International Journal on Advances in Systems and Measurements

2022 vol. 14 nr. 3&4

The *International Journal on Advances in Systems and Measurements* is published by IARIA.
ISSN: 1942-261x
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Systems and Measurements
Volume 15, Number 3 & 4, 2022

Leonel Sousa, INESC-ID/IST, TU-Lisbon, Portugal
Arvind K. Srivastav, NanoSonix Inc., USA
Grigore Stamatescu, University Politehnica of Bucharest, Romania
Raluca-Ioana Stefan-van Staden, National Institute of Research for Electrochemistry and Condensed Matter, Romania
Pavel Šteffan, Brno University of Technology, Czech Republic
Chelakara S. Subramanian, Florida Institute of Technology, USA
Sofiene Tahar, Concordia University, Canada
Muhammad Tariq, Waseda University, Japan
Roald Taymanov, D.I.Mendeleyev Institute for Metrology, St.Petersburg, Russia
Francesco Tiezzi, IMT Institute for Advanced Studies Lucca, Italy
Wilfried Uhring, University of Strasbourg // CNRS, France
Guillaume Valadon, French Network and Information and Security Agency, France
Eloisa Vargiu, Barcelona Digital - Barcelona, Spain
Miroslav Velev, Aries Design Automation, USA
Dario Vieira, EFREI, France
Stephen White, University of Huddersfield, UK
Shengnan Wu, American Airlines, USA
Qingsong Xu, University of Macau, Macau, China
Xiaodong Xu, Beijing University of Posts & Telecommunications, China
Ravi M. Yadahalli, PES Institute of Technology and Management, India
Yanyan (Linda) Yang, University of Portsmouth, UK
Shigeru Yamashita, Ritsumeikan University, Japan
Patrick Meumeu Yomsi, INRIA Nancy-Grand Est, France
Alberto Yúfera, Centro Nacional de Microelectronica (CNM-CSIC) - Sevilla, Spain
Sergey Y. Yurish, IFSA, Spain
David Zammit-Mangion, University of Malta, Malta
Guigen Zhang, Clemson University, USA
Weiping Zhang, Shanghai Jiao Tong University, P. R. China

International Journal on Advances in Systems and Measurements
Volume 15, Numbers 3 & 4, 2022

CONTENTS

Continuous Monitoring of Gender Equality

The development of a counter-account system for the hospitality industry

Anais Heeren
Hotel Management
Zuyd University of Applied Sciences
Maastricht, the Netherlands
lectoraat.okb@zuyd.nl

Martijn Zoet, Eric Mantelaers
Finance and Control & Accountancy
Zuyd University of Applied Sciences
Maastricht, the Netherlands
Martijn.zoet@zuyd.nl

Abstract **- Sustainability reporting has become common practice in organizations. A factor that is associated with the rise of sustainability reporting is greenwashing. To counter greenwashing practices, counter-accounting, referring to the use of information produced by actors outside a given organization or industry, aims to help verify statements made by organizations. Although counter-accounting already exists in the toolbox of the auditor, it is mainly applied in an ad-hoc manner and rarely continuous. In this study, a continuous monitoring system for counter-accounts is proposed to measure gender inequality. The usefulness of the system is demonstrated by analyzing 22075 job appointment announcements in the hospitality sector and 55 statements regarding organizations' efforts and goals in the field of workforce equality. The presented results shed light on occupational gender segregation and provide a basis from which more continuous counter-accounting systems can be developed. In this paper, we re-address and - present our earlier work, yet we extend the previous study with additional research by increasing the appointment announcement dataset from 1000 instances to 22075 and by adding an additional form of usefulness validation for the developed system. In addition, we provide a more detailed description of the related literature, findings, and results.**

Keywords - continuous monitoring; hospitality sector; sustainability reporting; counter-accounting; continuous auditing; job titles; occupational gender segregation

I. Introduction

Over the last few decades, social and environmental challenges have led to a push in organizations' sustainability-related activities. With regard to this topic, we extend our previous work presented in [1]. The consensus regarding organizations' shared responsibility to act in accordance with sustainable purposes for all stakeholders resulted in increased regulations and policies [2] and consequently, also growth in organizations' sustainability reporting. For instance, in 2020, 92% of the S&P 500 companies published sustainability reports or disclosures as opposed to only 20% in 2011 [3]. Such reports are an important part of the conversation between organizations and their stakeholders.

Simultaneously, parallel to the increased sustainability communication, skepticism has grown toward the authenticity of such reporting [4]. A large body of research addresses how some organizations use such sustainability disclosures to "greenwash" in order to develop a more sustainable image and "window dress" corporate behavior [5][6]. Hence, greater emphasis has been placed on ensuring the reliability of corporate sustainability reporting. The growing awareness of the (lack of) reliability of such reports has, for instance, translated into the implementation of a standardized reporting framework with principles to define the content and the quality of these reports, as proposed by the Global Reporting Initiative (GRI) [7].

Despite this development, there still remains tension between corporate sustainability discourse and practice [5]. To address the critique regarding the credibility and reliability of sustainability reporting and to restore confidence in such disclosures, auditors and assurance providers are introduced to verify the statements made by the reporting organizations [7]. However, given the questioned honesty of corporate disclosures, it becomes increasingly more worthwhile to explore other accounts of organizational activities such as "counter-accounts", that are, contrary to voluntary published corporate reports, outside the control of the organization that is the subject of the account [8]. Counter-accounting through media such as the internet and social media contributes to verifying the organization's legitimacy as it provides an alternative representation of an organization with the aim to rectify otherwise harmful or undesired practices [6][8]. As previous research suggests, the use of counter-accounts should be further explored [6]. Specifically, to systematically include such counter-accounts when challenging organizations' operations, auditors are in need of an appropriate toolbox existing of a continuous monitoring system [9].

Whereas a continuous monitoring system would be useful to review organizations' disclosures with regard to each and every one of the United Nations' 17 Sustainable Development Goals (SDGs), this paper proposes a continuous monitoring system that addresses the fifth goal, gender equality. Specifically, an application of the system is provided for the hospitality industry, providing insights into occupational gender segregation by mapping the differences in the job titles fulfilled by male and female executives. The research question addressed in this study is the following: "How can a counter-account monitoring system for gender equality in the hospitality industry be designed?".

The remaining part of the paper is structured as follows. Section II provides a literature review on counter-accounts, occupational gender segregation, and available monitoring systems. In Section III, a description of the research method is presented. Section IV gives insight into the data collection and Section V describes the data analysis procedure. Section VI shows the results from the job title labels, the system architecture, the usefulness evaluation, and the gender quality results across job titles. Lastly, Section VII concludes the paper.

II. Literature

In 2015, all United Nations Member States endorsed the 2030 Agenda for Sustainable Development, a roadmap for peace and prosperity for both people and the planet, with 17 SDGs at its core [10]. After having defined these integrated, universal goals for sustainable development, the next step toward achieving them was to set specific targets for each goal, which were then in turn further broken down into measurable indicators. However, the incompleteness of the indicators, even after more than three years into the program, make tracking the progress toward meeting the SDGs challenging [11].

To assess whether the actions taken by countries and organizations to reach the SDGs are effective and in correspondence with their own reporting, scholars have argued the need to explore new Information, Communication, and Technology (ICT) in combination with multiple data sources to provide a common, continuous, and transparent representation of their efforts [11].

A. Counter-accounts

As sustainability becomes an indispensable topic on corporate agendas, growing skepticism toward the authenticity of organizations' sustainability reporting arises. This calls for effective monitoring and auditing in this environment in order to ensure the trust and credibility of the information contained in such reports.

With the acceleration of available, real-time information flows, the "archival audit", where the auditor evaluates organizations' yearly reports, is complemented if not replaced by a more real-time evaluation called "continuous auditing" [12]. Together with continuous monitoring, which is described as an ongoing management process to monitor internal controls, continuous auditing aims to provide the organization with a reasonable level of objective assurance [13].

Aside from introducing a continuous monitoring mechanism to provide assurance on these reports and the organizations behind them, scholars have argued the need to explore other accounts, or "counter-accounts", that are outside the control of the respective organization [8]. Counter-accounts are defined as accountings that challenge the representation established by the subject organization and contribute to critically assessing the organization's corporate accountability or lack thereof [6].

B. Occupational gender segregation

Over the past decades, one of the most pressing social issues is inequality [14]. Previous research linked inequality to the emergence of free-market capitalism in which, at a macro-level, organizations and those working for them remain rather invisible. Viewing organizations as rational entities with neutral structures and practices is particularly problematic when addressing inequality as they play a vital role in people's daily lives, such as employment and other opportunities that in turn impact their social and economic status [15][16]. Therefore, equality plays a central role in many of the SDGs.

Inequality is shown in a broad range of forms, however, this paper focuses on gender equality, the fifth United Nations SDG. A recent study commissioned by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs [17], shows not only a difference in the share of employment between working-age men (79%) and working-age women (67%) but also that those women who are employed, are on average paid 14.1% less per hour compared to their male counterparts.

The gender gap, with its key dimension being the gender pay gap, has a considerable impact on individuals' socioeconomic status since gender equality contributes to both economic growth and sustainable development [17]. Evidently, aside from being listed as one of the United Nations SDGs, gender equality is also addressed by the European Commission in the 2020-2025 Gender Equality Strategy, which strives for equal access to the economy across genders. In addition, the European Parliament in 2021 called for a new gender pay gap action plan, addressing women's accessibility to study and work in male-dominated sectors, more flexible work arrangements, and improved wages in female-dominated sectors.

According to Blackburn and Jarman [18], the topic of occupational gender segregation has been at the heart of the gender inequality debate due to its significant role in the gender pay gap and career constraints. Occupational gender segregation refers to a phenomenon in which occupations are stereotyped according to gender. For example, as shown in a study by He et al. [19], women are more represented in occupations that are characterized by high warmth and low competence. For example, statistics show that women in the European Union, in general, tend to be overrepresented in service industries, and professional fields like the arts and humanities, whereas they tend to be excluded from fields like science, technology, engineering, and mathematics [17]. These occupational stereotypes have a widespread effect on how workers are distributed across different jobs. For example, with respect to people's career choices, they may opt for an occupation whose stereotyped attributes are corresponding with their self-perceptions. Similarly, regarding prejudice and discrimination, preconceived notions about individuals, that are not based on reason or reality, may lead to unjust and differential treatment of certain social groups [19]. As a result, the gendered division of the labor market largely explains the gender pay gap due to women's overrepresentation in lower-paid sectors and women's difficulty accessing other, higher-paid sectors [17].

The presence of occupational gender segregation and its role in gender inequality has been widely addressed [18][19][20], yet there remains little work on the monitoring and auditing process of occupational gender segregation. Therefore, this paper proposes a continuous monitoring system for counter-accounts that allows auditors to map both the current state of gender division across different jobs and their evolution over time.

C. Continuous monitoring systems

Monitoring statements and claims organizations communicate on the one hand, and monitoring counter-accounts related to the statements on the other hand can provide valuable insights and prevent greenwashing and/or brownwashing [8]. Multiple studies have focused on analyzing such statements and their counter-accounts. For example, Perkiss et al. [6] analyze counter-accounts and responses by various groups to challenge Nestlé on its sustainability actions. Although data analysis is conducted during these studies, commonly the research is performed once, and is singular problem-oriented, meaning that a specific study focuses on one organization and/or one problem, and executes the analyses once for the purpose of the study [21]. At the end of the studies, there is no information system in place that continuously monitors organizations' claims and related external data to compare both. Some examples of previous studies in which continuous monitoring systems have been developed concern the analysis of stock prices [22] and political analyses [23]. Also, research has been conducted that focus on extracting the right information from texts to be able to conduct the analyses [22]. Information systems that *"enable independent parties to provide assurance on a subject matter, using a series of reports, issued simultaneously with or a short period of time after, the occurrence of events underlying the subject matter"* are called Continuous Monitoring Systems. In general, such systems are used automatically to monitor internal controls within business processes [22].

However, two changes in the current business environment force organizations to start continuous monitoring of external sources. The first change is that organizations are exposed to increased requirements in terms of regulations and business objectives that require managing and monitoring the entire value chain [24]. Second, organizations more and more have to deal with actors that provide counter-accounts through the monitoring of external sources that provide statements about the organization. The focus of this study is automated continuous monitoring of gender equality across job titles and the challenges that occur. To overcome these challenges a system architecture is proposed and its application is presented. Similar to previous research, we consider Named Entity Recognition (NER) as the basis of our counter-accounting system [22][25].

III. Methodology

The research problem addressed in this study concerns the establishment of an automated, continuous monitoring system for gender equality in job titles. The need for such a system is a consequence of the growing need for organizations to comply with the various laws and standards implemented in the field of sustainability reporting. To be able to continuously monitor gender equality in an automated manner, an information technology artifact must be built. A research methodology that focuses on the development and performance of artifacts, specifically with the intention to improve them, is design science research [26][27]. Therefore, a design science approach is employed to perform this study. Although different authors propose different approaches to executing design science research, consensus exists on the following method to execute such a study: awareness of the problem, suggestion, development, evaluation, and conclusion. As stated by Kuechler et al. [28]: *"the research process frequently iterates between development and evaluation phases rather than flowing in waterfall fashion from one phase into the next."* Design research is therefore a continuous cycle of building and evaluation, making it practically impossible to measure all elements of the designed artifact and/or related theories in one study [27]. Regarding the counter-accounting system, multiple elements need to be built and evaluated. Examples of these components are the job classification component, the job analysis component, and the overall system (architecture). The main task of the counter-accounting system is to show how job titles are distributed across gender. If the system is unable to perform this task it is useless. Therefore, in this study, we perform the first iteration of the building and evaluation cycle. Specifically, we focus on the mutual exclusivity and completeness of the job titles (goal 1), the development of a system that can map and monitor how those job titles are distributed across gender (goal 2), and the use and usefulness of the system in relation to the organizations' current report content (goal 3).

The first goal, '*create a mutually exclusive and complete list of job titles*', requires the development of a framework in which job titles are listed and classified. To achieve this goal, a research approach is needed in which job titles are identified, compared, and standardized. Therefore, an inductive approach is chosen. Through grounded theory-based data collection and analysis [29][30], we search for job titles and their relationship to each other. This study design will strengthen the validity of the results as it enables the possibility to compare and combine existing knowledge with observations from practice [29]. The inductive approach is suitable for this research objective given that an industry-wide standard for job titles is not available or adhered to by hospitality organizations. Therefore, based on the job titles used in the appointment announcements posted on the internet, a standardized list of job titles can be induced. Furthermore, grounded theory-based data collection is selected because, to the knowledge of the authors, most research around job titles focuses on the perspective of job titles mentioned in job advertisements and human resources systems rather than the perspective of job appointment announcements. In total 81 iterations of open coding and axial coding with multiple cycles of ordinal comparison have been conducted. The entire research process is visualized in Figure 1, with the above-described procedure for the first goal depicted in the upper section; 'Goal 1'.

To reach the second goal, '*the development of a system that can map and monitor how those job titles are distributed across gender*', a research technique is needed that can compare job titles that are more characteristic to males or females than others in a collection of job appointments. A technique that is used to [31] "*extract words and phrases that are more characteristic of a category than others in a text corpus*" is Scattertext [32][33][34][35][36]. Based on Scattertext we created an experimental setup (pilot) of the system (architecture) to see if it is possible to create a report that can counter account organizations' claims based on open source.

Although the start of design science research is *"the identification of a challenging problem or opportunity in an interesting application environment,"* [28] which in this case is the need for organizations to comply with the various laws and standards implemented in the field of sustainability reporting, this study aims to complete a third research goal, namely, '*to measure the usefulness of the artifact (Scattertext architecture) in providing answers to these new laws and standards*.' Specifically, to evaluate the use and usefulness of the developed Scattertext system, its results should be compared against the organizations' current reporting content, which it aims to counteract. To accomplish this goal, data needs to be extracted from the hotel chains' annual/integrated/sustainability reports. Therefore, an inductive approach is chosen. Through grounded theory-based data collection and analysis [29], we search for statements regarding workforce equality and their relationship to each other. This data analysis process consisted of two coding cycles. Specifically, the report analysis involved 1 iteration following a cycle of (1) open coding and (2) axial coding. During the open coding cycle, 55 statements regarding organizations' efforts and goals in the field of workforce equality have been identified and coded as such. The axial coding cycle focused on identifying relationships among the statements and forming unique categories within the topic of workforce equality that are addressed by organizations in their reports. As can be seen in Figure 1, following the first research goal, the results of the coding cycles ('GOAL 3') are compared to the results produced by the current version of the system ('GOAL 2') to evaluate the usefulness of the system in relation to the current reporting content.

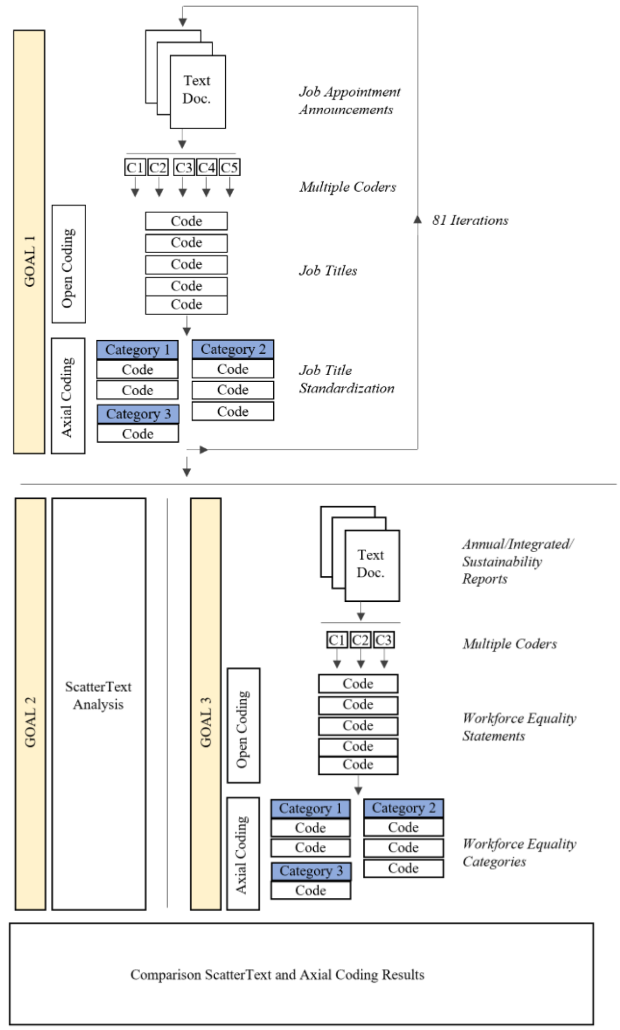


Figure 1. Full Research Process Visualization

IV. DATA COLLECTION

Grounded theory states that the first selection of respondents and documentation is based on the phenomenon studied by a group of individuals, organizations, information technology, or community that best represents this phenomenon [29]. With the first goal of this study being the creation of mutually exclusive, complete, and useful labels for job titles, the currently used job titles must be collected. In our case, we collected job titles from two sources, with the main source being Hospitality.net and the second source being LinkedIn. From Hospitality.net we collected appointment announcements from which the job titles could be derived. The collected data has been posted between the 19th of April 2011 and the 25th of August 2022. The data collection started with a batch collection of 1652 announcements forming the base of the job title dataset. Subsequently, daily new appointment announcements were collected, leading to a total of 22075 announcements.

Design research states that after the artifact is built, it needs to be measured on predefined elements. In this research, the focus is on usefulness. Since usefulness is considered a prerequisite for the intent to adopt a specific system [37]. Usefulness in this study is defined as the degree to which the counter-accounts monitoring system can measure and either confirm or counter-act the statements made by the hotels. Therefore, in this research, documents are needed in which hotel chains make formal statements about gender equality in their hotels. Organizations provide such information in their annual reports, integrated reports, and/or sustainability reports. We have collected a total of 77 reports. The unit of analysis in these documents is every statement made about equality in the workforce.

V. DATA ANALYSIS

Data analysis for the first goal: *'create mutually exclusive, complete, and useful categories'*, was conducted in 81 iterations following two cycles of coding, namely (1) open coding and (2) axial coding with multiple cycles of ordinal comparison. The coding procedure has been completed by a group of five researchers, each having a hospitality education background. During the first cycle, the open coding, job titles were identified and extracted from the collected appointment announcements and coded accordingly. This resulted in a total of 6201 different job titles. Due to the extent of the list, the complete list of job titles has not been added to the paper, however, a snapshot can be found in Table I.

TABLE I. SNAPSHOT JOB TITLE LIST FROM HOSPITALITY.NET ANNOUNCEMENTS

Job Title
CEO
CFO
General manager
Hotel manager
Director
Director of sales
Chef de Cuisine
Senior Vice President
Spa Director

After the open coding was finished, axial coding has been applied. The collection of job titles revealed a wide variety and inconsistency among them. To combat this disparity and allow for comparison, the second cycle of coding, axial coding, is performed to identify more precise categories and relationships within the data. During axial coding, different job titles describing the same 'job' have been coded to represent this. For example, the role of *'Chief executive officer'* is described in appointment announcements as *'CEO'*, *'New company CEO'*, *'Chief executive officer'*, and multiple other variations as can be seen in the example in Table II. The standardized labels were determined using both the hermeneutic skills of the researchers involved and the job description information that described the job content. Whereas the determination of standardized labels was straightforward for many job titles, such as the '*Chief Executive Officer'* presented in the example (Table II), also less obvious axial codings have been encountered. For example, the difference between an area manager, multi-property manager, group manager, and regional manager. In some cases, an area manager had the area Australia whereas others had the area Washington DC. These differences had to be discussed and registered to allow for consistent, mutually exclusive labeling. In these cases, the coding families of grounded theory have been applied to standardize the job titles, specifically, the dimension family type. Moreover, certain job titles were decided to be removed from the dataset. The reason for omitting certain titles was due to them being non-management level jobs such as 'A*ssociate*', '*Secretary*', and '*Mixologist*', award appointments, or appointments of jobs that are unique to certain hotels like the '*Duck Master*' at the Peabody in Memphis [38].

TABLE II. EXAMPLE JOB TITLE STANDARDIZATION

Standardized Job Title	**Job Title**
Chief Executive Officer	CEO
	Interim CEO
	Chief Executive
	Acting CEO
	CO-CEO
	Chief exe. Officer New Company CEO

To create mutually exclusive, complete, and useful labels a procedure had to be adopted among the coders. Specifically, after the first round of open and axial coding, coded parts were discussed among the coders to understand the process and agree on how certain elements had to be coded. Each of the five coders coded all elements and for those elements that were not consistently coded, they discussed their reasoning until they agreed upon a certain label. The unique coding as well as the business rules leading to this coding have been codified in patterns for a rule-based matcher engine. An example of such a pattern included in the matcher engine is '*director of people and culture*' leading to the label '*Director of Human Resources'*: *{"label": "Director of Human Resources", "pattern": [{"LOWER": "director"},{"LOWER": "of"}, {"LOWER": "people"},{"LOWER": "and"},{"LOWER": "culture"}]}*. After this first round, 80 additional rounds of coding have been conducted in which the rule-based matcher engine was used as an additional coder. The process of this worked as follows. First, the rule-based matcher engine coded the new job title, which could lead to two results. The first result would be the matcher engine assigning a standardized job title label, whereas the second result would be labeling the element with a '999' classification. In the first case, the coders would compare the coding of the rule-based matcher engine to what they would have coded the job titles as. In the second case, the rule-based matcher engine can not propose a label for that certain job title, and the coders must label it themselves. After the coders have agreed upon the label, a new rule is added to the pattern file and fed to the matcher engine. After 81 rounds of coding, the matcher engine no longer returned any '999' classifications, meaning the current job titles in the dataset could be successfully labeled to one of the standardized job titles.

With regards to the second research goal: '*the development of a system that can map and monitor how those job titles are distributed across gender*', Figure 3 and Figure 4 show a visualization (ScatterText) of the job title dataset. A Scattertext system shows what words and phrases are more characteristic of a category than others. The X-axis and Y-axis indicate the term frequency in male and female job appointments, respectively. For instance, the upper-left area shows the job titles frequently appointed to women, while the lower-right area shows the frequently appointed job titles to men. Job titles that are presented in the middle are evenly distributed amongst male and female executives. To determine the gender of the person to whom the job title was appointed, the gender indication words in the appointment announcement are considered (*e.g., he/she/him/her)*.

Finally, to evaluate the use and usefulness of the developed system, data needs to be extracted from the hotel chains' annual/integrated/sustainability reports. This data analysis process consisted of two coding cycles. Specifically, the report analysis involved 1 iteration following a cycle of (1) open coding and (2) axial coding. During the open coding cycle, 55 statements regarding organizations' efforts and goals in the field of workforce equality have been identified and coded as such. For example, a statement saying *"We aim to increase representation of people of color in executive positions from 20.5% to 25% by 2025"* has been recoded to '*Racial Equality*', whereas a statement like "*The percentage of female leaders shall be the same as the percentage of female employees*" has been coded to '*Man/Woman Parity*'. A snapshot of the axial coding of the statements can be found in Table III.

VI. RESULTS

The result of the research is discussed in four sections. First, the results of the patterns and labels are discussed. Next, the current system architecture is explained. Following that, the system's usefulness to counteract on the organizations' formal statements is discussed. Lastly, the overall results on the male/female parity across job titles are reviewed.

A. Pattern and Label Results

As described in the Data Analysis Section, in order to create standardized, mutually exclusive job title labels, the business rules defining labels had to be codified in patterns. These patterns include a word or a combination of words that are assigned a predefined label. The 81 rounds of coding resulted in a total of 1266 patterns which were fed to the

TABLE III. AXIAL CODING EXAMPLE OF WORKFORCE EQUALITY STATEMENTS

Label	Statement
Man/Woman Parity	"We aim to increase representation of people of color in executive positions from 20.5% to 25% by 2025."
Racial Equality	"Our talent management team is working with leadership to identify ways to help high-performing Black employees advance within their careers and to broaden our sources of Black recruitment."
Man/Woman Parity	"We are very close to reaching the goal of gender equality in top management roles: 46% female executives."
Racial Equality	"We promote pluralism of origins and seeks diversity through recruitment and career management."
Man/Woman Parity	"At the executive levels, we have achieved 100% gender equity and below our executive levels we have achieved 95% compared to the external marketplace of 80%."
L.G.B.T.Q.I.A.+	"We support LGBT-owned businesses and spent $4+ million with these businesses in 2020."
Equal Compensation	"Ratio of average base salary for women to average base salary for men: 87,5%."
Disability Equality	"We now employ 47 differently abled persons across business, and continues to guide industry initiatives to create inclusive workplaces."
Man/Woman Parity	"Currently, more management roles are held by men (56 percent) than women (44 percent), and more than two-thirds (70 percent) of top management positions are occupied by men."

matcher engine. These patterns included a total of 323 unique job titles. An example of the patterns can be found in Table IV.

Whereas the current pattern list covers the job titles presently occurring in the data, it should be considered that new titles, such as '*vice president sustainability*' or '*vice president happiness*', and new title variations, like the example in Table II, will appear as the data collection continues over time. A process is in place to facilitate such pattern list adjustments. Specifically, when a new title arises, the matching engine will simply return a '999' classification, after which the coders can decide on a label for the job title and add it to the pattern list.

TABLE IV. PATTERN LIST EXAMPLE

Label	**Pattern**
Area Director of Sales and Marketing	area director of marketing
Area Director of Communications	apac media relations
Chairman of the Board of Directors	chairperson
General Manager	hotel manager
Senior Meeting and Events Manager	senior weddings department man.

B. *System Architecture*

Part of the design science research process is the development of a system architecture in which the artifact's structure is presented [39]. Having described the analysis of different system elements, the architecture to ground the counter-accounting system for gender equality as proposed in this study is visualized in Figure 2. The architecture includes three different sections (Collection, Cleaning, and Analysis & Visualization), which are each supported by a different tool (Selenium, Spacy, and ScatterText). The application of each section is discussed.

1) Collection of counter-account data: First, counter-account data, that can counter the statements made by the organization concerning gender equality, need to be identified and collected. For this application, external sources, that publish information on individual management-level appointments, should be consulted. The hotel chains themselves mainly report individual appointments on a board level, e.g., CEO, CFO, and COO as these are mandatory for investor relationships. In their annual reports, the organizations do not report on individual management appointments, therefore, other sources are needed that may serve as counter-account data. Within the hospitality industry, multiple sources, of different types, report on individual management appointments. The first type of

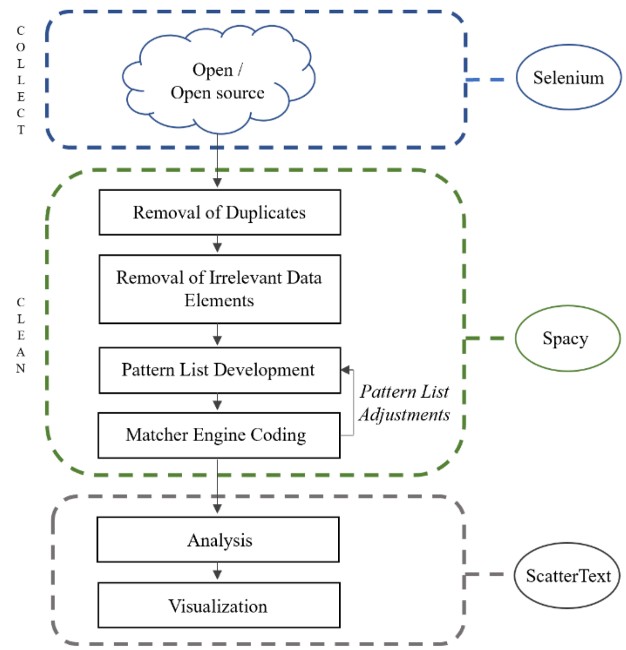


Figure 2. System Architecture: Continuous Monitoring System for Counter-Accounts

sources are websites that collect and publish management appointments such as Hospitality.net [40] and Lodging [41]. Related to this first type are professional network platforms like LinkedIn on which people indicate their current jobs. The third type of website provides global travel industry news and also reports on job appointments such as Hotel Management [42] and Travel News Asia [42][43]. The last type of website is tourism and newspaper websites, for example, Romania Journal [44] and Aruba Today [45]. The appointments per website can be collected by hand or using a variety of tools. For this specific research, Selenium, a tool for automated web scraping is employed to extract useful information from the identified sources, being the appointment announcements on employees' new positions. Important to notice when applying these tools is to respect the privacy policies and the robot.txt files.

