Communication

Using Machine Learning for Enhancing the Understanding of Bullwhip Effect in the Oil and Gas Industry

Ana L. Sousa 1,*, Tiago P. Ribeiro 2, Susana Relvas 3 and Ana Barbosa-Póvoa 3

1 CERENA, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal
2 Tal Projecto Lda., 1350-252 Lisboa, Portugal; tpribeiro@gmail.com
3 CEG-IST, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal; susana.relvas@tecnico.ulisboa.pt (S.R.); apovoa@tecnico.ulisboa.pt (A.B.-P.)

* Correspondence: ana.margarida.sousa@tecnico.ulisboa.pt; Tel.: +351-218-417-000

Received: 30 June 2019; Accepted: 4 September 2019; Published: 6 September 2019

Abstract: Several suppliers of oil and gas (O & G) equipment and services have reported the necessity of making frequent resources planning adjustments due to the variability of demand, which originates in unbalanced production levels. The occurrence of these specific problems for the suppliers and operators is often related to the bullwhip effect. For studying such a problem, a research proposal is herein presented. Studying the bullwhip effect in the O & G industry requires collecting data from different levels of the supply chain, namely: services, upstream and midstream suppliers, and downstream clients. The first phase of the proposed research consists of gathering the available production and financial data. A second phase will be the statistical treatment of the data in order to evaluate the importance of the bullwhip effect in the oil and gas industry. The third phase of the program involves applying artificial neural networks (ANN) to forecast the demand. At this stage, ANN based on different training methods will be used. Further on, the attained mathematical model will be used to simulate the effects of demand fluctuations and assess the bullwhip effect in an oil and gas supply chain.

Keywords: artificial neural networks; bullwhip effect; oil and gas industry; research proposal; supply networks; machine learning

1. Introduction

The behavior problems experienced by the oil and gas (O & G) suppliers find one plausible explanation on the bullwhip effect, which is a well-known distribution channel problem [1]. Considering this, the scope of the research proposal is to develop an investigation of the bullwhip effect in the O & G supply network, giving an insight overview of this phenomenon in this system.

Suppliers must adjust to the workload associated with the large demand peaks [2], as well as the variation of cyclical periods of low and high demand [3]. Such variations lead both to challenging industrial management and sustainability problems and to increased technical and maintenance issues [4], fostered by over or underproduction. In fact, current industrial panorama is shaped by ecological and social sustainability concerns [5,6], which have been ubiquitous in the fourth industrial revolution, or Industry 4.0 [7,8], in which concepts such as the bullwhip effect have been articulated.

Production and financial data from different levels of the oil and gas supply chain shall be uniform, embracing a wide market share throughout the supply chain and easy access.

It is significant to observe that ongoing and accelerating digitalization and the digital transformation of energy and specifically O & G industries allow gathering and making available...
a profuse amount of multivalent data, including financial, production, and sustainability insights. Thence, there is an immense potential for employing data analytics techniques and machine learning in particular for enhancing the knowledge about well-known phenomena; however, some elements require further studies, such as the bullwhip effect.

Within the oil and gas industry, it is a well-known fact that the demand variability increases from downstream to upstream as in a usual supply chain [9,10]. This behavior is similar to the phenomenon known as the bullwhip effect, which has been vastly studied in the retail industry [11–13]. However, in the O & G industry, the analysis of the bullwhip effect considering the multiple-level supply network has not been deeply studied.

Oil and gas supply networks can become complex, as they encompass multiple players and intricate systems. As a result, a large amount of data shall be examined to evaluate the phenomenon. Also, taking into consideration that too-simple models have been a major hindrance for past bullwhip effect studies [14], this is clearly a case for using the current artificial intelligence capabilities. Cutting edge techniques, such as the recent developments in artificial neural networks will provide the means to infer the meaningful relations within an oil and gas supply network. Thus, the availability of network simulation and forecasting tools will be pivotal for studying the bullwhip effect.

Beyond learning from data, the use of artificial neural networks (ANN) will allow attaining mathematical solutions with a generalization capacity. This is yet another very significant advantage of implementing ANN-based analysis methodologies, since the results of the proposed research program will not cease in the bullwhip effect for one specific case, but rather are expected to provide tools for the community that can be efficiently applied to similar problems. Developing industry-specific or integrating broader [15] risk analysis tools is a possible outcome for the research output.

2. Literature Review

2.1. Bullwhip Effect

2.1.1. Overview

The industrial era brought the need for operations management and supply chain management development. Over time, technology, the economy, and the social landscape have been responsible for changing it.

