stochastic Optimization (SO) is used. Stochastic optimization considers uncertainty in

* Corresponding authors.
E-mail addresses: flgc1@alu.ua.es (F.L. Garcia-Castro), ruben.ruiz@ua.es (R. Ruiz-Femenia).

https://doi.org/10.1016/j.jclepro.2023.137612
Received 22 January 2023; Received in revised form 19 April 2023; Accepted 28 May 2023
Available online 31 May 2023

Table 1
Topics studied in various recent publications.

Article	Energy prices	Social dimension	Carbon regulation	Correlation	Scenario reduction
Baringo and Conejo (2013)				✓	✓
Das and Shaw (2017)		✓	✓		
Jahani et al. (2018)				✓	
Fattahi et al. (2018)					✓
Dehghani et al. (2020)					
Delberis et al. (2021)	✓			✓	
Xu et al. (2022)			✓		✓
Garcia-Castro et al. (2022)			✓		✓
Krishnan et al. (2022)		✓			
This manuscript	✓	✓	✓	✓	✓

the parameters by generating a number of scenarios and corresponding probabilities. It then optimizes the expected outcome of the objective of all scenarios. There are different methods to generate these scenarios, with one of the most prominent ones being the Monte Carlo method, which uses randomness to generate new samples. Elsayed et al. (2022) used the Monte Carlo technique to sample demand and supply scenarios for a blood supply chain. Another sampling method is Latin Hypercube Sampling, which using an a priori discretization of the parameter space is able to achieve a certain level of sampling-accuracy using fewer samples than the Monte Carlo method, at the cost of higher complexity. (Fattahi et al., 2018) presented a multi-stage stochastic problem, in which they employed Latin Hypercube Sampling to generate demand scenarios.

The present manuscript investigates uncertainty in energy and carbon prices, since both of these quantities have undergone substantial changes in the recent past, not least as a result of the EU energy crisis (ecoinvent, 0000). Since the exact probability distribution of neither energy nor carbon prices is known, a straight-forward application of the previously mentioned sampling methods was not expected to produce realistic pricing scenarios. Therefore, an AutoRegressive Integrated Moving Average (ARIMA) model that best fits the historical price data was identified, and could subsequently be used to forecast realistic future pricing scenarios. ARIMA models have been widely used in different contexts, for example by Fanoodi et al. (2019) in combination with neural networks to predict blood platelet demand.

When considering uncertainty in more than one parameter, a less considered topic is the correlation among these parameters. Not taking the correlation into account when generating scenarios treats both uncertain variables as if they were independent, which introduces a new source of structural uncertainty into the model. Lately, several authors have taken correlation of the data into consideration and pointed out its implications in their analysis. Delberis et al. (2021) proposed the use of correlated scenarios for energy, power demand and spot prices to optimize the parameters of a contract for a large electricity consumer in Brazil. They used a Seasonal ARIMA (SARIMA) model and considered risk management. In 2020, Dehghani et al. presented a design of a photovoltaic supply chain in Iran under correlated uncertainty using a robust optimization method and Jahani et al. (2018) studied the redesign of a supply chain network considering correlated scenarios in demand and prices, modeled as two correlated Geometric Brownian Motion (GBM) processes. The GBM is linked to the normal distribution, for which correlated samples can be generated using explicit formulas. Their results indicate that ignoring the effect of correlation leads to profit overestimation.

A main limitation of stochastic simulations is the fast increasing number of variables and equations for a growing number of scenarios. However, the underlying probability space must be adequately discretized, and the number of scenarios cannot be chosen arbitrarily small. A reasonable trade-off between the accuracy of the stochastic method and its complexity needs to be found. Xu et al. (2022) presented an eco-friendly closed-loop supply chain under uncertainty in carbon price and demand. To reduce the number of scenarios they applied a forward and reverse scenario reduction technique. Garcia-Castro et al. (2022) presented a cooperative sustainable supply chain design under uncertainty in carbon dioxide allowance prices and compared the reduction of scenarios using the k-means clustering technique and the Scenario Reduction algorithm (ScenRed), a powerful scenario tree reduction algorithm for multivariate data (Dupacova et al., 2002). In this manuscript the Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) algorithm is used to reduce the number of required scenarios. Classical clustering algorithms such as k-means or DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering methods perform well on datasets having visually separable clusters, while the BIRCH algorithm can handle highly connected, dense datasets.

In the present manuscript, a new methodology for generating correlated time series, that does not require a priori knowledge of the corresponding probability distributions, is presented. This methodology is then applied to generate scenarios with correlated energy and carbon prices. The cap-and-trade system is implemented to create a direct link between economic and environmental performance and to reduce structural uncertainty in the model. The social impact of the network and its risk robustness were also analyzed. Finally, the results were compared with the design obtained form a deterministic approach, and the design from a stochastic optimization that does not consider correlation of parameters. This comparison helps to better understand the impact of correlation in a stochastic optimization.

Table 1 presents a summary of the topics that have been studied in recent publications. In this manuscript, all the mentioned topics will be addressed conjointly.

The next section presents the methodology, which is divided into four subsections: The mathematical formulation of the supply chain model; the proposed method to generate correlated scenarios and reduce the total number of scenarios needed; the social assessment; and risk management applied to the resulting supply chain design. In the last two sections, numerical results and conclusions that can be drawn from them are presented.

2. Methodology and mathematical formulation

The presented mathematical model integrates the calculation of benefit and cost associated with a certain supply chain network. It also includes the determination of the Global Warming Potential (GWP), accounting for emissions associated to energy use, raw material consumption and transportation, as well as the carbon cap-and-trade system, which links the emissions to the total profit of the network. The model includes uncertainty through a number of scenarios for the carbon and energy prices, taking into consideration the correlation of the historical data. For the obtained design, the social corresponding impact and risk are calculated using different indicators. Each of the mentioned topics will be described in detail in the followings subsections.

2.1. Supply chain model

The supply chain network is described by different blocks of equations, including economic assessment, mass balances and capacity constraints, environmental impact assessment as well as the corresponding constraints to the cap-and-trade regulations.

In the following equations, the subscript $s \in \mathrm{S}$ denotes the scenario, $\mathrm{T} = \{0, \ldots, n_T - 1\}$ is the set of n_T timesteps, $t \in \mathrm{T}$ denotes a certain time period and $p \in \mathrm{P}$ stands for the plant of interest. Warehouses and markets are indicated by the subscripts $w \in \mathrm{W}$, resp. $k \in \mathrm{K}$. For chemicals and technologies, the respective subscripts are $j \in \mathrm{J}$ and $i \in \mathrm{I}$.

If in time period $t \in \mathrm{T}$ the capacity of technology $i \in \mathrm{I}$ is extended at plant $p \in \mathrm{P}$, the binary variable $\mathrm{Y}_{\mathrm{P}p,i,t}$ takes value one, otherwise it is zero. Analogously, the binary variable $\mathrm{Y}_{\mathrm{W}w,t}$ takes value one, if the capacity of warehouse $w \in \mathrm{W}$ is extended in timestep $t \in \mathrm{T}$.

2.1.1. Mass balance constraints

The variable $\mathrm{PU}_{p,j,t,s}$ designates the amount of chemical $j \in \mathrm{J}$ that is purchased by plant $p \in \mathrm{P}$. $F_{p,i,j,t,s}$ denotes the flow rate of chemical j due to production or consumption in technology $i \in \mathrm{I}$ at plant $p \in \mathrm{P}$. Lastly, $\nu_{i,j}$ represents the material balance for chemical j in technology i, and $\xi_{\mathrm{P}j,p,w,t,s}$ denotes the amount of chemical j transported from plant p to warehouse $w \in \mathrm{W}$. The mass balances

$$\mathrm{PU}_{p,j,t,s} - \sum_{w \in \mathrm{W}} \xi_{\mathrm{P}j,p,w,t,s} = \sum_{i \in \mathrm{I}} \operatorname{sign} \nu_{i,j} \cdot F_{p,i,j,t,s} \tag{1}$$

must be satisfied for all $p \in \mathrm{P}$, $j \in \mathrm{J}$, $t \in \mathrm{T}$ and $s \in \mathrm{S}$. Each technology $i \in \mathrm{I}$ has certain main products, denoted $\mathrm{JM}_i \subset \mathrm{J}$, that imply further mass balances,

$$\mathrm{F}_{p,i,j,t,s} = |\, \nu_{i,j} \,|\, \mathrm{F}_{p,i,\tilde{j},t,s}, \tag{2}$$

for all $\tilde{j} \in \mathrm{JM}_i$. The variable $\theta_{j,w,t,s}$ represents the inventory of chemical $j \in \mathrm{J}$ stored at warehouse $w \in \mathrm{W}$, which in each timestep is modified due to the amount of incoming product $\xi_{\mathrm{P}j,p,w,t,s}$, and the amount of product $j \in \mathrm{J}$ transported from the warehouse to a market $k \in \mathrm{K}$, $\xi_{\mathrm{W}j,w,k,t,s}$. In Europe, transportation is mostly done via trucks powered by diesel. Therefore, and to simplify the model, only this type of transportation is considered. The following equation describes the warehouse mass balances, assuming that $\theta_{j,w,-1,s} := \underline{\theta}_{j,w}$ is a given initial value.

$$\theta_{j,w,t,s} + \sum_{k \in \mathrm{K}} \xi_{\mathrm{W}j,w,k,t,s} = \theta_{j,w,t-1,s} + \sum_{p \in \mathrm{P}} \xi_{\mathrm{P}j,p,w,t,s}, \tag{3}$$

for all $j \in \mathrm{J}$, $w \in \mathrm{W}$, $t \in \mathrm{T}$ and $s \in \mathrm{S}$. The amount of chemical j that a market k receives from all warehouses $w \in \mathrm{W}$, must equal $\mathrm{SA}_{j,k,t,s}$, being the amount of chemical j sold at this market, i.e.