2) Cleaning of counter-account data: The second section of the system includes the cleaning of the collected counter-account data, existing of multiple steps. First, duplicate appointment announcements need to be removed. Those occur in the dataset when an appointment announcement is published on multiple sources. For example, one specific appointment [46] was announced at Hospitality.Net, TravelNewsAsia, AsianTravelTips, and Business Today. Consequently, after merging the collected data, those duplicates need to be removed. Aside from removing duplicates, other data-cleaning measures may be taken. For this system, it was decided to omit non-management appointments, award appointments, and appointments of hotel-specific positions. Remaining with unique, management-level appointment announcements, the next cleaning step involves the coding procedure to bring the large number of variations in the job titles back to a manageable, standardized set of job titles. This step is important to allow for the gender comparison analysis across job titles. As described in detail in the Data Analysis Section, the coding procedure involved various rounds of coding resulting in the development of a pattern list. This pattern list is fed to a rule-based matcher engine. To automate this process, a tool called Spacy is used [25]. When the matcher engine returns an unlabeled job title, the pattern list can be adjusted by either adding a new label in the case of a new job title or labeling the title with an existing label in the case of a new title variation. This procedure accommodates the completeness of the job title labels over time. The last step is to remove all privacy-related information.

3) Analysis and Visualization: After having the data collected and cleaned, the following element of the system involves the comparative analysis of job titles fulfilled by males and females. To visually present the counter-account data, the system makes use of ScatterText [32]. The interactive scatter plot allows for comparison between job titles fulfilled by male and female executives. The longitudinal nature of the dataset allows for the evaluation of changes over the past 10 years and the potential identification of trends. The covered period per plot can thus be manually determined, however, for this research the gender comparison was performed per year. The ScatterText visualizations for 2011 (Figure 3) and 2022 (Figure 4) are presented below to review the most apparent changes over the past 10 years.

C. System Usefulness

Having developed the artifact and presented its system architecture, the usefulness of the system should be evaluated. The outcome of the counter-account analysis can be of value to auditors and assurance providers in their assessment of the reporting organizations' credibility and reliability. For this application, this refers to the verification of statements made about gender equality efforts in the hospitality industry. As discussed in previous literature, equality has a broad range of dimensions, of which many relate to one another. To assess the extent to which the counter-account data could counter-act the reported equality statements, the statements are first codified according to their topic. Specifically, for this study, the hotel chains' statements on workforce equality were collected and analyzed following the two folded coding cycles as lined out in the Data Analysis Section. The results show that, within the hotel industry, the currently reported information on workforce equality could be categorized using six different labels as specified in Table V.

Considering the different categories of statements, it can be established that the developed counter-accounting system mainly has counter-acting power regarding the '*Man/Woman Parity*' statements. For example, organizations frequently report figures on man/woman representation across the workforce of which some also specify the man/woman representation on management levels. The developed counter-accounting system could not only provide insights into the man/woman representation but also shed light on what kind of positions are fulfilled by men and women. For example, when an organization proudly reports on achieving gender parity in management positions, the counter-accounting system may show how there are indeed an equal number of appointments to man and woman executives, but, there may remain occupational gender segregation. Specifically, the counter-account system may identify cases where organizations only reach their gender parity goals by promoting women in stereotyped positions, like '*Spa Manager'*, '*Sales and Marketing Manager*', and '*Human Resources Manager*'. In sum, the systems' usefulness currently focuses on the '*Man/Woman Parity*' statements, painting a detailed picture of the real-life gender parity across job titles, however, its usefulness could potentially be expanded to different categories by collecting and analyzing different counter-account data.

TABLE V. CLASS LABELS FOR WORKFORCE EQUALITY STATEMENTS IN THE HOTEL INDUSTRY (2022).

Label
Man/Woman Parity
Racial Equality
Disability Equality
L.G.B.T.Q.I.A.+ Equality
Equal Compensation
Other

D. System Results for Man/Woman Parity

As discussed, the system results could be viewed per organization and per year, however, for this study, the counter-accounting system is applied to the full dataset of appointment announcements in the hospitality industry. The results show an industry overview per year on how job titles are fulfilled by men or women. In this section, the results shown in Figure 3 (2011) and Figure 4 (2022) will be reviewed in more detail.

As can be observed in Figure 3 and Figure 4, with several job titles positioned more towards either the y-axis (female appointments) or the x-axis (male appointments), some of the stereotypes in occupational gender segregation could be identified. For example, in line with arguments in from He et al. [19] regarding women's presence in occupations characterized by high warmth and low competence and evidence showing women tend to be excluded from fields like science, technology, engineering, and mathematics [17], results in Figure 3 show job titles like 'S*ales and Marketing Manager*' and '*Spa Director*' to be more frequently fulfilled

Figure 3. ScatterText Visualization of Gender Segregation on Job Title Level for Hospitality Industry (2011)

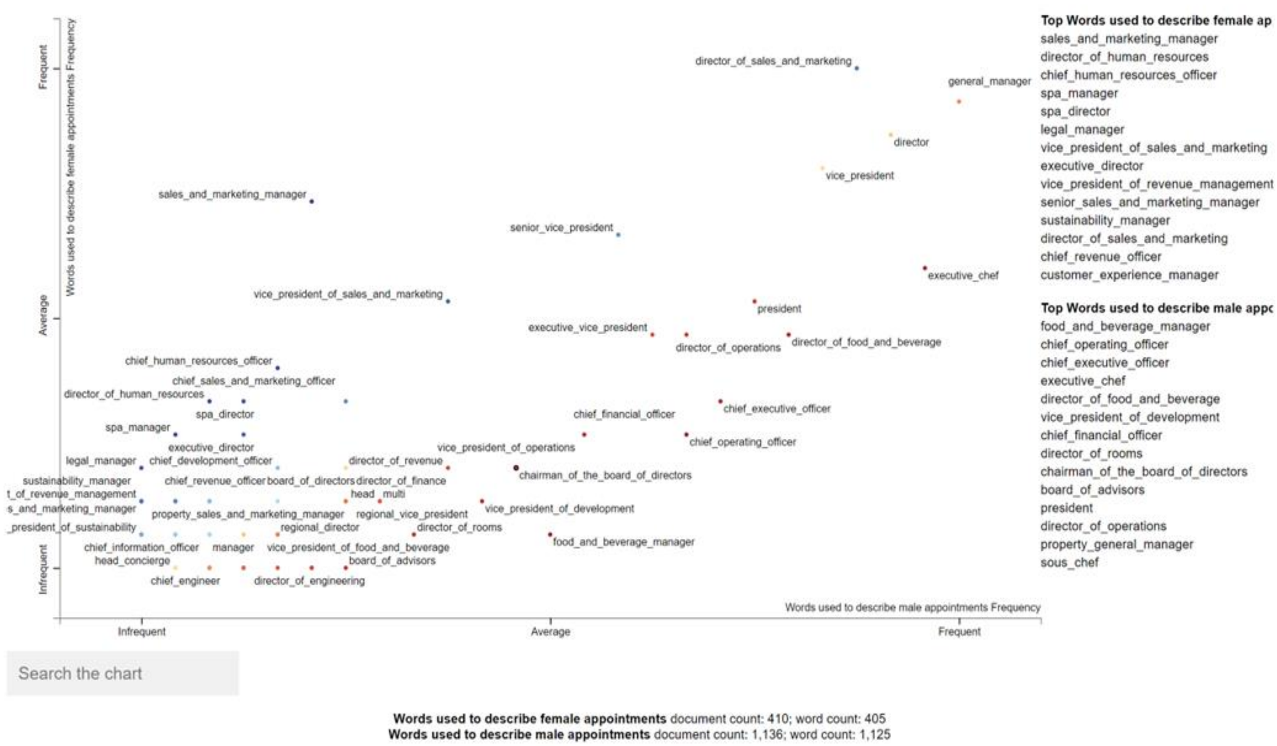


Figure 4. ScatterText Visualization of Gender Segregation on Job Title Level for Hospitality Industry (2022)

by females. On the other hand, job titles like '*Chief Operating Officer*', '*Chief Executive Officer'*, '*Director of Food and Beverage*', '*President*', and '*Chairman of the Board of Directors*' are substantially more often fulfilled by males. These results correspond with some of the preconceived stereotypes of gender division in the hospitality workforce, however, for some job titles, an equal gender division was already established ten years ago. For example, as shown in Figure 3, the position of '*General Manager*' is approximately equally fulfilled by male and female executives.

When reviewing the current gender division in the hospitality workforce in Figure 4, some changes can be observed when comparing the division of titles with those in Figure 3. Some of the stereotypes can still be observed, of which some are even more pronounced. For example, aside from titles like '*Sales and Marketing Manager*' and '*Spa Director*', titles such as '*Director of Human Resources'* became more frequently appointed to female executives. The position of '*Executive Chef*' has been significantly more frequently fulfilled by Male Executives.

Despite the continued stereotyped gender division of some positions, a general trend could be observed, mostly for high-level positions, showing how certain job titles are moving towards a more linear relationship between male and female appointments. Some examples are positions such as '*President'*, '*Vice President*', and '*Chief Operating Officer'*, which are more equally balanced in gender in Figure 4 as opposed to 10 years ago. Ideally, continued monitoring would show a similar trend for all positions, striving for a less gender-segregated workforce.

A final interesting element to consider when reviewing the system results in Figure 4, is the gender division of 'new' positions. For instance, relatively new positions such as '*Vice President of Sustainability'* and '*Chief Sustainability Officer*' do not appear to be appointed to any gender specifically. Mapping the gender division on a continued basis allows for tracking changes in occupational gender segregation within organizations or industries.

VII. CONCLUSION

With the growing awareness of social and environmental challenges as well as the changing regulatory environment, sustainability reporting has become common practice. This development accelerates the need to develop appropriate monitoring tools with the aim to provide a reasonable level of objective assurance on the claims made in such reports. In this study, we aim to answer the following research question: "*How can a counter-account monitoring system for gender equality in the hospitality industry be designed?"* To answer the research question, a system has been designed and tested. The purpose of such a system is to review organizations' gender diversity policies and monitor the progress of their action plan. Our paper offers several theoretical and practical contributions.

The first result of the study consists of the standardization of job titles used in the hospitality industry. In this study, a large array of job titles have been identified. The study has shown that multiple titles are used to indicate the same job. This array of job titles can be confusing to the public, applicants, and to employers. It is necessary to address this vagueness and the overwhelming amount of job titles. The second result is a pilot system architecture for the counter-accounting system on gender equality. The third result is a categorization of workforce equality topics reported upon in hotel chains' integrated, annual, or sustainability reports. The topics of the statements found in the reports could be grouped into six different categories. This categorization could form a foundation for a standardized reporting outline for workforce equality in the hotel industry. The fourth and final result of the study is a view of the industry's gender division in different positions for the past 10 years.

With regards to organizations' gender equality reporting, it can be seen that within their own documentation, organizations (can) choose their own level of granularity on which to report. Most organizations subject to our analysis choose to report on a low level of granularity, for example, the male/female ratio of executives. This gives them the ability to steer and control communication with external stakeholders. Our research has shown that based on the analysis of counter-accounts, a higher level of granularity with respect to reporting can be achieved, which may in turn change the narrative of a specific hotel chain, positively or negatively. We need to challenge not only hotel chains, but organizations in general, to start reporting on a higher level of granularity and provide a complete narrative. In addition, hotel chains may also choose to include counter-accounts in their reporting to provide context to these counter-accountants.

The research has several limitations, both on the subject matter of hospitality and on the actual counter-account system. First, the current use of the counter-accounting system is limited to challenging workforce equality statements made about man/woman parity. Future research should be devoted to extending the current system so that it can also counter-act statements made on other workforce equality topics, such as Racial Equality, Disability Equality, L.G.B.T.Q.I.A.+ Equality, and Equal Compensation. Another direction for future research is to focus on collecting more and different job titles to improve the current system, which is an integrated part of the application. Lastly, future research could focus on additional sectors to expand the results.

REFERENCES

[1] A. M. Y. Heeren, M. M. Zoet, and E. Mantelaers, "Continuous Monitoring of Counter-Accounts in Hospitality," Proceedings of the Fourteenth International Conference on Information, Process, and Knowledge Management (eKNOW 22) IARIA, Porto, Portugal, June 2022, pp. 23-28, ISSN: 2308-4375, ISBN: 978-1-61208-986-7.

[2] C. van Aartsen, R. Bauer, T. Bauer, and M. Olaerts, "Sustainability Embedding in Dutch Listed Companies. VBA Journal, vol. 37, no. 149, pp. 32-38, October 2021, doi: 10.2139/ssrn.3957250.

[3] Governance & Accountability Institute, "Sustainability reporting in focus 2021," 2021, Retrieved: April 2022 from:

https://www.gainstitute.com/fileadmin/ga_institute/images/FlashReports/2021/Russell-1000/G_A-Russell-Report-2021-Final.pdf?vgo_ee=1GwDh4sTgtYj0oCrAM%2BSEBwUnRnlmwiuCIJkd9A7F3A%3D.

[4] T. P. Lyon and A. W. Montgomery, “The Means and End of Greenwash. Organization & Environment, vol. 28, no. 2, pp. 223-249, March 2015, doi: 10.1177/1086026615575332.

[5] C. H. Cho, M. Laine, R. W. Roberts, and M. Rodrigue, “Organized Hypocrisy, Organizational Facades, and Sustainability Reporting,” Accounting, Organizations and Society, vol. 40, no. 40, pp. 78–94, January 2015, doi: 10.1016/j.aos.2014.12.003.

[6] S. Perkiss, C. Bernardi, J. Dumay, and J. Haslam, “A Sticky Chocolate Problem: Impression Management and Counter Accounts in the Shaping of Corporate Image,” Critical Perspectives on Accounting, vol. 81, pp. 102229, December 2021, doi: 10.1016/j.cpa.2020.102229.

[7] O. Boiral, I. Heras-Saizarbitoria, and M. C. Brotherton, “Assessing and Improving the Quality of Sustainability Reports: The Auditors’ Perspective”, Journal of Business Ethics, vol. 155, no. 3, pp. 703-721, March 2019, doi: 10.1007/s10551-017-3516-4.

[8] E. Vinnari and M. Laine, “The Moral Mechanism of Counter Accounts: The Case of Industrial Animal Production,” Accounting, Organizations, and Society, vol. 57, pp. 1-17, February 2017, doi: 10.1016/j.aos.2017.01.002.

[9] J. Kuhn, J. Courtney, and B. Morris, “A Theory of Complex Adaptive Inquiring Organizations: Application to Continuous Assurance of Corporate Financial Information,” The Knowledge Engineering Review, vol. 30, no. 3, pp. 265–296, May 2015, doi: 10.1017/S0269888913000428.

[10] United Nations, “The 2030 Agenda and the Sustainable Development Goals: An opportunity for Latin America and the Caribbean,” January 2018, Retrieved: May 2022, from https://repositorio.cepal.org/handle/11362/40156.

[11] R. Saner, L. Yiu, and M. Nguyen, “Monitoring the SDGs: Digital and Social Technologies to ensure Citizen Participation, Inclusiveness, and Transparency. Development Policy Review, vol. 38, no. 4, pp. 483-500, February 2019, doi: 10.1111/dpr.12433.

[12] M. G. Alles, A. Kogan, and M. A. Vasarhelyi, “Feasibility and Economics of Continuous Assurance,” Auditing, vol. 21, no. 1, pp. 125-138, March 2002, doi: 10.2308/aud.2002.21.1.125.

[13] The Institute of Internal Auditors, “Global technology audit guide continuous auditing: Implications for assurance, monitoring, and risk assessment”, 2005.

[14] A. B. Atkinson, Inequality: What can be done?, Cambridge (Mass.) Harvard University Press, 2015, pp. 384.

[15] J. M. Amis, J. Mair, and K. A. Munir, “The organizational reproduction of inequality,” Academy of Management Annals, vol. 14, no. 1, pp. 195–230, 2020.

[16] A. B. Atkinson, T. Piketty, and E. Saez, “Top Incomes in the Long Run of History,” Journal of Economic Literature, vol. 49, pp. 3–71, March 2011, doi: 10.1257/jel.49.1.3.

[17] European Parliament, “Reporting obligations regarding gender equality and equal pay,” 2021, Retrieved: May 2022 from https://www.europarl.europa.eu/RegData/etudes/STUD/2021/698641/IPOL_STU(2021)698641_EN.pdf.

[18] R. M. Blackburn and J. Jarman, Occupational gender segregation. Social Research Update, vol. 16, Spring 1997.

[19] J. C. He, S. K. Kang, K. Tse, and S. M. Toh, “Stereotypes at Work: Occupational Stereotypes Predict Race and Gender Segregation in the Workforce,” Journal of Vocational Behavior, vol. 115, pp. 103318, December 2019, doi: 10.1016/j.jvb.2019.103318.

[20] P. Valet, “Social Structure and the Paradox of the Contented Female Worker: How Occupational Gender Segregation Biases Justice Perceptions of Wages. Work and Occupations, vol. 45, no. 2, pp. 168-193, January 2018, doi: 10.1177/0730888417753048.

[21] G. Goncharenko, “In the spotlight: Rethinking NGO accountability in the #MeToo era,” Critical Perspectives on Accounting, March 2021, in press.

[22] X. Sibande, R. Gupta, R. Demirer, and E. Bouri, “Investor Sentiment and (Anti) Herding in the Currency Market: Evidence from Twitter Feed Data,” Journal of Behavioral Finance, pp. 1-17, May 2021, doi: 10.1080/15427560.2021.1917579.

[23] A. Bermingham and A. Smeaton, “On Using Twitter to Monitor Political Sentiment and Predict Election Results,” Proc. Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 11), AFNLP Press, November 2011, pp. 2-10.

[24] European Commission, “Proposal for a Directive of the European Parliament and of the Council amending Directive 2013/34/EU, Directive 2004/109/EC, Directive 2006/43/EC and Regulation (EU) No 537/2014, as regards corporate sustainability reporting, COM (2021) 189 final,” 2021b, Retrieved: May 2022 from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0189. Accessed on June 26, 2021.

[25] A. Fantechi, S. Gnesi, S. Livi, and L. Semini, “A SpaCy-based Tool for Extracting Variability from NL Requirements,” Proc. ACM International Systems and Software Product Line (SPLC 21), ACM Press, September 2021, pp. 32-35, doi: 10.1145/3461002.3473074.

[26] A. Hevner and S. Chatterjee, Design research in information systems, 1ste ed. Springer New York, 2010.

[27] A. Hevner, S. March, J. Park, and S. Ram, “Design Science in Information,” MIS Quarterly, vol. 28, no. 1, pp. 75-105, March 2004, doi: 10.2307/25148625.

[28] B. Kuechler and V. Vaishnavi, “On theory development in design science research: anatomy of a research project.” European Journal of Information Systems, vol. 17, no. 5, pp. 489-504, 2017, doi: 10.1057/ejis.2008.40.

[29] J. M. Corbin and A. Strauss, “Grounded Theory Research: Procedures, Canons, and Evaluative Criteria,” Qualitative Sociology, vol. 13, no. 1, pp. 3-21, March 1990, doi: 10.1007/BF00988593.

[30] L. K. Nelson, “Computational Grounded Theory: A Methodological Framework. Sociological Methods & Research, vol. 49, no. 1, pp. 3-42, November 2017, doi: 10.1177/0049124117729703.

[31] F. Jubair, N. A. Salim, O. Al-Karadsheh, Y. Hassona, R. Saifan, and M. Abdel-Majeed, "Sentiment Analysis for Twitter Chatter During the Early Outbreak Period of COVID-19," 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2021, pp. 257-262, doi: 10.1109/ISRITI54043.2021.9702837.

[32] J. S. Kessler, “Scattertext: a browser-based tool for visualizing how corpora differ,” ArXiv, April 2017, unpublished.

[33] M. U. S. Khan, A. Abbas, A. Rehman, and R. Nawaz, "HateClassify: A Service Framework for Hate Speech Identification on Social Media," in IEEE Internet Computing, vol. 25, no. 1, pp. 40-49, 1 Jan.-Feb. 2021, doi: 10.1109/MIC.2020.3037034.

[34] R. Kalaiarasan, T.K. Agrawal, J. Olhager, M. Wiktorsson, and J. B. Hauge, “Supply chain visibility for improving inbound logistics: a design science approach.” International Journal of Production Research, pp. 1-16, 2022, doi: 10.1080/00207543.2022.2099321

[35] S. Yan Eureka Ho and P. Crosthwaite, “Exploring stance in the manifestos of 3 candidates for the Hong Kong Chief Executive election 2017: Combining CDA and corpus-like insights,”

Discourse & Society, vol. 29, no. 6, pp. 629-654, 2018, doi: 10.1177/0957926518802934.

[36] K. Robinson and V. Mago, "Birds of prey: identifying lexical irregularities in spam on Twitter," Wireless Networks, vol. 28, pp. 1189–1196, 2022, doi: 10.1007/s11276-018-01900-9.

[37] N. Prat, I. Comyn-Wattiau, and J. Akoka, "A taxonomy of evaluation methods for Information Systems artifacts," Journal of Management Information Systems, vol. 32, no. 3, pp. 229–267, 2015.

[38] "The Peabody ducks are marching with a new Duckmaster," Hospitality Net, 10-Feb-2012. [Online]. Available: https://www.hospitalitynet.org/news/4054822.html. [Accessed: 18-Aug-2022].

[39] P. Offermann, O. Levina, M. Schönherr, and U. Bub, "Outline of a design science research process," In Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology, pp. 1-11, May 2009.

[40] "Hospitality net," Hospitality Net. [Online]. Available: https://www.hospitalitynet.org/. [Accessed: 06-Sep-2022].

[41] "LODGING," LODGING Magazine. [Online]. Available: https://lodgingmagazine.com/. [Accessed: 06-Sep-2022].

[42] "Hotel Management Homepage: Hotel Management," Hotel Management. [Online]. Available: https://www.hotelmanagement.net/. [Accessed: 15-Oct-2022].

[43] "Travel News Asia" Travel News Asia. [Online]. Available: https://www.travelnewsasia.com/. [Accessed: 15-Oct-2022].

[44] "The Romania Journal," The Romania Journal. [Online]. Available: https://www.romaniajournal.ro/. [Accessed: 06-Sep-2022].

[45] "Aruba today," Aruba Today. [Online]. Available: https://www.arubatoday.com/. [Accessed: 06-Sep-2022].

[46] "Jeff Nicholson has been appointed chief marketing officer (CMO) at Duetto," Hospitality Net, 06-Sep-2022. [Online]. Available: https://www.hospitalitynet.org/appointment/79024749.html. [Accessed: 10-Sep-2022].

Maturity Level Continuum Paradigm of Internal Audit Function

Based on pilots performed during the period 2008 - 2022

Ed Curfs
Future-Proof Auditor
Zuyd University of Applied Sciences
Sittard, the Netherlands
ed.curfs@zuyd.nl

Eric Mantelaers
Future-Proof Auditor
Zuyd University of Applied Services
Sittard, the Netherlands
eric.mantelaers@zuyd.nl

Martijn Zoet
Future-Proof Financial
Zuyd University of Applied Services
Sittard, the Netherlands
martijn.zoet@zuyd.nl

***Abstract*— Continuous (control) monitoring, continuous auditing, continuous assurance, continuous reporting and the Continuum Paradigm are topics in which research has been performed for more than 30 years. The possibilities and challenges of these elements have been researched frequently. However, there are limited studies focused on the holistic view. Several pilots, financial as well as non-financial data, in relationship with the Audit Maturity Model have been performed. Based on the existing studies, we defined the following research question: What is the actual status regarding the Continuum Paradigm? Based on a review of pilots and plotting the results of the pilots on the Audit Maturity Model by a focus group, further insight is provided regarding the actual status. In this paper, we re-address and -present our earlier work, yet we extended the previous research with additional research by adding one pilot and extended the participants to plot the pilots with 13 members of the following three professions (1) internal audit, (2) external audit and (3) business. In addition, we provide a more detailed description of the related literature, findings, and results. The overall conclusion is that the average maturity level for Continuous Monitoring and Continuous Assurance reaches nearly stage 3: 'Maturing and no research has been performed based on a holistic and fully integrated continuous process'.**

Keywords – Audit Maturity Model; Continuous (Control) Monitoring; Continuous Auditing; Continuous Assurance; Continuous Reporting; Continuum Paradigm; Financial and Non-financial Data; Sustainability Reporting.

I. Introduction

Scientific papers in the context of the Continuum Paradigm in nature have often been descriptive in recent years. Common concepts within this research area are Continuous Auditing (CA), Continuous Assurance (CAss), Continuous Monitoring (CM), Continuous Control Monitoring (CCM) and Continuous Reporting (CRe). This whole is now referred to as the Continuum Paradigm [1] [2]. The definition of the Continuum Paradigm is a holistic and pragmatical CA maturity model, which facilitates the assessment of CA capabilities [1] [2].

Only a few studies have been conducted in practice within Continuum Paradigm, combined with results of pilots. What is the actual status regarding the Continuum Paradigm? Based on earlier research, the aim is to provide insight in the actual status.

Maturity models [1] [2] [3] can be helpful to provide insight in the actual status. Maturity models enable organizations to assess their current situation and provide handholds for improving and future research. An example of the maturity model is the Audit Maturity Model (AMM) proposed by, Vasarhelyi, Alles, Kuenkaikaew and Littley [3]. For the purpose of our research, we assessed and plotted the outcomes of a number of selected pilots on a maturity model. Two pilots of 2008 have been selected, research of 2015 has been used as reference and one description of an Internal Audit Function of 2022 has been prepared and added to perform research to identify trends within the Continuum Paradigm.

The current study extends previous research by adding the results from three pilots instead two. The background of the participants has been extended with experts in the area of internal audit, external audit and business. The number of participants has been extended from three to fifteen.

The remainder of this paper is structured as follows. Section two summarizes the results of the literature review. The research method is described in section three. In section four, the data collection is described. The results are described in section five. Section six describes areas for future research. Finally, the paper concludes in section seven with the conclusion.

II. Literature Review

The need for ongoing, timely assurance of data and information utilizing CA, CAss, CM, CCM and CRe is becoming more apparent.

In the last decades, Vasarhelyi, Kuenkaikaew, Alles and Willems [3] performed research in the area of CA, CAss, CM, CCM and CRe. However, this research was mainly related to financial data. Nowadays management needs additional data to provide assurance of non-financial data, driven by regulation, e.g., regarding climate change and the appetite of stakeholders to be timely informed.

Due to these developments the interests for CA, CAss, CM, CCM and CRe grow. When assessing the reliability of data produced by the system, the auditor will review confidentiality, integrity, and availability and how the data is ensured by the system of internal controls. IT and the AMM can be used to allocate the current status of CA, CAss, CM, CCM and CRe. The allocation provides insight in the actual level of auditing. This is relevant information to guide research and further developments of these elements.

There are several ideas of what continuous concepts and systems are, and how they work. Each of the concepts has their own definition. CM is "the process and technology used to detect compliance and risk issues associated with an organization's financial and operational environment" [5]. The financial and operational environment consists of people, processes, and systems, working together to support efficient and effective operations. Controls are put in place to address risks within these components. Through CM of the operations and controls, weak or poorly designed or implemented controls can be corrected or replaced – thus enhancing the organization's operational risk profile. Investors, governments, the public and other stakeholders continue to increase their demands for additional effective corporate governance and business transparency.

The most widely accepted definition for CA is the one released in 1999 by CICA/AICPA and reads as follows: "a methodology for issuing audit reports simultaneously with, or a short period of time after, the occurrence of the relevant events" [6]. The definition for CAss released by Vasarhelyi is therefore "an aggregate of objectively provided assurance services, derived from continuous online management information structures—the objective of which is to improve the accuracy of corporate information processes. These same services may also provide different forms of attestation including point-in-time, evergreen, and continuous" [7]. CAss and CRe are closely linked. There is no CAss without monitoring and intense measuring of the data and data sources.

The AMM classifies the audit evolution into four stages, which are traditional audit, emerging, maturing, and continuous audit. Per stage seven domains have been considered: objective, approach, IT / data access, audit automation, audit and management sharing, management of audit functions, and analytical methods [8].

The first domain is related to a "level of internal audit organization providing financial reports and monitoring internal controls including the task that is undertaken by CA systems". The second domain is related to a "method of audit review, frequency and technique including the extent to which audit outputs shift from being periodic to being undertaken continuously". The third "domain IT / data access" is related to the level and frequency of access to the information system, firm's data systems and general data. The fourth domain "audit automation" is related to the automated level of auditing, usage of technology to assist the audit review cycle. The fifth domain "audit and management" sharing is related to an internal audit department shares systems and resources with management and between the different functions, e.g., finance, legal, tax, compliance, sourcing, production, research and development, sales, distribution and supply chain. There is access to the data and the system is utilized together. The sixth domain "management and audit function" is related to the degree of cooperation between financial audit and IT audit, collaboration with other compliance departments. The seventh domain "analytic methods" is related to the level of technical sophistication of analytical procedure that an internal auditor performs, techniques, and details. The general purpose of maturity model is to provide guidance for a sustainable implementation and growth for organizations [9]. See Table XI: Audit Maturity Model. To improve the readiness of the article, all relevant abbreviations are presented in Table I below.

TABLE I. OVERVIEW RELEVANT ABBREVIATIONS

Abbreviation	*Description*
AMM	Audit Maturity Model
CA	Continuous Auditing
CAss	Continuous Assurance
CM	Continuous Monitoring
CCM	Continuous Control Monitoring
CRe	Continuous Reporting

III. RESEARCH METHOD

The goal of this study is to create an overview of the actual status of the separate elements of the Continuum Paradigm. Maturity models are a well-known instrument to support the improvement of functional domains.

A focus group is a group interview involving a small number of participants who have other common experiences. The focus group should be based on the group of individuals that best represents the phenomenon studied. A focus group existing three professions of (1) internal auditors, with a broad experience in internal auditing, (2) external auditors, with a broad experience in external auditing and (3) business and senior managers with background in auditing and with a broad experience on business rules has been established. For each profession 5 members have been selected and were invited to participate in this research.

The focus group exists of 15 members. Response has been received by 13 members. Only one member was a female. The average years of working experience of the focus group was 24 years. The average years of experience in auditing was 16.9, of which 9.17 in external audit and 7.75 in internal audit.

Before a focus group is conducted, a number of topics need to be addressed: (1) the purpose of the exercise, (2) the selection of the participants, (3) the number of participants, (4) the protocol of the focus group, (5) the AMM model, (6) the protocol for plotting the pilots on the AMM model and (7) useful pilots for research. These have been addressed by a core team of three researchers.

Based on the research performed in the past, there are different AMM models available. First, we needed to select which AMM model could be used as reference model for this study. The AMM as described by Mantelaers & Zoet [2] and the AMM as described by Vasarhelyi [3] have been selected as starting point. Both AMM's have been compared. The AMM of Vasarhelyi has been used intensively in research articles since 1990. This AMM has also been used as reference for one other similar study (Metcash's). For that reason, it has been decided to use the AMM of Vasarhelyi.