By the industrial revolution, products were quite simple, markets were locally oriented, communication was slower, as the transportation speed also was, and innovation cycles were longer [16], allowing the supply chains to be simpler and more vertically integrated. After World War II, the production paradigm changed. Seizing technological developments that enabled the production of more innovative, complex, and customized products, companies adapted to provide the best service to their clients. Productivity and efficiency were the main concepts that supported the production paradigm. While compressing the speed of operations is the main method for increasing productivity, efficiency is reached by reducing the use of resources and increasing asset utilization. To achieve a large scale and reduce costs, it was necessary to standardize products and production systems.

During the 1950s and 1960s, Forrester focused his research on the operational behavior of companies and published several studies about Industrial Dynamics [1,17], becoming known as the first author who described the bullwhip effect. The effect of lack of intercoordination among supply chains and the unawareness of managers to deal with the complex interactions results in an increased variability of placed orders at suppliers, comparing to the received orders given to customers. To face the challenges of manufacturing both more complex products and supplying them to global markets, most companies focused on optimizing their internal processes. However, the bullwhip effect was not perceived as a major issue. In the 1950s, supply chain structures were still quite simple and vertically integrated [16]. In this period, the simulation of Industrial Dynamics was restricted due to the technological limitations of analogue computers [18].
Due to the technological development and increasingly complex processes, it became impossible to keep the full process inside a single company. More specialized and smaller companies have grown skills to manufacture semi-finished products with greater cost efficiency. In globalized markets, to cope with longer geographical distances, big companies generally choose to decentralize, fragmentating or outsourcing some operations [19]. This decomposition changed the supply chains into supply networks [14]. As the focus on intra-organizational operations management has shifted to inter-organizational, the bullwhip effect has become one major target of analysis [16]. For instance, in 1989, Sterman proposed the Beer Game experiments, which demonstrated the severe impact of the bullwhip effect [20].

In the same decade, the appearance of information technology systems transformed companies’ management operations, since it allowed much cheaper and faster communications, and more accurate planning [21]. This technological development also enabled additional globalization, since it allowed an improved interconnectivity between geographically isolated operations [22]. Offering more variability and enhanced product customization were ongoing milestones in company strategies. During the 1990s, the bullwhip effect was comprehensively studied using empirical data. For instance, Pampers, HP, and Barilla [10,23] were some of the studied companies.

During the 21st century, research on supply chain management has also focused on the supply chain’s networked structure. Such a structure of modern supply chains forms a barrier to collaboration [24], due to the players’ behavior [25,26]. Lee et al. concluded that it is challenging to reach agreements between the parties in the supply chain to implement the bullwhip mitigation solutions in order to improve the overall performance [10,21,27]. Although companies should focus their strategies on cooperation and supply chain synchronization [28,29], since information sharing is regarded as the main principle to prevent this effect [12,30–35], the managers tend to be hesitant regarding collaborative programs [36]. Thus, the quantity and quality of shared information remains low. Nevertheless, Tsay and Lovejoy concluded that some mitigation solutions developed under such principle have resulted in counterproductive results [19]. The main causes for such a lack of cooperation may be due to a lack of knowledge, lack of competence, or lack of trust.

Recently, some studies have been developed to understand the bullwhip phenomenon in the oil and gas industry. However, while most studies consider a simple supply chain, the oil and gas industry systems are a multilevel supply network. One example can be found in Shizheng et al., who developed their study to quantify the bullwhip effect in a supply chain based on a single petroleum distribution system [37]. Within the same study field, Huang et al. implemented a method to control the bullwhip effect in two Chinese petrochemical companies: Liaoyang Petrochemical and Shanghai Baosteel Yichang Steelstrip [38]. Jacoby simulated the cost of the bullwhip effect in the industry [39]. Zhang and Zhang analyzed the bullwhip effect in China’s processed oil supply chain, using a dynamic analysis system and programming a model to mitigate this effect, and the authors concluded that the delay time is the main reason for the existence of the bullwhip effect [40]. Lastly, Sherhart assessed a specific bullwhip effect problem at British Petroleum [41].

However, few articles offer strategies that specifically mitigate the bullwhip effect in the oil and gas industry. The aforementioned studies (Huang et al. [38], Jacoby [39], Zhang and Zhang [40], Sherhart [41]) may be only regarded as useful developments for modeling the bullwhip effect in specific oil and gas companies.

2.1.2. Performance Measures, Models, and Methods

The bullwhip effect refers to a demand distortion in a supply chain, which propagates and amplifies as one moves up in the supply chain, from the buyer to the supplier ([10,28,29,35,42–45]).