$$\mathrm{SA}_{j,k,t,s} = \sum_{w \in \mathrm{W}} \xi_{\mathrm{W}j,w,k,t,s}, \quad \forall (j,k,t,s) \in \mathrm{J} \times \mathrm{K} \times \mathrm{T} \times \mathrm{S}\,. \tag{4}$$

2.1.2. Capacity constraints

The total mass flow of chemicals in any technology $i \in \mathrm{I}$ at plant $p \in \mathrm{P}$ cannot exceed the corresponding capacity of this technology at the plant, $\mathrm{Q}_{\mathrm{P}p,i,t}$, i.e.

$$\sum_{j \in \mathrm{JM}_i} \mathrm{F}_{p,i,j,t,s} \le \mathrm{Q}_{\mathrm{P}p,i,t}, \quad \forall (p,i,t,s) \in \mathrm{P} \times \mathrm{I} \times \mathrm{T} \times \mathrm{S}\,. \tag{5}$$

However, these capacities are eventually extended by $\gamma_{\mathrm{P}p,i,t}$, satisfying

$$\underline{\gamma}_{\mathrm{EP}p,i,t} \cdot \mathrm{Y}_{\mathrm{P}p,i,t} \le \gamma_{\mathrm{P}p,i,t} \le \overline{\gamma}_{\mathrm{EP}p,i,t} \cdot \mathrm{Y}_{\mathrm{P}p,i,t}, \quad \forall (p,i,t) \in \mathrm{P} \times \mathrm{I} \times \mathrm{T}, \tag{6}$$

where $\underline{\gamma}_{\mathrm{EP}p,i,t}$ and $\overline{\gamma}_{\mathrm{EP}p,i,t}$ are fixed lower, resp. upper expansion limits.

The capacity in timestep t equals the capacity in the previous timestep plus the capacity expansion carried out, that is

$$\mathrm{Q}_{\mathrm{P}p,i,t} = \mathrm{Q}_{\mathrm{P}p,i,t-1} + \gamma_{\mathrm{P}p,i,t}, \quad \forall (p,i,t) \in \mathrm{P} \times \mathrm{I} \times \mathrm{T}, \tag{7}$$

with a given initial capacity $\mathrm{Q}_{\mathrm{P}p,i,-1} := \underline{\mathrm{Q}}_{\mathrm{P}p,i}$ for all $p \in \mathrm{P}$ and all $i \in \mathrm{I}$. Similarly, the capacity of any warehouse $w \in \mathrm{W}$, $\mathrm{Q}_{\mathrm{W}w,t}$ is expanded by $\gamma_{\mathrm{W}w,t}$, defined within fixed bounds

$$\underline{\gamma}_{\mathrm{EW}w,t} \cdot \mathrm{Y}_{\mathrm{W}w,t} \le \gamma_{\mathrm{W}w,t} \le \overline{\gamma}_{\mathrm{EW}w,t} \cdot \mathrm{Y}_{\mathrm{W}w,t}, \quad \forall (w,t) \in \mathrm{W} \times \mathrm{T}, \tag{8}$$

where $\underline{\gamma}_{\mathrm{EW}w,t}$ and $\overline{\gamma}_{\mathrm{EW}w,t}$ denote the lower, resp. upper expansion limit of the warehouse. Similar to plants, also for warehouses an initial capacity $\mathrm{Q}_{\mathrm{W}w,-1} := \underline{\mathrm{Q}}_{\mathrm{W}w}$ is defined, and the expanded capacity reads

$$\mathrm{Q}_{\mathrm{W}w,t} = \mathrm{Q}_{\mathrm{W}w,t-1} + \gamma_{\mathrm{W}w,t}, \quad \forall (w,t) \in \mathrm{W} \times \mathrm{T}\,. \tag{9}$$

The parameter τ_w describes how many times per timestep the warehouse can be entirely filled and emptied. This imposes further conditions on the amount of chemicals that can be transported to the markets,

$$\frac{1}{\tau_w} \sum_{j \in \mathrm{J}} \sum_{k \in \mathrm{K}} \xi_{\mathrm{W}j,w,k,t,s} \le \mathrm{Q}_{\mathrm{W}w,t}, \quad \forall (w,t,s) \in \mathrm{W} \times \mathrm{T} \times \mathrm{S}\,. \tag{10}$$

2.1.3. Objective function

The next equation defines the expected Net Present Value $\mathbb{E}(\mathrm{NPV})$, which will be maximized during the solution of the model,

$$\mathbb{E}(\mathrm{NPV}) = \sum_s \mathbb{P}(s) \cdot \mathrm{NPV}(s), \tag{11}$$

with $0 < \mathbb{P}(s) < 1$ denoting the probability of scenario $s \in \mathrm{S}$. The net present value of a scenario s is the sum of cashflows in all timesteps,

$$\mathrm{NPV}_s = \mathrm{C}_{\mathrm{flow}\,t=1,s} + \sum_{t \in \mathrm{T}, t>1} \frac{\mathrm{C}_{\mathrm{flow}\,t,s}}{(1+\imath)^{t-1}} - \chi \sum_{\substack{j \in \mathrm{J} \\ k \in \mathrm{K} \\ t \in \mathrm{T}}} \gamma_{j,k,t} \cdot (\overline{\sigma}_{j,k,t} - \mathrm{SA}_{j,k,t,s}) \tag{12}$$

with $\imath$ being the interest rate. The net present value for each scenario includes a penalization term in case of unsatisfied demand, which was set as to a fraction χ of the profit that was not realized. For any $t \in \mathrm{T}$, the difference between the Fraction of Total Depreciable Capital, $\mathrm{FTDC}_{t,s}$ and the net earnings, defines the cashflow,

$$\mathrm{C}_{\mathrm{flow}\,t,s} = \begin{cases} \mathrm{NETE}_{t,s} - \frac{\mathrm{FCI}}{n_T}, & 0 \le t < n_T - 1 \\ \mathrm{NETE}_{t,s} - (1 - |\,\mathrm{T}\,| \cdot \mathrm{SV}) \frac{\mathrm{FCI}}{n_T}, & t = n_T - 1, \end{cases} \tag{13}$$

where $|\,\mathrm{T}\,| = n_T$ denotes the cardinality of the set T, i.e. the number of elements contained in T, and n_T is the last timestep. The parameter SV denotes the Salvage Value fraction of the supply chain network. Its net earnings are calculated for any $t \in \mathrm{T}$, $s \in \mathrm{S}$ as difference between total income and costs, including taxes,

$$\begin{aligned} \mathrm{NETE}_{t,s} \quad =& (1 - \mathrm{TAX}) \sum_{\substack{j \in \mathrm{J} \\ k \in \mathrm{K}}} \gamma_{j,k,t} \cdot \mathrm{SA}_{j,k,t,s} + \mathrm{NET}_{\mathrm{CO2}\,t,s} \\ & - \sum_{\substack{p \in \mathrm{P} \\ j \in \mathrm{J}}} \lambda_{p,j,t} \cdot \mathrm{PU}_{p,j,t,s} - \sum_{\substack{p \in \mathrm{P} \\ i \in \mathrm{I} \\ j \in \mathrm{J}}} \mathrm{F}_{p,i,j,t,s} \cdot \delta_{\mathrm{P}p,i,j,t} \\ & - \sum_{\substack{j \in \mathrm{J} \\ w \in \mathrm{W} \\ k \in \mathrm{K}}} \frac{\xi_{\mathrm{W}j,w,k,t,s}}{\tau_w} \delta_{\mathrm{W}j,w,t} \\ & - \sum_{\substack{j \in \mathrm{J} \\ p \in \mathrm{P} \\ w \in \mathrm{W}}} \xi_{\mathrm{P}j,p,w,t,s} \cdot \mathrm{PE}_{t,s} \cdot \mathrm{d}_{\mathrm{P}p,w} \cdot 0.014 \\ & - \sum_{\substack{j \in \mathrm{J} \\ w \in \mathrm{W} \\ k \in \mathrm{K}}} \xi_{\mathrm{W}j,w,k,t,s} \cdot \mathrm{PE}_{t,s} \cdot \mathrm{d}_{\mathrm{W}w,k} \cdot 0.014 - \mathrm{DEP}_{t,s} \end{aligned} \tag{14}$$

where $\mathrm{DEP}_{t,s}$ is the depreciation term

$$\mathrm{DEP}_{t,s} = \frac{1 - \mathrm{SV}}{|\,\mathrm{T}\,|}\,\mathrm{FCI}, \tag{15}$$

that is calculated using the total Fixed Cost Investment (FCI),

$$\begin{aligned} \mathrm{FCI} \quad &= \sum_{\substack{p\in \mathrm{P}\\ i\in \mathrm{I}\\ t\in \mathrm{T}}} \alpha_{\mathrm{P}p,i,t}\cdot \gamma_{\mathrm{P}p,i,t} + \beta_{\mathrm{P}p,i,t}\cdot \mathrm{Y}_{\mathrm{P}p,i,t} \\ &+ \sum_{\substack{t\in \mathrm{T}\\ w\in \mathrm{W}}} \alpha_{\mathrm{W}w,t}\cdot \gamma_{\mathrm{W}w,t} + \beta_{\mathrm{W}w,t}\cdot \mathrm{Y}_{\mathrm{W}w,t}\,. \end{aligned} \tag{16}$$

Here, α_{P} and β_{P} stand for the variable and fixed costs of running a plant $p \in \mathrm{P}$, whereas α_{W} and β_{W} denote the variable and investment cost for a warehouse $w \in \mathrm{W}$. The cost of transportation is proportional to the amount of product transported, $\xi_{\mathrm{P}j,p,w,t,s}$ resp. $\xi_{\mathrm{W}j,w,k,t,s}$, the price of energy, PE, and the distance, $\mathrm{d}_{\mathrm{P}p,w}$, resp. $\mathrm{d}_{\mathrm{W}w,k}$. A truck approximately consumes 0.014 liters of diesel to transport one ton of goods over a distance of one kilometer (Ministerio de Formento, 2019).