The next step was to define what pilots could be used to perform this study. Research articles during the period 1990 until 2021 have been selected by the core research team using the following separate and combinations of the key words: audit, auditing, assurance, combined, control, continuous, data, external, financial, integrated, internal, maturity, model, monitoring, non-financial, pilot(s) and studies. The results have been reviewed by the core research team resulting in seven useful articles. Based on the review of the articles there are limited articles published containing sufficient detailed data and information to make it possible to rank and plot the results in the AMM of Vasarhelyi. The content of the selected articles, the level of detail of the data, level of detail of description of the data collection have been investigated. Based on the defined seven sections of the AMM and the data in the articles, we searched for relationships and references. In case that there were sufficient relationships and reference these articles were selected to rank and plot the results in the AMM.

The core team of three researchers requested a Head of Internal Audit of a South African listed firm to describe the current status of the Internal Audit Function. Adding the current situation of a listed firm provides the option to identify trends regarding the Continuum Paradigm over the period of nearly 15 years.

In the article of Hardy and Laslett, the results of the study have been plotted in the AMM [9]. The aim of this paper is to report on the implementation of CA and CM at Metcash Limited, an Australian wholesale distribution and marketing company (hereafter, Metcash). The results for this organization are so far notable: over 100 fully automated tests performed daily, a fully integrated exception management system, advancement from data to predictive analytics, and the use of visualization technologies for enhanced reporting. The results of this study have been used as reference to compare the results.

Every participant followed the same protocol, each starting with an introduction and explanation of the purpose and procedure of the meeting. After the introduction, ideas were generated, shared, discussed, and refined by the participants. Furthermore, the participants were invited to submit secondary data regarding CA, CAss, CM, CCM and CRe in the AMM.

Based on the pre-work, two pilots and the actual description of the Internal Audit Function of a South African listed firm have been identified by the researchers useful for plotting the results on the AMM. Each participant plotted the pilots individually. The results have been collected and the average results have been calculated for all three pilots. The average results of the three pilots have been compared with the reference pilot of Metcash's. The results have been shared and discussed with the focus group and the core team of three researchers. During this meeting conclusions have been defined and agreed.

The applicable corporate governance code for South African Listed Firms is King IV [12]. In this code the Combined assurance model has been introduced. A combined assurance model incorporates and optimizes all assurance services and functions so that, taken as a whole, these enable an effective control environment, support the integrity of information used for internal decision-making by management, the governing body and its committees, and support the integrity of the organization's external reports. This model has no impact on the internal control system of an organization, as this should be risk based. For that reason, there is no major difference between an internal audit function for a South African listed firm versus a non-South African listed firm.

IV. DATA COLLECTION

Per pilot further information will be provided as well as the reason why the data has (not) been used for further research.

A. *Reference Pilot: Metcash's*

The goal of the Metcash pilot (the subject) was to provide key lessons relating to the adoption and implementation of CA and CM because of its advanced maturity. A maturity assessment of Metcash's CA/CM activities was conducted [3]. The results are mapped onto the four stages and seven dimensions.

This research has been performed during 2015 and the articles provided detailed information how to plot the results of a pilot or study [9]. For that reason, the outcome of this study has been used as reference to provide insight in the gap and deviations with other pilots or studies.

Reference research three Pilots: SAPSECURE, CAMAP and BAGHEERA-S™.

The goal of this study of the three studies (1) SAPSECURE. (2) CAMAP and (2) Bagheera-S™ is to collect evidence from actual implementations for the need of CA and CM [9]. SAPSECURE was developed to permit auditors to review SAP security settings on a regular basis. It may also be used to provide answers to questions such as, ''Who can create a vendor, enter an invoice, and pay it?'' SAPSECURE was implemented and tested in a large public-sector organization. The design of CAMAP is based on surveillance of financial transaction data with the intention of profiling and identifying users that violate Segregation of Duties. This CA/CM solution provides an automated, independent mechanism for monitoring key business processes within an organization. Bagheera-S™ has the capability to report on three key business processes: (1) Payroll / Human Resources, (2) Procurement and (3) Finance. The outcome of the three studies SAPSECURE, CAMAP and BAHEERA-S™ could not be used to plot the results on the AMM as data was missing with regard to approach, IT/Data Access, audit and management sharing and management of the audit function.

B. Pilot: VODAFONE

The American Institute of Certified Public Accountants (AICAP) published a booklet Audit Analytics and Continuous Audit. In this booklet reference is made to the Vodafone Iceland pilot: Implementing Continuous Monitoring [6]. The scope of the project was to implement an IT application called, 'exMon', for revenue assurance as revenue leakage is a known issue in the telecom industry, to decrease the time required to process the financial closing month-end, fraud detection and to enhance the quality of the Customer Relationship Management. The data of the Vodafone Iceland project was limited and for that reason this could not be used to prepare an AMM rating.

C. Pilot: Siemens

One of the main advantages of this pilot was to test how CA would move from concept to implementation. Vasarhelyi predicted both that ERP-enabled firms are the environments most suited to first deploy CA, and that the course of the implementation would begin with automation of existing audit procedures and then, once the feasibility and value added has been demonstrated, move on to re-engineering the audit to make it extra CA ready [4].

At Siemens several hundred procedures regarding Audit Action Sheets are created (these sheets are detailed descriptions of internal audit working papers), which describe in considerable detail what the internal auditor is supposed to test for in each SAP system environment. After examination of 25-30 Audit Action Sheets, twelve were chosen as representative of the challenges on automating and reengineering. The Audit Action Sheets are related to inadequate protection for SAP access. Testing one of the major general IT controls, logical access, which is non-financial data. In Visual Basic a prototype has been developed. The Siemens experience indicates that in environments characterized by highly automated business processes, CA can be defined as a process that continually tests controls based upon criteria prescribed by the internal auditor and identifies exceptions for the internal auditor to perform additional procedures. During this project the CM of internal control settings into the CA concepted model have been achieved. This included the treatment of transactional level data (non-financial).

The adoption of CM of automated business process control settings at the mode of CCM was a novel contribution to the project. This approach could not be utilized systematically in the past because of the extent of automation of business process controls was extremely limited. The current high level of business process automation and specific Robotic Process Automation (RPA) in leading companies such as Siemens makes this feasible and attractive [10]. The ongoing broad advances of business solutions across many industries will support the process as well as the implementation of the Continuum Paradigm.

D. Pilot: HSP

The HSP project is based on modeling processes required data at a highly disaggregate level, far below the level of account balances that are used in the standard audit analytical procedures. Due to fact that there was access to the full richness of the dataset, it was feasible to create the process-based audit models using as benchmarks. Continuity Equations (CE) has been defined by Rogers as stable probabilistic models of highly disaggregated business processes [11]. The CE defined is related to the following strictly enforced business rule of the procurement process: no deliveries are to be accepted without a cross reference to a purchase order. The existence of a deterministic relationship between the counts of purchase orders sent and of the shipments received can be tested. The HSP experience indicates that for CE systems of this level of complexity require powerful statistical techniques, which allow for dynamic set of CEs with multiple time lags and feedback loops. The experience also made clear that CEs to become an essential component in the future CA systems, they will have to be sufficiently easy to implement. This means that generic CE models developed in the laboratory must be generally applicable to different firms and processes.

E. Pilot: South African Listed Firm

The Head of Internal Audit of a South African Listed Firm and three Internal Auditors have been interviewed. Based on these interviews a description has been made of the actual status of the Internal Audit Function by the core team.

The description and data collected has been reviewed by an independent researcher to verify the correctness and completeness of the description.

New IT developments such as RPA, Data Analytics, Process Mining have been taken into account. This provides the option to extend the research and align with the actual status of the Internal Audit Function.

Based on the results provided by the independent researcher the core team of three researchers decided to add this pilot and requested the participants to plot in addition the South African Listed Firm.

The South Africa Listed Firm was founded in 1936. The firm operates globally. There are an Internal Audit Function and an Audit and Risk Committee established.

The combined assurance approach as described in the KING IV corporate governance code [12] has been adopted.

At the Internal Audit department 17 FTEs are employed. The main part of the auditors is based in South Africa. Several auditors are based in Europe and in the United States. The average years of experience of the internal auditor is more than 15 years. There are several specialists, two IT auditors, one sustainability auditor and one forensic auditor member of the Internal Audit team. The other are operational auditors with a background in at least two areas, e.g., finance and IT, production and technical, technics and supply chain, compliance and legal, etc.

IT Controls have been defined for the areas: Business Development, Entity-level Controls, Legal Compliance, Risk Management, Expenditure, Financial Reporting, Fixed

Assets, HR and Payroll, Inventory, Revenue, IT Cobit, IT SAP Basis (around 60 in total), IT ICS Security, IT Cloud Computing, IT Network, IT Local Applications, IT RPA, Treasury, Taxation, Internal Audit Professional Standards, Sustainability People and Sustainability Planet.

In total around 800 internal controls are defined, implemented, and reviewed. These Internal Controls have been appointed to process owners of each operational unit.

With regard to logical access there are several standard SAP reports prepared monthly. These reports are reviewed by management and Internal Audit reviews these reports. Continuous Auditing has not been implemented, however several internal controls are automated monitored (via exception reporting and RPA).

In addition, for each audit engagement is defined what SAP transactions and the related data can be used for data analytics. The Internal Controls to be tested and audited are defined for each audit engagement. Based on the audit charter of Sappi IAD reasonable assurance is provided and stated in the management reports (used for the annual financial statements as well as the integrated reporting and sustainability reporting – global and regional level).

RPA started during 2018. In the main time there are around 20 RPAs designed and implemented. For example, RPA for (1) travel age analysis, (2) exception identification for payroll, (3) exception identification for procurement, e.g., recorded purchase invoices without adequate three-way match.

Based on the experiences RPA is a very powerful tool and improves the quality of the processes. Specific exception reports as these are created on a daily base and followed up by the staff members.

For each Internal Audit engagement, a scope letter and final report is prepared. In the scope letter the operational risks and related internal controls are recorded (and these are to be audited). The financial performance of the entity (mill, sales office) is audited and standard part of the engagement. Every quarter the results are presented to the Audit and Risk Committee (e.g., planning of engagements versus budget, number of audits finding total and per process, audit coverage, costs, education, staffing and special topics).

F. Plotting Pilots Results on AMM

The focus group used the data of the Siemens Project, HSP project and the South African Listed Firm to complete the AMM as the provided data was sufficient to make plotting possible. Three of the five selected internal auditors, five selected external auditors (two are working for the Big-4 audit firm and three for non-Big-4 audit firm) and five selected senior managers provided their results on time, resulting in a response rate of nearly 87 %. The outcome is presented in the Table II, which contains an overview of the results per pilot, the reference pilot and average for the seven sections of the AMM.

One internal auditor was not able to plot the domain 7, analytical methods, for Siemens, and the domains 5 and 6, audit and management sharing and management of the audit function of HSP. Based on the guidance the maturity level 1 has been plotted.

Three external auditors were not able to plot some domains. The domains that could not be plotted per participant were audit automation (domain number 4) for Siemens, objective, approach, IT / data access and audit automation (domain number 1, 2, 3, and 4) for HSP and audit and management sharing and analytical methods for the South African listed firm.

The participants of the business were able to plot all values of the 3 pilots based on the data provided.

Based on the analysis of the description of the pilots and the results plotted, it can be concluded that there is a tendency that the Continuum Paradigm is slightly embedded. Some improvements have been achieved since 2008 in the domains of objectives, management of the audit function and analytical methods.

Siemens is one of the global Leading Internal Audit Departments and is seen within global Internal Audit community as an example how to establish a best-in-class Internal Audit Function. This could be an explanation for the rating deviations between Siemens and HSP.

Table II shows the overall results of the 13 participants per pilot and per domain and the case Metcash of 2015 as reference. Table III represents the results of the internal auditors. The results of the external auditors are presented in Table IV. Table V provides the outcome of the business.

TABLE II. RESULTS PILOTS AND AVERAGE AMM

	Pilot 1: Siemens	*Pilot 2: HSP*	*Pilot 3: South Africa Listed Firm*	*Average results*
	2008	*2008*	*2022*	
Participants	*13*	*13*	*13*	
Objectives	3.3	3.2	3.7	3.4
Approach	3.9	3.0	3.4	3.5
IT/Data access	3.9	3.5	3.7	3.7
Audit automation	4.1	3.1	3.6	3.6
Audit and management sharing	3.6	3.2	3.1	3.3
Management of the audit function	3.7	2.8	4.1	3.3
Analytical methods	3.4	3.4	4.0	3.6
Average per pilot	3.7	3.2	3.7	3.5

TABLE III. RESULTS PILOTS INTERNAL AUDITORS

	Pilot 1: Siemens	*Pilot 2: HSP*	*Pilot 3: South Africa Listed Firm*	*Average results*
Internal Auditors	*2008*	*2008*	*2022*	
Participants	*3*	*3*	*3*	*4*
Objectives	2.3	3.3	3.0	3.0
Approach	2.7	2.3	2.7	2.8
IT/Data access	3.0	3.7	3.0	3.3
Audit automation	3.0	3.3	2.3	3.0
Audit and management sharing	3.3	2.0	2.2	2.8
Management of the audit function	2.0	1.7	3.0	2.3
Analytical methods	2.3	3.0	3.0	3.0
Average per pilot	2.7	2.8	2.7	2.9

TABLE IV. RESULTS PILOTS EXTERNAL AUDITORS

	Pilot 1: Siemens	*Pilot 2: HSP*	*Pilot 3: South Africa Listed Firm*	*Average results*
External Auditors	*2008*	*2008*	*2022*	
Participants	*5*	*5*	*5*	*4*
Objectives	2.3	1.9	2.8	2.6
Approach	3.0	1.9	2.4	2.7
IT/Data access	3.0	2.2	2.4	2.8
Audit automation	2.6	2.0	3.0	2.8
Audit and management sharing	2.3	2.4	2.2	2.6
Management of the audit function	3.2	2.0	3.0	2.7
Analytical methods	2.5	1.7	2.5	2.6
Average per pilot	2.7	2.0	2.6	2.7

TABLE V. RESULTS PILOTS BUSINESS

	Pilot 1: Siemens	*Pilot 2: HSP*	*Pilot 3: South Africa Listed Firm*	*Average results*
Business	*2008*	*2008*	*2022*	
Participants	*5*	*5*	*5*	*4*
Objectives	2.3	1.9	2.1	1.9
Approach	2.4	2.1	2.2	2.0
IT/Data access	2.3	1.9	2.5	2.0
Audit automation	2.9	1.6	2.0	1.8
Audit and management sharing	2.2	2.1	2.0	1.9
Management of the audit function	2.3	2.0	2.5	1.8
Analytical methods	2.2	2.7	2.9	2.3
Average per pilot	2.4	2.0	2.3	1.9

The core team of researchers was also interested if there were major deviations between the three professions internal auditors, external auditors and business. The expectation of the core team of researchers was that the results of the internal auditors will be higher than the rating of external auditors and business. Based on the results and the analysis this seems to be indeed the case.

The results of each category have been collected and the average results per category have been calculated. The results are presented in Table VI. About HSP there is a small gap between the rating of internal auditors versus the external auditors and business.

Tables VII, VIII and IX present a table of the results of the three pilots Siemens, HSP and the South African Listed Firm as plotted per profession and per domain.

TABLE VI. RESULTS 3 PILOTS INTERNAL EXTERNAL BUSINESS

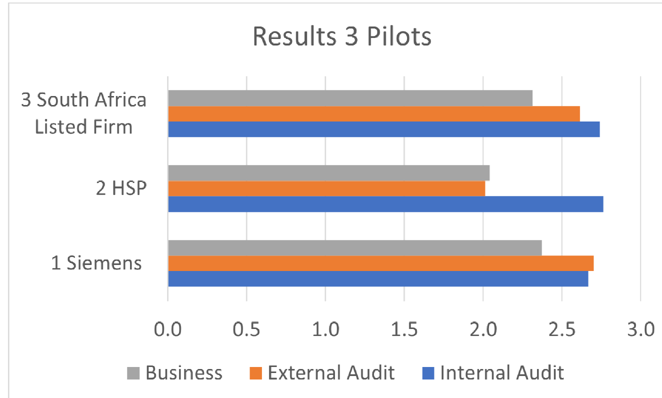


TABLE VII. RESULTS OF SIEMENS 2008

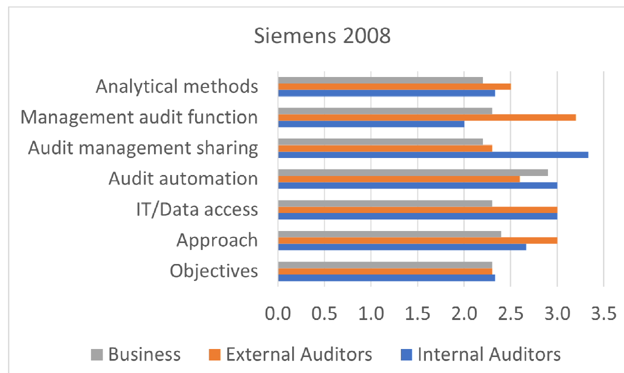


TABLE VIII. RESULTS OF HSP 2008

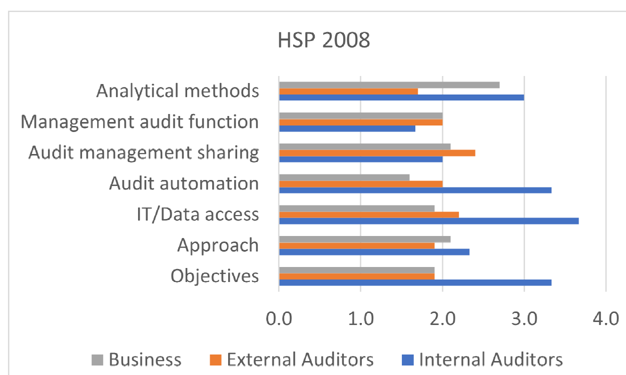


TABLE IX. RESULTS OF SOUTH AFRICAN LISTED FIRM 2022

Table X provides insight in the elements that have been in scope of the pilot. The first column of Table III refers to the main scope of the pilot, Assurance, Audit or Monitoring. The second column of Table III relates to the part of the Continuum Paradigm in scope of the pilot, e.g., CA, CAss, CM, CCM and CRe and in addition related elements such as process data, general IT controls, Internal Control System and Enterprise Risk Management system. The elements process data, general IT controls, Internal Control System and Enterprise Risk Management system are added as these are relevant for the external audit to decide the final level of assurance of the audit. The elements process data and general IT controls are defined further in detail. In case 'In scope' is mentioned the related element was part of the pilot. All the other elements were not part of the pilot and for that reason no research has been performed to the Continuum Paradigm based on a holistic view.

TABLE X. SCOPING PILOTS ELEMENTS CONTINUOUS

Scope	Part	IT or Data	*Results of the Pilots*		
			Pilot 1 Siemens	*Pilot 2 HSP*	*Pilot 3 SA Listed Firm*
Assurance	Cre				
	Cass				
Audit	CA		In scope	In scope	In scope
Monitoring	CM		In scope	In scope	In scope
	Process Data	Financial Data			In scope
		Non-Financial Data	In scope	In scope	In scope
	General IT Controls	Logical Access	In scope		
		Physical Access			
		Back-up & Recovery			
		Change Management			
	Internal Control System		In scope	In scope	In scope
	Enterprise Risk Management System				

V. RESULTS

The overall outcome of plotting three pilots including the reference pilot is that the overall average maturity level is nearly stage 3: Monitoring. The level of stage 4: Continuous Audit has still not been achieved yet.

In total 15 participants have been invited to plot the three pilots in the AMM. Participants from internal audit profession, external audit profession and business with a previous background in auditing have been invited. For each profession 5 participants have been selected. The response rate of the participants was nearly 87 %.

The first conclusion is that the rating of the internal auditors, external auditors and business are close, however in general, the rating of the internal auditors is higher than the one of the external auditors and business. A possible explanation could be that the own profession is less critical and more positive of its own performance and results. The second conclusion is there is tendency that the maturity level of the Internal Audit function improved. This could be explained by the fact that the profession of auditing further developed, further precise and detailed guidance has been implemented and that the IT developments and IT solutions available at the market support the Internal Audit function to implement CM, CCM and CRe.

Regarding the Pilot HSP there is a small gap between the outcome from the internal auditors versus the outcome external auditors and business. The AAM overall results of the pilots Siemens and South African Listed Firm are per domain closer.

As the description of the South African Listed Firm is the most recent one, it includes new IT technology like RPA, data analytics and process mining. This results in a higher rating, however, it seems that CA, CAss, CM, CCM and CRe are still not fully embedded in the Internal Audit function of the South African Listed Firm.

There are limited pilots and studies with sufficient detailed data and information available to plot the results and outcome on the AMM. The analysis of the three projects makes clear that the research on CM and CA is still scarce and in maturation and mainly related to non-financial data.

All pilots used for this experiment are related mainly to CM of non-financial data. CAss and CRe are not part of the selected pilots. No pilots or studies based on actual implementations have been set up to perform research in the fields of CA, CAss, CM, CCM, and CRe as one holistic and fully integrated process. To achieve a successful implementation of the continuous concept it requires an integrations and alignment of all elements of the chain from the start of selection of the data until providing CAss and deliver CRe.

VI. FUTURE RESEARCH

There is an increased need and growing pressure from stakeholders for receiving a continuous flow of assured data and information, specific non-financial data will become relevant. Compliance, new laws and regulations will become applicable requesting assurance services of financial and non-financial data. Some examples are: (1) the development of Corporate Sustainability Reporting, the EU regulation applicable as of 2025 regarding Environmental Social and Governance (ESG), (2) the monitoring of General Data Protection Regulation and current (3) IT developments, e.g., RPA, data mining, data analytics. All these new developments included reporting reliability of financial data as well as non-financial data.

CCM can be applied to achieve insight in the existing level of assurance, to make the internal audit function efficient and possibly, even more effective. Auditors are skeptical about reliability of automated controls. There is a need for general audit and assurance models which fit system-based auditing approaches and a Continuous Integrated Assurance Concept for financial as well as non-financial data.

Further research is needed regarding new IT developments, e.g., blockchain solution, data analytics, RPA, etc. to increase the reliability of the data, financial as well as non-financial data, and the overall level of trust. The outcome could be used to fine-tune the AMM and provide guidance for further development of the Internal Audit function.

New guidance and auditing standards should be defined and implemented in close cooperation and partnership with the developers, producers, users and end users of CA, CAss, CM, CCM and CRe to control and manage the levels of expectation. Further research is needed to get insight in the bottlenecks of why CA, CAss, CM, CCM and CRe has yet not as a fully holistic process been implemented.

VII. Conclusion

The overall outcome was that the average maturity level for CM and CA reaches nearly stage 3: Maturing. The overall average maturity level for CAM and CA increased slightly during the period 2008 and 2022.

No research was performed based on a holistic and fully integrated process including the status and maturity level of the Line of Assurance at the client.

Limited research has been performed on CCM, CM, CA, CAss and CRe of financial data.

A certain minimum level of the Line of Assurance at the client will be required to implement CCM, CM, CA, CAss and CRe. Combined Assurance has not a major impact on the way or level of implementation CCM, CM, CA, CAss and CRe.

The holistic and fully integrated process could be very helpful for organizations accountable for reporting of sustainability, non-financial data across the organizations, e.g., usage of CO^2 in the supply chain process or energy process, wastewater management, employees (own) and contractors Lost Time Injury Frequency Rate (LTIFR). Research could be performed in combination with automated audit standards and assurance standards to achieve CAss.

To meet the upcoming requirements of CCM, CM, CA, CAss and CRe a comprehensive holistic and fully integrated approach would need to address Line of Assurance testing, CCM testing, continuous internal control testing, continuous data testing, continuous transaction testing and continuous assurance testing.

The AMM could support academics by research and the business by further development of new IT concepts and IT solutions.

References

[1] E. Curfs, E. Mantelears, and M. Zoet, "How to Plot Current Pilots on the Audit Maturity Model? The Continuum Paradigm," in Proc. eKNOW 2022, pp. 12-17.

[2] E. Mantelaers and M. Zoet, “Association for Information Systems AIS Electronic Library (AISeL) Continuous Auditing: A Practical Maturity Model,” 2018. [Online]. Available: https://aisel.aisnet.org/mcis2018/40

[3] M. A. Vasarhelyi, M. Alles, S. Kuenkaikaew, and J. Littley, “The acceptance and adoption of continuous auditing by internal auditors: A micro analysis,” International Journal of Accounting Information Systems, vol. 13, no. 3, pp. 267–281, Sep. 2012, doi: 10.1016/j.accinf.2012.06.011.

[4] M. A. Vasarhelyi, S. Kuenkaikaew, J. Littley, and K. Katie Williams, “Continuous Auditing technology adoption in leading internal audit organizations,” 2008. [Online]. Available: https://www.researchgate.net/publication/229004600

[5] G. Locke and P. D. Gallagher, “Guide for Applying the Risk Management Framework to Federal Information Systems,” 2010. [Online]. Available: http://csrc.nist.gov/publications.

[6] AICPA, Audit Analytics and Continuous Audit: Looking Toward the Future. 2015.

[7] M. A. Vasarhelyi, “Chapter 12 Concepts in continuous assurance.”

[8] M. G. Alles, A. Kogan, and M. A. Vasarhelyi, “Putting Continuous Auditing Theory into Practice: Lessons from Two Pilot Implementations,” 2008. [Online]. Available: www.businessfinancemag.com

[9] C. A. Hardy and G. Laslett, “Continuous auditing and monitoring in practice: Lessons from Metcash’s business assurance group,” Journal of Information Systems, vol. 29, no. 2, pp. 183–194, Jun. 2015, doi: 10.2308/isys-50969.

[10] K. Singh, P. J. Best, M. Bojilov, and C. Blunt, “Continuous auditing and continuous monitoring in ERP environments: Case studies of application implementations,” Journal of Information Systems, vol. 28, no. 1, pp. 287–310, 2014, doi: 10.2308/isys-50679.

[11] C. Rogers, “Continuous monitoring of business controls: A pilot implementation of a continuous auditing system at Siemens Discussant’s comments,” International Journal of Accounting Information Systems, vol. 7, no. 2. Elsevier Inc., pp. 167–169, 2006. doi: 10.1016/j.accinf.2006.03.006.

[12] M.E. King, KING IV, Report on Corporate Governance for South Africa 2016, Available: King IV Report on Coproate Governance for South Africa (ymaws.com)

TABLE XI. AUDIT MATURITY MODEL

Satge	*1*	*2*	*3*	*4*
Domain	*Traditional audit*	*Emerging*	*Maturing*	*Continuous audit*
(1) Objectives	- Assurance on the financial reports presented by management	- Effective control monitoring	- Verification of the quality of controls and operational results	- Improvements in the quality of data - Creation of a critical meta-control structure
(2) Approach	- Traditional interim, and year-end audits	- Traditional approach with some key monitoring processes	- Usage of alarms as evidence - Continuous control monitoring	- Audit by exception
(3) IT/Data access	- Case by case basis - Data is captures during the audit process	- Repeating key extractions on cycles	- Systematic monitoring of processes with data capture	- Complete data access - Audit data warehouse. Production, finance, benchmarking and error history
(4) Audit automation	- Manual processes & separate IT audit	- Audit management software - Work paper preparation software	- Automated monitoring module - Alarm and follow up process	- Continuous monitoring and immediate response - Most of audit automated
(5) Audit and management sharing	- Independent and adversarial	- Independent with some core monitoring shared	- Shared systems and resources with natural process synergies	- Purposeful parallel systems and common infrastructures
(6) Management of the audit function	- Financial organization supervises audit and Matrix to Board of Directors	- Some degree of coordination between the areas of risk, auditing and compliance - IT audit works independently	- Internal Audit and IT audit coordinate risk management and share automatic audit processes - Auditing links financial data to operational processes	- Centralized and integrated with risk management, compliance and SOX / layer with external audit
(7) Analytical methods	- Financial ratios	-Financial ratios at sector level / account level	- KPI level monitoring - Structural continuity equations -Monitoring at transaction level	- Corporate models of the main sectors of the business - Early warning system

2D Virtual Learning Environments for Tertiary Education

Gerhard Hube, Kevin Pfeffel, Nicholas H. Müller
Faculty of Economics and Business Administration and
Faculty of Computer Science and Business Information Systems
University of Applied Sciences Würzburg-Schweinfurt
Würzburg, Germany
e-mail: gerhard.hube@fhws.de
e-mail: kevin.pfeffel@fhws.de
e-mail: nicholas.mueller@fhws.de

Abstract — **In the last two years, university teaching has been strongly influenced by online formats, mainly by video conference systems. Beyond that, there are also some practical examples for the use of immersive environments in higher education, mainly focused on the usage of virtual reality (VR) or augmented reality (AR) environments. However, this study aims to see if immersive 2D environments are also holistically suitable for teaching in terms of presence, participation, collaboration and active learning for higher education, as they can offer advantages over video conferencing systems, but are not as costly as VR and AR solutions. A Master's program at the University of Applied Sciences Würzburg-Schweinfurt was chosen for the study. The selected course was held completely in an immersive 2D environment over one semester. Accompanying the course, subjects were asked to complete the Online Learning Environment Survey (OLLES) questionnaire weekly for analysis. In addition, qualitative interviews were conducted with the subjects afterwards. Thereby a descriptive analysis of the questionnaires takes place. All dimensions of the OLLES questionnaire achieve high to very high values. In doing so, the interviews provide insights into the reasons for the ratings. From a purely descriptive point of view, it can therefore be assumed that the used immersive 2D environment is holistically suitable as a learning environment in the tertiary sector.**

Keywords - Virtual Learning Environments; Online Teaching; Tertiary Education; 2D Environments; Desktop virtual reality

I. Introduction

This contribution is based on the IARIA conference contribution "Suitability of Immersive 2D Environments for Tertiary Education using the Gather Environment as an Example" published in June 2022 [1]. University teaching has been heavily influenced by online teaching over the past two years as a result of the COVID 19 pandemic measures. Besides the isolated usage of VR or AR environments [2], primarily the classic video conferencing tools such as "zoom", "GoToWebinar" or "Cisco Webex" were used, according to their market shares [3]. All of the classic video conferencing tools use video and audio transmission in a simple representation of the participants on the screen of the end device. Due to the continuous and long lasting use of these systems, signs of fatigue and weariness could be observed, often referred to as "zoom fatigue" [4] [5] also and especially for university students within online courses [6]. Nevertheless, it can be assumed that online communications and events will continue to some extent after the COVID 19 pandemic [7]. Therefore, alternatives or additions to classic video conferencing systems such as VR should also be analyzed, in order to check their suitability, especially for online university lectures. A first pilot study showed higher spatial and social presence for VR group meetings in comparison with video conference systems [8]. In contrast to video conferencing systems, the representation of the participant in VR is integrated into a virtual world and allows to explore and interact within a dynamic virtual environment [9].

In this introduction, some definitions and explanations of the basic terms are given. These are VR and immersion, for example. Additionally, the status quo of VR in education, as well as virtual learning environments (VLE) will be discussed. Section 2 shows the related works for VR and VLE in higher education and especially in tertiary education. Section 3 presents the virtual learning environment gather.town and their specific software features, which are used in the study and also the measuring instrument OLLES [10] for analyzing the different dimensions. Section 4 resumes the results, which are then discussed in detail in Section 5 with some limitations. Section 6 forms the end of the paper and contains the conclusion with the main results and future studies.