Sucky (2009) divided the research on bullwhip effect into six categories [14], which are described in the Table 1. Based on the research developed in this paper, the collected references were categorized according to Sucky’s classification.
Table 1. Categorization of the bullwhip effect studies.

<table>
<thead>
<tr>
<th>Categories</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantification of the bullwhip effect</td>
<td>[10,46–52]</td>
</tr>
<tr>
<td>Identifying the causes of the bullwhip effect</td>
<td>[44,53]</td>
</tr>
<tr>
<td>Observation studies in some industries</td>
<td>[54–56]</td>
</tr>
<tr>
<td>Methods to reduce the bullwhip effect</td>
<td>[57–61]</td>
</tr>
<tr>
<td>Simulation of the system behavior</td>
<td>[62–65]</td>
</tr>
<tr>
<td>Experimental validation of the bullwhip effect</td>
<td>[30,41,66–68]</td>
</tr>
</tbody>
</table>

To perform a statistical survey about the metrics applied to study the bullwhip effect, a wide set of publications was gathered, and the information was treated in order to provide an overview about the relevant supply chain contributions. Criteria for selecting the reference works included its publication during the past 20 years on Web of Science (WoS) indexed journals with “bullwhip effect” and “supply chain” keywords. After an initial group of articles was gathered, only the ones with explicit and extensive performance metrics descriptions were included in the collection herein presented in this section. Considering the information provided in the set synthetized in Table 2, it was possible to summarize the main performance metrics used to study the bullwhip effect (Figure 1).

Table 2. Performance metrics per reference.

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order rate variance ratio</td>
<td>[43,48,61,64,69–79]</td>
</tr>
<tr>
<td>Amplitude rate cost ratio</td>
<td>[80]</td>
</tr>
<tr>
<td>Amplification ratio</td>
<td>[81]</td>
</tr>
<tr>
<td>Ratio inventory</td>
<td>[71,75]</td>
</tr>
<tr>
<td>Ratio backlog inventory</td>
<td>[64]</td>
</tr>
<tr>
<td>Variance ratio fill rate</td>
<td>[75,77]</td>
</tr>
<tr>
<td>Ratio inventory integrated squared error</td>
<td>[69]</td>
</tr>
<tr>
<td>Ratio root mean square costs</td>
<td>[61]</td>
</tr>
<tr>
<td>Fill rate</td>
<td>[82]</td>
</tr>
<tr>
<td>Costs order rate variance ratio</td>
<td>[43,73,83]</td>
</tr>
<tr>
<td>Costs</td>
<td>[84]</td>
</tr>
<tr>
<td>Ratio inventory stock</td>
<td>[70,74]</td>
</tr>
<tr>
<td>Out size stock out number</td>
<td>[70]</td>
</tr>
</tbody>
</table>

Figure 1. Performance metrics to study the bullwhip effect.
As Figure 1 illustrates, the most common performance metric suggested in the literature is the order rate variance ratio [85], which was proposed by Chen et al. [48]. This indicator is given by the following Equation (1) [29]:

\[
\text{Order rate variance ratio} = \frac{\sigma_o^2}{\mu_o} \cdot \frac{\sigma_d^2}{\mu_d}
\]

(1)

where \( \sigma_o^2 \) is the orders’ variance, \( \sigma_d^2 \) the variance of the market demand, and \( \mu_o \) and \( \mu_d \) are the respective mean values.

2.2. Artificial Neural Networks

2.2.1. Background

Artificial intelligence (AI), also referred as machine intelligence, is a wide scope, cross-disciplinary term deemed to encompass all machine behavior that is inspired by human cognitive functions. AI is based on algorithms, but cannot be accurately defined as mathematical reasoning, since its most notorious ability is the capability for adaptation with insufficient knowledge and resources [86]. Such ability is not provided by Mathematics, and can only be found in the current definition of “intelligence”.

The roots of AI go back to the World War II [87,88], from which it has grown consolidated upon several disciplines. While many classifications and clustering have been proposed for such techniques, the most common taxonomy distinguishes between, at least, seven different disciplines [89–93]. Those, which can also be referred as sub-disciplines, sub-fields, or branches, are machine learning (ML), expert systems, fuzzy logic, robotics, computer vision, natural language processing, and speech recognition. Furthermore, several techniques have achieved notability through wide employment. That is the case of deep learning techniques, artificial neural networks, decision trees and random forests, among many others, applied within the machine learning framework, but also evolutionary and genetic algorithms, support vector machines, or case-based reasoning, to name a few.

Among the aforementioned disciplines, machine learning is the one with the widest employment for engineering problems, and especially for the oil and gas industry [89,94]. ML aims to conceive algorithms that improve automatically through experience [93]. Yet, for doing so, machine learning techniques typically sort from a comprehensive set of possible algorithms, initially finding the best suited for the data, given certain performance criteria. ML can be classified as supervised learning, unsupervised learning, and reinforcement learning.