2.1.4. Environmental impact assessment

There are different indicators to quantify the environmental impact of a supply chain. In this work, the global warming potential will be addressed. Emissions due to the consumption of raw materials are taken into consideration by multiplying the amount of raw material j purchased, $\mathrm{PU}_{p,j,t,s}$, by the Life Cycle Impact Assessment (LCIA) indicator $\mathbb{I}_{\mathrm{RM}j}$, which denotes the amount of CO_2-equivalent emissions per ton of chemical j being consumed. Moreover, emissions associated with energy consumption are calculated using the LCIA indicator $\mathbb{I}_{\mathrm{EN}}$, related to the amount of CO_2 emissions per ton of fuel oil consumed. This factor will be multiplied by the parameter EN_i, that describes the amount of energy required to produce one ton of material using technology i, and the amount of chemical being produced. Lastly, the indicator $\mathbb{I}_{\mathrm{TR}}$ is used to quantify CO_2 emissions associated with transportation of one ton of chemicals over a distance of one kilometer. Transportation between plants and warehouses as well as between warehouses and markets is also accounted for, and the corresponding distances are denoted $\mathrm{d}_{\mathrm{P}p,w}$, resp. $\mathrm{d}_{\mathrm{W}w,k}$. The global warming potential for one timestep $t \in \mathrm{T}$ and scenario $s \in \mathrm{S}$ is calculated as

$$\begin{aligned} \mathrm{GWP}_{t,s} \quad &= \sum_{\substack{p\in \mathrm{P}\\ j\in \mathrm{J}}} \mathrm{PU}_{p,j,t,s}\cdot \mathbb{I}_{\mathrm{RM}j} + \sum_{\substack{p\in \mathrm{P}\\ i\in \mathrm{I}\\ \tilde{j}\in \mathrm{JM}_i}} \mathrm{F}_{p,i,\tilde{j},t,s}\cdot \mathrm{EN}_i\cdot \mathbb{I}_{\mathrm{EN}} \\ &+ \sum_{\substack{p\in \mathrm{P}\\ j\in \mathrm{J}\\ w\in \mathrm{W}}} \xi_{\mathrm{P}j,p,w,t,s}\cdot \mathrm{d}_{\mathrm{P}p,w}\cdot \mathbb{I}_{\mathrm{TR}} \\ &+ \sum_{\substack{j\in \mathrm{J}\\ w\in \mathrm{W}\\ k\in \mathrm{K}}} \xi_{\mathrm{W}j,w,k,t,s}\cdot \mathrm{d}_{\mathrm{W}w,k}\cdot \mathbb{I}_{\mathrm{TR}} \end{aligned} \tag{17}$$

leading to the total global warming potential

$$\mathrm{GWP}_s = \sum_{t\in \mathrm{T}} \mathrm{GWP}_{t,s}, \quad \forall s \in \mathrm{S}\,. \tag{18}$$

2.1.5. Carbon regulation

Governments are implementing different regulations to achieve a reduction of the total carbon dioxide emissions over time. In Europe, supply chains are subject to the European emissions trading cap-and-trade system. This system reduces the maximum feasible amount of emissions $\mathrm{MAX}_{\mathrm{CO2}t,s}$ over the years. In the present model, the initial amount of CO_2 allowances is set to 2×10^8 kg. An annual reduction rate of the CO_2 allowances by 2.2% per year is estimated, according to the EU ETS emissions right phase 4 (European-Commission, 2015). The different companies can trade the available emissions. For this reason, the buying and selling price of CO_2 allowances, $\mathrm{PRICE}_{\mathrm{CO2}t,s}$ and $\mathrm{COST}_{\mathrm{CO2}t,s}$, are highly volatile and should be considered as an uncertain parameter. For simplification, the ideal case that $\mathrm{PRICE}_{\mathrm{CO2}} = \mathrm{COST}_{\mathrm{CO2}}$ at all timesteps and for all pricing scenarios is assumed.

According to the price and amount of carbon allowances being traded in any scenario s at every timestep t, the net income related to the trading of emissions, $\mathrm{NET}_{\mathrm{CO2}}$, can be defined as

$$\mathrm{NET}_{\mathrm{CO2}t,s} = \mathrm{PRICE}_{\mathrm{CO2}t,s}\cdot \mathrm{sales}_{\mathrm{CO2}t,s} - \mathrm{COST}_{\mathrm{CO2}t,s}\cdot \mathrm{buy}_{\mathrm{CO2}t,s}. \tag{19}$$

The net amount of CO_2-allowances being traded influences the upper bound for the global warming potential,

$$\mathrm{sales}_{\mathrm{CO2}t,s} + \mathrm{GWP}_{t,s} \le \mathrm{MAX}_{\mathrm{CO2}}t, s + \mathrm{buy}_{\mathrm{CO2}t,s}, \quad \forall t \in \mathrm{T}, \forall s \in \mathrm{S}\,. \tag{20}$$

The cap-and-trade system creates a link in the model between the environmental and economic performance, since $\mathrm{NET}_{\mathrm{CO2}t,s}$ is included in the calculation of net earnings of the network.

2.2. Scenario generation and reduction

To consider uncertainty in the model, correlated scenarios for CO_2 allowance prices and energy prices were introduced. Firstly, using the historical data for the carbon prices, the best fitting parameters for an ARIMA model using the Akaike/Bayesian Information Criteria (AIC/BIC) were identified. Using this ARIMA model, 1000 possible future carbon price scenarios were generated (Shumway and Stoffer, 2000; Brockwell and Davis, 2016). Afterwards, the methodology presented below in Section 2.2.2 was used to generate energy pricing scenarios that are correlated to these carbon pricing scenarios in the same extent as the historical data for both prices. Details on ARIMA models and its parameters will be described in Section 2.2.1.

Solving the stochastic model with a very large number of scenarios can easily exceed the capacity of state-of-the-art computers. In order to overcome this limit, the deterministic model was solved for all 1000 correlated scenarios, before applying the BIRCH algorithm to cluster the scenarios according to their economic (NPV) and environmental (GWP) performance. The BIRCH algorithm generates clusters of different sizes and forms, and it requires the desired number of clusters as an input parameter. This number of clusters can never be generic and should be decided upon using some measure of suitability of the resulting clustering to the data. In order to find a suitable number of scenarios, the S-dbw index was used, which measures the goodness of approximation of the clustering.

2.2.1. AutoRegressive Integrated Moving Average (ARIMA) models

Future values of historical time series may be forecasted using numerical models. One class of such models are ARIMA models, which are defined by three input parameters $p, d, q \in \mathbb{N}$. They can be fitted against the historical data, meaning that several internal parameters of the model are adjusted to the data. A scoring in form of the AIC/BIC is computed, measuring the probability that the model at hand can reproduce the historical data. By comparing the scoring for different p, d, q, a best-fitting model can be identified.

The parameter d describes the number of times, that the time series at hand is integrated, meaning that instead of the data points themselves, dth order differences of the data are considered. When chosen appropriately, this preprocessing helps to remove any seasonal trend from the data. The parameter p denotes the number of lagged observations used in an autoregression of the data points. Instead of the integrated data points themselves, one computes weighted averages of several previous observations. Finally, approximation quality of the model can be improved by using an appropriate number of moving average terms, described by parameter q. For further details on ARIMA models, the reader is referred to Shumway and Stoffer (2000).

2.2.2. Generation of correlated scenarios

In this subsection, the process of generating scenarios for two time series which are correlated to a certain extent will be described. Not only the historical correlation between the CO_2 prices, $\mathrm{PRICE}_{\mathrm{CO2}}$, and the prices of energy, PE, was studied, but also their individual correlations with the historical interest rates ι. The latter were lower than 0.1 in both cases, and therefore the focus will be on the correlation of $\mathrm{PRICE}_{\mathrm{CO2}}$ and PE. The corresponding historical values are shown in Fig. 1 (Statista Research Department, 2022; ADAC, 2022).

A linear interpolation was applied to the historical CO_2 prices and the historical energy prices to obtain daily values. The mean, variance,

Table 2
Left to right: Dataset name, mean value, variance, standard deviation and correlation factor for/between the historical time series for CO_2 and energy prices. CO_2 prices are given in €/ton, energy prices are given in €/MW.

X	$\overline{X}$	Var(X)	$\sigma(X)$	Cov(PRICE_{CO2}, X)	$\varrho(\text{PRICE}_{CO2}, X)$
PRICE_{CO2}	0.01700	0.0003	0.0170	0.0003	1.0000
PE	1.2300	0.0262	0.1620	0.0013	0.5000

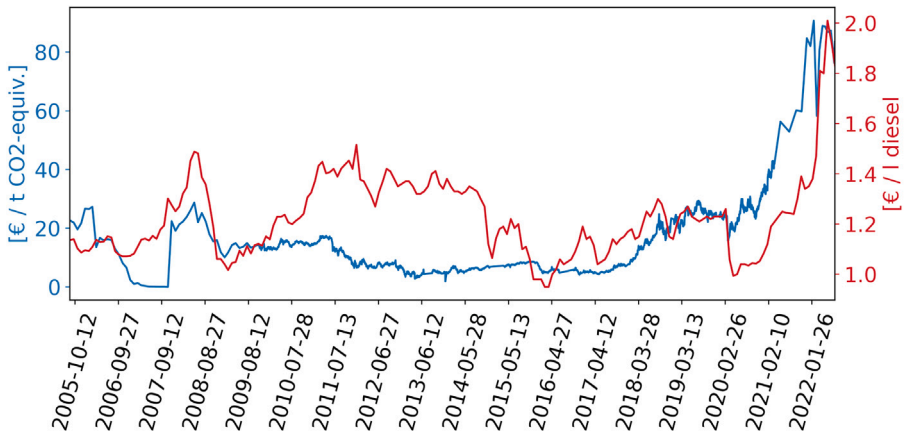


Fig. 1. Historical prices of energy and carbon-dioxide allowances. Left axis (blue): price of allowances.

standard deviation, and correlation factors between these time series were computed. The results are shown in Table 2.