A. VR/immersion

VR can be distinguished between immersive VR (I-VR) including additional devices like a head mounted display (HMD) and non-immersive VR on the screen of some end devices, also declared as desktop VR (D-VR) [11] [12] [13] [14]. Di Natale [15] proposes a tripartition. He differentiates at the poles between non-immersive systems such as desktop VR (D-VR) and immersive systems such as HMD or specially designed rooms with projected walls (CAVE). In between, he places semi-immersive systems such as AR or wide-field-displays. While the definition for VR seems to be clear in literature, the term of immersion is a multifaceted concept without clarification [2]. On the one side, immersion is viewed as a kind of objective characteristics in terms of technical systems and affordances [16] or a psychological subjective characterized by one's perception of presence and interaction [17]. While Bergstrom [18] defines immersion as an objective property of the platform environment and presence as a subjective feeling, it seems that the term of immersion started to become synonymous with "presence" [19]. Despite the strict separation between non-immersive and immersive VR, recent studies tend to consider immersion as a kind of continuum from highly immersive or high-end for I-VR and low immersive or low-end for desktop VR systems (D-VR) [19] [20] [21]. This is probably because there can be some kind of immersion and spatial presence on desktop VR systems as well.

B. Virtual learning environment (VLE)

Another keyword often used in connection with virtual learning is virtual learning environment (VLE). This term includes a wide range of systems like simple web pages, learning management systems like MOODLE but also three-dimensional learning environments like Second Life or OpenSim [22]. Reisoğlu [23], following Zuiker [24], defines the term "3D Virtual Learning Environment" (3DVLE) and describes it as platforms for virtual worlds with avatars as representatives and the ability to communicate via audio or text, such as Second Life or OpenSim. Other authors use the term of "immersive 3D virtual world" or "immersive 3D virtual environment" for similar systems to describe computer based simulated environments, in which users are able to immerse themselves through avatars [25] [26]. We will follow the wording of "immersive 3D/2D virtual environment" to describe desktop VR with different levels of immersion. If 3D-like representations are used in the desktop environment, we assign them to an immersive 3D desktop environment and, in the case of a two-dimensional representation, to an immersive 2D desktop environment. Within this paper we do not include learning management platforms (LMS) for distribution of contents, messages, notices and communication via forums and chats, like e.g., Moodle although they are included in the term of virtual learning environment (VLE) [27] [28]. We want to focus on low immersion desktop solutions that provide the ability to move, interact, collaborate and communicate in a kind of virtual environment using an avatar. The aim is to use them for online master lectures at universities.

In this paper, the related works are presented below. This is followed by the method section, in which the learning environment used is presented in detail. In addition, the measurement instruments used are explained. Afterwards the results are given. This is followed by a discussion of the results, the limitations of the study and finally the conclusion.

II. Related Work

In the following sections, the state-of-the-art, several studies on VLE in specific topics are discussed. But there is a research gap regarding the basic suitability of such virtual environments in higher education. In particular, usage in the tertiary education sector for the implementation of regular courses, and not just for individual specific and short learning units, does not seem to have been sufficiently analyzed.

A. VR in education

The high-end immersive VR seems to fascinate and inspire people in their first reaction, probably because of the high level of immersion and appearance [29] [30]. Especially in terms of education, there were several announcements about groundbreaking improvements by the usage of immersive VR, like increasing memory capacity or making better decisions [31]. Wu et al. [32] reported that I-VR-lectures are more effective than non-immersive environments and Gao [33] assumes better learning outcomes because I-VR is more engaging than traditional methods. A meta-study found that the majority of studies on immersive learning environments from 2014-2019 used AR or VR applications, although all forms of immersion in learning and education were explicitly included. Among other things, the study shows the need for more research on less immersive learning environments with higher narrative and greater challenge [2]. Although the level of immersion in desktop VR systems is not as intense as fully immersive VR technologies, it is not the case that higher immersion and presence directly lead to better learning performance [20]. Johnson-Glenberg [19] discovered that the main effect for better learning is not the level of immersion between 2D or 3D virtual environment but the level of embodiment. The study compared the learning outcomes between groups learning with a low immersion platform on a desktop and a high immersive platform with an HMD (I-VR). The low embodied I-VR group performed significantly worse than the desktop group with high level embodying. Radianti [34] states that immersive VR technologies are particularly used in education, even if their level of maturity still seems questionable and there are several research gaps. Hamilton [14] found in his literature review that in most I-VR studies between 2013 and 2019, there was a significant benefit of using I-VR in education. However, he also restricts that most studies used short interventions and were mainly focused on scientific topics such as biology or physics. Additionally, there are still limitations while using immersive VR. Besides higher costs for immersive VR, above all cyber sickness in terms of e.g., headache, blurred vision or dizziness are effects of using HMD technologies [35]. This is one reason why such systems should be used only for a limited span of time [36]. Due to this and considering the specific requirements and accommodations for university lectures, desktop VR applications appear to be more suitable for online education [10] [21] [37].

B. Immersive VR (I-VR) in higher education

There are several studies on the impact of mainly immersive VR (I-VR) in higher education. Chien et al. [38] stated that a VR environment increases the motivation and critical thinking skills. Tepe [39] concluded that a VR environment increases performance and professional skill development. Other studies also showed several positive effects on the academic success and motivation [40] [41]. Wen-Yu Lee [42] discovered higher scores in science concepts for sixth-grade students learning with I-VR systems in comparison to students without the help of immersive systems. In the field of higher education, a meta-study analyzed studies on desktop-based virtual environments, games and simulations in particular. They concluded that these virtual tools could be effective in improving learning outcomes [13]. Mystakidis et al. [43] conducted a literature review analyzing the outcomes of distant learning and their effect on various criteria of "deep and meaningful learning" such as cognitive, social or affective aspects for K-12 high school students. As a result, positive outcomes were found, especially in terms of performance, satisfaction, cooperation and motivation. Although it is also emphasized that insufficient didactic quality cannot be compensated by online formats. In a metastudy on the effects of immersive VR on students' academic performance, Akgün [44] concluded that there are many positive effects on students' abilities, such as an increase in motivation and other positive contributions to learning. Despite these positive results, the study also determined that there are still technical and health problems to be solved.

C. Virtual learning environment (VLE) in higher education

In addition, studies with desktop VR in higher education detected better performance achieved in groups using desktop VR. However, dependent from the individual spatial ability [45], Reisoğlu [23] analyzed studies between 2000 and 2015 on 3D virtual learning environments (3DVLEs) and various aspects such as platforms used, research topics and achievements. He found that the Second Life platform was the most used platform and that studies on 3DVLEs peaked around 2012 for simulation and learning support. He concluded some overall positive emotional and cognitive achievements on presence, satisfaction, communication skills and engagement. Coffey [26] also analyzed the second life platform against a normal computer surface for comparing the impact on intercultural sensitivity and reveals significant gains with the usage of a virtual environment. Another study analyzed the effects of collaborative learning in virtual environments with the use of 3D avatars in a virtual learning environment (VLE). The results showed that regardless of a collaborative group or an individual group, learning improved, but participation in a collaborative group had a significant positive effect on academic achievement and satisfaction in higher education [46]. In a systematic literature review on "simulation games", it was discovered that better results in terms of declarative knowledge, procedural knowledge and knowledge retention could be achieved through the use of desktop-based immersive environments for the education of trainees [47].

D. VR/VLE in tertiary education:

One of the early publications on "desktop 3D learning environments" without the use of head-up displays in tertiary education comes from Charles Sturt University [48]. Here it is already pointed out that a desktop application is easier for the users and reduces physical and psychological stress compared to immersive virtual worlds with head-mounted displays. A combination of learning management system with Moodle and 3D desktop environment with OpenSim was used in a study to design and evaluate a VLE for teaching with undergraduate students. There were effects on learning skills and understanding of sociocultural aspects that have a strong impact on social interaction when students participate and collaborate in common tasks and activities [30]. Collaboration and interaction seemed to be a high demanded factor influencing VLE systems, either by students as well as academic staff [49]. A special form of 3D virtual learning environment is used for analyzing dental students' performance. When comparing stereoscopic 3D vision with passive circular polarized glasses to 2D vision on screen, significantly better results and higher appreciation for the 3D vision were found [50]. Another specific anatomy medical study about the role of stereopsis in virtual and mixed reality conducted that virtual and mixed reality is inferior to physical models [51].

Overall, there are several studies of desktop VR (D-VR) respectively VLE for specific topics, often computer science or medicine [35] [50] [51] [52]. These studies include various intensities of immersion, but still lack an evaluation of the overall and holistic suitability of 2D desktop learning environments for higher education, including the new immersive 2D environments that have appeared in the last three years.

There are many different forms of virtual learning environments that are used in one way or another. The difference between the individual environments lies especially in the level of immersion. Fig. 1 shows an overview of the different virtual learning environments and their classification on the level of immersion.

III. METHOD

In the following we present the immersive learning environment gather.town, in which the course took place and the measuring instrument OLLES, which was used for the assessment. In addition, qualitative interviews were

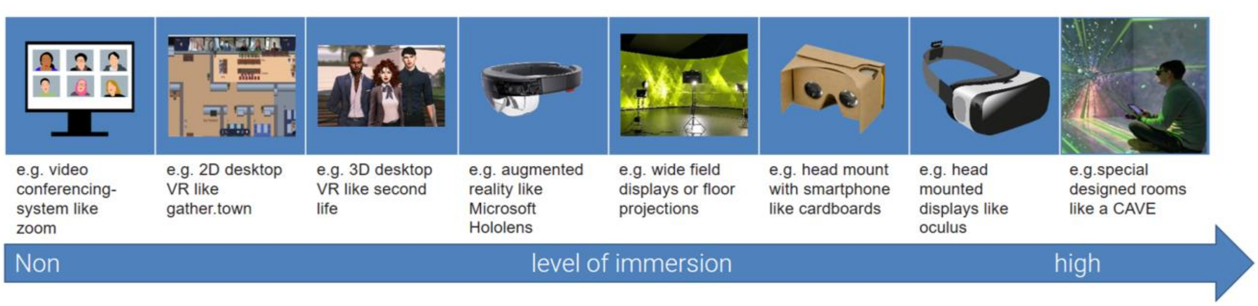


Figure 1. Overview of different virtual learning environments according to the level of immersion.

subsequently conducted with some of the subjects, which will also be presented here.

A. Immersive 2D environment gather.town

The software gather.town [53] was used as an immersive 2D environment. This is a web conferencing software, which allows to create a complete virtual replica of the teaching building. Within this virtual space, users can move around using avatars and interact with each other and their environment, similar to real life. If the avatars now walk around in the virtual environment and then meet each other at a certain distance, the camera and the microphone of the computers are automatically switched on, and the users have the opportunity to communicate. The graphical user interface is quite simple and it does not demand any special requirements to run on a variety of computers. In preparation, the entire real seminar building was recreated in the gather.town environment and the following virtual environment settings and software features were used:

1) Podium:

The podium is the classic teaching situation (see Fig. 2). Within the gather.town environment, all students and the tutor are in one large room. The tutor stands in front at the lectern, while the students take their places at the tables. All students can see, hear and of course communicate with each other via camera and microphone. It is possible to share the screen to provide lecture slides or other content to all participants in the plenum area. In this way, the tutor can use lecture slides in addition to a verbal execution of the learning topic, as they would be used in a real teaching situation.

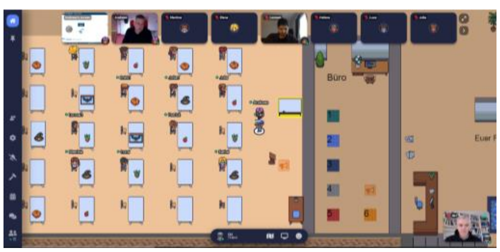


Figure 2. This is the podium. You can see a classic teaching situation in a shared space.

2) Workshops:

Workshops are smaller rooms that provide fewer seats than the large seminar rooms. Here, there are tables with seats and a whiteboard (see Fig. 3). Thus, the users have the possibility to do smaller group work. They can use the table for meetings via the camera, or the whiteboard for joint work or screen sharing for presentation.

3) Whiteboards:

The whiteboard (see Fig. 4) provides an opportunity for collaborative work. To do this, the whiteboard must first be activated. After that, all users who access the whiteboard at the same time can work together on it. This means that all users get write permissions and can interact with the whiteboard. In addition, a video and audio function for communication is available for the workgroup to discuss and exchange while working on the board.

4) Group discussion:

This is a room that is designed in such a way that a pro and a con side can sit opposite each other and participate in a group discussion by means of the camera (see Fig. 5). The whole setting is accompanied by possible viewers but would also be monitored by a jury that rules the discussion and evaluates the individual arguments.

Figure 3. Here you can see a small workshop room with several seats and a whiteboard in the room.

5) Break rooms:

In the break rooms, users can stay between the individual seminars and have the opportunity to play various card games at a game table, making music or watching videos (see Fig. 6). In another break room, users have the opportunity to get on a yoga mat. A 10-minute instructional video is then played so users can join in on the yoga session from home.

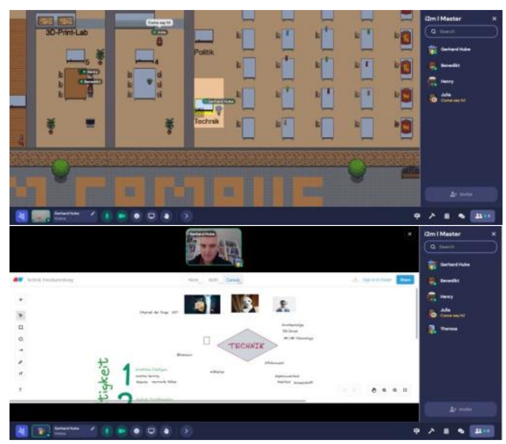


Figure 4. In the upper part of the picture, you can see the whiteboard placed in the room. Below you can see the view when using the interactive whiteboard.

6) Other Interactive Objects:

Within the environment, other interactive objects are stationed in the individual rooms or corridors. In the entrance area, for example, there is a blackboard, on which the timetable can be viewed. Next door, there is a tutorial that once again describes the functionality of the gather.town environment in a video. There is also a bookcase. If you use it, you get a web window within the gather.town environment, which leads you to the online catalog of the university (see Fig. 7). There the literature search can be accomplished.

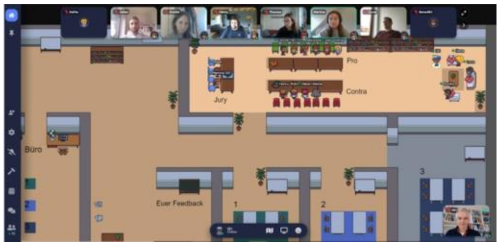


Figure 5. This is a group discussion room, where users sit across from each other in teams and a jury sits in the middle.

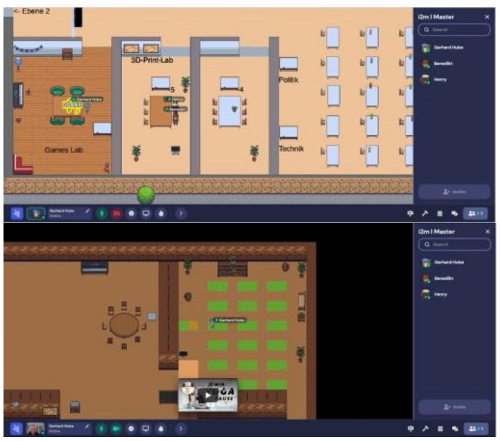


Figure 6. Here you can see the break rooms, where multiple users can gather and share interactive applications like a gaming table or a yoga room where a yoga tutorial is played as a video as soon as you step onto one of the green mats.

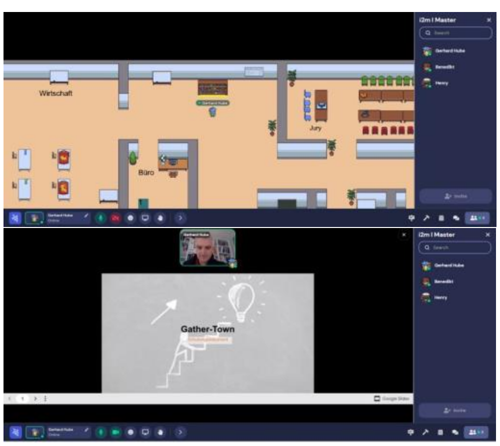


Figure 7. In the upper part of the picture, you can see a bookshelf, which stands freely in the room. Below is the view when you use the bookshelf. This is the online catalog of the university.

B. *Measuring instrument*

The OLLES questionnaire [10] in its modified 35-item form was used as the measurement instrument. The OLLES questionnaire is a web-based survey instrument for use in online learning environments in tertiary education. In this context, the OLLES questionnaire provides inferences about students' perceptions of interaction opportunities within an online environment in terms of economy and efficiency. The dimensions of the OLLES are Student Collaboration (SC), Computer Competence (CC), Active Learning (AL), Tutor Support (TS), Information Design and Appeal (IDA), Material Environment (ME) and Reflective Thinking (RT). In addition, questions about general computer use and Internet use were also recorded. All items were measured using a 5-point Likert scale.

For the qualitative interviews, a separate questionnaire was developed, which can be viewed in full in the appendix. First, an introductory question was asked in order to lead the test persons into the interview situation in a relaxed manner and to check whether they could still remember the seminar well within the virtual learning environment gather.town. Building on this, at least one question was asked about each dimension of the OLLES to develop a deeper understanding of why one of the dimensions had performed well or poorly. In addition, the questions of the questionnaire still investigate whether the subjects prefer face-to-face classes, a virtual learning environment such as gather.town or classic video conferencing software such as Zoom and why this is so. Finally, the questionnaire examines whether the virtual learning environment gather.town was also used outside the actual seminar and, if so, for what other purposes and, questions are asked about the highlights and the low of the software used.

C. *Experimental procedure*

Even before the first seminar, all test persons were familiarized with the gather.town environment. Especially the basic functions were tested, so that everybody knows them and can use them independently. In addition, the OLLES questionnaire was introduced, since this was used in its original language English, but the test persons were not native English speakers.

There were a total of four measurement time points. The seminar duration was always from 8:15 am to 13:15 pm. From the start of the test, the seminar was always first held in the gather.town environment and at all four measurement times the entire questionnaire was completed online directly afterwards.

The qualitative interviews were collected with a time lag after the actual seminar, but they were conducted within gather.town. An appointment was made with a respondent within gather.town, where the interview was conducted and the audio track was recorded. The audio track was then transcribed, analyzed and interpreted.

D. *Sample*

All data were collected at the University of Applied Sciences Würzburg-Schweinfurt within the seminar “trend analysis and innovation assessment” of the master study program "Innovation for small and medium Enterprises". A total of 17 subjects participated in the study. However, there were not measured values from all subjects at all four measurement time points. From two subjects there were only three measured values and from four subjects there were only two measured values. This is still sufficient to form an arithmetic mean. Nevertheless, one subject was excluded from the final analysis because he produced outlier values on three dimensions. This leaves n = 16 valid subjects for the final analysis. The average age of the subjects is 24.44 years, with a minimum of 22 years and a maximum of 30 years. Of the n = 16 subjects, 7 are female and 9 are male.

Five randomly selected subjects were used for the qualitative interviews. Afterwards, it was checked to what extent the answers of the subjects overlapped or whether new insights could still be gained with further surveys, but a feeling of saturation set in. Therefore, n = 5 interviews were considered sufficient. Of the n = 5 subjects, 2 are female and 3 are male.

IV. Results

The results section is divided into different areas. First, there is a purely descriptive part, in which the mean values of the OLLES questionnaire are considered. After that there is a statistical part, where the Wilcoxon signed-ranked test was used to find out if there are differences between different measuring times. Finally follows the part, in which the results of the qualitative interview are reported.

A. Analysis of the OLLES Questionnaire

In the case of computer use, it was found that all subjects use their computers daily or at least several times a week. In the case of Internet use, it was found that all subjects used the Internet on a daily basis.

When tested for normal distribution with respect to the dimensions of the OLLES, Student Collaboration (SC), Computer Competence (CC), Active Learning (AL), Tutor Support (TS), Information Design and Appeal (IDA), Material Environment (ME) and Reflective Thinking (RT), all were found to be normally distributed. Those descriptive values can be seen in Tab. 1.

Then, the Wilcoxon signed-ranked test was used to examine whether there were differences between the individual measurement points and thus whether there was a change in the evaluation with regard to the repetition of the use of the gather.town environment.

Since a normal distribution could not be determined for all variables, even after the exclusion of six subjects with partly missing values, the Wilcoxon test was used. Here, all requirements were met.

There were only significant differences between measurement time point 3 and measurement time point 4 for the dimensions Student Collaboration (Exact Wilcoxon Test: $z = -2.09$, $p = .037$, $n = 12$) and Material Environment (Exact Wilcoxon Test: $z = -2.41$, $p = .016$, $n = 12$). Otherwise, there were no other significant differences between measurement time points.

B. Analysis of the qualitative Interviews

A complete overview of the guideline interview can be found in the appendix and can be referred to for better understanding. Question 1 revealed that all subjects could still remember the seminar and the use of gather.town well to very well. Question 2 revealed that the majority perceived collaboration within gather.town as practical, fun, relaxed and easy. Group work in particular was rated very positively. The whiteboard function, on the other hand, was sometimes perceived negatively, as it did not always function correctly from a technical point of view. This was also confirmed by the query. All subjects found that there were sufficient opportunities for successful collaboration, although, as already mentioned, the whiteboard was sometimes replaced by external software in the form of Miro. Question 3 showed that although there were sometimes technical problems in using gather.town, the use itself was always understandable and simple and therefore it did not represent a technical hurdle. Question 4 showed that subjects found teaching within the gather.town environment motivating. Upon further inquiry, it turned out that this was due in particular to a higher degree of interactivity. For example, simply by moving or controlling one's own avatar. In addition, the virtual learning environment was also perceived as varied and interesting,

TABLE I. OLLES

Descriptive Analysis					
Dimension	***Mean Value***	***Standard Error of the Mean***	***Standard Deviation***	***Minimum Value***	***Maximum Value***
Student Collaboartion (SC)	3.76	0.11	0.42	3.10	4.60
Computer Competence (CC)	4.57	0.11	0.44	3.55	5.00
Active Learning (AL)	3.64	0.13	0.46	2.70	4.60
Tutor Support (TS)	4.10	0.12	0.55	3.20	4.80
Information Design and Appeal (IDA)	3.73	0.12	0.47	2.93	4.80
Material Environment (ME)	3.84	0.07	0.28	3.50	4.45
Reflective Thinking (RT)	3.19	0.16	0.62	2.25	4.10

since it is a diverse world. Question 5 and the related query revealed that the tutor's contact and accessibility was good, and enough opportunities were given for feedback and further questions were answered quickly. Question 6 revealed that the working slides as well as the sources were partly a bit outdated and would need optimization. An inconsistent design was also pointed out. Nevertheless, it was not perceived as particularly negative. On the other hand, the feedback showed that the gather.town environment was initially perceived as taking some getting used to, but after a period of getting used to it was evaluated with positive attributes such as entertaining, appealing and varied. In particular, the real proximity and thus easy navigation, as well as the possibilities for decorating and designing were mentioned positively. Nevertheless, a somewhat unprofessional impression remained. Question 7 showed that subjects rated their learning success within gather.town as good. This was also due in particular to the high level of interactivity, the richness of variety and the motivating aspect. Nevertheless, it was already apparent here that all test subjects prefer face-to-face teaching, but would prefer a virtual learning environment such as gather.town to classic video conferencing software such as Zoom. This was also confirmed in question 8, where all subjects preferred gather.town to Zoom. The most frequently mentioned point was the constant availability, since one could log in 24/7 within gather.town and did not have to send links by e-mail for a meeting. In addition, it was said that the exchange among each other worked better and there were several opportunities to collaborate. In addition, there are aspects like a higher individuality, a small gamification approach, higher activity and better design possibilities. One response should still be highlighted, as one respondent also made the point that the avatars created more closeness to fellow students than simple tiles. That this is a particularly important point was then shown in question 9, where all respondents answered that they prefer classroom teaching. In particular, the proximity to the person sitting next to them, the contact itself, but also the additional body language were cited as reasons. In addition, face-to-face teaching is more interactive, it is easier to work together and there are no connection problems. Question 10 then showed that the gather.town environment was also used by the subjects outside the actual seminar. Mainly for group work of other seminars, but also for private meetings such as vacation planning. The environment was also used for a Christmas party. Finally, question 11 and the two follow-up questions showed that the subjects particularly appreciated the fact that they did not have to register and could get started straight away. They also liked the conversation circle function, where you only took part in a conversation if you were within a certain radius. This gave a real-world feel. The usability beyond the seminar and the design options were also rated very positively. If something was rated as bad, it was mainly technical problems in the form of connection problems and the technical problems with the whiteboard function.

V. Discussion

In the discussion section, the OLLES questionnaire scores are first discussed in relation to the findings from the qualitative interviews. Each dimension of the OLLES questionnaire is analyzed individually. Subsequently, the limitations of this study will be discussed.

A. Overall

Repeated measurement of user ratings of the gather.town environment showed that there was virtually no difference. Although a meta-study by Merchant et al. [13] found small effects in simulation studies in terms of number of sessions, these were measures of learning outcome and not an assessment of the immersive environment as in this study. Therefore, it can be assumed that a one-time survey after the first unit or even after the last unit is quite sufficient.

In the dimensions of computer use and Internet use, the subjects indicated that they use this on a daily basis. In addition, the gather.town environment and all basic functions were sufficiently explained before the start of the study. Thus, we assume that there were no poor ratings for the environment due to possible lack of technical skills.

All dimensions of the OLLES questionnaire reach high to very high scores. From a purely descriptive point of view, it can therefore be assumed that the gather.town environment is holistically suitable as a learning environment in the tertiary sector. Nevertheless, the individual dimensions will be examined below.

B. Student Collaboration (SC)

The Student Collaboration (SC) dimension asks in particular about the frequency of communication between students. This includes the question of help and feedback as well as the mutual exchange of information and resources. As already mentioned, studies have shown that collaboration [43] [46] [49] and communication [23] [49] have positive effects on users within a VLE. Therefore, this is an important factor for learning. It can be assumed that high values were achieved here in the evaluation, since gather.town provides enough possibilities, especially through the functions whiteboard, workshops, group discussion and informal encountering, that this can also be used profitably. This assumption can also be further supported in part by the interview results, since from the subjects' point of view, the simple and fast group work in particular was decisive for good collaboration. This could also be due to the fact that group formation is similar to a face-to-face event and the individual groups can then move individually to their own meeting rooms. Whiteboards in particular, on the other hand, had technical problems more often and thus certainly led to a point deduction in the rating. Nevertheless, it became apparent that there were enough possibilities for the test persons to collaborate successfully.

C. Computer Competence (CC)

The dimension Computer Competence (CC) asks in particular about the assessed competence of one's own computer and Internet use and also the ability to solve minor problems oneself. Since the highest values were achieved here, this further supports the assumption that all subjects had more than sufficient technical skills to use the gather.town environment to its full extent. This was also confirmed by the interviews. Although there were sometimes technical problems with the connection, there were no fundamental problems in understanding how to use it. On the contrary, everything was very easy and intuitive to use.

D. Active Learning (AL)

The Active Learning (AL) dimension specifically asks about the motivation created, as well as the feedback received through the activities or the teaching unit within the environment itself. Again, various studies already showed that motivation [38] [40] [41] [43] [44] is a crucial factor in the use of VLE's. That there was increased motivation was confirmed by the interviews. The motivation arose primarily through increased interactivity. For the test persons, it was clearly more motivating to walk through the virtual environment by moving the avatar and not just to sit in front of the laptop. This also led to the environment being perceived as very varied. It was also mentioned here that a kind of fatigue nevertheless developed over a certain period of time. However, this was not evident in the Wilcoxon signed-ranked test, in which the individual values of the dimensions were compared across the individual measurement time points. Nevertheless, this could have led to a deduction of the score.

E. Tutor Support (TS)

The dimension Tutor Support (TS) asks in particular about the participation and accessibility of the tutor. In this respect, the response time to questions and feedback play an important role. Good communication [23] and interaction [49] lead to positively perceived VLEs. The second highest score was obtained for this dimension. This may be due to constant availability and timely communication, as the tutor himself was also always present and responsive within the environment. Therefore, from this perspective, the gather.town environment is well suited for interactive teaching. This assumption could also be confirmed by the interviews. All subjects felt that the tutor's accessibility was good and sometimes even saw advantages over a face-to-face lecture in the form of direct messages, which thus did not have to be put in front of all seminar participants. In addition, there was sufficient feedback and questions were also answered quickly.

F. Information Design and Appeal (IDA)

The dimension Information Design and Appeal (IDA) asks in particular how creative and original presented teaching materials are and whether graphics used are helpful and visually appealing. This mainly refers to the teaching slides presented as if they were in a presentation. Nevertheless, the colors and walking around within the environment can also have an impact on visual perception and lead to improved learning. In addition, there are the varied break rooms, so that there is also a fairly high rating here. The impression that the subjects evaluated not only the work materials per se, but also the design of the environment per se on this dimension was confirmed by the interviews. The work slides were perceived as outdated and somewhat confusing. The environment, on the other hand, took some getting used to at first, but after using it, the variety, the decoration and the discovery of little things were perceived as nice and fun. Perhaps this double assessment was due to the fact that, in this particular case, it was not always clear to the subjects what the individual question items referred to in this dimension. It could also play a role here that English is not the native language.

G. Material Environment (ME)

The dimension Material Environment (ME) asks in particular about the installation process and clarity in using the software. Since very high values were also achieved here, this further supports the point that all test subjects had more than sufficient technical skills to use the gather.town environment to its full extent. In addition, it can also be assumed that the environment is easy to learn and therefore has a high practical value. In general, it can be assumed that VLEs must be accessible and not have too many hurdles to ensure a successful learning environment. The interviews revealed that the usage was very understandable, simple and intuitive. Only the web interface was used, which is particularly easy to use.

H. Reflective Thinking (RT)

The dimension Reflective Thinking (RT) asks in particular how well subjects were able to learn within the online environment, but also for a comparison to a real classroom. Here, too, high scores were generated, but these are the lowest in comparison with the other dimensions. Nevertheless, it can be deduced that an online environment can be a sufficient substitute due to sufficient positively rated features, but real classrooms are still the most suitable form of teaching. This was also confirmed by the interviews. All subjects were motivated by the virtual learning environment and rated their learning success as good. Especially the interactivity and the richness of variety seemed to be conducive to learning. Nevertheless, all test persons also preferred face-to-face teaching. Probably the most crucial point that a virtual learning environment in the form of gather.town cannot copy is human proximity. The interviews showed that face-to-face teaching is primarily characterized by direct contact and closeness to the person sitting next to you, as well as to all other seminar participants. This again leads to more interactivity, better collaboration and simply closer togetherness among the test persons and thus also to a different feeling. However, since a virtual learning environment such as gather.town is always preferred to classic video conferencing software such as Zoom, it can be said that the closer the learning environment used resembles a real-world experience, the better it is accepted. This is also shown, for example, by the statement that a special highlight was the conversation function. You could walk towards other avatars and as soon as you were within a certain radius, the camera and microphone automatically started and you could start a conversation, while avatars outside the radius could not take part in the conversation. It was said here that this made the environment feel more real.