Machine learning is devoted to foster humanly behavior in machines. Specifically, it allows “teaching” or “training” computers how to compute tasks by giving examples on how it should be done [95]. Thus, it can be employed when there is profuse data (used as patterns or examples) on a certain phenomenon, but not necessarily an analytical relation or perfectly described explanation among variables.

Artificial neural networks (also referred as ANN) are machine learning techniques. In fact, those are the most powerful and oldest techniques [96]. Therefore, it is not a surprise that ANN also leads the number of known practical applications, and are applied to most fields of knowledge ([97,98]).

An ANN can be regarded as a mathematical model designed to perform a specific task, simulating the manner by which the brain computes information by employing processing units (the neurons). Despite being mathematically simple, ANN-based solutions combine different network topology and network algorithms together in order to provide more accurate results than those given by classical approaches, as the nonlinear multivariate regression. Furthermore, ANN-based solutions can be employed even without an extensive knowledge of the modeled function or numerical relations [99]. Nonetheless, prior data analysis is critical for planning optimal ANN-based solutions. As a matter of fact, the dataset shape has an indelible impact upon the solutions and approaches [100], making topological studies of paramount importance, especially when datasets are gathered from profuse information, as suggested in this research proposal.
The general ANN structure integrates several nodes disposed in connected vertical layers (an input layer, one or more hidden layers, and an output layer). Associated to each node, or neuron, in the hidden and output layers is a transfer (or activation) function, which can be linear or nonlinear. Such function receives a net input and transmits an output.

Most ANN are multilayered and feedforward. The latter means that data inserted in the input layer flows in the vertical direction only, so that each node is only connected to nodes belonging to layers disposed at the following level. Among multilayer feedforward networks, one common and useful learning algorithm is the back propagation (BP), proposed by Rumelhart et al., in 1986 [101]. BP allows constantly adjusting the network weight and threshold values back and forth in order to attain the minimum sum of a squared error. Further developments brought the back propagation adaptive algorithm (BPA) or enhanced convergence and precision even for few data, as attained by implementing the Marquardt algorithm [102]. However, there are many other learning algorithms suited for multilayer feedforward networks. The recent and non-iterative extreme learning machine (ELM) algorithm is an example, including its developments in convex incremental ELM (CI-ELM), incremental ELM (I-ELM), and mini-batch ELM (MB-ELM) [32,103–105]. Figure 2 shows a schematic representation of a multilayered ANN.

![Figure 2. Example of a multilayer artificial neural networks (ANN).](image)

Using ANN enables a very efficient use of computing power, providing a rather scarce, and surprisingly foreseeable [106], computing time to yield results from complex models. This is due to its massively parallel distributed structure, making them suitable to efficiently solve simple and complex problems. In fact, the ANN structure owes its power and versatility to its core characteristics of self-learning and generalization. The first means that the algorithm development based on external conditions enables ANN to adapt to several, and even evolving, data contexts, without the need to be “reshaped” or “reprogrammed”. The latter is deemed to express that once the learning process has been performed, the resulting algorithm can be less sensitive to input quality, being able to cover a profuse amount of possible inputs. This also provides a greater tolerance to data lacking quality or quantity.

On the other hand, one cannot disregard some shortcomings in ANN. Among the most relevant are the inability for necessarily reaching globally optimized algorithms, the slow convergence speed, and the lack of theoretical guidance for algorithm building (the number of hidden layers and the unit layer choice, for example), which leads to the need for profuse experiments [107]. Furthermore, practical hindrances emerge when dealing with very profuse raw data. Finding and filtering the most significant subdata is a critical task that precedes ANN application [108].
2.2.2. Developing ANN Solutions

Many practical issues arise when conceiving and developing ANN solutions. A brief summary of important steps and solutions that an ANN developer must take into account is given in the following lines.

Strategies must be implemented in order to assure that the input variables are independent and relevant, considering its effect upon the ANN accuracy [109,110]. Thus, input dimensionality reduction is a required step for which two approaches can be used, namely the feature selection and the feature extraction. Several algorithms allow fulfilling the input reduction. Among the most common are the linear correlation, the autoencoder or the orthogonal, and sparse random projections [110].

Once the input datasets handling has been defined, training, validation, and testing procedures must also be set. This means deciding the amount of data to be shared between those tasks. Usual relations for the former procedures are 50–25–25, 60–20–20, 70–15–15, or 80–10–10 [94].