It can be noticed, that PRICE_{CO2} and PE are correlated positively, and that the correlation is quite strong. Hence, scenarios will be generated in the following way: First, a suitable ARIMA model will be used to generate n_S scenarios for future CO_2 prices. Afterwards, the method described below to generate future energy prices in such a way that the correlation factor remains invariant will be applied. For this, let the historical values of two time series $x^n = \{x_i\}_{i=1}^n$ and $y^n = \{y_i\}_{i=1}^n$ be given and compute their corresponding correlation factor,

$$\varrho(x^n, y^n) \overset{Def.}{=} \frac{\text{Cov}(x^n, y^n)}{\sigma(x^n)\sigma(y^n)}. \tag{21}$$

Suppose now that x^n has been extended by m values. In this manuscript, x^n denotes the historical CO_2 prices and $x^{n+r} := \{x_1, \dots, x_n, x_{n+1}, \dots, x_{n+r}\}$ for all $r = 1, \dots, m$ denotes their extension by a forecast of an ARIMA model. Consequently, $y^n = \{y_1, \dots, y_n\}$ will be the historical data of the energy prices, and new values $y_{n+1}, \dots, y_{n+m}$ will be constructed iteratively.

Using x^{n+1}, a feasible value for y_{n+1} that leaves the correlation invariant can be found by solving the equation

$$\varrho(x^n, y^n) = \varrho(x^{n+1}, y^{n+1}). \tag{22}$$

In order to find a value y_{n+1} such that $y^{n+1} := \{y_1, \dots, y_n, y_{n+1}\}$ satisfies Eq. (21), one needs to solve

$$\varrho(x^n, y^n) = \frac{\text{Cov}(x^{n+1}, y^{n+1})}{\sigma(x^{n+1})\sigma(y^{n+1})} \tag{23}$$

that is

$$\varrho(x^n, y^n)\sigma(x^{n+1})\sigma(y^{n+1}) = \text{Cov}(x^{n+1}, y^{n+1}). \tag{24}$$

For simplicity of the subsequent calculations, the equation is rewritten in squared form, reading

$$\varrho(x^n, y^n)^2 \text{Var}(x^{n+1})\text{Var}(y^{n+1}) = \text{Cov}(x^{n+1}, y^{n+1})^2. \tag{25}$$

Next, note that

$$\text{Cov}(x^{n+1}, y^{n+1}) \overset{\text{Def.}}{=} \frac{1}{n+1}\sum_{i=1}^{n+1}(x_i - \overline{x^{n+1}})(y_i - \overline{y^{n+1}}) \tag{26}$$

as well as

$$\begin{aligned}\overline{y^{n+1}} &= \frac{1}{n+1}\sum_{i=1}^{n+1} y_i = \frac{n}{n+1}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) + \frac{1}{n+1}y_{n+1} \\ &= \frac{n}{n+1}\overline{y^n} + \frac{1}{n+1}y_{n+1}\end{aligned} \tag{27}$$

By inserting Eq. (27) into Eq. (26) it can be seen that

$$\text{Cov}(x^{n+1}, y^{n+1}) = A + y_{n+1}B \tag{28}$$

with

$$\begin{aligned}A &= \frac{1}{n+1}\sum_{i=1}^{n} x_i y_i - \frac{n}{n+1}\overline{y^n}\,\overline{x^{n+1}} \\ B &= \frac{1}{n+1}x_{n+1} - \frac{1}{n+1}\overline{x^{n+1}}.\end{aligned} \tag{29}$$

Analogously, inserting (27) into the definition of the variance,

$$\text{Var}(y^{n+1}) = \frac{1}{n+1}\sum_{i=1}^{n+1}(y_i - \overline{y^{n+1}})^2, \tag{30}$$

yields

$$\text{Var}(y^{n+1}) = Cy_{n+1}^2 + Dy_{n+1} + E \tag{31}$$

with coefficients

$$\begin{aligned}C &= \frac{n^2+n}{(n+1)^3} \\ D &= \frac{-2n}{(n+1)^2}\overline{y^n} \\ E &= \frac{1}{n+1}\sum_{i=1}^{n} y_i^2 - \frac{n^3+n^2}{(n+1)^3}\left(\overline{y^n}\right)^2.\end{aligned} \tag{32}$$

Finally, (28) and (31) can be combined with (25). This yields an equation that can be written in terms of the unknown value y_{n+1} and the coefficients A, B, C, D, E, reading

$$(A + By_{n+1})^2 - \varrho(x^n, y^n)^2\text{Var}(x^{n+1})(Cy_{n+1}^2 + Dy_{n+1} + E) = 0. \tag{33}$$

After introducing the new coefficient

$$F = \varrho(x^n, y^n)^2\text{Var}(x^{n+1}) \tag{34}$$

a quadratic equation in y_{n+1} is obtained, reading

$$y_{n+1}^2 + \frac{2AB - DF}{B^2 - CF}y_{n+1} + \frac{A^2 - EF}{B^2 - CF} = 0. \tag{35}$$

This equation can have 0, 1 or 2 solutions depending on the value of the discriminant in the formula

$$y_{n+1,1/2} = -\frac{2AB - DF}{2(B^2 - CF)} \pm \sqrt{\left(\frac{2AB - DF}{2(B^2 - CF)}\right)^2 - \frac{A^2 - EF}{B^2 - CF}}. \tag{36}$$

If (36) is not solvable, then there does not exist a feasible value y_{n+1} to extend the time series. If there are two solutions, the value that best matches the historical variance of the dataset y is taken. Proving that always at least one solution exists means showing that

$$\left(\frac{2AB - DF}{2(B^2 - CF)}\right)^2 - \frac{A^2 - EF}{B^2 - CF} \geq 0 \tag{37}$$

holds, and this task turns out to be highly nontrivial for generic data. In our numerical examples, it was always possible to solve (36). The calculations described are then repeated iteratively to obtain the values at subsequent timesteps $y_{n+2}, \ldots, y_{n+m}$. The result is an energy pricing scenario that is correlated to the forecasted CO_2 pricing scenario with the historical correlation factor.

One benefit of using correlated scenarios is that the network will only be optimized for realistic cases. As energy and CO_2 prices are assumed to be interrelated, generating correlated scenarios of these two parameters enables the model to exclude unlikely scenario realizations that would result in a more conservative supply chain design (i.e., with worse performance). Our approach, described above, works for any given input data, even if the underlying distributions differ or cannot be specified analytically, compared to other methods used to generate multivariate (log-)normally distributed variables. Another advantage is, that correlation can reduce the number of scenarios needed to cover the underlying probability space, which in turn drastically reduces the numerical complexity of the overall stochastic problem.

The proposed method has two potential limitations. Firstly, note that solving a stochastic optimization problem is a priori much harder than solving a deterministic optimization problem, and an application of the method to more complex supply chain models may result in an infeasible increase in computational resources required to solve them. This is due to the number of variables and equations increasing sharply with the number of scenarios. A second potential limitation is the number of scenarios needed to characterize the joint distribution of two or more correlated parameters. Note that, in general, the number of scenarios needed will be large for non-correlated parameters, and smaller if they are correlated. However, if the corresponding deterministic problem, which only includes one scenario, is still solvable, both limitations can be overcome by using decomposition techniques such as the Augmented Lagrangian Relaxation, which splits the problem into one-scenario subproblems. These subproblems are solved individually, i.e. deterministically, and their solutions are then linked in an iterative algorithm until all subproblems converge to the same network design.

2.2.3. BIRCH algorithm and S-dbw index

The Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) algorithm is a powerful data mining algorithm that is able to perform hierarchical clustering of large datasets. Instead of storing detailed information about the elements in each cluster, a cluster is only represented by a triple (N, LS, SS) with N being the number of points in the cluster, and LS resp. SS being the linear, resp. squared, sum of the cluster points. This is called a Clustering Feature (CF) representation of the cluster. The algorithm constructs a CF tree (i.e. a clustering) by first subsequently inserting the points into an existing leaf, or, if this would violate the given threshold, into a new leaf. Once all leafs are constructed, the algorithm tries to merge leafs without violating the threshold until the desired number of clusters is attained (Zhang et al., 1996).

The S-dbw validity index measures the goodness of a clustering by computing the sum of the mean dispersion of the datapoints inside each cluster and an inter cluster density. The best value achievable is 0. In this case, the clusters will have a high density and can be strictly separated from each other (Desgraupes, 2017; Arbelaitz et al., 2013).

2.3. Social assessment

A sustainable supply chain network has three dimensions, being the environmental impact, the economic viability and the social dimension. For the first two, there are different quantitative indicators that have been widely implemented during the last years. Nevertheless, the social dimension is very complex and usually difficult to quantify. In the literature, there are different qualitative and quantitative indicators but only some of them have been implemented and studied by other authors so far. The indicator that can be found most often in publications is the number of job positions generated by the supply network. The number of jobs created by the supply chain does not only include the ones related directly to production, but also those generated as an indirect or induced consequence of the supply chain design. The number of direct jobs $\mathrm{DJ}_{t,s}$ is calculated as

$$\mathrm{DJ}_{t,s} = R(VA_{t,s} - EBE_{t,s}) + A, \tag{38}$$

where VA_t, s is the value added and $EBE_{t,s}$ denotes the earnings before interest, tax, depreciation and amortization for each timestep t and scenarios s. The R value of $1.431 \cdot 10^{-}5$ and a regression coefficient A with value 2.539 for the chemical sector are considered. The reader is referred to the work of Chazara et al. (2017) for more details on these indicators.