I. Limitations

This study has some limitations, which are discussed below. The main limitation is that there is not yet a comparison group. The OLLES questionnaire is applied, resulting in a set of scores. These scores only reach their full significance when they are also put into relation. However, this limitation will be addressed by follow-up studies. At the time of publication, a second study had already been started.

This time two courses are running in parallel, with one course using the gather.town environment while the other course takes place in Zoom. Thus, in the follow-up study, the two teaching environments can be compared with each other, but also a comparison of the two gather.town courses can be carried out. This is interesting in that it cannot be assumed that very high scores will again be generated on the OLLES questionnaire. For instance, the scores in this study could be biased due to the effect that the test persons rate new and exciting interfaces better and this effect could wear out. It is also important to keep in mind that this study was conducted in the midst of the Corona pandemic and students had no choice but online teaching. This has changed again and there is also the option of real-world teaching again. This could result in the online teaching being rated significantly lower.

Another limitation is the small number of subjects, but this could not be implemented otherwise due to the small class size.

Furthermore, it should be noted that the qualitative interviews were only collected retrospectively for this study. This meant that there was a period of several months between the last teaching unit in gather.town and the survey of the qualitative interviews. Even though subjects reported that the complete teaching unit was well remembered for them, this remains a limitation. In the follow-up study, all data will be collected directly afterwards.

VI. Conclusion

This study was exploratory in nature with the primary goal of seeing if an immersive 2D environment is holistically suitable for teaching in terms of presence, participation, collaboration and active learning, and thus an enhancement over classic video transmission tools such as “zoom”, “GoToWebinar” or “Cisco Webex” and the like. As the main result of the study, the high scores of the OLLES questionnaire can be mentioned. In connection with the interviews, it can be said that an immersive 2D environment can be used holistically as a form of teaching and also has advantages over classic video transmission tools. As a practical implication, it can be deduced that the use of virtual learning environments in the tertiary sector, on the one hand, can be relatively easily deployed with existing software solutions and, on the other hand, are also well received and therefore offer added value for students. In future online seminars, instructors should therefore think carefully about what kind of learning environment they want to use.

Nevertheless, for now, only an overview of the use of an immersive 2D environment as a learning tool could be provided through this study. Group comparisons with other teaching formats could not yet be made. However, this is the next step in the research. At the time of publication, another survey has already started. Here the same teaching unit is being tested again in gather.town and at the same time another teaching unit is being tested in Zoom. Again, the OLLES questionnaire is used and additionally the Igroup Presence Questionnaire (IPQ). The IPQ is a scale for measuring the sense of presence experienced in a virtual environment (VE). The qualitative interviews will also be used again for data collection. Afterwards a comparison of the two forms of teaching can be made using t-tests.

The interviews further provided new insights. Here, it should be particularly emphasized that face-to-face teaching is always preferred. However, it also became apparent that an enriched virtual learning environment can lead to greater acceptance, more motivation and thus a better learning experience compared to classic video conferencing software. Therefore, in future experiments not only virtual 2D learning environments and classical videoconferencing software will be compared, but also an extension with a 3D virtual learning environment as well as I-VR environments is planned to be able to make a comparisons for this as well. Since it has been found that realism plays an important factor in the evaluation of virtual learning environments, this will also be used to explore, which factors contribute to a higher degree of realism. For example, the change from a 2D learning environment to a 3D learning environment with 3D avatars could be an improvement. This could best be explored by expanding the interview questions by asking more questions that specifically ask about a sense of reality. At the moment, there are many indications that hybrid forms of teaching and learning will be used in the future. However, the goal here should always be to provide the best possible teaching and learning experience for all involved.

References

[1] G. Hube, and K. Pfeffel, “Suitability of Immersive 2D Environments for Tertiary Education using the Gather Environment as an Example” The Fifteenth International Conference on Advances in Computer-Human Interactions (ACHI 2022) IARIA, June 2022, pp 44-52, ISSN: 2308-4138, ISBN: 978-1-61208-982-9

[2] D. Beck, L. Morgado, and P. O. Shea, “Finding the Gaps about Uses of Immersive Learning Environments: A Survey of Surveys” in: Journal of Universal Computer Science, vol. 26, no. 8, pp. 1043-1073, 2020. [Online]. Available from: https://repositorioaberto.uab.pt/bitstream/10400.2/10070/1/jucs_26_08_1043_1073_beck.pdf 2022.12.07

[3] Datanyze, LLC: zoom, top competitors of zoom, [Online]. Available from: https://www.datanyze.com/market-share/web-conferencing--52/zoom-market-share 2022.12.07

[4] G. Fauville, M. Luo, A. C. Muller-Queiroz, J. N. Bailenson, and J. Hancock, (2021) “Nonverbal Mechanisms Predict Zoom Fatigue and Explain Why Women Experience Higher Levels, than Men” SSRN Electronic Journal. [Online]. Available from: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3820035 code4557167.pdf?abstractid=3820035&mirid=1K 2022.12.07

[5] J. Rump and M. Brandt, “Zoom-Fatigue” 2020. [Online]. Available from: https://www.ibe-ludwigshafen.de/wp-content/uploads/2020/09/EN_IBE-Studie-Zoom-Fatigue.pdf 2022.12.07

[6] E. Peper, V. Wilson, M. Martin, E. Rosegard, and R. Harvey, “Avoid Zoom Fatigue, Be Present and Learn” International society for neurofeedback and research ISNR, 2021. [Online]. Available from: https://www.neuroregulation.org/article/view/21206/13976 2022.12.07

[7] S. Lund, A. Madgavkar, J .J. Manyika, S. S. Smit, K. Ellingrud, M. Robinson, and O. Meaney, “The future of work after COVID-19” McKinsey Global Institute, 2021 [Online]. Available from: https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-after-covid-19 2022.12.07

[8] F. Steinicke, N. Lehmann-Willenbrock, and A. L. Meinecke, "A First Pilot Study to Compare Virtual Group Meetings using Video Conferences and (Immersive) virtual reality" In Symposium on Spatial User Interaction (SUI '20). Association for Computing Machinery, New York, NY, USA, Article 19, pp. 1–2, 2020. DOI:https://doi.org/10.1145/3385959.3422699

[9] K. M. Chuah, C. J. Chen, and C. S. Teh, "Incorporating Kansei engineering in instructional design: designing virtual reality based learning environments from a novel, perspective," Themes in Science and Technology Education, 1(1), pp. 37-48, 2008.

[10] J. Clayton, "Development and Validation of an Instrument for Assessing Online Learning Environments in Tertiary Education: The Online Learning Environment Survey (OLLES)," doctoral thesis at Curtin university, 2007, [Online]. Available from: https://espace.curtin.edu.au/handle/20.500.11937/550 2022.12.07

[11] G. G. Robertson, S. K. Card, and J. Mackinlay, "Three views of virtual reality: nonimmersive virtual reality," Computer, 26(2), pp. 81, 1993. [Online]. Available from: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=192002 2022.12.07

[12] C. J. Chen, S. C. Toh, and M. F. Wan, "The theoretical framework for designing desktop virtual reality-based learning environments," Journal of Interactive Learning Research, 15(2), pp. 147-167, 2004.

[13] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis, "Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis," in: Computers & Education 70 pp. 29–40, 2014. [Online]. Available from: https://www.sciencedirect.com/science/article/pii/S0360131513002108 2022.12.07

[14] D. Hamilton, J. McKechnie, E. Edgerton, and C. Wilson, "Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design," in J. Comput. Educ. 8 (1), pp. 1–32, 2021. DOI: 10.1007/s40692-020-00169-2

[15] A. F. Di Natale, C. Repetto, G. Riva, and D. Villani, "Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research," in: Br. J. Educ. Technol. 51 (6), pp. 2006–2033, 2020. DOI: 10.1111/bjet.13030

[16] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," Philosophical Transactions of the Royal Society B: Biological Sciences, Vol. 36, pp. 3549-3557, 2009. https://doi.org/10.1098/rstb.2009.0138

[17] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," Presence: Teleoperators and Virtual Environments, 1998 Vol. 7, No. 3 p. 22, 1998. https://doi.org/10.1162/105474698565686

[18] I. Bergstrom, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater, "The plausibility of a string quartet performance in virtual reality," IEEE Xplore, 23(4), pp. 1352–1359, 2017. https://doi.org/10.1109/TVCG.2017.2657138

[19] M. C. Johnson-Glenberg, H. Bartolomea, and E. Kalina, "Platform is not destiny: Embodied learning effects comparing 2D desktop to 3D virtual reality," STEM experiences Journal of Computer Assisted Learning, 37(5), pp. 1263–1284, 2021. https://doi.org/10.1111/jcal.12567

[20] J. Zhao, T. Sensibaugh, B. Bodenheimer, T. P. McNamara, A. Nazareth, N. Newcombe, M. Minear, and A. Klippel "Desktop versus immersive virtual environments: effects on spatial learning, Spatial Cognition & Computation," 2020. DOI: 10.1080/13875868.2020.1817925

[21] M. N. Selzer, N. F. Gazcon, and M. L. Larrea, "Effects of virtual presence and learning outcome using low-end virtual reality systems," Displays, vol. 59, pp. 9-15, 2019. ISSN 0141-9382, https://doi.org/10.1016/j.displa.2019.04.002

[22] J. Khlaisang and N. Songkram, "Designing a Virtual Learning Environment System for Teaching Twenty-First Century Skills to Higher Education Students in ASEAN," Technology, Knowledge and Learning, 24, pp. 41-63, 2019. https://link.springer.com/article/10.1007/s10758-017-9310-7

[23] I. Reisoğlu, B. Topu, R. Yılmaz, T. K. Yılmaz, and Y. Göktaş, "3D virtual learning environments in education: a meta-review," in: Asia Pacific Educ. Rev. 18 (1), pp. 81–100, 2017. DOI: 10.1007/s12564-016-9467-0

[24] S. J. Zuiker, "Educational virtual environments as a lens for understanding both precise repeatability and specific variation in learning ecologies," British Journal of Educational Technology. 2012. doi:10.1111/j.1467-8535.2011.01266.x

[25] B. Dalgarno, M. J. W. Lee, L.Carlson, S. Gregory, and B. Tynan, "Institutional support for and barriers to the use of 3D immersive virtual worlds in higher education," in G. Williams, P. Statham, N. Brown & B. Cleland (Eds), Changing Demands, Changing Directions. Proceedings ascilite Hobart 2011 pp. 316-330, 2011. http://www.ascilite.org.au/conferences/hobart11/procs/dalgarnofull.pdf

[26] A. J. Coffey, R. Kamhawi,P. Fishwick, and J. Henderson, "The efficacy of an immersive 3D virtual versus 2D web environment in intercultural sensitivity acquisition," in: Education Tech Research Dev 65 (2), pp. 455–479, 2017. DOI: 10.1007/s11423-017-9510-9

[27] P. Alves, L. Miranda, and C. Morais, "The Influence of Virtual Learning Environments in Students' Performance," in: ujer 5 (3), pp. 517–527, 2017 DOI: 10.13189/ujer.2017.050325

[28] L. V. Magalhães and L. M. Li, "Web-Based Undergraduate Medical Education in a Virtual Learning Environment Using an Original Pedagogical Approach: an Observational Longitudinal Study," Revista Brasileira de Educação Médica, 2019. https://www.scielo.br/j/rbem/a/TcFDTxnxzMgrbLyCRPmhnXd/?format=pdf&lang=en

[29] S. S. Sundar, D. J. Tamul, and M. Wu, "Capturing "cool": Measures for assessing coolness of technological products," International Journal of Human-computer Studies, 72(2), pp. 169–180, 2014. doi:10.1016/j.ijhcs.2013.09.008

[30] D. Quesnel and B. E. Riecke, "Awestruck: Natural interaction with virtual reality on eliciting awe. In B. Thomas, R. J. Teather, & M. Marchal (Eds.), 2017 IEEE symposium on 3D user interfaces (3DUI): Proceedings," March 18- 19,2017, Los Angeles, CA, USA pp. 205–206, 2017. Piscataway, NJ: IEEE. doi:10.1109/3DUI.2017.7893343

[31] N. Elmquaddem, "Augmented and virtual reality in education. Myth or reality?" International Journal of Emerging Technologies in Learning, 14(3), pp. 234–242, 2019. https://doi.org/10.3991/ijet.v14i03.9289

[32] Y. Wu, Z. Yuan, D. Zhou, and Y. Cai, "A mobile Chinese calligraphic training system using virtual reality technology," AASRI Procedia, 5, pp. 200-208, 2013.

[33] Y. Gao, V. A. Gonzalez, and T. W. Yiu, "The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review," Computers & Education, 138, pp. 101–115, 2019. https://doi.org/10.1016/j.compedu.2019.05.003

[34] J. Radianti, T. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research," in Computers & Education 147, 2020. [Online]. Available from:

https://www.sciencedirect.com/science/article/pii/S03601315 19303276 08 2020.12.07

[35] T. M. Porcino, E. Clua, D. Trevisan, C. N. Vasconcelos, and L. Valente, "Minimizing cyber sickness in head mounted display systems: Design guidelines and applications," in IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH) pp. 1–6, 2017. doi:10.1109/SeGAH.2017.7939283 .

[36] L. Freina and M. Ott, "A Literature Review on Immersive virtual reality in Education: State Of The Art and Perspectives," in: LE@D - Laboratório de Educação a Distância e Elearning | Artigos em revistas / Papers in journals, [Online]. Available from: https://www.researchgate.net/publication/280566372_A_Literature_Review_on_Immersive_Virtual_Reality_in_Education_State_Of_The_Art_and_Perspectives 2022.12.07

[37] H.-M. Huang, U. Rauch, and S.-S. Liaw, "Investigating learners' attitudes toward virtual reality learning environments: based on a constructivist approach," Computers & Education, 55(3), pp. 1171-1182, 2010.

[38] S. Y. Chien, G. J. Hwang, M. S. Y. Jong, "Effects of peer assessment within the context of spherical video-based virtual reality on EFL students' English-Speaking performance and learning perceptions," Computers & Education, 146, 103751, 2020.

[39] T. Tepe, "Başa Takılan Görüntüleyiciler İçin Geliştirilmiş Sanal Gerçeklik Ortamlarının Öğrenme ve Buradalık Algısı Üzerine Etkilerinin İncelenmesi [Investigating The Effects of virtual reality Environments Developed for HeadMounted Display on Learning and Presence]". (Unpublished Doctoral Thesis). Hacettepe University. Institute of Education Sciences. Ankara, 2019.

[40] M. H. Kim, "Effects of Collaborative Learning in a Virtual Environment on Students' Academic Achievement and Satisfaction," Journal of Digital Convergence, 19(4), pp 1–8, 2021. https://doi.org/10.14400/JDC.2021.19.4.001 .

[41] B. Yildirim, E. Sahin-Topalcengiz, G. Arikan, and S. Timur, "Using virtual reality in the classroom: Reflections of STEM teachers on the use of teaching and learning tools," Journal of Education in Science, Environment and Health (JESEH), 6(3), pp. 231-245, 2020. DOI:10.21891/jeseh.711779

[42] S. W.-Y. Lee, Y.-T. Hsu, and K.-H. Cheng, "Do curious students learn more science in an immersive virtual reality environment? Exploring the impact of advance organizers and epistemic curiosity," Computers & Education, vol. 182, 2022, 104456, ISSN 0360-1315, https://doi.org/10.1016/j.compedu.2022.104456

[43] S. Mystakidis, E. Berki, and J. Valtanen, "Deep and Meaningful E-Learning with Social virtual reality Environments in Higher Education: A Systematic Literature Review," MDPI, Appl. Sci. 2021, 11, 2412. [Online]. Available from: https://doi.org/10.3390/app11052412 2022.12.07

[44] M. Akgün and B. Atıcı, "The Effects of Immersive virtual reality Environments on Students' Academic Achievement: A Meta-analytical and Meta-thematic Study," Participatory Educational Research, 9 (3), pp. 111-131, 2022. DOI: 10.17275/per.22.57.9.3

[45] E. A.-L. Lee and K. W. Wong, "Learning with desktop VR: Low spatial ability learners are more positively affected," Computers & Education, vol.79, pp. 49-58, 2014. ISSN 0360-1315, https://doi.org/10.1016/j.compedu.2014.07.010.

[46] D. Kim and D. Jo, "Effects on Co-Presence of a Virtual Human: A Comparison of Display and Interaction Types," in: Electronics 2022, 11, 367. [Online]. Available from: https://doi.org/10.3390/electronics11030367 2022.12.07

[47] T. Sitzmann, "A meta-analytic examination of the instructional effectiveness of computer-based simulation games," Pers. Psychol. 64, pp. 489–528, 2011. [Online]. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1744-6570.2011.01190.x 2022.12.07

[48] B. Dalgarno and J. Hedberg, "3D learning environments in tertiary education," paper published at 18th Annual Conference of thje Australasian Society for Computers in Learning in Tertiary Education (ASCILITE), Publisher: BMU, University Melbourne, 2001. [Online]. Available from: https://researchoutput.csu.edu.au/ws/portalfiles/portal/9632917/PID%26%2320%3B6456%26%2320%3Bpre-pub.pdf 2022.12.07

[49] R. Charanya and M. Kesavan, "Analysis of Factors Influencing the Virtual Learning Environment in a Sri Lankan Higher Studies Institution," 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE), 2019. DOI: 240-244 https://ieeexplore.ieee.org/document/8842719

[50] I. R. Boer, P. R. Wesselink, and J. M. Vervoorn, "Student performance and appreciation using 3D vs. 2D vision in a virtual learning environment," in: European Journal of Dental Education 20 (3), pp. 142–147, 2016. DOI:10.1111/eje.12152.

[51] B. Wainman, G. Pukas, L. Wolak, S. Mohanraj, J. Lamb, and G. R. Norman, "The Critical Role of Stereopsis in Virtual and Mixed Reality Learning Environments," in: Anatomical Sciences Education 13 (3), pp. 401–412, 2020. DOI:10.1002/ase.1928.

[52] Y.-P. Chao, H.-H. Chuang, L.-J. Hsin, C.-J. Kang, T.-J. Fang, H.-Y. Li, C.-G. Huang, T. Kuo, C. Yang, H.-Y. Shyu, S.-L. Wang, L.-Y. Shyu, and L.-A. Lee, "Using a 360° virtual reality or 2D Video to Learn History Taking and Physical Examination Skills for Undergraduate Medical Students: Pilot Randomized Controlled Trial," JMIR Serious Games. 9. e13124. 10.2196/13124, 2021.

[53] Gather Presence, Inc. (2022). gather.town. [Online]. Available from https://www.gather.town 2022.07.12

APPENDIX - GUIDELINE INTERVIEW

Question 1: You participated in the gather.town study, do you remember the virtual learning environment and the seminar well?

Question 2a: How did you feel about the collaboration within gather.town? Did you enjoy using the individual features like whiteboard, workshops and group discussions?

Question 2b: So there were enough opportunities for successful collaboration? Or were you still missing something specific?

Question 3: From the technical side, were there any ambiguities in using gather.town or was everything understandable from installation to use?

Question 4a: Did the teaching within gather.town motivate or demotivate you?

Question 4b: What do you think led you to be motivated / demotivated?

Question 5a: How did you feel about the tutor's contact/participation and accessibility?

Question 5b: Did you receive enough feedback and were questions also answered quickly?

Question 6a: Were the learning materials well prepared and understandable? This really only refers to the learning materials, i.e., mainly slides and materials or graphics?

Question 6b: Away from the learning materials how appealing or off-putting did you find the gather.town environment?

Question 7: How well or poorly do you rate your learning success within gather.town?

Question 8: You have also used video conferencing software such as Zoom. Which virtual learning environment would you prefer and why?

Question 9: Normally, teaching takes place in presence. Do you prefer face-to-face teaching or gather.town? Please give reasons for your decision.

Question 10: Have you used gather.town outside the actual seminar? And if so, why and for what?

Question 11a: Can you think of anything else you would like to say?

Question 11b: Was there anything that you found particularly good?

Question 11c: Was there anything that you found particularly bad?

Modeling and Sensitivity Analysis of Compressor Stations in Gas Transport Simulations

Anton Baldin
Fraunhofer Institute for Algorithms and Scientific Computing
Sankt Augustin, Germany
email: Anton.Baldin@scai.fraunhofer.de

Kläre Cassirer
Fraunhofer Institute for Algorithms and Scientific Computing
Sankt Augustin, Germany
email: Klaere.Cassirer@scai.fraunhofer.de

Tanja Clees
University of Applied Sciences Bonn-Rhein-Sieg and Fraunhofer Institute for Algorithms and Scientific Computing
Sankt Augustin, Germany
email: Tanja.Clees@scai.fraunhofer.de

Bernhard Klaassen
Fraunhofer Institute for Algorithms and Scientific Computing
Sankt Augustin, Germany
email: Bernhard.Klaassen@scai.fraunhofer.de

Igor Nikitin
Fraunhofer Institute for Algorithms and Scientific Computing
Sankt Augustin, Germany
email: Igor.Nikitin@scai.fraunhofer.de

Lialia Nikitina
Fraunhofer Institute for Algorithms and Scientific Computing
Sankt Augustin, Germany
email: Lialia.Nikitina@scai.fraunhofer.de

Sabine Pott
Fraunhofer Institute for Algorithms and Scientific Computing
Sankt Augustin, Germany
email: Sabine.Pott@scai.fraunhofer.de

***Abstract*—This paper describes the mathematical modeling of compressors used in gas transport networks. Compressors of various types (piston, generic, turbo), different levels of modeling (free, advanced), as well as their combinations into compressor stations (serial, parallel) are considered. Particular attention is paid to the questions of global convergence and stability of the result to the variations of starting point and other parameters of the solution procedure. Sensitivity Analysis and Principal Component Analysis for stationary gas transport problems are considered. A number of numerical experiments on realistic scenarios confirm the conclusions of the theoretical analysis.**

***Index Terms*—simulation and modeling; mathematical and numerical algorithms and methods; advanced applications; gas transport networks; sensitivity analysis; principal component analysis**

I. Introduction

This work is an extension of our conference paper [1], where gas compressors of piston and generic type have been considered. For the sake of completeness, we have added a review of the main results of the paper [2] on turbine compressors. Also, we present a detailed review of [3] results on Sensitivity Analysis (SA) and Principal Component Analysis (PCA) of gas transport simulations and supplement it with new numerical experiments.

In this paper, we will continue the study of globally converging methods for solving stationary network problems on the example of gas transport networks. In our earlier work [4], we introduced the concept of generalized resistivity of network elements and formulated stability conditions for the algorithm solving the corresponding network problems. Under these conditions, for arbitrary variation of the starting point, the solution procedure converges to the same or numerically close result. The approach is universal and can be applied also, e.g., for water transport and electric power networks [5]. In the works [2], [6], [7] we have considered in detail the modeling of gas compressors of the turbine type. For these compressors, individually calibrated characteristics and data resampling on a regular grid were used. In the present paper, we consider compressors of piston and generic type, which are characterized by the existence of analytical solutions and a simpler representation of control equations. This simulation extends our system MYNTS (Multi-phYsics NeTwork Simulator).

Globally convergent method of solution of stationary network problems in applications to electric networks was formulated in [8]. The method has been designed for piecewise linear systems of equations. The space of variables was subdivided to polyhedral cells, where the system has been represented by a non-degenerate linear mapping, with continuous connection on the boundary. The method converges from an arbitrary starting point to a unique solution, in a finite number of steps. Further, in [9], the method has been extended to the linear mappings that can be degenerate in finite cells. In [10], these systems have been represented in min-max or equivalent abs-normal form and several methods for their solution have been discussed. In application to smooth nonlinear systems, the piecewise linear mappings can be used as an approximation. However, the methods above increase the number of steps when the cells become smaller and will perform slowly in practice. The general methods for solution of non-linear systems are described in [11]. In particular, there is a mathematically strict but little-known result, that Newton method equipped with Armijo line search stabilizer provides the global convergence for solution of smooth non-degenerate systems. No methods are known well working for general smooth non-linear systems in the presence of degeneracy.

Modeling of gas transport networks is described in detail in [12], [13]. The real gas networks consist mainly of pipes, and their modeling is based on the nonlinear friction law, simplest by Nikuradse [14], more advanced by Colebrook-White [15] and Hofer [16]. The other parts of the modeling are empirical approximations for the equation of state of a real gas, the simplest by Papay [17], more complex AGA8-DC92 [18] and GERG2008 [19] standards. However, neither pipe nor state equations generally present an obstacle for solving gas transport problems. Rare exceptions are cycles of short pipes where unstable circulations can be excited and regions of phase transitions where jumps and folds of state equation can appear. Otherwise, the pipe and state equations are very smooth and typically solved with several iterations of stabilized Newton method. The problem is presented by compressors and regulators, the elements increasing and decreasing pressure in the network to the desired values. The corresponding element equations make the system piecewise degenerate [4], requiring a development of special methods for its solution. Possible ways for construction of such methods will be discussed in this paper.

In Section II, we recall the general concepts of element resistivity and describe their physical meaning in more detail. In Section III, we will look at compressors of piston, generic and turbine type. In Section IV, we describe Sensitivity Analysis and Principal Component Analysis of gas transport simulations. Section V presents the methods for solving nearly degenerate nonlinear systems. In Section VI, we will carry out a numerical solution of several realistic network problems, presenting the application of the above described methods.

II. Transport variables in stationary network problems

Network problems of a stationary type are described by a system of equations that includes linear Kirchhoff equations of the form $\sum Q_i = 0$, which describe the conservation of flows in network nodes, and equations of elements of the form $f(P_{in}, P_{out}, Q) = 0$, in the general case, nonlinear, introduced on each edge of the network graph. Here the transport variables $P_{in/out}$ are used – nodal variables for the input and output of the element, for gas networks – pressure values, Q – the flow through the element. In gas problems, flows are considered in different normalizations, which is indicated by the index: Q_m – mass flow, Q_ν – molar flow, Q_N – volumetric flow under normal conditions, $Q_{vol,in/out}$ – volumetric flow in input or output conditions (by default, input conditions are taken), etc. An element is called generalized resistive if its equation has derivatives of the following signature:

$$\partial f/\partial P_{in} > 0,\ \partial f/\partial P_{out} < 0,\ \partial f/\partial Q < 0. \quad (1)$$

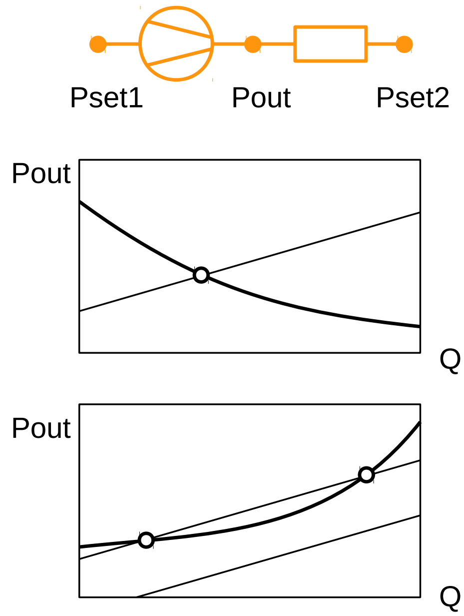


Fig. 1. On the top: a serial connection of compressor (circle) and resistor (rectangle); in the center: decreasing compressor $P_{out}(Q)$ characteristics (thick line) and increasing resistor $P_{out}(Q)$ characteristics (thin line) have a single intersection (stable case); at the bottom: increasing compressor $P_{out}(Q)$ characteristics (thick line) and increasing resistor $P_{out}(Q)$ characteristics (thin lines) can have multiple intersections or no intersection (unstable case). Image from [1].

The work [4] shows that stationary network problems in which all elements have a given signature have a unique solution that can be found by the standard stabilized Newton algorithm with an arbitrary choice of starting point. Technically, it also requires a supply with a set pressure P_{set} in each disconnected component of the graph, as well as a proper condition for the behavior of functions at infinity, which can be satisfied if there are linear continuations of the equations of elements outside the working region that have the signature (1). Also, the completely opposite signature is formally admissible, since the sign change of $f \to -f$ is admissible for stationary problems. To eliminate this trivial ambiguity, one can choose the sign of f, postulating the fulfillment of one of the conditions (1), for example, the first one.

The physical meaning of these conditions is illustrated in Figure 1. It shows the serial connection of the tested element (in this case the compressor, a circle) and a linear resistor (a rectangle). Pressure $P_{set1,2}$ is set at the free ends. The intermediate node must satisfy the equation

$$P_{out}(P_{set1}, Q) = P_{set2} + RQ, \quad (2)$$

graphically depicted in the central and lower parts of the figure. Here $R > 0$ is the resistance value, the corresponding line on the figure increases monotonically. If the tested element has the signature (1), then the function $P_{out}(P_{set1}, Q)$ decreases monotonically in Q, which corresponds to the central part of the figure. In this case, the intersection of lines exists and is unique. It can also occur outside this graph, when the above condition is met at infinity (continuation of the element's characteristic by a linearly strictly decreasing function outside the working region). In the case, if the signature (1) would be violated and the function $P_{out}(P_{set1}, Q)$ would increase in Q, then by choosing the parameters P_{set2} and R it is possible to achieve that the lines will have several intersections or no intersection. Even if the function $P_{out}(P_{set1}, Q)$ increases in Q only locally, a linear resistor can be fitted to it, which will give several solutions to the problem under consideration. It is also clear that a nonlinear resistor can also be used for this purpose, as long as its characteristic increases and has enough parameters for tuning.

Similarly, by connecting elements in reverse order, as well as considering their parallel connection, it can be shown that any violation of the condition (1) leads to a violation of the uniqueness of solution. If the signature is violated, then the tested element can be connected to an elementary resistive element in such a way that the equation will have several solutions or none. The case when the signature is satisfied for all elements and the system has a unique solution is, of course, more preferable in practical applications.

Compressors are the most complex elements in gas problems; several levels of modeling are used to represent them. The main purpose of introduction of these levels is the gradual sophistication of modeling, where the solution of a simple model is used as a starting point for the more complex one. Also, it allows to separate effects dependent on individual calibration of compressors from their basic representation.

Free model: is the simplest, formulated only in terms of transport variables, and is described by a piecewise linear formula of the form

$$\max(\min(P_{in} - P_L, -P_{out} + P_H, -Q + Q_H), \quad (3)$$
$$P_{in} - P_{out}, -Q) + \epsilon(P_{in} - P_{out} - Q) = 0,$$

where parameters P_L, P_H, Q_H define target values, for example, $P_H = SPO$ for specified output pressure, or upper and lower limits for other controlled values. This formula defines a polyhedral surface in the space of transport variables in the so-called *maxmin* representation [10]. Typical surface is shown below on Figure 2 left. The last term, controlled by small positive parameter ϵ, serves regularization and will be explained below.