Training efficiency can be impaired if the concentration of training data points is heterogeneous. Thus, several techniques can be employed for input normalization, such as the ones proposed by Lachtermacher and Fuller [111]; Pu and Mesbahi [112]; and Tohidi and Sharifi [113].

Some transfer functions can be employed for output and hidden layers. Among the most common, one can count the logistic (a sigmoid type), the hyperbolic tangent (also sigmoid), bilinear, identity, identity–logistic, bipolar, positive saturating linear, sinusoid and radial basis functions. Afterwards, output normalization may be considered in order to avoid numerical problems in the training phase [114,115].

Network arrangement must also be defined. Even if limiting the option to one type of ANN, such as the multilayer feedforward, important decisions about the behavior of nodes in the hidden layers will define the arrangement type. Common choices span from the multilayer perceptron network to radial-basis function network [94]. While no arrangement can be concluded better than any other, there is extensive literature about the adequacy of each of these networks for certain problems. Besides the previous choices, the connectivity criteria must be defined. Even if considering a feedforward network, only adjacent or distant layers may be set for connecting to each other.

2.2.3. Application to Supply Chain Management

A supply chain can be regarded as a complex network spanning several companies and sectors. As materials and products are acquired, processed, and sold among the network players, a flow of information is generated. Additionally, the companies involved keep extensive records of almost every aspect of its activities. Cumulatively, this is an ever-growing amount of information which, when wisely used, can provide analysts and managers with excellent tools for supply chain management.

However, as data becomes profuse, the difficulty of handling it increases. Furthermore, mathematical descriptions for the relations that link the network have not been widely established yet [116]. In this context, information can be used for extracting performance indicators.

Artificial neural networks are a possible answer for this problem [117], such that its use in supply chain management has boomed in the past years [107]. The ability of ANN to learn from abundant data allows modeling the behavior of supply chain networks [118–120]. Such mathematical output is extremely important, since it offers the possibility of forecasting the network behavior both for previous and new conditions as well as testing the effects of network manipulation and modifications.

Currently, ANN are employed in supply chain management mostly in three areas [107]. Those are optimization [121,122], forecasting [123], and decision support [124–128]. In the first case, ANN are used for performing optimization tasks within models deemed to solve logistic problems such as optimizing transport routes, schedules, equipment, and shop scheduling, or for enhancing warehouse management, especially in a context of real-time information updates. Forecasting plays a major role in uncertainty and risk mitigation. Unfortunately, management decisions are often made in a context where data is insufficient or unreliable; thus, other forecasting methods face significant inadequacy problems. That is the case for expert systems, statistical methods, and time series [107], but not for
well-trained ANN. Finally, decision support has benefited from ANN capacities for identification, classification, and self-organization tasks. Furthermore, ANN versatility allows the development of case-specific algorithms, which can support particular decision problems.

2.2.4. Application to Oil and Gas Industry

Considering the recent increase in using artificial intelligence techniques for studying virtually all types of engineering problems, with a special emphasis on artificial neural networks, it is with no surprise that recent developments both in oil and gas research and practice make extensive use of ANN. In fact, the size of databases generated either in geological analyses or in production control and optimization require the use of the most advanced and powerful big data analysis tools that science can provide.

Recently, many scientific journals devoted to O & G research published special issues for AI applications. Furthermore, many of the recently most cited articles within O & G research employ ANN. That is the case of Ahmadi [129] (production performance), Mahdiani and Khamehchi [130] (crude oil price), Mirzaei-Paiaman et al. [131] (production flow rate), Torabi et al. [132] (physical properties), and Arjun and Aneesh [133] (physical properties), to name a few. However, using ANN to study the bullwhip effect on the oil and gas industry is still not common. Benefiting from such resource is one further reason why this research program proposal suggests the development of ANN.

3. Research Objectives

3.1. Main Goals Delineation

The problem the proposed work aims to address can be reduced to the following specific research questions:

- Does the bullwhip effect occur in the petroleum industry?
- Does the suppliers’ level exhibit the highest demand variability in the supply network?
- Are smaller companies more susceptible to higher variability than larger companies?
- Is it possible to forecast the bullwhip effect using artificial neural network techniques?
- Is it possible to create mathematical models for O & G supply networks’ behavior, so that possible remedial measures for the bullwhip effect can be assessed prior to testing in real conditions?

Answering these questions will allow the researcher to comprehensively characterize the effects of demand variability onto a typical oil and gas supply network, draw conclusions about the existence of the bullwhip effect upon the interest group, distinguish such effects based on the company profile of that group, and ultimately, be able to model, forecast, and manipulate the network behavior, with a special focus on the equipment and service suppliers.