The number of indirect jobs is estimated by considering an employment multiplier for the chemical sector of $W = 2.89$ indirect jobs generated for each direct job (Economic Policy Institute, 2019)

$$\mathrm{IJ}_{t,s} = W\,\mathrm{DJ}_{t,s}\,. \tag{39}$$

The number of induced jobs is estimated by taking into consideration information on the region, such as the total number of jobs in the chemical sector JIA_p, the Gross Domestic Product due to household consumption GDP_p, average household size SF_p and size of the population Pop_p. These values for the different regions of the European supply chain network under study can be found in Table 3, and the equation describing the creation of induced jobs for each t, s reads

$$IdJ_{t,s} = \sum_{p \in \mathrm{P}} \frac{\mathrm{JIA}_p\,\mathrm{GDP}_p(\mathrm{DJ}_{t,s} + \mathrm{IJ}_{t,s})SF_p}{\mathrm{Pop}_p}. \tag{40}$$

2.4. Risk management

The optimization of supply chain networks under uncertainty provides valuable information for the decision makers, since the expected performance in multiple scenarios will outperform the results of a deterministic approach. The class of indicators that are most important for decision makers, and can mostly be found in literature, are risk management indicators. These indicators provide information regarding the economic robustness of the network. Four different indicators were studied: First, the expected Downside Risk (DR) for a target value of $\Omega = 190$ million euros which indicates the expected deviation from the target value for all scenarios s having a lower profit than the target value,

$$\mathrm{DR}(s, \Omega) := \mathbb{E}[\delta_\Omega(s)], \tag{41}$$

where for all $s \in \mathrm{S}$

$$\delta_\Omega(s, \Omega) := \begin{cases} \Omega - \mathrm{Profit}(s), & \text{if Profit} < \Omega \\ 0, & \text{if Profit}(s) \geq \Omega. \end{cases} \tag{42}$$

Second, the Value at Risk (VaR), that indicates the minimal profit that is obtainable with a certain probability $1 - \alpha$, i.e.

$$\mathrm{VaR}(s, \alpha) := \inf_{\ell \in \mathbb{R}} \{\ell : \mathbb{P}(\mathrm{Profit}(s) \leq \ell) \leq \alpha\}, \quad \forall s \in \mathrm{S}\,. \tag{43}$$

Third, the Conditional Value at Risk (CVaR), which indicates the expected profit for all scenarios s with profit lower than VaR,

$$\mathrm{CVaR}(s, \alpha) := \mathbb{E}[\mathrm{Profit}(s) | \,\mathrm{Profit}(s) \leq \mathrm{VaR}(s, \alpha)]. \tag{44}$$

Lastly, the worst case scenario provides a lower bound to the overall economic performance.

Table 3
Average household size SF_p, size of population Pop_p, gross domestic product due to household consumption GDP_p and number of jobs in the chemical sector JIA_p for each location in the supply chain network (United Nations and Social Affairs, 2019; Eurostat, 2008; theGlobalEconomy.com, 2021).

Location	Avgerage household size	Size of population	GDP due to household consumption [billion Euros]	number of jobs in the chemical sector
Germany	2.02	83783942	2141.9	362000
Italy	2.58	60461826	1214.6	119000
Spain	2.83	46754778	797.7	94000
Portugal	2.79	10196709	160.5	15072
Czech Republic	2.5	10708981	128.6	56600
Poland	2.84	37864611	382.5	91000

3. Results and discussion

The model was implemented in `Python 3.8` using `pyomo` and solved using IBMs `CPLEX` solver algorithm. The experiments were conducted on a Macbook Pro model 2020 running on OSX 11.6 with 16 GB RAM and a 2.6 GHz 6-Core Intel Core i7 processor. In the model of a sustainable supply chain, various trade-offs exist. One of them is between the economic and environmental performance, which has been previously studied, e.g. by Ruiz-Femenia et al. (2013). They solved a multi-objective model that maximizes the profit while minimizing the global warming potential using the ϵ-constraint method. As expected, the emissions are much higher when only focusing on the economic performance of the network. In this work, the environmental objective is indirectly optimized by quantifying its monetary impact on the supply chain and incorporating it into the objective function of the model. The resulting optimization problem can be solved either deterministically, i.e. for one specific pricing scenario, or stochastically, incorporating uncertainty. While stochastic models are usually harder to solve, the obtained designs are more resilient towards possible future changes of the uncertain parameters. In the model under study, uncertainty is incorporated through the use of scenarios for CO_2 and energy prices. To forecast the carbon price scenarios, an ARIMA model is used. According to the AIC/BIC criteria, the best fitting model for the historical data from August 2005 to August 2022 is given for parameters $p = 2, d = 1, q = 3$, and this model was used to generate 1000 carbon price scenarios. Afterwards, the proposed methodology to generate the correlated energy pricing scenarios is applied. Next, the deterministic Mixed Integer Linear Program (MILP) problem for each correlated scenario is solved, and the BIRCH algorithm is used to cluster the resulting economic (NPV) and environmental (GWP) performance. The BIRCH algorithm needs the desired number of clusters as an input parameter. In Fig. 2, the left plot shows the value of the S-dbw index for different numbers of scenarios. The slope of the resulting curve decreases sharply for a number of clusters lower than 30, then the slope flattens and for a number of clusters larger than 100 the index merely improves. In practice, the computation was parallelized with all 6 available processors solving one problem at a time. To allow for an even balancing among the processors, the number of scenarios should preferably be divisible by 6. Resulting from the foregoing discussion, it was decided to consider $102 = 17 \cdot 6$ scenarios in our computation.

The right plot in Fig. 2 shows a plot of the NPV and GWP of the original 1000 scenarios, as well as the centroids of the 102 clusters marked by an '×'. Table 4 shows the number of variables and constraints, as well as the overall solution time for the model with 102 scenarios. Fig. 3 exemplarily shows the carbon prices for the 102 reduced scenarios that are used in the computation.

The superstructure under study in this manuscript is a three-echelon petrochemical supply chain. The first echelon consists of production plants that are located all over Europe (Karzincbarcika, Leuna, Mantonva and Wloclaweck), which are able to process up to 18 chemicals using different technologies to produce 6 possible final products. The final products are acethaldehyde, acrylonitrile, isopropanol, cumene, phenol and acetone. Each plant has an initial capacity of 20 kt/year.

Table 4
Statistics for the solution of the model under study.

Number of variables	1108750
Number of inequality constraints	360620
Number of equality constraints	518470
Runtime (wall-clock)	27195 s

To each plant there is a corresponding warehouse, and all plants can deliver their final products to any of these warehouses. Each warehouse has an initial inventory of zero tons and an expansion limit between 5 and 400 kt per year. The last echelon are the markets. The superstructure counts with four European markets (Leuna, Neratovice, Sines and Tarragona) where the products can be sold. The model considers a time horizon of ten years.

The network obtained from the stochastic computation shows, that all plant and warehouse expansions are made during the first time period. The plant undergoing the largest expansion is the one in Karzincbarcika, followed by the one in Wloclaweck. The main reason for this capacity distribution is the operation cost, which is lowest for these two plants. More than 30% of the total capacity expansions of both plants is dedicated to expanding the use of the technology 'One step oxidation of ethylene' to produce acetaldehyde. Another 30% is dedicated to the technology 'Reaction of benzene and propylene' to produce cumene, which is a raw material required in the production of phenol and acetone. The warehouse experiencing the biggest expansion is the one in Mantova, since even tough the products are produced in other plants, this warehouse is located in the geographical center of the whole supply chain. The markets of Sines and Tarragona will mainly be supplied by the warehouse in Mantova, which is the closest and therefore ensures minimal CO_2 emissions. For acethaldehyde, the stochastic model decides to send more product from the plant in Karzincbarcika, which has the highest production capacity, to the warehouse in Mantova than to its own warehouse being located in Karzincbarcika. The same applies for the markets in Leuna and Neratorive, which will be provided by the warehouses in Leuna and Wloclaweck, and Karzincbarcika and Wloclaweck, respectively. Nevertheless, and to deal with possible changes in carbon prices and energy prices, there are also transportation links between the warehouse of Mantova and all four markets. This decision provides the network with more flexibility and makes it more robust. Fig. 4 represents the supply chain network for acethaldehyde, which is the product with the highest demand, exemplarily for scenario 0 at timestep 5.

In order to determine the impact of the stochastic simulation on the resulting design, the network obtained from the stochastic model was fixed and exposed to all original 1000 correlated pricing scenarios. The same procedure was repeated for a supply chain design obtained from a deterministic approach. Table 7 presents the economic and environmental performance of the resulting supply chain network for both stochastic and deterministic approach.

The results indicate that, on average, the stochastic model performs better than the deterministic model, with a difference of 73670€and a reduction of more than 3 ktons of CO_2-equivalent in emissions. While these improvements may not be significant in relative terms,

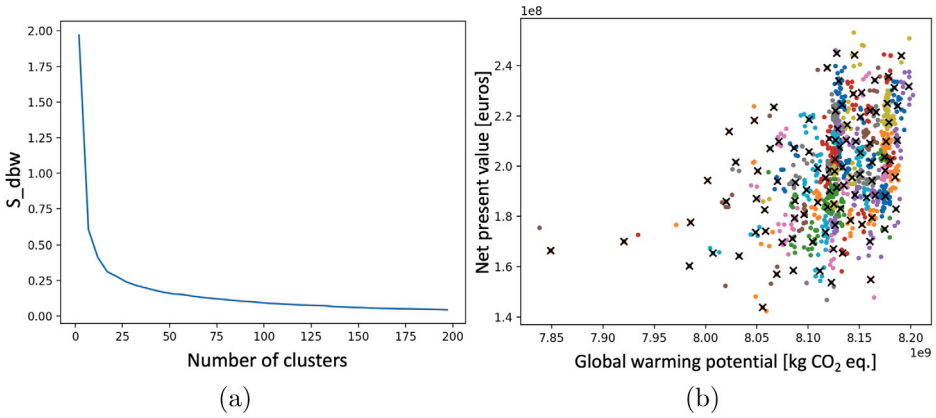


Fig. 2. (a) S-dbw index of the BIRCH clustering for different numbers of scenarios. (b) BIRCH clustering of the original 1000 scenarios into 102 clusters. The scenarios forming the centroids are marked with an '×'.