Advanced model: introduces additional internal variables for compressors: revolution number rev, adiabatic enthalpy increase H_{ad}, performance $Perf$, efficiency η, torque M_t, and additional equations:

$$P = \rho RTz/\mu,\ Q_m = Q_{vol}\rho_{in}, \quad (4)$$
$$H_{ad} = P_{in}/(\rho_{in}\alpha) \cdot ((P_{out}/P_{in})^{\alpha} - 1), \quad (5)$$
$$Perf = Q_m H_{ad}/\eta,\ M_t = Perf/(2\pi \cdot rev), \quad (6)$$
$$\alpha = (\kappa - 1)/\kappa,\ 0 < \alpha < 1,\ 0 < \eta < 1, \quad (7)$$

where the equation of state is written first with its parameters: density ρ, universal gas constant R, absolute temperature T, compressibility factor z, molar mass μ; the second is the relationship between the mass flow and the volumetric flow in the input conditions; the following are definitions of internal variables in terms of transport variables; $\kappa > 1$ is the adiabatic exponent.

The advanced model also introduces additional patches, inserted into the free formula (3) as follows:

$$\max(\min(P_{in} - P_L, -P_{out} + P_H, -Q + Q_H, \underline{f_1, ..., f_n}), \quad (8)$$
$$P_{in} - P_{out}, -Q) + \epsilon(P_{in} - P_{out} - Q) = 0.$$

The additional patches for various types of compressors are described below. The general strategy is to resolve all internal variables from the corresponding equations, obtain a formula in terms of transport variables, check its signature, and use it in the standard solution algorithm.

III. Gas compressors

Three different types of compressors are considered.

A. *Piston compressors*

Compressors of piston types are modeled by direct proportionality

$$Q_{vol} = V \cdot rev \quad (9)$$

with given constants η and V – compressor chamber volume. The control equation has the following patches:

$$f_1 = rev_{max} - rev \geq 0, \quad (10)$$
$$f_2 = M_{t,max} - M_t \geq 0, \quad (11)$$
$$f_3 = Perf_{max} - Perf \geq 0, \quad (12)$$
$$f_4 = rel_{max} - P_{out}/P_{in} \geq 0, \quad (13)$$
$$f_5 = \Delta P_{max} - (P_{out} - P_{in}) \geq 0, \quad (14)$$

with given constants rev_{max}, $M_{t,max}$, rel_{max}, ΔP_{max} and the function $Perf_{max}(rev)$ determined by the characteristics of the compressor drive.

Stability analysis: calculating the derivatives of f_i with respect to (P_{in}, P_{out}, Q_m) in the working region $0 < P_{in} \leq P_{out}$, $Q_m > 0$, $rev > 0$, we get the signatures given in Table I. In this case, the above formulas are used, as well as the stability of the equation of state: $\rho > 0$, $\partial\rho/\partial P > 0$. In particular, $rev = Q_m/(\rho_{in}V)$ has signature $(-0+)$, which implies the signature of f_1 in the table. $M_t = H_{ad}\rho_{in}V/(2\pi\eta)$ has signature $(*+0)$, where $* = \partial(H_{ad}\rho_{in})/\partial P_{in} < 0$ for $P_{out}/P_{in} < (1-\alpha)^{(-1/\alpha)} = \beta$. Thus, the signature f_2 is correct only if the compressor raises the pressure by no more than the factor β, with the value $\kappa = 1.29$ typical for natural gas, we get $\beta = 3.10408$. To eliminate the fold in the equation, f_2 should be replaced with $H_{ad}\rho_{in}|P_{in} \to \max(P_{in}, P_{out}/\beta)$. It is convenient to divide the expression f_3 by $(2\pi\, rev)$ and consider the signature $\tilde{f}_3 = M_{t,drv}(rev) - M_t$. As noted in [2], for drive equations to be stable it is necessary that $M_{t,drv}$ decrease with rev. Therefore, the first term in $\tilde{f}_3$ has the signature $(+0-)$, and the second already calculated $(+-0)$ in the region $P_{out}/P_{in} < \beta$, which gives the complete signature $(+--)$. Calculation of other derivatives is trivial. We also note that the presence of zeros in the signatures means that the

TABLE I
PATCH SIGNATURES OF PISTON COMPRESSOR [1]

patch	sgn	condition
f_1	$(+\ 0\ -)$	
f_2	$(+\ -\ 0)$	$P_{out}/P_{in} < \beta$
f_3	$(+\ -\ -)$	$P_{out}/P_{in} < \beta$, $\partial M_{t,drv}/\partial rev < 0$
f_4	$(+\ -\ 0)$	
f_5	$(+\ -\ 0)$	

TABLE II
PATCH SIGNATURES OF GENERIC COMPRESSOR [1]

patch	sgn	condition
f_1	$(+\ 0\ -)$	
f_2	$(+\ -\ 0)$	$\partial z_{in}/\partial P_{in} < 0$ or small
f_3	$(+\ -\ -)$	$\partial z_{in}/\partial P_{in} < 0$ or small

rule (1) is satisfied marginally, which is corrected by adding a regularizing ϵ-term to the element equation. Also, for the practical implementation of these formulas, it is necessary to introduce clamps, which force all variables to the working region: $Q_m \to \max(Q_m, 0)$, $P_{out}/P_{in} \to \max(P_{out}/P_{in}, 1)$, etc.

B. Generic compressors

Compressors of generic type can also be considered as an intermediate level of modeling (generic model). In this model, the variable rev is not introduced, and restrictions are imposed on other variables

$$f_1 = Q_{vol,max} - Q_{vol} \geq 0, \tag{15}$$
$$f_2 = H_{ad,max} - H_{ad} \geq 0, \tag{16}$$
$$f_3 = Perf_{max} - Perf \geq 0, \tag{17}$$

with constant $Q_{vol,max}$, $H_{ad,max}$ and $Perf_{max}$.

Stability analysis: calculating derivatives similarly, for $Q_{vol} = Q_m/\rho_{in}$ we have signature $(-0+)$, hence $(+0-)$ for f_1, see Table II. For $H_{ad} = RT_{in}z_{in}/(\mu_{in}\alpha)((P_{out}/P_{in})^\alpha - 1)$ we get $(*+0)$, where $* = \partial(z_{in}((P_{out}/P_{in})^\alpha - 1))/\partial P_{in} < 0$. For an ideal gas $z = 1$, hence, obviously, $* = -$. For natural gas z is a decreasing function of P, in this case also $* = -$. For some gases, such as hydrogen, z may increase with P, but it remains close to 1 and changes so slowly that the remaining decreasing dependence of H_{ad} on P_{in} dominates. Under these conditions, f_2 has signature $(+-0)$. For $Perf = Q_m H_{ad}/\eta$ the signature $(-++)$ under the same conditions on z_{in}, thus f_3 has the signature $(+--)$.

C. Turbocompressors

Detailed modeling of turbocompressors was done in [2], [6], [7]. For the sake of completeness, here we review the main results.

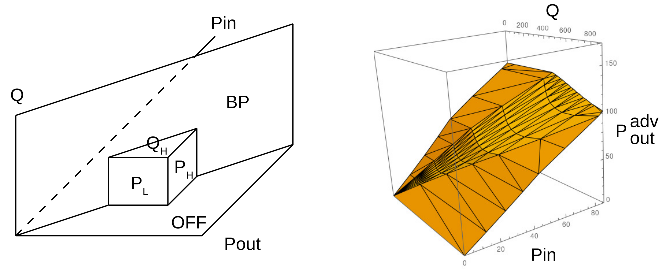


Fig. 2. Compressor modeling: 'free' model on the left; 'advanced' model for typical turbocompressor on the right. Images from [4], [7].

The equation of advanced patch has the form

$$f_1 = -P_{out} + P_{out}^{adv}(\hat{P}_{in}, \hat{Q}) \tag{18}$$
$$+ \min(P_{in} - P_{in,min}^{adv}, 0) + \max(P_{in} - P_{in,max}^{adv}, 0) \tag{19}$$
$$+ \min(-Q + Q_{max}^{adv}, 0) + \max(-Q + Q_{min}^{adv}, 0)), \tag{20}$$
$$\hat{P}_{in} = \min(\max(P_{in}, P_{in,min}^{adv}), P_{in,max}^{adv}), \tag{21}$$
$$\hat{Q} = \min(\max(Q, Q_{min}^{adv}), Q_{max}^{adv}). \tag{22}$$

The main part (18) is the equation $-P_{out} + P_{out}^{adv}(P_{in}, Q) = 0$, defined in rectangular region $P_{in} \in [P_{in,min}^{adv}, P_{in,max}^{adv}]$, $Q \in [Q_{min}^{adv}, Q_{max}^{adv}]$, with derivatives $\partial P_{out}^{adv}/dP_{in} > 0$, $\partial P_{out}^{adv}/dQ < 0$. Additional terms (19,20) carry out a continuous extension of this equation outside the region, with the same signs of derivatives. The resulting function $P_{out}^{adv}(P_{in}, Q)$ in typical case is shown in Figure 2 on the right. This surface represents the union of $rev \leq rev_{max}$ and $Perf \leq Perf_{max}$ patches converted into the space of transport variables. This conversion is done in the following way.

Step 1: calibration of the compressors and their drives with (bi-)quadratic functions:

$$H_{ad} = \sum_{ij} A_{ij} v_i^r v_j^q, \ \eta = \sum_{ij} B_{ij} v_i^r v_j^q, \tag{23}$$
$$Q_{vol,min} = \sum_i C_i v_i^r, \ Perf_{max} = \sum_i D_i v_i^r, \tag{24}$$
$$v^r = (1, rev, rev^2), \ v^q = (1, Q_{vol}, Q_{vol}^2), \tag{25}$$

with constant coefficient matrices A, B, C, D.

Step 2: definitions (4-6) are partially resolved until formulas in the (Q_m, ρ_{in}, H_{ad}) coordinate system are obtained, independent of temperature and gas composition:

$$Q_m = Perf_{max}\eta/H_{ad}, \ \rho_{in} = Q_m/Q_{vol}, \tag{26}$$

Step 3: definitions (4-6) are finally resolved until formulas in the (Q_m, P_{in}, P_{out}) coordinate system are obtained, taking into account temperature and gas composition:

$$P_{in} = EOS_{inv}(\rho_{in}), \ z_{in} = P_{in}/(\gamma\rho_{in}), \tag{27}$$
$$P_{out} = P_{in}(H_{ad}\alpha/(\gamma z_{in}) + 1)^{1/\alpha}.$$

Here $\gamma = RT_{in}/\mu$ and the equation of state $\rho = EOS(P)$ is inverted to determine P_{in}.

Step 1 is performed once in the calibration procedure using fitting methods. Step 2 is also performed once, while the domain of definition of functions on the (Q_{vol}, rev) plane is triangulated, the corresponding formulas are applied at the nodes, and linear interpolation in triangles is used for the continuous representation. Only step 3 is performed during simulations, as part of the procedure that determines the temperature and gas composition.

Stability analysis: consists in checking that the normals to the triangles used to represent the functions point to the octant corresponding to the correct signature. It suffices to carry out such a check in the coordinate system corresponding to step 2, while step 3 consists in a monotonic reparametrization of the axes that preserves the signature of the normal.

Further details: in order for the compressor to work properly, the flow through it must satisfy certain restrictions. On the $Q_{vol} = Q_{vol,min}$ line, the bypass regulator (rbp1,2 on Figure 4 bottom) is activated at the compressor, through which the flow can circulate, thereby ensuring the minimum required flow through the compressor (surge line). Since in this case the total flow through the compressor and the regulator may be less than this minimum value, this boundary is modeled by continuing the surface shown in Figure 2 on the right side along the Q axis towards small values. On the other side, there is a line $\eta = \eta_{min}$ (choke line), which is modeled by the continuation of the surface along the Q axis towards larger values. Since the compressor does not work efficiently on this part of the surface, the working point in this region is accompanied by a warning, and getting there should be avoided by adjusting the parameters. In addition, the continuation of the biquadratic functions into this region would lead to the formation of folds, and their replacement by the ruled continuation solves this problem. Switches similar to the surge line also occur on the $rev = rev_{min}$ line. At the same time, the location of the working point on this line and simultaneously on $Perf = Perf_{max}$ patch also leads to the formation of folds, which for typical compressors are outside the physical area and are eliminated by cutting.

The characteristics of the compressor drive can also depend on the ambient temperature T_{amb}, which can be modeled by constructing the biquadratic formula $Perf_{max}(rev, T_{amb})$. In this case, the resulting dependence in step 2 is linear in $Perf_{max}$, which leads to the possibility of precomputing the surface for three values of T_{amb} and then taking into account the dependence on T_{amb} using linear weighting with weights quadratically dependent on T_{amb}.

Compressor stations: are formed, in the simplest case, when individual compressor units are connected in parallel or in series. Examples of such combinations are shown in Figure 3a,b and in more detail in Figure 4 bottom. In this case, the type and controlled values of each compressor may be the same, they may also be different. The station also includes other elements, some of which (valves, shortcuts) have trivial functions. To ensure the stability of the solution algorithm (in particular, no cycles from shortcuts, indefinite pressure in the sequence of closed valves, etc.), such elements are eliminated in the preprocessing procedure. There are also other elements: coolers that affect temperature distributions inside and outside of stations and bypass-regulators providing a non-zero flow through the compressor. Parallel and serial connections of compressors can also create solution ambiguities, detailed below.

IV. Sensitivity and Principal Component Analysis

Below, Sensitivity Analysis and Principal Component Analysis of gas transport simulations will be considered.

A. Sensitivity Analysis

Determining the sensitivity of a model to changing parameters is a standard tool for in-depth analysis of simulation results. Usually it consists in the computation of sensitivity matrix $S^x_{ij} = \partial x_i/\partial p_j$, defined in terms of partial derivatives, along with the Jacobi matrix $J_{ij} = \partial y_i/\partial x_j$ and the sensitivity of the equations in terms of the parameters $S^y_{ij} = \partial y_i/\partial p_j$. Here x are model variables that also represent the simulation result, p are model parameters, y are equations in terms of which the model is formulated. With a small number of parameters, the derivatives can be found using the numerical differentiation algorithm:

```
init: simulate p → x
for j=1,num(p) do
   variate parameter p̃_j = p_j + dp_j
   simulate p̃ → x̃
   differentiate S^x_ij = (x̃_i − x_i)/dp_j
done
```

or a similar algorithm with the choice of the central difference scheme. The differentiation step dp_j must be chosen reasonably, it must be small enough that the variation of dx_i can be considered linear, and large enough that this variation exceeds the numerical error of the simulation result. It is usually sufficient to perform several variations of different orders and verify that these properties are satisfied, using 1D plots $\tilde{x}_i(dp_j)$ for several key variables. The 'for' loop in this algorithm can be parallelized on several processors, providing a significant speedup for the analysis procedure. Alternatively, if there is an access to derivatives of equations with respect to variables and parameters, the sensitivity matrix can be found by solving the linear system

$$\sum_j (\partial y_i/\partial x_j)(\partial x_j/\partial p_k) + \partial y_i/\partial p_k = 0, \quad (28)$$

$$\sum_j J_{ij} S^x_{jk} + S^y_{ik} = 0, \quad (29)$$

which does not require additional simulations. However, often the simulation algorithm is encapsulated inside a software module where information about derivatives is not available. Also, sometimes the simulation algorithm is not a classical solution of a unified system of equations, but includes external iterations, relaxation procedures and/or calls to external 'black box' software modules. In this case, only the numerical differentiation algorithm described above remains for SA.

B. Principal Component Analysis

In the case of a large number of variables, equations and/or parameters, it makes sense to use PCA. In this approach, several components can be calculated, represented as linear combinations of variables that make the main contribution to the relations of the considered model. Technically, Singular Value Decomposition (SVD) of the respective matrices is performed for this purpose. Here we illustrate the application of PCA/SVD to the Jacobi matrix: $J = u\lambda v^T$, with a diagonal matrix λ and orthogonal matrices u and v. The largest eigenvalues λ correspond to the strongest dependencies. In this case, the corresponding left eigenvectors, columns of the u matrix, correspond to most rapidly changing combinations of y equations, and the right eigenvectors, columns of the v matrix, correspond to combinations of x variables with the strongest dependencies. To study the stability of numerical solution algorithms, it is also useful to know the smallest eigenvalues λ, which correspond to the weakest dependencies. In this case, the left eigenvectors represent linear combinations of equations that change little at a normalized change of all variables, and the right eigenvectors represent combinations of variables on which all equations weakly depend.

In the presence of small eigenvalues, algorithms for solving nonlinear systems, such as Newton's method, lose their convergence [11]. Mathematically, at a zero eigenvalue, the solutions of a linearized problem lose their uniqueness – the solutions either disappear or a continuous set of equivalent solutions appears. In practice, the solution of nonlinear systems is made with some given accuracy, which defines the ball of admissible solutions in the y-space, $|y| < tol_y$, using the l^2-norm. The preimage of this ball in x-space is the error ellipsoid, $dx = J^{-1}y$. The SVD of J-matrix determines the sizes of the semiaxes of the ellipsoid $|\delta x_i| = tol_y/\lambda_i$, while the right eigenvectors determine the orientation of the ellipsoid. For small eigenvalues, a strongly prolate ellipsoid arises corresponding to large x-errors, and in the $\lambda \to 0$ limit, to indifferent directions in the x-space.

Note that the error ellipsoid with large semiaxes is an indicator of instability in the solution of the problem. With variations in the problem statement, such as the choice of a starting point, parameters of iterative procedures, as well as with small variations in the free parameters of the modeling, the number of performed iterations may change. In this case, the end point makes jumps inside the y-ball, in practice can be considered random. As a result, the solution of the problem randomly changes within the x-ellipsoid. Large random variations of the solution indicate the ambiguity, which manifests both at the modeling level and in the physical system itself.

Application of PCA/SVD methods in gas transport problems was carried out in our recent work [3]. Here we will review the main results, now including the implementation details and modes of usage of these methods.

Implementation of PCA for large scale problems: large problems are usually represented by sparse matrices. The technical difficulty of SVD for large sparse matrices is that the result is generally a dense matrix, which makes computation time and memory requirements problematic. In special cases, when it is required to calculate not all, but only a few largest or smallest eigenvalues and the corresponding eigenvectors, there are algorithms that can keep the problem in a sparse class and represent the result in an economical form.

We use *Mathematica V12* for the described calculations, where the standard `SingularValueDecomposition` method is available. For full decomposition this method converts sparse matrices to dense by default. In a special case, if not all eigenvalues are required, but a small number of the highest ones, the method works in the sparse class. However, our analysis does not require the highest, but the lowest values, with which, according to the description, the method does not work.

This problem can be resolved as follows. Computing the product $J^T J$, we get a sparse symmetric positively semi-definite matrix. Its eigenvectors coincide with the columns of the v-matrix, and its eigenvalues coincide with the squares of the SVD eigenvalues. For this matrix, in *Mathematica V12* one can use the `Eigensystem` method, which works for the sparse case and allows one to restrict the computation to several lowest eigenvalues and corresponding eigenvectors. The same method, Arnoldi iterations, is used in *Arpack* system. To find u-vectors, one can find the product JJ^T and repeat the decomposition.

Note that for nonlinear systems, the linearization-based error estimate gives only an approximate result, more accurate for small δx and less accurate for large ones. For piecewise linear systems, this estimate is valid up to the cell boundary, where exact linearity is maintained.

Usage of PCA: according to the analysis, the result of simulation turns out to be not just a point in the space of variables, but a point with an error ellipsoid. This makes it possible to classify variables into more / less accurately defined and possibly completely inaccurately defined due to degenerations in the system. Other types of analysis should also take into account this information. For example, in the above considered SA, it may turn out that some variables undergo strong changes, jumps with a continuous change in parameters. This may not mean that these variables are highly sensitive to parameters, but that they change randomly within the error ellipsoid.

Another example of using PCA is the comparison of *simulation vs experiment*. The measurements have their own error ellipsoid, simulation and experiment results are compatible if these ellipsoids intersect. In practice, it is necessary to check the correspondence also at the 2-3-σ level and increase the ellipsoids by the appropriate factor.

Likewise, when comparing simulation results from *two different solvers*, one must check the intersection of their error ellipsoids. In the case if it is known in advance that the modeling used in the two solvers is the same, then it is enough to check that the $|y| < tol_y$ condition is satisfied for both of them, then both solutions will automatically belong to the common error ellipsoid. If it is not known whether

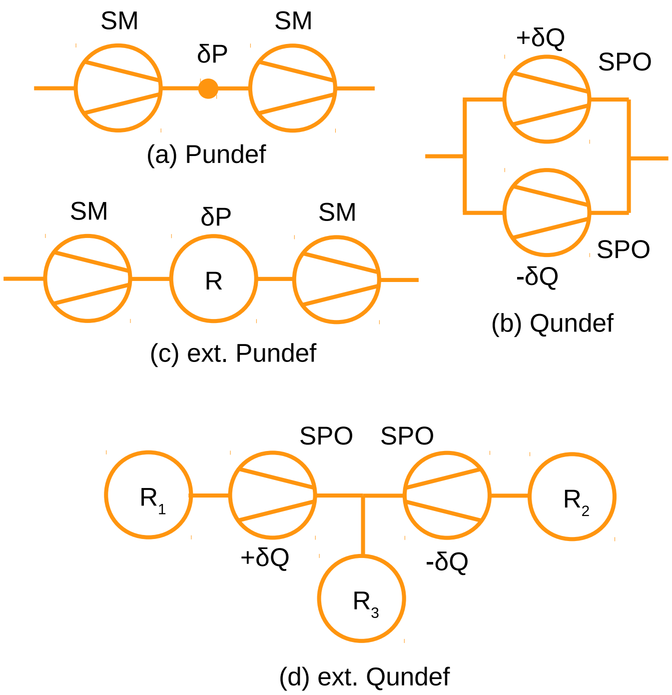


Fig. 3. Instabilities in gas transport problems: (a) undefined intermediate pressure in serial connection of SM-compressors; (b) undefined flow balance in parallel connection of SPO-compressors; (c,d) extended versions of the instabilities. Images from [3].

the modeling is the same, then cross-validation should first be carried out, substituting the answer of one solver into the evaluation function of another. If the $|y| < tol_y$ condition is violated, then it is possible to find out which y-components have large deviations, thus, which aspects of modeling differ between the solvers.

V. Solving nearly degenerate systems

As shown by the theoretical analysis and the numerical experiments below, stationary gas transport problems can possess instabilities. They are associated with the modeling of control elements, such as the compressors. While the advanced modeling is usually stable, the source of instabilities is the free model.

The simplest types of instabilities were discussed in [3] and shown here in Figure 3. In case (a), two serially connected SM compressors are considered. For such compressors, the control equation (3) is set to the value of specified mass flow, $Q_H = SM$. Since the flow through the compressors is the same, the equation $Q = SM$ is applied twice, which leads to the degeneration of the system. Because one equation is actually wasted, one continuous degree of freedom appears in the solution of the system. This degree of freedom corresponds to an undefined value of pressure between the compressors, which is not constrained by any equations. In case (b) there are two parallel SPO compressors. For such compressors, the equation for output pressure is doubled, $P_{out} = SPO$, the flow balance dQ through the compressors is undefined. In these two cases, the instability is localized within the compressor station. Instability can also go beyond the station. In case (c) there is a resistive subsystem between the two SM compressors, in which the pressure values appear to be undetermined. In case (d), two SPO compressors with a common output produce flow imbalance over wide areas of the network.

The situation is complicated by the following factors. The compressors do not have to be of the predefined type SPO/SPI/SM. They may formally belong to another type, but be on one of the faces of the surface for free model Figure 2, a face that actually corresponds to the conflicting type. Thus, the conflict may include faces corresponding to $P_H = POMAX$ and $P_L = PIMIN$; $Q_H = QMAX$ and $Q_L = QMIN = 0$, etc. The presence of the advanced part of the modeling does not solve the problem, since the advanced model (8) contains all the faces of the free model (3), which continue to generate conflicts on solutions. Other control elements that have similar modeling (regulators, flaptraps, etc.), as well as nodes with a fixed pressure or flow (Pset, Qset) can also participate in the conflicts. It is clear that it is impossible to foresee all conflicting combinations of faces in the control equations, due to combinatorial reasons. We have to look for special methods for solving degenerate systems that could be applied to this case. The potentially useful methods have been briefly listed in [3], now we consider them in full detail.

Regularization: is performed by the ϵ-term in free model equation (3). The reason for its introduction is that the equation at $\epsilon = 0$ satisfies the signature condition (1) only marginally, some derivatives vanish. The geometric interpretation of this is that the normals to the faces of the polyhedron shown on Figure 2 left are directed strictly along the axes, although they should be directed inside the octant described by the condition (1). Such marginality leads to degeneracy of the Jacobi matrix, ambiguity of solutions, bad condition numbers, and other troubles for the numerical solution procedure. The introduction of a regularizing ϵ term formally eliminates this problem by making the condition (1) strictly satisfied. At the same time, adjusting this parameter represents a compromise between the physical accuracy and the numerical stability of the solution procedure. In practice, the values $\epsilon = 10^{-6}...10^{-3}$ are tolerable, meaning the relative violation of, e.g., SPO-condition, up to 0.1%, simultaneously keeping the convergence rate near 100%.

Relaxed Armijo rule: it is clear that part of the problem is related to the line search algorithm [11]. According to this algorithm, in the process of system solving, a step is taken along the Newtonian direction $dx_N = -J^{-1}y$, not completely, but so that the residual of equations $|y|$ made a sufficient reduction. In [7], it was proposed to relax this requirement, allowing a controlled small increase of residual in nearly degenerate cases. Indeed, in practice this leads to a drastic improvement in the stability of the algorithm and a decrease in the number of iterations. The drawback is that, in rare cases, Newtonian iterations go in cycle. Theoretically, the conditions for the convergence of Newtonian iterations [11] require the exact fulfillment of the Armijo rule. Thus, the rule relaxation is an empirical method that often improves convergence in degenerate problems, but does not work in 100% of cases.

Topological reduction: the method described in [7] makes it possible to contract parallel and serial stations, eliminating the degenerate degrees of freedom contained in them, which leads to an increase in the stability of the solution algorithm. However, this method only works for local conflicts and cannot resolve the extended ones.

Dynamical problems: algorithms for integrating dynamical systems used to obtain stationary states can be more stable than the straightforward solution of stationary problems. Indeed, *Pundef*-conflicts can be resolved, since the intermediate pressure values are determined by the starting point and the integration process. These values are still arbitrary, but this arbitrariness does not complicate the process of finding a solution. *Qundef*-conflicts can be resolved similarly by including the corresponding kinetic terms in the equations.

Homotopic methods: this class of methods [20] considers the deformation of equations from some simple solvable form to the required form, in which the trajectory of the solution is continuously tracked. For example, the linear deformation $y = y_0(1-\alpha) + y_1\alpha$ can be used, with $\alpha \in [0,1]$. The problem with such methods is the possible appearance of a fold in the equations, leading to a loss of the solution. In our particular case, when using regularization, the required equations are guaranteed to have a unique solution. As an auxiliary form, one can use a linear non-degenerate system, which also has a unique solution. However, this does not mean that the mixed system will also have a unique solution. It is easy to construct examples of matrices that have a determinant of one sign, for which a linear combination with positive coefficients has a determinant of another sign, or equal to zero. In this case, it makes sense to use the structure of the system, for example, do not touch the Kirchhoff equations and deform only the equations of the elements. It is clear that only singular equations need to be deformed, since the rest keep the Jacobi matrix nondegenerate and present no problems. Then this method becomes equivalent to our ϵ-regularization with a gradual decrease in the ϵ parameter. Theoretically, this method in its last stage is unstable, since an increasingly degenerate system is being solved. Our numerical experiments presented below show that the described homotopy method works as a stabilizer, although not in 100% of cases.

Note also that the linear deformation of equations described above is only the simplest version of the homotopic method. The work [20] offers tools to overcome $rank\ J = n-1$ degeneracy by extending J from $n \times n$ to $n \times (n+1)$ full-rank matrix. However, our system typically have multiple zero eigenvalues. The described methods do not seem to work immediately for this case, at higher order degeneracy requiring an increasingly complex parameterization.

Piecewise linear systems: for such systems, the space of variables is divided into polyhedral cells, in each the system is linear, with continuous connection at the boundaries. For such systems, there are methods that allow finding a solution in a finite number of steps, see [8]–[10] and also [20] Chap. 12-15. It is noteworthy that some of these methods work even if the system is piecewise degenerate. In algorithm [9], it is proposed to follow the Newtonian direction in non-degenerate cells, or opposite one, dependently on the sign of Jacobian. In degenerate cells, it is proposed to follow the right annulator of the system matrix. Each time, the movement should stop at the border of the next cell. Such algorithm has theoretically guaranteed convergence [9]. For our applications, this algorithm should be extended to nonlinear systems. The annulators can be found using fast procedures (*Mathematica, Arpack*).

Pseudo-inverse: a related SVD-based algorithm is the procedure for finding the Newtonian step $dx_N = -J^{-1}y$ with elimination of zero eigenvalues, see Chap. 2.6 in [21]. In practice, to find the pseudo-inverse J^{-1}, after SVD, all non-zero eigenvalues are replaced by inverse $\lambda \to \lambda^{-1}$, and zero ones remain unchanged: $0 \to 0$. If the problem would be globally degenerate, this procedure would provide sliding along a trajectory orthogonal to the annulators, ignoring degenerate combinations of equations and unstable variables, concentrating only on changing the essential variables in the system. It is not known whether there is a generalization of this algorithm to the nonlinear piecewise degenerate case, that is of interest for our applications.

Among the described methods, the empirical approach with ϵ-regularization, relaxed Armijo rule and optional error calculation by PCA method works satisfactorily in practice. Among the most promising approaches, we consider the extension of the described methods to dynamic problems.

VI. Numerical experiments

A number of numerical experiments have been performed to test the performance of above described methods.

A. Test networks

We use the stabilized Newton algorithm described in [4] to solve the gas transport problems on the following networks. A small network N1 has been created by us to test functionality of various elements, shown in Figure 4 on the top. A medium size network ME, created in frames of the project *MathEnergy*, containing all features of realistic networks, shown in Figure 4 on the center. A set of large scale realistic networks N85, given to us by our industrial partner for benchmarking of the methods. The main parameters of all test networks are contained in the Table III.

B. Test of piston and generic compressors

Since their implementation is a main contribution of the paper, we have tested the stability and performance of the solver for the networks containing such elements. In particular, the network N1 has 100 nodes and 111 edges, of which 4 compressors are organized into two compressor stations c1|2 and c3|4 with individual compressors connected in parallel, as shown in Figure 4 at the bottom. Compressors in station c1|2 are configured as piston ones, in station c3|4 as generic ones.