Considering the former, the main goal of the research proposed can be summarized as to characterize the bullwhip effect on the oil and gas industry, and to develop a tool to model it.

3.2. Scientific and Social Relevance

As highlighted in the historical brief regarding the state of the art of the bullwhip effect, its study has gained an increasing relevance as supply chains have evolved to supply networks, as this has led to the processes, stakeholders, and their connections becoming much more complex and the information flow achieving unprecedented dimensions.

Therefore, understanding the bullwhip effect in depth and being able to promote validated remedial measures is a stepping stone toward securing industry efficiency, keeping the consumer prices as low and stable as possible, and avoiding product unavailability or excessive stocks, which frequently come with high economic and environmental costs.
Considering the reach and impact of oil and hydrocarbons’ sub-products in our society, stability and cost effectiveness are paramount for our main goals, spanning from families’ economic stability to global economic growth, energetic efficiency, and environmental and sustainability goals.

From a scientific perspective, the ability to model the bullwhip effect in complex supply networks, based on collected data, is a remarkable advance, since it will add to the currently limited available models. Furthermore, proposing and validating remedial measures for this industry can bring real innovation in supply networks management and foster further scientific advances.

4. Methodology and Planning

4.1. Methodology

A data-driven work methodology is proposed for the current research idea. It shall encompass a problem definition, solution development, and solution testing, as extensively described by Sage and Armstrong [134]. In a broader sense, the suggested methodology can be regarded as a pragmatist epistemology, since it will be based on real data collection—a form of empiricism—to perform advanced analyses and achieve logical inferences, using the principles or rationalism.

Among those stages, the problem definition has been briefly addressed in this proposal, and shall be enhanced with the research program’s initial literature review. Solution development and testing will be nested under the ANN sub-program. In accordance, the following major specific steps for work methodology are defined in Figure 3.

![Diagram](image)

**Figure 3.** Methodology to perform this research program.

Despite the previous suggestions, which were deliberately kept broad in scope and far-reaching in order to foster researchers’ freedom, it is important to notice that the work methodology must be continuously reassessed throughout the program’s duration, taking into consideration the attained partial results as well as the literature review that will be undertaken.

4.2. Work Phases

Table 3 provides a time schedule for the completion of the proposed research program. The tasks have been clustered in five functional groups, namely performing the literature review and data gathering planning; data gathering and analysis; artificial neural network; the writing and submission of articles; and other dissemination activities.

While the first group is self-explanatory and accounts for the continuous research and planning activities that must take place throughout the program, the remaining steps must proceed in order.

Therefore, data gathering and its critical analysis must commence from the first moment, since it is the basis for the remaining steps. Soon after, the ANN model development must take place, but it may extend for four semesters, given its complexity. Also, at the beginning of the second year of work, the first paper shall be written, using both the literature review and the preliminary data analysis.
From this stage on, ANN model development, validation, and implementation run continuously, while the results analysis with papers and thesis writing shall occur whenever relevant results are achieved.

<table>
<thead>
<tr>
<th>Table 3. Research program plan: Time schedule.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year</strong></td>
</tr>
<tr>
<td><strong>Semester</strong></td>
</tr>
<tr>
<td><strong>Literature Review and Data Gathering Planning</strong></td>
</tr>
<tr>
<td>Literature review and data gathering planning</td>
</tr>
<tr>
<td><strong>Data Gathering and Analysis</strong></td>
</tr>
<tr>
<td>Collecting the data</td>
</tr>
<tr>
<td>Statistical analysis</td>
</tr>
<tr>
<td><strong>Artificial Neural Network</strong></td>
</tr>
<tr>
<td>Developing the model</td>
</tr>
<tr>
<td>Validating the model</td>
</tr>
<tr>
<td>Implementing the model to the case study</td>
</tr>
<tr>
<td><strong>Writing and Submission of Articles</strong></td>
</tr>
<tr>
<td>Paper I—Quantitative analysis on the bullwhip effect in a supply chain network</td>
</tr>
<tr>
<td>Paper II—The bullwhip effect: a case study in the oil and gas industry</td>
</tr>
<tr>
<td>Paper III—Oil and gas supply chain management based on artificial neural network</td>
</tr>
<tr>
<td>Writing up the thesis</td>
</tr>
<tr>
<td><strong>Other dissemination activities</strong></td>
</tr>
<tr>
<td>Oral presentation at a conference</td>
</tr>
<tr>
<td>Poster presentation at a conference</td>
</tr>
</tbody>
</table>

Legend: ★ Critical tasks.

The last group of tasks related to disseminating the work through conferences may be left for the scheduled second half of the research program.