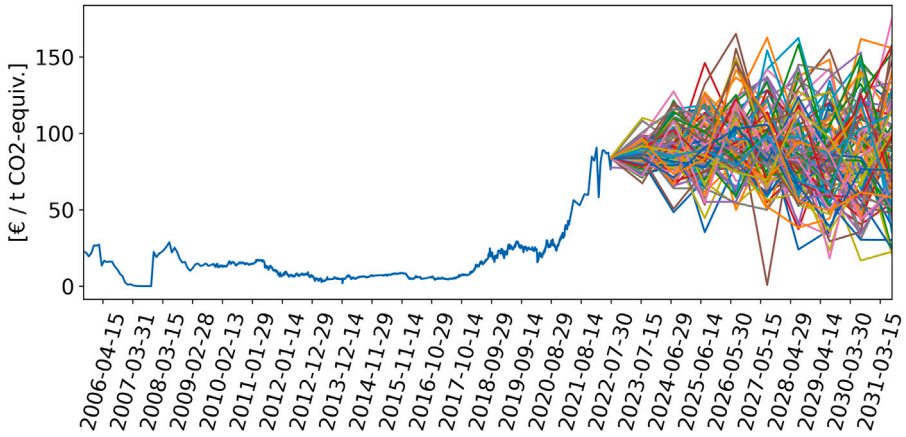


Fig. 3. Historical data, as well as the 102 reduced forecasted scenarios for the carbon price.

Table 5
Expected net present value, global warming potential and number of constraints for the stochastic and deterministic model network.

Indicator	Stochastic model	Deterministic model	Units
Expected net present value	199565301	199491631	Euros
Expected global warming potential	8129607929	8133004126	kg CO_2-eq.
Number of constraints	879090	9174	

Table 6
Expected global warming potential, as well as sum of carbon and energy costs for all timesteps, for models obtained minimizing carbon resp. energy costs.

Objective	Minimization of carbon cost	Minimization of energy cost	Units
Global warming potential	3641353555	11372014890	kg CO_2-eq.
Sum of carbon costs for all timesteps	309557678	492806927	Euros
Sum of energy costs for all timesteps	275121422	175986840	Euros

any reduction in carbon footprint is crucial because companies are investing substantial amounts of money to discover ways to mitigate their carbon emissions. Therefore, any contribution to sustainability goals outweighs the additional computational effort required to solve the stochastic optimization model. Moreover, the moderated gains of the stochastic model with respect to the deterministic design obtained in our case study are expected to increase for more complex supply chains or under correlated uncertainty in a larger number of parameters. To show the increased numerical difficulty of the problem, the number of constraints is also stated in the table. The stochastic model required almost 100 times the number of constraints appearing in the deterministic approach (see Table 5).

Another trade-off in the model is between the reduction of carbon cost and energy cost, both very relevant for a sustainable supply chain design. Two deterministic simulations for one specific energy and carbon pricing scenario were carried out. In the first one the carbon cost is minimized, while in the second one the objective was to minimize the energy cost of the supply chain design. Table 6 presents the obtained results. It is clearly visible that the carbon cost and global warming potential will be higher when minimizing energy costs.

Fig. 5 shows the expected cost and CO_2 distribution of the stochastic network. The largest contributor to the overall costs is due to raw material purchases. A reduction of the related costs can be achieved by investigating alternative production technologies and materials, which

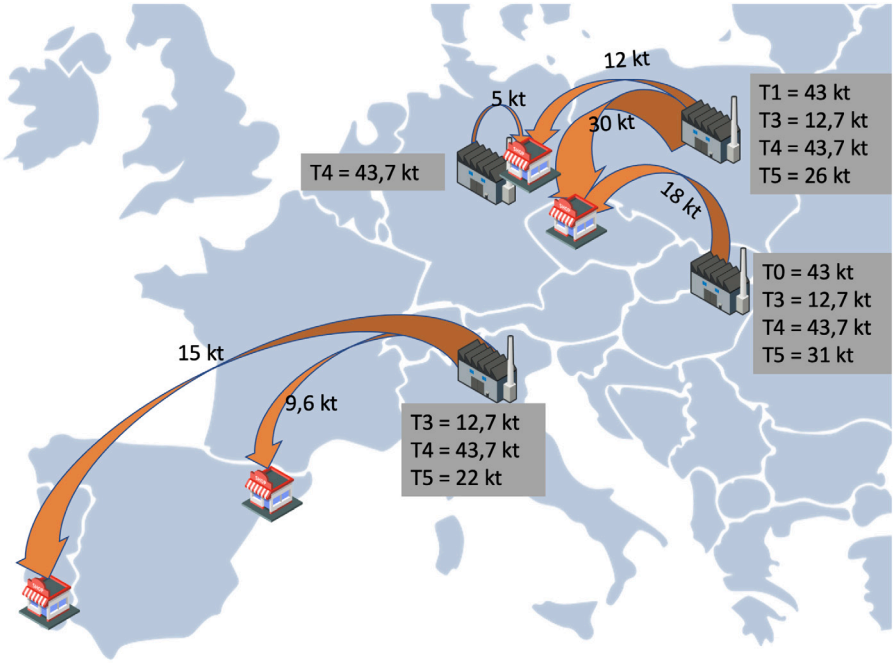


Fig. 4. Supply chain network obtained with the stochastic model for product acethaldehyde for carbon and energy price scenario 0 and timestep 5. The arrows represent the amount of product transported from a certain warehouse to a market. The gray boxes contain information regarding the plant capacity expansion for different technologies. T0 refers to technology 'one step oxidation of ethylene', T3 is 'Hydration of propylene', T4 is 'Reaction of benzene and propylene' and T5 is 'Oxidation of cumene'.

may lead to higher investment cost in the short and long term. The second-largest contributor to cost is the purchase of carbon allowances. The emissions of the overall supply chain may be reduced for example by the usage of more efficient technologies, and green raw materials. In the short term, this may lead to an increase in raw material and production costs, together with higher investment costs. The model also accounts for emissions due to energy consumption and transportation, and the use of renewable energy sources may have a significant positive impact on the overall global warming potential of the supply chain. At the moment, the lack of availability and higher prices of clean energy would imply an increase in energy costs, while reducing emissions as well as the need to buy allowances. Nowadays, sustainability goals in the public and private sector are becoming more important, while changes in prices and regulations are frequent. This shows, that uncertainty and correlation of uncertain quantities need to be considered not only for the 'here-and-now' decisions, but also when monitoring and adjusting a supply chain to match the economic and environmental goals.

Another trade-off occurs between the environmental and social performance, but since the number of created jobs is proportional to the economic performance, it will be indirectly maximized when maximizing the profit. The value added and the number of job positions are proportional, meaning that a larger VA leads to a larger number of jobs. Hence, the number of jobs created is calculated a posteriori from the supply chain design obtained. When exposing the network obtained with the stochastic model to the 1000 possible scenarios, it would generate an average of 726557 jobs (direct, indirect and induced).

To obtain a better understanding of the risk associated with the obtained supply chain network, the cumulative distribution function is represented in Fig. 6. It can be seen that the worst case scenario attains a NPV of 142.3 million euros. A target value of 190 million euros was used, a value that is approx. 5% smaller than the expected net present value. The expected downside risk, that represents the average deviation from this target value in case of a worse economic performance is 12.2 million euros, implying a mean deviation of less than 7%. Regarding the Value at Risk (VaR), the obtained NPV for a probability of 5% is 168.2 million euros, and the associated Conditional Value at Risk (CVaR) is 160.1 million euros.

In order to further investigate the benefits of incorporating correlations of uncertain parameters into the model, an additional stochastic simulation using uncorrelated scenarios was carried out. It was assumed, that the stochastic behavior of both carbon and energy price may be characterized using 100 scenarios each, and in both cases the pricing predictions were generated using the best-fitting ARIMA model to the historical data. By combining these independent scenarios, only a fraction of the resulting 10000 scenarios will show correlation between energy and carbon prices. This way, relevant statistical information from the historical conjoint behavior of both quantities is not represented adequately in the scenarios. Analogously to the correlated simulation described above, the scenarios were then clustered using the BIRCH algorithm into 102 clusters, and a stochastically optimized supply chain design was determined. This network design was then exposed to 1000 possible future pricing scenarios, leading to two conclusions: First, a claim made in the introduction can be corroborated, since the emissions caused by the network based on correlated scenarios are 3.28 ktons of CO_2-equivalent lower than with the network based on independent pricing scenarios. Second, it was observed that the computation of the network based on independent pricing scenarios took more than 5 times longer. This is due to the fact, that the correlation in the pricing scenarios reduces the feasible region of the stochastic problem, which in turn simplifies its solution.

4. Conclusions

In this manuscript, a methodology to generate correlated time series is presented, that can be applied regardless of the underlying probability distributions of the historical data. The method is able to maintain the historical correlation factor among both quantities of interest. It was applied in the context of stochastic optimization of a supply chain under uncertain and correlated energy and carbon prices. The obtained network was compared to a network obtained from a deterministic approach, which is widely used in literature. The results show, that through different investment decisions regarding expansion and transportation links, the stochastic network based on correlated uncertainty is in average more flexible to possible changes in carbon

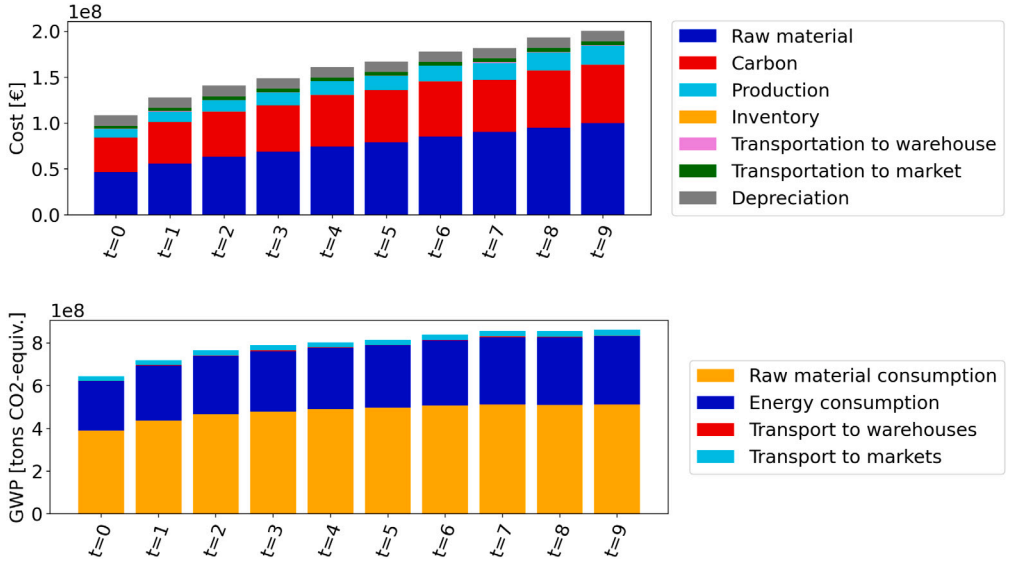


Fig. 5. Top: Expected distribution of cost for the supply chain design obtained through stochastic optimization and using correlated scenarios. Bottom: Expected distribution of CO_2 emissions of the same network.