TABLE III
PARAMETERS OF TEST NETWORKS [2]

network	tot.num.	nodes	edges	pipes	compressors	regulators	Psets	Qsets
N1	1	100	111	34	4	4	2	3
ME	1	437	482	370	20	24	3	164
N85/L	23	3232-3886	3305-3974	2406-2835	1-7	59-77	6-7	625-843
N85/H	62	2914-3818	2989-3952	1498-1937	16-42	59-107	5-9	328-505

Values P_H, Q_H are set to unreachable high values, thereby activating the f_i patches described above.

The procedure consists of several phases with a gradual increase in the modeling level. At first, the compressors are set to fulfill the main target values, e.g., $P = P_H$, then the modeling level (3) is used, taking into account additional conditions, then the modeling level (8) is taken. The solution procedure described in [5] consists of the translation phase of the system from the network description language to the language understood by the numerical solver, and the actual numerical solution phase. In this test, approximately the same results are obtained if turbocompressors are used instead of piston/generic ones. These numerical experiments show that the inclusion of piston and generic compressors in the system does not lead to any divergences or slowdown of the solution procedure, which is a direct consequence of the implementation of the stability criteria described above.

We also performed numerical experiments with test networks from N85 set. This set contains 85 networks with complexity up to four thousand nodes and up to 42 compressors. Among them are multiple piston and generic compressors, in parallel and series connections. We have found that the presence and placement of such compressors does not affect performance in any way, and this is consistent with the convergence conditions we developed.

C. Test of ϵ-homotopy

The dependence of the convergence of network problems on the ϵ-parameter was investigated in [3] on N85 networks for free compressor setting. In the present work, we have carried out such a study for advanced compressors. In addition, we have included an external iterative loop in the solution algorithm, the so-called mix-iteration, in which the gas composition, temperature, parameters of the refined equation of state, etc. are calculated. Here we have taken the opportunity to explore a simple homotopy algorithm by inserting parameter division into the mix-iteration: $\epsilon \to \epsilon/q$, $q > 1$. The simulation results are shown in Tables IV and V. In accordance with the above considerations, as the ϵ-parameter decreases, the problem becomes more and more degenerate. In total across all N85 networks, the number of divergent scenarios and the computation time increase with decreasing ϵ-parameter. For the homotopy algorithm, we performed two numerical experiments. Both started at $\epsilon = 10^{-3}$ and ran 10 mix-iterations. In one experiment, $q = 2$ was chosen and the homotopy was carried out up to the value $\epsilon \sim 10^{-6}$. The algorithm is slightly more stable, leaving divergent 2 scenarios out of 85, compared to simply fixing ϵ to a final value, which makes divergent 3 scenarios. In the second experiment, $q = 4$ was chosen and the division was carried out up to the deeper value $\epsilon \sim 10^{-9}$. Here, the homotopy algorithm turns out to be much more stable, although it still leaves divergent 10 scenarios out of 85. Since the value of $\epsilon = 10^{-6}$ is sufficient for practical purposes, we conclude that the described homotopy algorithm stabilizes the solution of the problem, with moderate success.

D. Sensitivity Analysis of N1 network

It is carried out by us in order to test the algorithm described above and the software module based on it. For this test, in each of the two compressor stations c1|2 and c3|4, one compressor was configured as specified output pressure (SPO), the other as specified mass flow (SM). The results are shown in Figure 5. In case (a), input pressure in n99 supply is changed for 1 bar. The resulting variation of pressure does not propagate beyond the nearest compressor station c3|4. In case (b), SPO in c1 compressor is changed for 1 bar. The change propagates further to the network. The maximum dP is reached in the consumer n76. The increase of dP is due to nonlinearity of pipe friction law. In case (c), SM in c4 compressor is changed for 2%. The result corresponds to a different balance of flows in the network, while the pressure is changed insignificantly.

Thus, the application of SA to the N1 network shows that of the three parameter variations considered, the resulting pressure in the network is sensitive to only one parameter, SPO in c1 compressor.

E. Principal Component Analysis of ME network

It is carried out to test the performance of the method on a larger network than in our earlier study [3]. This early study looked at N1 network with compressors in free SPO mode. There, PCA revealed two small eigenvalues that varied proportionally to the ϵ parameter and corresponded to conflicts within the stations. A larger small eigenvalue was also found, corresponding to an extended conflict between compressor stations. In the current work, for the ME network, we obtain

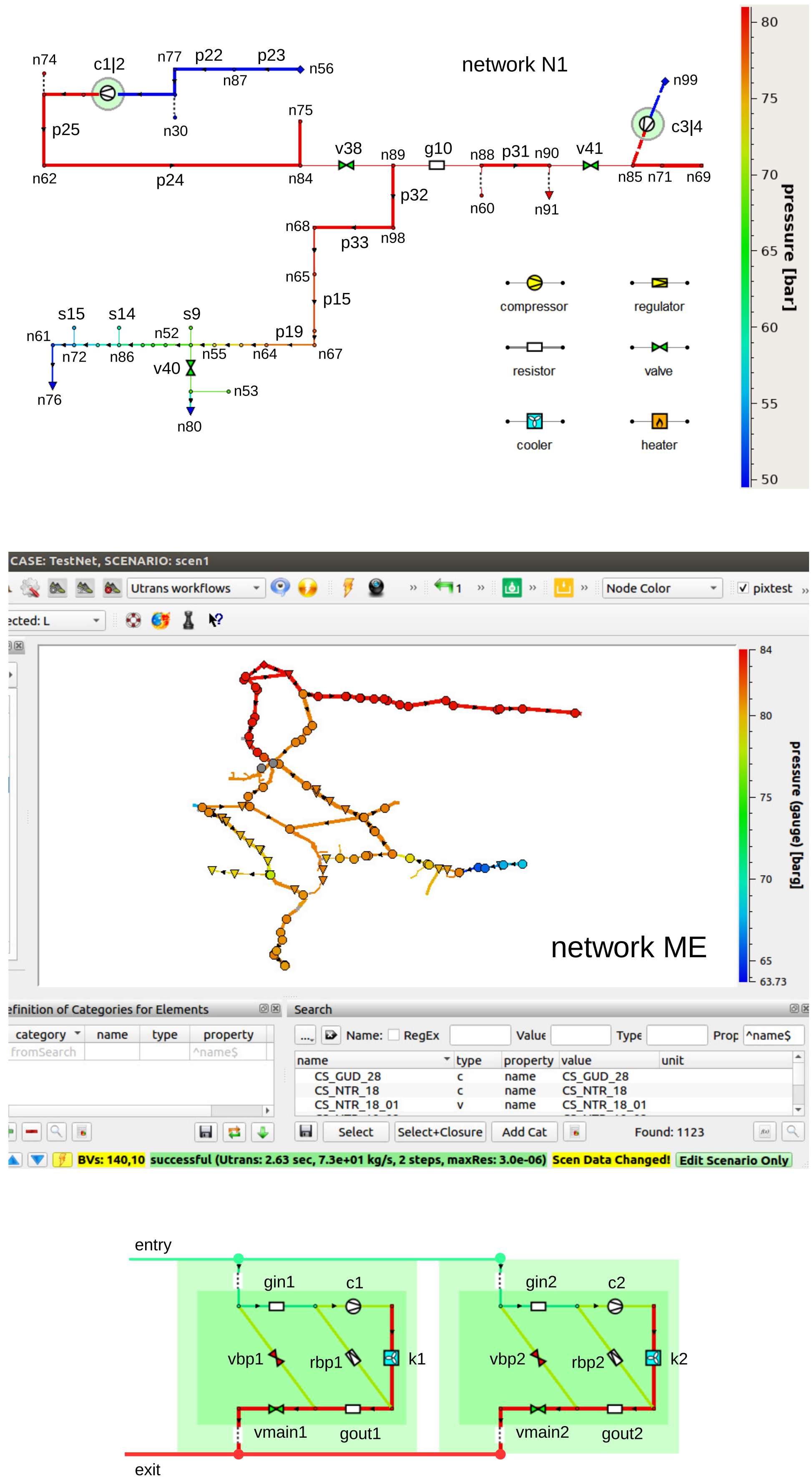


Fig. 4. On the top: test network N1; in the center: test network ME; at the bottom: the structure of parallel compressor station. Images from [2], [4].

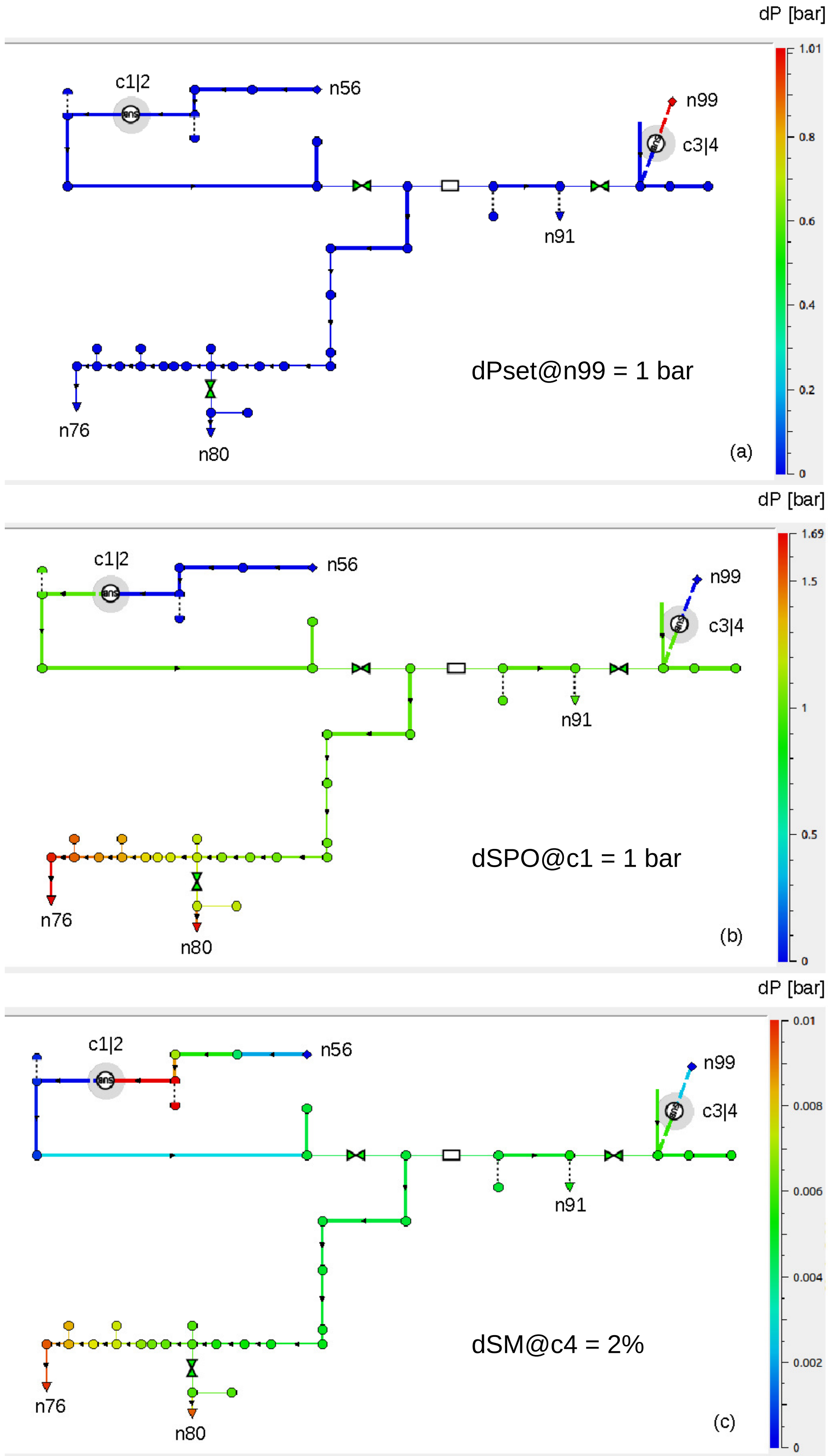


Fig. 5. Sensitivity Analysis, for N1 network.

TABLE IV
CONVERGENCE FOR FIXED ϵ, N85 NETWORKS

ϵ	total div.	total time*
10^{-3}	1	84
10^{-6}	3	117
10^{-9}	40	533

* in minutes, for Intel i7 / CPU 2.6 GHz / RAM 16 GB computer.

TABLE V
CONVERGENCE FOR ϵ-HOMOTOPY, N85 NETWORKS

q	total div.	total time*
2	2	87
4	10	98
$\epsilon = 10^{-3}/q^n,\ n = 0...10$		

* in minutes, for Intel i7 / CPU 2.6 GHz / RAM 16 GB computer.

the results presented in Tables VI-VIII. Initially, the value of the parameter $\epsilon = 10^{-3}$ was selected, all compressors were set to free SPO mode.

The results obtained indicate another type of conflict, that is related to pipes. The lowest eigenvalue corresponds to a cycle of two pipes with lengths and diameters $L_{p0618} = 118m$, $D_{p0618} = 0.6m$, $L_{p0643} = 120m$, $D_{p0643} = 0.9m$, possessing a zero flow. The next largest eigenvalue corresponds to a cycle of three pipes with lengths $L_{p0110} = 136m$, $L_{p0765} = 1917m$, $L_{p0766} = 1967m$, common diameter $D = 0.8m$ and a small flow $m \sim 0.1$ kg/s. The left eigenvectors correspond to an indefinite flow circulating in cycles. The right eigenvectors for a given associated with pipes conflict are the same as the left ones. They correspond to the sum of equations taken over the cycle, which in leading order have the form [4]: $P_1|P_1| - P_2|P_2| = R_q Q|Q|$. When such a sum is taken, the P-terms disappear, leaving only the sum of the Q-terms (Kirchhoff's second law). The reason for the degeneracy is the low resistance R_q for short and thick pipes. Also, due to the quadratic dependence on Q, additional suppression occurs for small flows. Note that the complete expression for the friction equation in pipe also contains a laminar term. This term is linear in Q, but its value is insufficient to stabilize the system at low flows. Possible ways to solve this problem can be: artificial increase of the laminar term, similar to ϵ-regularization; replacing short pipes with shortcuts followed by contraction [6]; topological reduction of serial and parallel connections of pipes [7].

At higher eigenvalues, there are similar pipe-conflicts, while the conflicts associated with compressors and regulators also begin to occur. As the regularizing parameter decreases to $\epsilon = 10^{-6}$, one of these conflicts descends from the high eigenvalues $\lambda = 4.29 \cdot 10^{-4} \to 4.29 \cdot 10^{-7}$, proportionally following the change in ϵ by three orders of magnitude. At the same time, the eigenvalues and vectors of pipe-conflicts change little. This new conflict corresponds to the compressor and regulator connected by a series of 4 pipes with a total length of 35 km. On the solution, the compressor and regulator are closed, have zero flow. In this case, the pressure in the intermediate segment is undefined. Thus, the described conflict belongs to the extended type (c), shown on Figure 3.

In the simulations under consideration, the accuracy parameter $tol_y = 10^{-5}$ is used. The found eigenvalues correspond to the error in the x-space according to the formula $|\delta x_i| = tol_y/\lambda_i$. In our simulations, the range of operating values of pressures and flows is in the order of $P = 0{-}100$ bar, $Q_m = 0 - 100$ kg/s. Now it is clear that the first eigenvalue gives an error exceeding this range, and the corresponding degree of freedom is completely undefined. For the next eigenvalue, the error is about 4% and further decreases to $< 1\%$. Note also that for higher eigenvalues, the eigenvectors are distributed over a large number of elements, and the error in each element is even smaller.

Thus, an application of PCA to the ME network appears to be useful, bringing additional knowledge about the new type of conflicts, confirming the conclusions about the ϵ-regulated instabilities, and making a differentiated estimation of the simulation error for different degrees of freedom present in the solution.

TABLE VI
THE LOWEST EIGENVALUE AND RELATED PRINCIPAL COMPONENTS, FOR ME NETWORK

λ_1	$\|\delta x_1\|$	$\delta Q_{m,p0618}$	$\delta Q_{m,p0643}$	eq_{p0618}	eq_{p0643}
$2.56 \cdot 10^{-8}$	390	-0.707	-0.707	-0.707	-0.707

TABLE VII
THE SECOND LOWEST EIGENVALUE AND RELATED PRINCIPAL COMPONENTS, FOR ME NETWORK

λ_2	$\|\delta x_2\|$	$\delta Q_{m,p0110}$	$\delta Q_{m,p0765}$	$\delta Q_{m,p0766}$	eq_{p0110}	eq_{p0765}	eq_{p0766}
$2.46 \cdot 10^{-6}$	4.06	0.589	0.571	0.571	0.589	0.571	0.571

TABLE VIII
THE NEXT LOWEST EIGENVALUES, FOR ME NETWORK

λ_3	$\|\delta x_3\|$	λ_4	$\|\delta x_4\|$	λ_5	$\|\delta x_5\|$	λ_6	$\|\delta x_6\|$
$1.38 \cdot 10^{-5}$	0.725	$2.10 \cdot 10^{-5}$	0.477	$2.59 \cdot 10^{-5}$	0.386	$2.63 \cdot 10^{-5}$	0.380

VII. CONCLUSION

In this work, modeling of piston and generic type gas compressors was carried out. The signatures of the derivatives of the control equation are analyzed, the ranges of parameter values are identified, under which the conditions for the stable operation of the algorithm for solving stationary network problems are satisfied. After the practical implementation of the modeling, in numerical experiments on realistic gas networks, the convergence of the solution algorithm is shown.

In addition, an extended review was made of results on turbine type compressors, methods of Sensitivity and Principal Component Analysis applied to gas transport networks, methods for solving nearly degenerate nonlinear systems. A number of new numerical experiments were carried out on realistic scenarios of a stationary type, representing the application of these methods. In particular, it is shown that regularization methods, relaxed line search and ϵ-homotopy have a stabilizing effect on the solution procedure. Sensitivity and Principal Component Analysis show areas of increased system responsiveness to parameter variations and solution procedure setups. In addition to the already known conflicts associated with the compressor stations, there are flow ambiguities in cycles of short pipes. All the effects observed in numerical experiments are in agreement with the theoretical analysis.

Our future plans include extending the described methods to dynamic problems.

ACKNOWLEDGMENT

The work has been supported by the project TransHyDE-Sys, grant 03HY201M.

REFERENCES

[1] A. Baldin et al., “Solving Stationary Gas Transport Problems with Compressors of Piston and Generic Type”, in Proc. of INFOCOMP 2022, International Conference on Advanced Communications and Computation, pp. 1-5, IARIA, 2022.
[2] A. Baldin et al., “AdvWarp: a transformation algorithm for advanced modeling of gas compressors and drives”, in Proc. of SIMULTECH 2021, International Conference on Simulation and Modeling Methodologies, Technologies and Applications, pp. 231-238, SciTePress, 2021.
[3] A. Baldin et al., “Principal component analysis in gas transport simulation”, in Proc. of SIMULTECH 2022, International Conference on Simulation and Modeling Methodologies, Technologies and Applications, pp. 178-185, SciTePress, 2022.
[4] T. Clees, I. Nikitin, and L. Nikitina, “Making network solvers globally convergent”, Advances in Intelligent Systems and Computing, vol. 676, 2018, pp. 140-153.
[5] T. Clees et al., “MYNTS: Multi-phYsics NeTwork Simulator”, in Proc. of SIMULTECH 2016, International Conference on Simulation and Modeling Methodologies, Technologies and Applications, pp. 179-186, SciTePress, 2016.
[6] T. Clees, I. Nikitin, L. Nikitina, and L. Segiet, “Modeling of gas compressors and hierarchical reduction for globally convergent stationary network solvers”, International Journal On Advances in Systems and Measurements, vol. 11, 2018, pp. 61-71.
[7] A. Baldin, T. Clees, B. Klaassen, I. Nikitin, and L. Nikitina, “Topological reduction of stationary network problems: example of gas transport”, International Journal On Advances in Systems and Measurements, vol. 13, 2020, pp. 83-93.
[8] J. Katzenelson, “An algorithm for solving nonlinear resistor networks”, Bell System Technical Journal, vol. 44, 1965, pp. 1605-1620.
[9] M. J. Chien and E. S. Kuh, “Solving piecewise-linear equations for resistive networks”, International Journal of Circuit Theory and Applications, vol. 4, 1976, pp. 1-24.
[10] A. Griewank, J.-U. Bernt, M. Radons, and T. Streubel, “Solving piecewise linear systems in abs-normal form”, Linear Algebra and its Applications, vol. 471, 2015, pp. 500-530.
[11] C. T. Kelley, Iterative Methods for Linear and Nonlinear Equations, SIAM, Philadelphia, 1995.
[12] J. Mischner, H. G. Fasold, and K. Kadner, System-Planning Basics of Gas Supply, Oldenbourg Industrieverlag GmbH, 2011 (in German).

[13] M. Schmidt, M. C. Steinbach, and B. M. Willert, “High detail stationary optimization models for gas networks”, Optimization and Engineering, vol. 16, 2015, pp. 131-164.

[14] J. Nikuradse, “Laws of flow in rough pipes”, NACA Technical Memorandum 1292, Washington, 1950.

[15] C. F. Colebrook and C. M. White, “Experiments with fluid friction in roughened pipes”, in Proc. of the Royal Society of London, Series A, Mathematical and Physical Sciences, vol. 161, num. 906, 1937, pp. 367-381.

[16] P. Hofer, “Error evaluation in calculation of pipelines”, GWF-Gas/Erdgas, vol. 114, num. 3, 1973, pp. 113-119 (in German).

[17] J. Saleh, Fluid Flow Handbook, McGraw-Hill 2002.

[18] DIN EN ISO 12213-2: Natural Gas - Calculation of Compression Factor, European Committee for Standardization, 2010.

[19] O. Kunz and W. Wagner, “The GERG-2008 wide-range equation of state for natural gases and other mixtures: an expansion of GERG-2004”, J. Chem. Eng. Data, vol. 57, 2012, pp. 3032-3091.

[20] E. L. Allgower and K. Georg, Introduction to Numerical Continuation Methods, SIAM, Philadelphia, 2003.

[21] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes in C, Cambridge University Press 1992.

Distributionally Robust Chance-Constrained Zero-Sum Games with Moments Based and Statistical Based Uncertainty Sets

Nguyen Hoang Nam
L2S, Centrale Supelec
University Paris Saclay
91190 Gif-sur-Yvette, France
hoang-nam.nguyen3@centralesupelec.fr

Vikas Vikram Singh
Department of Mathematics
Indian Institute of Technology Delhi
New Delhi, 110016, India
vikassingh@iitd.ac.in

Abdel Lisser
L2S, Centrale Supelec
University Paris Saclay
91190 Gif-sur-Yvette, France
abdel.lisser@centralesupelec.fr

Monika Arora
Department of Mathematics
Indraprastha Institute of Information Technology Delhi
New Delhi, 110016, India
monika@iiitd.ac.in

***Abstract* - We consider a two-player zero-sum game with random linear chance constraints whose distributions are known to belong to moments based uncertainty sets or statistical distance based uncertainty sets. The game with chance constraints can be used in various applications, e.g., risk constraints in portfolio optimization, resource constraints in stochastic shortest path problem, renewable energy aggregators in the local market. We propose a reformulation of the chance constraints using distributionally chance-constrained optimization framework. We show that there exists a saddle point equilibrium of the game, which is the optimal solution of a primal-dual pair of second-order cone programs. As an application, we present a competition of two firms in financial market to simulate our theoretical results.**

Keywords-Distributionally robust chance constraints; Zero-sum game; Saddle point equilibrium; Second-order cone program.

I. INTRODUCTION

This paper is an extended version of [1], presented at the Seventeenth International Conference on Internet and Web Applications and Services (ICIW), from June 26 to June 30, 2022 in Porto, Portugal.

Equilibrium is an important notion in game theory, in which there is no incentive for any player to deviate unilaterally. The researches in the literature usually focus on sufficient conditions for the existence of an equilibrium point and its characterization. The first notion of equilibrium was introduced in the book Researches into the Mathematical Principles of the Theory of Wealth by Cournot in 1838 [2]. In 1951, Nash [3] showed that there exists an equilibrium point in a finite strategic game, which is known as a Nash equilibrium nowadays. The theory of Nash equilibrium is especially hard when it deals with practical applications with random payoffs and strategy sets. In order to deal with random payoffs, the most common way is using the expectation function, which is equivalent to study deterministic payoffs. In many real life applications, the strategy sets are restricted by random linear constraints, which are called chance constraints. The distribution of random factors in chance constraints can be known exactly or unknown, which leads to different approaches to define a game. In known distribution case, the true distribution of random factors is usually assumed to be elliptically distributed, which includes many known distributions, e.g., Gaussian distribution, Laplace distribution, Kotz distribution or Pearson distribution. Otherwise, in unknown distribution case, the true distribution of random factors is assumed to belong to an uncertainty set, where only a partial information of the distribution is known due to historical data and we call these games as distributionally robust chance-constrained games. A two-player zero-sum game is modeled using continuous strategy sets, where the sum of two players' payoffs is zero. Consequently, it is defined using a single payoff function, where one player plays the role of maximizer and another player plays the role of minimizer. More commonly, a zero-sum game is introduced with a payoff matrix, where the rows and the columns are the actions of player 1 and player 2, respectively. A Saddle Point Equilibrium (SPE) is the solution concept to study the zero-sum games and it exists in the mixed strategies [4].

In the conference paper [1], we considered a two player zero-sum game with continuous strategy set, where the payoff function has a special form and the strategies of each player are modeled using random linear constraints reformulated as distributionally robust chance constraints. We proposed an SOCP reformulation of distributionally robust chance constraints under two uncertainty sets based on the partial information about the mean vectors and covariance matrices of the random constraint vectors. We showed the existence of an SPE and characterized it as the optimal solution of a primal-dual pair of SOCPs. The conference paper has some shortcomings, e.g., the payoff function has a quadratic form, the uncertainty sets are mainly constructed based on moments from historical data and it lacks of numerical results which allow us to compare different uncertainty sets. As an extended version of [1], our contribution of this paper is as follows:

- We study a more general framework as compared to [1] by considering two types of uncertainty sets based on either the partial information on the mean vectors and covariance matrices of the random constraint vectors (moments based uncertainty sets) or the statistical distance between their true distribution and a nominal distribution (statistical based uncertainty sets). We show that in both cases, there exists an SPE of the game and an SPE problem is equivalent to a primal-dual pair of SOCPs.
- As an application, we present a competition problem of two firms in financial market and we show our numerical results using randomly generated data to compare different uncertainty sets considered in the paper.

We keep the same form of payoff function as considered in the conference paper, since we need a different game model for different form of payoff function, which would break the uniformity of our paper. We might consider this point in future works.

The rest of the paper is organized as follows. We present related works in Section II. The definition of a distributionally robust zero-sum game is given in Section III. Section IV presents the reformulation of distributionally robust chance constraints as second order cone constraints under different uncertainty sets. Section V outlines a primal-dual pair of SOCPs whose optimal solutions constitute an SPE of the game. Section VI presents a competition of two firms in financial market as and shows numerical results. Conclusion and future works are given in Section VII.

II. Related work

In this section, we introduce previous studies on chance-constrained games. Dantzig and later Adler showed the equivalence between linear programming problems and two-player zero-sum games [5] [6]. Charnes [7] generalized the zero-sum game considered in [4] by introducing linear inequality constraints on the mixed strategies of both the players and called it a constrained zero-sum game. An SPE of a constrained zero-sum game can be obtained from the optimal solutions of a primal-dual pair of linear programs [7]. Singh and Lisser [8] considered a stochastic version of constrained zero-sum game considered by Charnes [7], where the mixed strategies of each player are restricted by random linear inequality constraints, which are modelled using chance constraints. When the random constraint vectors follow a multivariate elliptically symmetric distribution, the zero-sum game problem is equivalent to a primal-dual pair of Second-Order Cone Programs (SOCPs) [8]. Nash equilibrium is the generalization of SPE and it is used as a solution concept for the general-sum games [3] [9]. Under certain conditions on payoff functions and strategy sets, there always exists a Nash equilibrium [10]. The general-sum games under uncertainties are considered in the literature [11]–[15], which capture both risk neutral and risk averse situations. To the best of our knowledge, the distributionally robust chance-constrained approach has been widely studied in the literature but still not completed in game setup. In this paper, we want to apply different approaches in the literature to define uncertainty sets in a distributionally robust chance-constrained game and compare the performance of these approaches by simulation using randomly generated data models.