The following sections will detail the major aspects that should be developed at each phase.

4.2.1. Literature Review and Data Gathering Planning

The first step is to perform a deeper literature review, which will support the identification of systemic and behavioral factors that have an impact on the bullwhip effect in the oil and gas supply chain. Also, a clear definition of the oil and gas supply chain and all the stakeholders involved will be set up, in order to understand how each entity interacts with the others and the impact of each entity on the bullwhip effect.

In this stage, the definition of which parameters should be gathered in order to characterize the oil and gas supply network should be established.

4.2.2. Data Gathering and Analysis

During this phase, the available financial and production data will be collected from the New York Stock Exchange database [135,136] as well as in companies [137–143] other stock markets [144], projects, and national authorities’ reports.

It is important to stress that all the significant companies involved in the oil and gas industry are either listed in stock markets or are publicly detailed, with thorough public reporting. Therefore, data gathering is a feasible task using information available on the internet. Beyond financial data, which allows assessing whether orders and production were matched by sales, companies’ reports provide
other important information on its business context, directly focusing on production and stocking when significant.

Furthermore, most companies, joint ventures, and large projects voluntarily provide, in their websites and to the media, information about their activity, which can be used for redundancy and checking.

Very significant and growing digitalization efforts, undertaken by oil and gas companies throughout recent times [145,146], assure not only a very wide availability of data, but also data quality. That is mostly due to the employment of data analytics techniques for data filtering, structuring, and consolidation [93,147,148]. Furthermore, data analytics, including several tools for big data analysis, can be successfully used with ease by the researchers to extract high-quality information from the aforementioned data sources.

It is also noteworthy that aggregation on each level of the supply network could be paramount for simulating it with available data. This means that each network level is considered as a collection of nodes, each one with similar data structure. At every level, financial flow is gathered, using the revenue for demand estimation and production costs for assessing offer.

In order to select which companies should be studied, it will be necessary to verify which ones are directly involved in the oil and gas industry, and then verify whether they have a set of public data within a common period and common nature. Another selection criterion is to ensure that the companies have not been merged, separated, acquired, or undergone significant organizational changes in the period under study. This principle is paramount to guarantee that information about production or demand data is not biased. One further relevant need is closely analyzing the suppliers’ chain level. In fact, equipment and service suppliers may interact with several other chain levels, as depicted in Figure 4; however, their most profuse activity is likely to be provided to upstream enterprises, leading to a higher vulnerability to the bullwhip effect. Therefore, companies must be analyzed to figure, through their client base and financial data, whether they can be fitted on the upper level of this supply chain. The companies should be arranged at the respective level of the supply chain to which they belong, as defined in Figure 4.

![Figure 4. Oil and gas supply chain.](image)

Table 4. describes and summarizes the companies on each supply chain level.

<table>
<thead>
<tr>
<th>Supply Chain Levels</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equipment and services suppliers</td>
<td>The population includes companies that provide services and equipment to support all the oil and gas activities, such as drilling, exploration, gathering, storing, and processing the oil and gas.</td>
</tr>
<tr>
<td>Upstream</td>
<td>Crude oil and natural gas production</td>
</tr>
<tr>
<td>Midstream</td>
<td>Transportation of oil and gas</td>
</tr>
<tr>
<td>Downstream</td>
<td>Refine and process crude oil and natural gas to be sold to the consumers</td>
</tr>
</tbody>
</table>

This phase characterizes each company, using the number of employees and revenue, to understand its size. This information will allow selecting the significant companies for the study, further investigating whether the bullwhip effect has a different level of importance for different company sizes, as well as assuring a consistent business size along the chain nodes.
In fact, the oil and gas companies belong to a supply network, as shown in Figure 5. As noted in significant research [14,24], it would be interesting to study the correlations between them, and to evaluate if it is possible to study this phenomenon as a supply network, instead of in relation to supply chain behavior, in case significant correlations among companies are found.

![Oil and gas supply network](image)

**Figure 5.** Oil and gas supply network (example with some companies).

Figure 5 has been populated with some significant companies to better illustrate this research idea with real cases and provide concrete suggestions for a preliminary investigation.

Based on the collected series, clients’ demands and orders submitted to suppliers shall be analyzed in order to determine the coefficient of variation in demand and supply, and evaluate the cyclical patterns of production and demand.

The statistical treatment of the gathered data will allow developing a bullwhip effect model that can be applied to the oil and gas industry. The aim of such analysis is to determine whether the higher level in the supply network is more variable than the lower level. Further on, the attained mathematical model will be used to simulate the effects of demand fluctuations and assess the bullwhip effect within an oil and gas supply network.