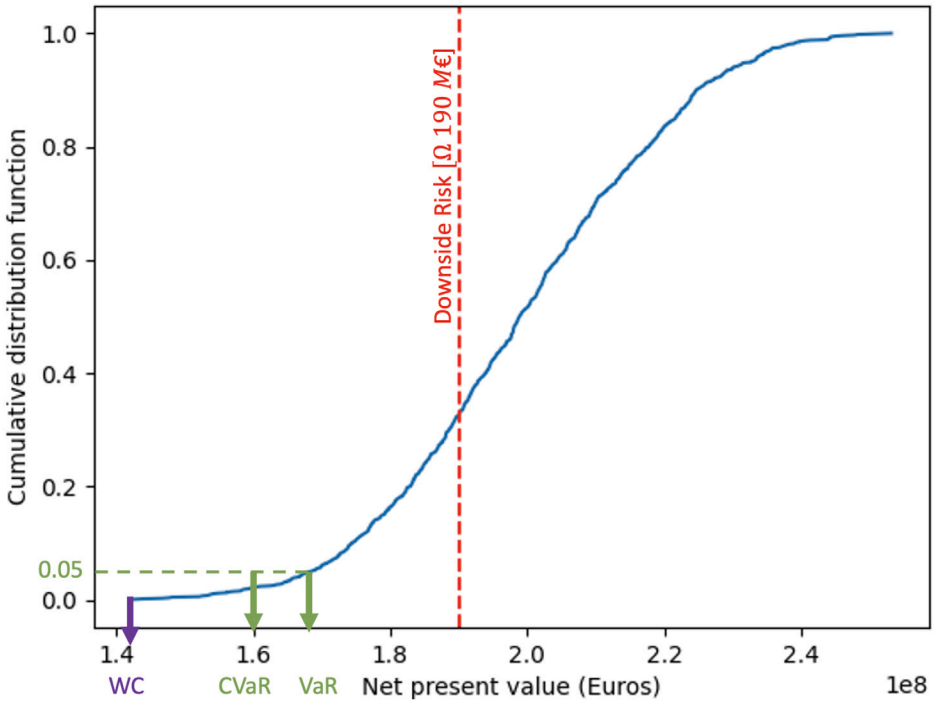


Fig. 6. Cumulative distribution function of the net present value for the supply chain network obtained from the stochastic model. The red dotted line represents the downside risk for a target value of 190 million euros. The green arrows represent the value at risk and conditional value at risk. The violet arrow indicates the worst case.

Table 7
Expected net present value, global warming potential and runtime for the stochastic model with correlated scenarios and the stochastic model with uncorrelated scenarios.

Indicator	Correlated scenarios	Uncorrelated scenarios	Units
Expected net present value	199565301	199558713	Euros
Expected global warming potential	8129607929	8132893011	kg CO_2-eq.
Runtime	27195	183655	s

and energy prices, ultimately leading to a reduction in emissions by more than 3 ktons of CO_2-equivalent.

The number of scenarios required to adequately characterize the probabilistic behavior of one or more parameters is usually a limiting factor in stochastic optimization. A combination of the methodology with data mining methods allowed for a reduction of the required number of correlated scenarios by more than 90%, with only a minor loss in statistical information. A trade-off between carbon and energy costs has been identified, therefore an analysis of the distribution of cost and emissions of the obtained stochastic network obtained was carried out. It can be concluded that the performance of the network may be improved through investments in green raw materials and efficient technologies and processes, as well as through the use of renewable energy. All of these decisions have a trade-off between short-term profitability and long-term investment, and the ability to achieve sustainability goals. In further work, it may be of interest to consider correlation among energy and carbon prices together with possible sustainability strategies, in order to reduce energy and carbon costs associated with the resulting supply chain design. Additionally, this helps to prevent delays in the decision-making process for sustainability investments, which otherwise could have an adverse effect on the network performance.

To further understand the impact of correlation on the resulting supply chain design, a stochastic optimization with uncorrelated combinations of scenarios for energy and carbon prices was carried out. Both stochastic designs were compared, showing that the network obtained using correlated scenarios has a better environmental performance. Moreover, the computation of the network using scenarios of independent carbon and energy prices was harder to compute, further highlighting the positive impact of considering correlation among both parameters. Therefore, if historical real-world data shows correlation between two or more parameters, this should always be taken into account for better model performance.

Sets

I	set of technologies, indexed by i
J	set of chemicals, indexed by j
K	set of markets, indexed by k
P	set of plants, indexed by p
S	set of scenarios, indexed by s
T	set of timesteps, indexed by t
W	set of warehouses, indexed by w

Parameters

$\alpha_{\mathrm{P}p,i,t}$	variable investment coefficient associated with technology i at plant p in timeperiod t [€/ton]
$\alpha_{\mathrm{W}wt}$	variable investment coefficient associated with warehouse w at timestep t [€/ton]
$\beta_{\mathrm{P}p,i,t}$	fixed investment term associated with technology i at plant p in timeperiod t [€]
$\beta_{\mathrm{W}w,t}$	fixed investment term associated with warehouse w at timeperiod t [€]
$\mathrm{COST}_{\mathrm{CO2}t,s}$	cost of CO_2 allowances at time t in scenario s [€/ kg of CO_2-equiv.]
$\delta_{\mathrm{P}p,i,j,t}$	production cost per unit of main product j manufactured with technology i at plant p at timestep t [€/ton]
$\delta_{\mathrm{W}j,w,t}$	inventory cost per unit of main product j at warehouse w in timeperiod t [€/ton]
$\mathrm{d}_{\mathrm{P}p,w}$	distance between plant p and warehouse w [km]
$\mathrm{d}_{\mathrm{W}w,k}$	distance between warehouse w and market k [km]
EN_i	primary energy requirements per ton manufactured with technology i [tons of fuel oil equivalent / ton]
$\underline{\mathrm{Q}}_{\mathrm{P}p,i}$	initial capacity of technology i at plant p [tons]
$\underline{\mathrm{Q}}_{\mathrm{W}w}$	initial capacity warehouse w [tons]
$\gamma_{j,k,t}$	price of final product j at market k in timestep t [€/ton]
$\mathbb{I}_{\mathrm{EN}}$	emissions of CO_2 equivalent per ton of fuel oil consumed [Kg CO_2-equiv. / ton of fuel oil equivalent]
$\mathbb{I}_{\mathrm{RM}j}$	emissions of CO_2 equivalent per ton of chemical consumed raw material j [Kg CO_2-equiv. / ton]
$\mathbb{I}_{\mathrm{TR}}$	emissions of CO_2 equivalent per km and ton transported [Kg CO_2-equiv. / (ton · km)]
ι	yearly interest rate [dimensionless]
$\underline{\theta}_{j,w}$	initial amount of product j at warehouse w at timestep $t = 0$ [tons]
$\lambda_{p,j,t}$	price of raw material j at plant p in timeperiod t [€/ton]
$\mathrm{MAX}_{\mathrm{CO2}t,s}$	emissions cap of CO_2 allowances at time t in scenario s [kg of CO_2-equiv.]
χ	penalty factor for demand insatisfaction
$\mathrm{PRICE}_{\mathrm{CO2}t,s}$	price of CO_2 allowances at time t in scenario s [€/ kg of CO_2-equiv.]
$\mathbb{P}_s$	probability of scenario s [dimensionless]
$\underline{\gamma}_{\mathrm{EP}p,i,t}$	lower bound for expansion of technology i at plant p during timestep t [tons]
$\underline{\gamma}_{\mathrm{EW}w,t}$	lower bound for expansion of warehouse w at timestep t [tons]
$\overline{\gamma}_{\mathrm{EP}p,i,t}$	upper bound for expansion of technology i at plant p during timestep t [tons]
$\overline{\gamma}_{\mathrm{EW}w,t}$	upper bound for expansion of warehouse w at timestep t [tons]
SV	salvage value fraction of the network [dimensionless]
TAX	tax rate [dimensionless]
τ_w	turn over ratio of warehouse w [dimensionless]

Variables

$\mathrm{buy}_{\mathrm{CO2}t,s}$	amount of CO_2 allowances bought at timestep t for scenario s [Kg CO_2-equiv.]
$\mathrm{C}_{\mathrm{flow}t,s}$	cashflow at timestep t for scenario s [€]
$\mathrm{DEP}_{t,s}$	depreciation term for timestep t in scenario s [€]
ENPV	expected net present value [euro]
$\mathrm{F}_{p,i,j,t,s}$	input/output flow of chemical j for technology i at plant p during timestep t for scenario s [tons]
FCI	total fixed cost investment [€]
$\mathrm{FTDC}_{t,s}$	fraction of the total depreciable capital that must be paid in period t for scenario s [€]
$\mathrm{GWP}_{t,s}$	global warming potential at timestep t for scenario s [Kg CO_2-equiv.]
$\theta_{j,w,t,s}$	inventory of chemical j kept in warehouse w in period t for scenario s [tons]
$\mathrm{NET}_{\mathrm{CO2}t,s}$	net income related to the trading of CO_2 allowances [€]
$\mathrm{NETE}_{t,s}$	net earnings at timestep t for scenario s [€]
NPV_s	net present value for scenario s [euro]
$\mathrm{PU}_{p,j,t,s}$	purchase of chemical j produced in plant p during period t in scenario s [tons]
$\mathrm{Q}_{\mathrm{P}p,i,t}$	capacity of technology i at plant p during timeperiod t [tons]
$\gamma_{\mathrm{P}p,i,t}$	capacity expansion of technology i at plant p during timeperiod t [tons]
$\gamma_{\mathrm{W}w,t}$	capacity expansion of warehouse w at timeperiod t [tons]
$\mathrm{Q}_{\mathrm{W}w,t}$	capacity of warehouse w at timeperiod t [tons]
$\mathrm{SA}_{j,k,t,s}$	sales of product j on market k at period t for scenario s [tons]
$\mathrm{sales}_{\mathrm{CO2}t,s}$	amount of CO_2 allowances sold at timestep t for scenario s [Kg CO_2-equiv.]

$\xi_{P_{j,p,w,t,s}}$ amount of chemical j being transported from plant p to warehouse w at timestep t for scenario s [tons]

$\xi_{W_{j,w,k,t,s}}$ amount of chemical j being transported from warehouse w to market k at timestep t for scenario s [tons]

$Y_{P_{p,i,t}}$ binary variable: 1 if manufacturing capacity of technology i at plant p is expanded during timestep t, else 0

$Y_{W_{w,t}}$ binary variable: 1 if capacity of warehouse w is expanded during timestep t, else 0

CRediT authorship contribution statement

Florencia Lujan Garcia-Castro: Writing – original draft, Conceptualization, Methodology, Software, Formal analysis. **Ruben Ruiz-Femenia:** Writing – review & editing, Conceptualization, Supervision. **Raquel Salcedo-Diaz:** Writing – review & editing. **Jose A. Caballero:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge financial support to the Conselleria de Innovacion, Universidades, Ciencia y Sociedad Digital of the Generalitat Valenciana, Spain, under project PROMETEO/2020/064 and project PID2021-124139NB-C21.

References

ADAC, A.D.A.C., 2022. Spritpreise 2000 bis 2010. URL: https://www.adac.de/verkehr/tanken-kraftstoff-antrieb/deutschland/kraftstoffpreisentwicklung/.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J., Perona, I., 2013. An extensive comparative study of cluster validity indices. Pattern Recognit. 46 (1), 243–256. http://dx.doi.org/10.1016/j.patcog.2012.07.021, URL: https://www.sciencedirect.com/science/article/pii/S003132031200338X.

Baringo, L., Conejo, A., 2013. Correlated wind-power production and electric load scenarios for investment decisions. Appl. Energy 101, 475–482. http://dx.doi.org/10.1016/j.apenergy.2012.06.002, URL: https://www.sciencedirect.com/science/article/pii/S0306261912004400, Sustainable Development of Energy, Water and Environment Systems.

Brockwell, P., Davis, R., 2016. Introduction to Time Series and Forecasting. Springer Switzerland.

Chazara, P., Negny, S., Montastruc, L., 2017. Quantitative method to assess the number of jobs created by production systems: Application to multi-criteria decision analysis for sustainable biomass supply chain. Sustain. Prod. Consum. 12, 134–154. http://dx.doi.org/10.1016/j.spc.2017.07.002, URL: https://www.sciencedirect.com/science/article/pii/S2352550917300234.

Das, R., Shaw, K., 2017. Uncertain supply chain network design considering carbon footprint and social factors using two-stage approach. Clean Technol. Environ. Policy 19, http://dx.doi.org/10.1007/s10098-017-1446-6.

Dehghani, E., Jabalameli, M., Naderi, M., Safari, A., 2020. An environmentally conscious photovoltaic supply chain network design under correlated uncertainty: A case study in Iran. J. Clean. Prod. 262, 121434. http://dx.doi.org/10.1016/j.jclepro.2020.121434, URL: https://www.sciencedirect.com/science/article/pii/S0959652620314815.

Delberis, A.L., Niemeyer, D., Teixeira, P., 2021. Free contract environment for big electricity consumer in Brazil considering correlated scenarios of energy, power demand and spot prices. Electr. Power Syst. Res. 190, 106828. http://dx.doi.org/10.1016/j.epsr.2020.106828, URL: https://www.sciencedirect.com/science/article/pii/S0378779620306283.

Desgraupes, B., 2017. Clustering Indices. University Paris Ouest, Lab Modal'X, URL: https://cran.r-project.org/web/packages/clusterCrit/vignettes/clusterCrit.pdf.

Dupacova, J., Gröwe-Kuska, N., Römisch, W., 2002. Scenario Reduction in Stochastic Programming; an Approach Using Probability Metrics. Math. Program. http://dx.doi.org/10.1007/s10107-002-0331-0.

ecoinvent, 0000. Ecoinvent database, URL: https://ecoinvent.org.

Economic Policy Institute, 2019. Updated employment multipliers for the U.S. economy. URL: https://www.epi.org/publication/updated-employment-multipliers-for-the-u-s-economy/.

Elsayed, N., Taha, R., Hassan, M., 2022. Measuring blood supply chain performance using Monte-Carlo simulation. IFAC-PapersOnLine 55 (10), 2011–2017. http://dx.doi.org/10.1016/j.ifacol.2022.10.003, URL: https://www.sciencedirect.com/science/article/pii/S2405896322020122, 10th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2022.

European-Commission, 2015. EU ETS Handbook.

European-Union, 2016. The EU emissions trading system (EU ETS). URL: https://ec.europa.eu/clima/sites/clima/files/factsheet_ets_en.pdf.

Eurostat, 2008. CEFIC and chemical and engineering news in European business: facts and figures.

Fanoodi, B., Malmir, B., Jahantigh, F.F., 2019. Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. Comput. Biol. Med. 113, 103415.

Fattahi, M., Govindan, K., Keyvanshokooh, E., 2018. A multi-stage stochastic program for supply chain network redesign problem with price-dependent uncertain demands. Comput. Oper. Res. 100, 314–332. http://dx.doi.org/10.1016/j.cor.2017.12.016, URL: https://www.sciencedirect.com/science/article/pii/S0305054817303167.

Garcia-Castro, F., Ruiz-Femenia, R., Salcedo-Diaz, R., Caballero, J., 2022. Design of a cooperative sustainable three-Echelon supply chain under uncertainty in CO2 allowance. ACS Sustain. Chem. Eng. 10 (34), 11098–11105. http://dx.doi.org/10.1021/acssuschemeng.2c01270.

Jahani, H., Abbasi, B., Alavi Fard, F., Talluri, S., 2018. Supply chain network redesign with demand and price uncertainty. Int. J. Prod. Econ. 205, 287–312.

Khorshidvand, B., Soleimani, H., Sibdari, S., Esfahani, M., 2021. A hybrid modeling approach for green and sustainable closed-loop supply chain considering price, advertisement and uncertain demands. Comput. Ind. Eng. 157, 107326. http://dx.doi.org/10.1016/j.cie.2021.107326, URL: https://www.sciencedirect.com/science/article/pii/S0360835221002308.

Krishnan, R., Arshinder, K., Agarwal, R., 2022. Robust optimization of sustainable food supply chain network considering food waste valorization and supply uncertainty. Comput. Ind. Eng. 171, 108499. http://dx.doi.org/10.1016/j.cie.2022.108499, URL: https://www.sciencedirect.com/science/article/pii/S0360835222005174.

Ministerio de Formento, 2019. Report from Observatorio de Costes de Transporte de Mercancias por Carretera.

Mitra, K., Gudi, R.D., Patwardhan, S.C., Sardar, G., 2008. Supply chain planning under uncertainty: A chance constrained programming approach. IFAC Proc. Vol. 41 (2), 10570–10575. http://dx.doi.org/10.3182/20080706-5-KR-1001.01790, URL: https://www.sciencedirect.com/science/article/pii/S1474667016406622, 17th IFAC World Congress.

Ruiz-Femenia, R., Guillén-Gosálbez, G., Jiménez, L., Caballero, J., 2013. Multi-objective optimization of environmentally conscious chemical supply chains under demand uncertainty. Chem. Eng. Sci. 95, 1–11. http://dx.doi.org/10.1016/j.ces.2013.02.054, URL: https://www.sciencedirect.com/science/article/pii/S0009250913001644.

Shumway, R., Stoffer, D., 2000. Time series regression and ARIMA models. In: Time Series Analysis and Its Applications. Springer New York, New York, NY, pp. 89–212. http://dx.doi.org/10.1007/978-1-4757-3261-0_2.

Statista Research Department, 2022. Preisentwicklung von CO2-Emissionsrechten im europäischen Emissionshandel (EU-ETS) von 2005 bis 2021. URL: https://de.statista.com/statistik/daten/studie/1304069/umfrage/preisentwicklung-von-co2-emissionsrechten-in-eu/.

theGlobalEconomy.com, 2021. Household consumption, percent of GDP - country rankings. URL: https://www.theglobaleconomy.com/rankings/household_consumption/.

United Nations, D.o.E., Social Affairs, P.D., 2019. World population prospects: The 2019 revision.

World Bank Group, 2019. State and trends of carbon pricing 2019. URL: https://documents1.worldbank.org/curated/en/191801559846379845/pdf/State-and-Trends-of-Carbon-Pricing-2019.pdf.

Xu, Z., Shaligram, S., Elomri, A., 2022. An eco-friendly closed-loop supply chain facing demand and carbon price uncertainty. Ann. Oper. Res. 1–27. http://dx.doi.org/10.1007/s10479-021-04499-x.

Zhang, T., Ramakrishnan, R., Livny, M., 1996. BIRCH: An efficient data clustering method for very large databases. In: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data. SIGMOD '96, Association for Computing Machinery, New York, NY, USA, pp. 103–114. http://dx.doi.org/10.1145/233269.233324.