### 4.2.3. Artificial Neural Network

The next phase of the research program consists of developing an ANN model to forecast the demand. Artificial neural networks can accommodate nonlinear data to determine functional relationships between empirical data, even when the underlying relationships are unknown or difficult to describe.

The choice of ANN methods will depend not only on the researchers’ proficiency, experience, and resources, but also on its suitability for a specific problem. Model conception and ANN adaptation will be shaped by the dimensions of the input and output variables, and continuously reassessed as preliminary results are inspected.

At this stage, different artificial neural networks based on different methods shall be used, and a comparative study shall be carried out in order to minimize the prediction error.

A brief of suggested ANN algorithms can be found in Section 2.2.1, while several practical issues are dealt with in Section 2.2.2, spanning some of the most common decisions that developers must take.

The last steps of the program will be devoted to generalized results and recommendations for suitable supply networks. This will also allow suggesting interesting and relevant future works to the scientific community.
5. Conclusions

Oil equipment and services suppliers are at the end of a complex supply network, and experience a bullwhip effect with oil demand fluctuations. However, no extended and comprehensive studies have been found yet on this issue, embodying a significant research gap. The relevancy of the research work herein proposed is further highlighted by the emerging need to mitigate uncertainty and inefficiency in an industry recently affected by a downturn and facing a fierce pressure to meet sustainability requirements.

Considering the latest developments and increasing computing power of ANN, there are now sufficiently mature resources to conduct the proposed research. Furthermore, data availability in the oil and gas industry has increased significantly in recent years due to higher accountability requirements and an industry wide effort toward digitalization.

The work outcome is deemed to be innovative and generalizable for similar problems, and is expected to offer solutions with scientific and social impact. Both theoretical and managerial applications are expect to be attained. Among the former, a better understanding and clarification of the bullwhip effect through real data is, necessarily, an advance in supply chain theory. Secondly, conclusions about the ability of ANN for forecasting the bullwhip effect could provide a sound scientific advance, as would conclusions about the oil and gas supplier’s susceptibility to the bullwhip effect.

On the other hand, managerial applications can include developing an ANN-based tool for modeling the oil and gas supply network and drawing validated proposals for mitigation measures.

Concerning the proposed study limitations, one shall highlight that, as a research idea, the current proposal is necessarily limited by the researcher freedom it is deemed to cultivate, as well as by the multitude of investigation paths it is supposed to offer. Yet, other important limitations lie in the inherent complexity and uncertainty on bullwhip effect analysis based on real data. Data pool idiosyncrasies will, undoubtedly, limit research outcome scope. Furthermore, attaining a pertinent, reliable, and robust analysis tool for modeling through machine learning cannot be regarded as a guaranteed result for the research program.

Data gathering and ANN development tasks shall be closely controlled, their planning shall be prioritized, and redundancy shall be pursued.


Funding: This research was funded by Fundação para a Ciência e a Tecnologia, I. P, grant number SFRH/BD/131005/2017.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References


41. Sherrard, E. The Bullwhip Effect: Recognizing the Phenomenon and Mitigating It Using the Theory of Constraints Illustrated by a Case Study from British Petroleum; Baylor University: Waco, TX, USA, 2013.
42. Croson, R.; Donohue, K. Impact of POS Data Sharing on Supply Chain Management: An Experimental Study. Prod. Oper. Manag. 2003, 12, 1–11. [CrossRef]
51. Zhou, L.; Disney, S.M. Bullwhip and inventory variance in a closed loop supply chain. OR Spectr. 2006, 28, 127–149. [CrossRef]
64. Ingalls, R.G.; Foote, B.L.; Krishnamoorthy, A. Reducing the bullwhip effect in supply chains with control-based forecasting. *Int. J. Simul. Process Model.* 2005, 1, 90–110. [CrossRef]
70. Chaharsooghi, S.K.; Heydari, J. LT variance or LT mean reduction in supply chain management: Which one has a higher impact on SC performance? *Int. J. Prod. Econ.* 2010, 124, 475–481. [CrossRef]
94. Abambres, M.; Rajana, K.; Tsavdaridis, K.D.; Ribeiro, T.P. Neural Network-Based Formula for the Buckling Load Prediction of I-Section Cellular Steel Beams. *Computers* 2018, 8, 2. [CrossRef]
111. Lachtermacher, G.; Fuller, J.D. Backpropagation in Time-series Forecasting. *J. Forecast.* 1995, 14, 381–393. [CrossRef]


148. Yoo, Y. It is not about Size: A Further thought on Big Data. J. Inf. Technol. 2015, 30, 63–65. [CrossRef]