III. The model

We consider a two player zero-sum game, where each player has continuous strategy set. Let $C^1 \in \mathbb{R}^{K_1\times m}$, $C^2 \in \mathbb{R}^{K_2\times n}$, $d^1 \in \mathbb{R}^{K_1}$ and $d^2 \in \mathbb{R}^{K_2}$. We consider $X = \{x \in \mathbb{R}^m \mid C^1x = d^1,\, x \geq 0\}$ and $Y = \{y \in \mathbb{R}^n \mid C^2y = d^2,\, y \geq 0\}$ as the strategy sets of player 1 and player 2, respectively. We assume that X and Y are compact sets. Let $u : X \times Y \to \mathbb{R}$ be a payoff function associated to the zero-sum game and we assume that player 1 (resp. player 2) is interested in maximizing (resp. minimizing) $u(x, y)$ for a fixed strategy y (resp. x) of player 2 (resp. player 1). For a given strategy pair $(x, y) \in X \times Y$, the payoff function $u(x, y)$ is given by

$$u(x, y) = x^{\mathrm{T}}Gy + g^{\mathrm{T}}x + h^{\mathrm{T}}y, \tag{1}$$

where $G \in \mathbb{R}^{m\times n}$, $g \in \mathbb{R}^m$ and $h \in \mathbb{R}^n$. The first term of (1) results from the interaction between both the players whereas the second and third term represents the individual impact of player 1 and player 2 on the game, respectively. The strategy sets are often restricted by random linear constraints, which are modeled using chance constraints. The chance constraint based strategy sets appear in many practical problems, e.g., risk constraints in portfolio optimization [16]. In this paper, we consider the case, where the strategies of player 1 satisfy the following random linear constraints,

$$(a^1_k)^{\mathrm{T}}x \leq b^1_k,\ k = 1, 2, \ldots, p, \tag{2}$$

whilst the strategies of player 2 satisfy the following random linear constraints

$$(a^2_l)^{\mathrm{T}}y \geq b^2_l,\ l = 1, 2, \ldots, q. \tag{3}$$

Let $\mathcal{I}_1 = \{1, 2, \ldots, p\}$ and $\mathcal{I}_2 = \{1, 2, \ldots, q\}$ be the index sets for the constraints of player 1 and player 2, respectively. For each $k \in \mathcal{I}_1$ and $l \in \mathcal{I}_2$, the vectors a^1_k and a^2_l are random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We consider the case, where the only information we have about the distributions of a^1_k and a^2_l is that they belong to some uncertainty sets $\mathcal{D}^1_k$ and $\mathcal{D}^2_l$, respectively. The uncertainty sets $\mathcal{D}^1_k$ and $\mathcal{D}^2_l$, are constructed based on the partially available information on the distributions of a^1_k and a^2_l, respectively. Using the worst case approach, the random linear constraints (2) and (3) can be formulated as distributionally robust chance constraints given by

$$\inf_{F^1_k \in \mathcal{D}^1_k} \mathbb{P}\left((a^1_k)^{\mathrm{T}}x \leq b^1_k\right) \geq \alpha^1_k,\ \forall\ k \in \mathcal{I}_1, \tag{4}$$

and

$$\inf_{F^2_l \in \mathcal{D}^2_l} \mathbb{P}\left((-a^2_l)^{\mathrm{T}}y \leq -b^2_l\right) \geq \alpha^2_l,\ \forall\ l \in \mathcal{I}_2, \tag{5}$$

where α^1_k and α^2_l are the confidence levels of player 1 and player 2 for kth and lth constraints, respectively. Therefore, for a given $\alpha^1 = (\alpha_k^1)_{k\in\mathcal{I}_1}$ and $\alpha^2 = (\alpha_l^2)_{l\in\mathcal{I}_2}$, the feasible strategy sets of player 1 and player 2 are given by

$$S_{\alpha^1}^1 = \Big\{x \in X| \inf_{F_k^1\in\mathcal{D}_k^1} \mathbb{P}\{(a_k^1)^{\mathrm{T}}x \leq b_k^1\} \geq \alpha_k^1,\ \forall\ k \in \mathcal{I}_1\Big\}, \tag{6}$$

and

$$S_{\alpha^2}^2 = \Big\{y \in Y| \inf_{F_l^2\in\mathcal{D}_l^2} \mathbb{P}\{(-a_l^2)^{\mathrm{T}}y \leq -b_l^2\} \geq \alpha_l^2,\ \forall\ l \in \mathcal{I}_2\Big\}. \tag{7}$$

We call the zero-sum game with the strategy set $S_{\alpha^1}^1$ for player 1 and the strategy set $S_{\alpha^2}^2$ for player 2 as a distributionally robust zero-sum game. We denote this game by Z_α. A strategy pair $(x^*, y^*) \in S_{\alpha^1}^1 \times S_{\alpha^2}^2$ is called an SPE of the game Z_α at $\alpha = (\alpha^1, \alpha^2) \in [0,1]^p \times [0,1]^q$, if

$$u(x, y^*) \leq u(x^*, y^*) \leq u(x^*, y),\ \forall\ x \in S_{\alpha^1}^1,\ y \in S_{\alpha^2}^2. \tag{8}$$

IV. Reformulation of distributionally robust chance constraints

We consider five different uncertainty sets based on the partial information about the mean vectors and covariance matrices of the random constraint vectors a_k^i, $i = 1, 2$, $k \in \mathcal{I}_i$ and four different uncertainty sets based on the statistical distance between the distribution of a_k^i and a nominal distribution. For each uncertainty set, the distributionally robust chance constraints (4) and (5) are reformulated as second-order cone (SOC) constraints.

A. Moments Based Uncertainty Sets

We consider five moments based uncertainty sets defined as follows.

1) Uncertainty set with known mean and known covariance matrix: In some situations, we do not know exactly the true distribution of the random constraint vectors a_k^i, for all $k \in \mathcal{I}_i$, $i = 1, 2$. We can only obtain some information of the underlying distribution from historical data. For example, by observing a sufficiently large number of data, we deduce the values of mean vector and covariance matrix of a_k^i approximated by the sample mean μ_k^i and the sample covariance matrix Σ_k^i. We consider an uncertainty set, which includes all distributions F_k^i with mean vector μ_k^i and covariance matrix Σ_k^i defined as follows

$$\mathcal{D}_k^{1,i}\left(\mu_k^i, \Sigma_k^i\right) = \left\{ F_k^i \;\middle|\; \begin{array}{c} \text{The distribution of } x \text{ is } F_k^i \\ E\left[x\right] = \mu_k^i \\ Cov[x] = \Sigma_k^i \end{array} \right\}, \tag{9}$$

We assume that for each $i = 1, 2$ and $k \in \mathcal{I}_i$, the true distribution of a_k^i belongs to the uncertainty set $\mathcal{D}_k^{1,i}\left(\mu_k^i, \Sigma_k^i\right)$ and the matrix Σ_k^i is a postive definite matrix. This uncertainty set has been widely considered in the literature, e.g., [17]. We present an SOC reformulation of the constraints (4) and (5) by the following lemma.

Lemma 1. *The constraints* (4) *and* (5) *are equivalent to* (10) *and* (11)*, respectively, given by*

$$(\mu_k^1)^T x + \sqrt{\frac{\alpha_k^1}{1-\alpha_k^1}}||(\Sigma_k^1)^{\frac{1}{2}}x||_2 \leq b_k^1,\quad \forall\ k \in \mathcal{I}_1, \tag{10}$$

$$-(\mu_k^2)^T y + \sqrt{\frac{\alpha_k^2}{1-\alpha_k^2}}||(\Sigma_k^2)^{\frac{1}{2}}y||_2 \leq -b_k^2,\quad \forall\ k \in \mathcal{I}_2. \tag{11}$$

Remark 1. *An SOC constraint is the set of points $x \in \mathbb{R}^n$ such that the following inequality holds*

$$\|Ax + b\|_2 \leq c^T x + d,$$

where $A \in \mathbb{R}^{m\times n}$ is an $m \times n$ real matrix, $b \in \mathbb{R}^m$ is an $m \times 1$ real vector, $c \in \mathbb{R}^n$ is an $n \times 1$ real vector and $d \in \mathbb{R}$ is a real number, $\|\cdot\|_2$ denotes the Euclidean norm. It is clear that (10) *and* (11) *are equivalent to SOC constraints. An SOC reformulation is useful since optimization problems with SOC constraints can be solved efficiently by known algorithms in polynomial time.*

Proof. Using the one-sided Chebyshev inequality, we have

$$\inf_{F_k^1\in\mathcal{D}_k^{1,i}(\mu,\Sigma)} \mathbb{P}\big\{(a_k^1)^{\mathrm{T}}x \leq b_k^1\big\} = \begin{cases} 1 - \frac{1}{1+\frac{((\mu_k^1)^{\mathrm{T}}x - b_k^1)^2}{\left(x^{\mathrm{T}}\Sigma_k^1 x\right)}}, \\ \text{if } \ (\mu_k^1)^{\mathrm{T}}x \leq b_k^1, \\ \\ 0, \ \text{otherwise}. \end{cases}$$

The bound of one-sided Chebyshev inequality can be achieved by a two-point distribution given by equation (2) of [18]. For the case $(\mu_k^1)^{\mathrm{T}}x > b_k^1$,

$$\inf_{F_k^1\in\mathcal{D}_k^{1,i}(\mu,\Sigma)} \mathbb{P}\big\{a_k^1 x \leq b_k^1\big\} = 0,$$

which makes constraint (4) infeasible for any $\alpha_1 > 0$. Therefore, for $x \in S_{\alpha_1}^1$, the condition $(\mu_k^1)^{\mathrm{T}}x \leq b_k^1$ always holds and the constraint (4) is equivalent to

$$1 - \frac{1}{1 + ((\mu_k^1)^{\mathrm{T}}x - b_k^1)^2/(x^{\mathrm{T}}\Sigma_k^1 x)} \geq \alpha_k^1.$$

The above inequality can be reformulated as (10). Similarly, we can show that (5) is equivalent to (11). □

2) Uncertainty set with known mean and unknown covariance matrix: For all $i = 1, 2$ and $k \in \mathcal{I}_i$, we consider the case, where the mean vector of the random vector a_k^i is known exactly (approximated by the sample mean μ_k^i) but the covariance matrix is unknown due to several reasons, e.g., the lack of data. We assume that it is only known to belong to a positive semidefinite cone defined with a linear matrix inequality as follows

$$Cov[a_k^i] \preceq \gamma_k^i \Sigma_k^i,$$

where $\gamma_k^i > 0$ is a strictly positive real number, Σ_k^i is a positive definite matrix, for the given matrices B_1 and B_2, $B_1 \preceq B_2$ implies that $B_2 - B_1$ is a positive semidefinite matrix. In practical applications, we usually approximate the matrix Σ_k^i by the sample covariance matrix. The parameter γ_k^i is used in controlling the uncertainty level, i.e., high value of γ_k^i implies a large number of distributions in the uncertainty set, which deals uncertain factors in a more secure way. We consider un uncertainty set, which includes all distributions F_k^i with mean vector μ_k^i and covariance matrix satisfied the above constraint as follows

$$\mathcal{D}_k^{2,i}\left(\mu_k^i, \Sigma_k^i\right) = \left\{ F_k^i \;\middle|\; \begin{array}{c} \text{The distribution of } x \text{ is } F_k^i \\ E\left[x\right] = \mu_k^i \\ Cov[x] \preceq \gamma_k^i \Sigma_k^i \end{array} \right\}. \tag{12}$$

This uncertainty set is considered in [19]. We assume that for each $i = 1, 2$ and $k \in \mathcal{I}_i$, the true distribution of a_k^i belongs to the uncertainty set $\mathcal{D}_k^{2,i}\left(\mu_k^i, \Sigma_k^i\right)$. We present an SOC reformulation of the constraints (4) and (5) by the following lemma.

Lemma 2. *The constraints* (4) *and* (5) *are equivalent to* (13) *and* (14)*, respectively, given by*

$$(\mu_k^1)^T x + \sqrt{\gamma_k^1}\sqrt{\frac{\alpha_k^1}{1-\alpha_k^1}}||(\Sigma_k^1)^{\frac{1}{2}} x||_2 \le b_k^1, \quad \forall\ k \in \mathcal{I}_1, \tag{13}$$

$$-(\mu_k^2)^T y + \sqrt{\gamma_k^2}\sqrt{\frac{\alpha_k^2}{1-\alpha_k^2}}||(\Sigma_k^2)^{\frac{1}{2}} y||_2 \le -b_k^2, \quad \forall\ k \in \mathcal{I}_2. \tag{14}$$

Proof. Based on the structure of uncertainty set (12), the constraint (4) can be written as

$$\inf_{(\mu,\Sigma)\in\mathcal{U}_k^1}\ \inf_{F_k^1\in\mathcal{D}_k^{1,i}(\mu,\Sigma)} \mathbb{P}\left\{(a_k^1)^{\mathrm{T}} x \le b_k^1\right\} \ge \alpha_k^1,$$

where

$$\mathcal{U}_k^1 = \left\{(\mu,\Sigma)\ \middle|\mu = \mu_k^1, \Sigma \preceq \gamma_k^i \Sigma_k^i\right\}.$$

Here, the inner infimum is taken over all distributions with same value of mean vector and covariance matrix. The outer infimum is taken over all couples (μ, Σ) satisfying the conditions in (12). Using the similar arguments as in the Lemma 1, the constraint (4) is equivalent to

$$\frac{b_k^1 - (\mu_k^1)^{\mathrm{T}} x}{\max\limits_{\Sigma\preceq\gamma_k^1\Sigma_k^1} \sqrt{x^{\mathrm{T}}\Sigma x}} \ge \sqrt{\frac{\alpha_k^1}{1-\alpha_k^1}}. \tag{15}$$

The above inequality (15) can be reformulated as (13). Similarly, we can show that (5) is equivalent to (14). □

3) Uncertainty set with unknown mean and unknown covariance matrix: For all $i = 1, 2$ and $k \in \mathcal{I}_i$, we consider the case, where both mean vector and covariance matrix of a_k^i are unknown. From historical data, we obtain the sample mean μ_k^i and the sample covariance matrix Σ_k^i. We deal the uncertainty level in a secure way by assuming that the mean vector and the covariance matrix of a_k^i are not exactly the same as its sample mean and sample covariance matrix. The mean vector lies in an ellipsoid of size $\gamma_{k1}^i \ge 0$ centered at μ_k^i defined by the following constraint

$$\left(\mathbb{E}[a_k^i] - \mu_k^i\right)^\top \left(\Sigma_k^i\right)^{-1}\left(\mathbb{E}[a_k^i] - \mu_k^i\right) \le \gamma_{k1}^i,$$

and the covariance matrix of a_k^i lies in a positive semidefinite cone defined as follows

$$Cov[a_k^i] \preceq \gamma_{k2}^i \Sigma_k^i.$$

where $\gamma_{k2}^i > 0$ and Σ_k^i is a positive definite matrix. The parameters γ_{k1}^i and γ_{k2}^i are used in controlling the uncertainty level. If $\gamma_{k1}^i = 0$, the mean vector is exactly the same as its sample mean. We consider un uncertainty set, which includes all distributions F_k^i with mean vector and covariance matrix satisfied the above constraints as follows

$$\mathcal{D}_k^{3,i}(\mu_k^i, \Sigma_k^i) = \left\{ F_k^i \;\middle|\; \begin{array}{l} \text{The distribution of } x \text{ is } F_k^i \\ \\ \left(\mathbb{E}[x] - \mu_k^i\right)^\top \left(\Sigma_k^i\right)^{-1} \\ \times \left(\mathbb{E}[x] - \mu_k^i\right) \le \gamma_{k1}^i, \\ \\ Cov[x] \preceq \gamma_{k2}^i \Sigma_k^i \end{array} \right\}, \tag{16}$$

The uncertainty set (16) is considered in [20]. We assume that for each $i = 1, 2$ and $k \in \mathcal{I}_i$, the true distribution of a_k^i belongs to the uncertainty set $\mathcal{D}_k^{3,i}\left(\mu_k^i, \Sigma_k^i\right)$. We present an SOC reformulation of the constraints (4) and (5) by the following lemma.

Lemma 3. *The constraints* (4) *and* (5) *are equivalent to* (17) *and* (18)*, respectively, given by*

$$(\mu_k^1)^T x + \left(\sqrt{\frac{\alpha_k^1}{1-\alpha_k^1}}\sqrt{\gamma_{k2}^1} + \sqrt{\gamma_{k1}^1}\right)\left\|\left(\Sigma_k^1\right)^{\frac{1}{2}} x\right\|_2 \le b_k^1, \quad \forall\ k \in \mathcal{I}_1, \tag{17}$$

$$-(\mu_k^2)^T y + \left(\sqrt{\frac{\alpha_k^2}{1-\alpha_k^2}}\sqrt{\gamma_{k2}^2} + \sqrt{\gamma_{k1}^2}\right)\left\|\left(\Sigma_k^2\right)^{\frac{1}{2}} y\right\|_2 \le -b_k^2, \quad \forall\ k \in \mathcal{I}_2. \tag{18}$$

Proof. Based on the structure of the uncertainty set (16), the constraint (4) can be written as

$$\inf_{(\mu,\Sigma)\in\tilde{\mathcal{U}}_k^1}\ \inf_{F_k^1\in\mathcal{D}_k^{1,i}(\mu,\Sigma)} \mathbb{P}\left\{a_k^1 x \le b_k^1\right\} \ge \alpha_k^1,$$

where

$$\tilde{\mathcal{U}_k^1} = \left\{(\mu,\Sigma) \;\middle|\; \begin{array}{l} \left(\mu - \mu_k^1\right)^\top \left(\Sigma_k^1\right)^{-1}\left(\mu - \mu_k^1\right) \le \gamma_{k1}^1, \\ \Sigma \preceq \gamma_{k2}^1 \Sigma_k^1. \end{array} \right\}.$$

Using the similar arguments as in the Lemma 1, the constraint (4) is equivalent to

$$\frac{b_k^1 + v_1(x)}{\sqrt{v_2(x)}} \geq \sqrt{\frac{\alpha_k^1}{1-\alpha_k^1}}, \tag{19}$$

where

$$v_1(x) = \begin{cases} \min_{\mu} -\mu^{\mathrm{T}} x \\ \text{s.t.} \quad \left(\mu - \mu_k^1\right)^{\top} \left(\Sigma_k^1\right)^{-1} \left(\mu - \mu_k^1\right) \leq \gamma_{k1}^1, \end{cases}$$
$$v_2(x) = \begin{cases} \max_{\Sigma} \; x^{\mathrm{T}} \Sigma x \\ \text{s.t. } \Sigma \preceq \gamma_{k2}^1 \Sigma_k^1. \end{cases} \tag{20}$$

Let $\beta \geq 0$ be a Lagrange multiplier associated with the constraint of optimization problem (20). By applying the KKT conditions, the optimal solution of (20) is given by $\mu = \mu_k^1 + \frac{\sqrt{\gamma_{k1}^1}\Sigma_k^1 x}{\sqrt{x^{\mathrm{T}}\Sigma_k^1 x}}$ and the associated Lagrange multiplier is given by $\beta = \sqrt{\frac{x^{\mathrm{T}}\Sigma_k^1 x}{4\gamma_{k1}^1}}$. Therefore, the corresponding optimal value $v_1(x) = -(\mu_k^1)^{\mathrm{T}} x - \sqrt{\gamma_{k1}^1}\sqrt{x^{\mathrm{T}}\Sigma_k^1 x}$. Since, $u^{\mathrm{T}}\Sigma u \leq u^{\mathrm{T}}\gamma_{k2}^1\Sigma_k^1 u$, then, $v_2(x) = \gamma_{k2}^1 x^{\mathrm{T}}\Sigma_k^1 x$. Therefore, using (19), (4) is equivalent to (17). Similarly, we can show that (5) is equivalent to (18). ☐

4) Polytopic uncertainty set: For all $i = 1, 2$ and $k \in \mathcal{I}_i$, we consider the case, where both mean vector and covariance matrix of the random vector a_k^i are unknown. From historical data, we consider M samples i.i.d of the random vector a_k^i. We obtain M sample means $\mu_{k1}^i, \ldots, \mu_{kM}^i$ and M sample covariance matrix $\Sigma_{k1}^i, \ldots, \Sigma_{kM}^i$, where Σ_{kj}^i is positive definite, for any $j = 1, \ldots, M$. We consider polytopes $U_{\mu_k^i} = \text{Conv}(\mu_{k1}^i, \mu_{k2}^i, \ldots, \mu_{kM}^i)$ and $U_{\Sigma_k^i} = \text{Conv}(\Sigma_{k1}^i, \Sigma_{k2}^i, \ldots, \Sigma_{kM}^i)$, where Conv denotes the convex hull. We assume that the mean vector and the covariance matrix of a_k^i are known to belong to polytopes $U_{\mu_k^i}$ and $U_{\Sigma_k^i}$, respectively. We consider an uncertainty set, which includes all distributions F_k^i defined as follows

$$\mathcal{D}_k^{4,i}\left(\mu_k^i, \Sigma_k^i\right) = \left\{ F_k^i \;\middle|\; \begin{array}{c} \text{The distribution of } x \text{ is } F_k^i \\ E[x] \in U_{\mu_k^i} \\ Cov[x] \in U_{\Sigma_k^i} \end{array} \right\}. \tag{21}$$

The uncertainty set (21) is considered in [17]. We assume that for each $i = 1, 2$ and $k \in \mathcal{I}_i$, the true distribution of a_k^i belongs to the uncertainty set $\mathcal{D}_k^{4,i}\left(\mu_k^i, \Sigma_k^i\right)$. We present an SOC reformulation of the constraints (4) and (5) by the following lemma.

Lemma 4. *The constraints* (4) *and* (5) *are equivalent to* (22) *and* (23), *respectively, given by*

$$(\mu_{kj}^1)^T x + \sqrt{\frac{\alpha_k^1}{1-\alpha_k^1}} ||(\Sigma_{kw}^1)^{\frac{1}{2}} x||_2 \leq b_k^1,$$
$$\forall\; j = 1, \ldots, M, \; w = 1, \ldots, M, \; k \in \mathcal{I}_1, \tag{22}$$

$$-(\mu_{kj}^2)^T y + \sqrt{\frac{\alpha_k^2}{1-\alpha_k^2}} ||(\Sigma_{kw}^2)^{\frac{1}{2}} y||_2 \leq -b_k^2,$$
$$\forall\; j = 1, \ldots, M, \; w = 1, \ldots, M, \; k \in \mathcal{I}_2. \tag{23}$$

Remark 2. *Lemma 4 shows that the constraint* (4) *(resp.* (5)*) is equivalent to a system of M^2 constraints in* (22) *(resp.* (23)*).*

Proof. Based on the structure of uncertainty set (21), the constraint (4) can be written as

$$\inf_{(\mu,\Sigma)\in\hat{\mathcal{U}}_k^1} \; \inf_{F_k^1\in\mathcal{D}_k^{1,i}(\mu,\Sigma)} \mathbb{P}\left\{(a_k^1)^{\mathrm{T}} x \leq b_k^1\right\} \geq \alpha_k^1,$$

where

$$\hat{\mathcal{U}}_k^1 = \left\{(\mu, \Sigma) \;\middle|\; \mu \in U_{\mu_k^1}, \Sigma \in U_{\Sigma_k^1}\right\}.$$

Using the similar arguments as in the Lemma 1, the constraint (4) can be reformulated as

$$\frac{\min_{\mu\in U_{\mu_k^1}} \left(b_k^1 - \mu^{\mathrm{T}} x\right)}{\max_{\Sigma\in U_{\Sigma_k^1}} \sqrt{x^{\mathrm{T}}\Sigma x}} \geq \sqrt{\frac{\alpha_k^1}{1-\alpha_k^1}}. \tag{24}$$

The above inequality (24) can be reformulated as

$$\frac{b_k^1 - (\mu_{kj}^1)^{\mathrm{T}} x}{\sqrt{x^{\mathrm{T}}\Sigma_{kw}^1 x}} \geq \sqrt{\frac{\alpha_k^1}{1-\alpha_k^1}},$$
$$\forall\, j = 1, \ldots, M, \, w = 1, \ldots, M, \, k \in \mathcal{I}_1,$$

which is equivalent to (22). Similarly, we can show that (5) is equivalent to (23). ☐

5) Uncertainty set with componentwise bounds on mean vector and covariance matrix: For all $i = 1, 2$ and $k \in \mathcal{I}_i$, we consider the case, where the mean vector and the covariance matrix of a_k^i are unknown. We obtain from historical data, a sample mean vector μ_k^i and a sample covariance matrix Σ_k^i. We do not approximate the mean vector and the covariance matrix of a_k^i by its sample mean vector and sample covariance matrix, but we deal the uncertainty level by a more secure way. For each $j = 1, \ldots, m$, we assume that the j^{th}-component of the mean vector of a_k^i lies in a ball of radius $\epsilon_{\mu,k}^i(j) \geq 0$, centered at the j^{th}-component of the sample mean vector μ_k^i, which can be reformulated as follows

$$\mu_k^i - \epsilon_{\mu,k}^i \leq \mathbb{E}[a_k^i] \leq \mu_k^i + \epsilon_{\mu,k}^i,$$

where $\epsilon_{\mu,k}^i = \left(\epsilon_{\mu,k}^i(1), \ldots, \epsilon_{\mu,k}^i(m)\right)$ is an $m \times 1$ vector and the above inequalities are understood componentwise. Similarly, for each $j = 1, \ldots, m$ and $w = 1, \ldots, m$, we assume that the $(j, w)-$ entry of the covariance matrix of a_k^i lies in a ball of radius $\epsilon_{\Sigma,k}^i(j, w) \geq 0$, centered at the $(j, w)-$ entry of the sample covariance matrix Σ_k^i, which can be reformulated as follows

$$\Sigma_k^i - \epsilon_{\Sigma,k}^i \leq \text{Cov}[a_k^i] \leq \mu_k^i + \epsilon_{\Sigma,k}^i,$$

where $\epsilon_{\Sigma,k}^i = \epsilon_{\Sigma,k}^i(j, w)_{1\leq j,w\leq m}$ is an $m \times m$ matrix. Let $\mu_{k-}^i = \mu_k^i - \epsilon_{\mu,k}^i, \mu_{k+}^i = \mu_k^i + \epsilon_{\mu,k}^i, \quad \Sigma_{k-}^i = \Sigma_k^i - \epsilon_{\Sigma,k}^i,$

and $\Sigma^i_{k+} = \Sigma^i_k + \epsilon^i_{\Sigma,k}$. We consider an uncertainty set, which includes all distributions F^i_k defined as follows

$$\mathcal{D}^{5,i}_k(\mu^i_k, \Sigma^i_k) = \left\{ F^i_k \;\middle|\; \begin{array}{l} \text{The distribution of } x \text{ is } F^i_k \\ \mu^i_{k-} \leq \mathbb{E}[x] \leq \mu^i_{k+}, \\ \Sigma^i_{k-} \leq \text{Cov}[x] \leq \Sigma^i_{k+} \end{array} \right\}, \tag{25}$$

Since Σ^i_k is a positive definite matrix, we can take $\epsilon^i_{\Sigma,k} > 0$ such that for any matrix H, if $\Sigma^i_{k-} \leq H \leq \Sigma^i_{k+}$, then H is a positive definite matrix. We define a set of vectors S^1_k as follows

$$S^1_k = \left\{ \mu \in \mathbb{R}^m \,|\, \mu(j) = \mu^1_{k-}(j) \,\text{or}\, \mu^1_{k+}(j),\, \forall\, j = 1, \ldots, m \right\},$$

where $\mu(j)$ is the $j^{\text{th}}-$ component of μ, $\mu^1_{k-}(j)$ is the $j^{\text{th}}-$ component of μ^1_{k-}, and $\mu^1_{k+}(j)$ is the $j^{\text{th}}-$ component of μ^1_{k+}. For example, if $\mu^1_{k-} = (1,2)^{\text{T}}, \mu^1_{k+} = (5,6)^{\text{T}}$, then S^1_k is a set of 4 vectors $\left\{(1,5)^{\text{T}}, (1,6)^{\text{T}}, (2,5)^{\text{T}}, (2,6)^{\text{T}}\right\}$. We define a set of covariance matrix T^1_k as follows

$$T^1_k = \left\{ \Sigma \,|\, \Sigma(j,w) = \Sigma^1_{k-}(j,w) \,\text{or}\, \Sigma^1_{k+}(j,w),\, 1 \leq j, w \leq m \right\},$$

Similarly, we define a set of vectors S^2_k and a set of covariance matrix T^2_k. The uncertainty set (25) is considered in [17]. We assume that for each $i = 1, 2$ and $k \in \mathcal{I}_i$, the true distribution of a^i_k belongs to the uncertainty set $\mathcal{D}^{5,i}_k\left(\mu^i_k, \Sigma^i_k\right)$. We present an SOC reformulation of the constraints (4) and (5) by the following lemma.

Lemma 5. *The constraints* (4) *and* (5) *are equivalent to* (26) *and* (27)*, respectively, given by*

$$(\mu^1)^T x + \sqrt{\frac{\alpha^1_k}{1-\alpha^1_k}} ||(\Sigma^1)^{\frac{1}{2}} x||_2 \leq b^1_k, \quad \forall\, \mu^1 \in S^1_k,\ \Sigma^1 \in T^1_k,\ k \in \mathcal{I}_1, \tag{26}$$

$$-(\mu^2)^T y + \sqrt{\frac{\alpha^2_k}{1-\alpha^2_k}} ||(\Sigma^2)^{\frac{1}{2}} y||_2 \leq -b^2_k, \quad \forall\, \mu^2 \in S^2_k,\ \Sigma^2 \in T^2_k,\ k \in \mathcal{I}_2. \tag{27}$$

Remark 3. *Note that S^1_k is a set of 2^m vectors and T^1_k is a set of 2^{m^2} matrix. Then, Lemma 5 shows that the constraint* (4) *is equivalent to a system of $2^m \times 2^{m^2}$ constraints in* (26)*, for any $k \in \mathcal{I}_1$ and the constraint* (5) *is equivalent to a system of $2^n \times 2^{n^2}$ constraints in* (27)*, for any $k \in \mathcal{I}_2$.*

Proof. Based on the structure of the uncertainty set (25), the constraint (4) can be written as

$$\inf_{(\mu,\Sigma)\in\bar{\mathcal{U}}^1_k} \inf_{F^1_k \in \mathcal{D}^{1,i}_k(\mu,\Sigma)} \mathbb{P}\left\{a^1_k x \leq b^1_k\right\} \geq \alpha^1_k,$$

where

$$\bar{\mathcal{U}}^1_k = \left\{ (\mu, \Sigma) \;\middle|\; \begin{array}{l} \mu^1_{k-} \leq \mu \leq \mu^1_{k+}, \\ \Sigma^1_{k-} \leq \mu \leq \Sigma^i_{k+}. \end{array} \right\}.$$

Using the similar arguments as in the Lemma 1, the constraint (4) is equivalent to

$$\frac{b^1_k + v_1(x)}{\sqrt{v_2(x)}} \geq \sqrt{\frac{\alpha^1_k}{1-\alpha^1_k}}, \tag{28}$$

where

$$v_1(x) = \begin{cases} \min\limits_{\mu} -\mu^{\text{T}} x \\ \text{s.t. } \ \mu^1_{k-} \leq \mu \leq \mu^1_{k+}, \end{cases}$$

$$v_2(x) = \begin{cases} \max\limits_{\Sigma} \ x^{\text{T}} \Sigma x \\ \text{s.t. } \Sigma^1_{k-} \leq \Sigma \leq \Sigma^i_{k+}. \end{cases}$$

Note that the objective functions $-\mu^{\text{T}} x$ and $x^{\text{T}} \Sigma x$ are linear functions w.r.t μ (resp. Σ). Then, it is clear that the optimal values $v_1(x)$ and $v_2(x)$ hold only when $\mu \in S^1_k$ and $\Sigma \in T^1_k$. Then, the constraint (4) can be reformulated as (26). Similarly, we can show that (5) is equivalent to (27). ☐

B. Statistical Distance Based Uncertainty Sets

In this section, we define uncertainty sets using a metric called $\phi-$divergence. For any $i = 1, 2$ and $k \in \mathcal{I}_i$, the decision makers (the two players in the game) believe that the true distribution of a^i_k oscillates around a Normal distribution of mean vector μ^i_k and covariance matrix Σ^i_k, where μ^i_k and Σ^i_k are sample mean vector and sample covariance matrix obtained from historical data. We assume that the true distribution of a^i_k lies in a ball of radius θ^i_k, centered at a nominal distribution ν^i_k and the distance between these two distributions is given by $\phi-$divergence metric. The nominal distribution ν^i_k is assumed to be Normal distributed of mean vector μ^i_k and covariance matrix Σ^i_k.

Definition 1. *The $\phi-$divergence distance between two measures μ and ν with densities f_μ and f_ν, respectively, with support in $\mathbb{R}^{r_i}$ is defined as follows:*

$$I_\phi(\mu, \nu) = \int_{\mathbb{R}^{r_i}} \phi\left(\frac{f_\mu(\xi)}{f_\nu(\xi)}\right) f_\nu(\xi) d\xi.$$

where $r_1 = m$ and $r_2 = n$.

There are different types of $\phi-$divergences distance, we refer to [21] and [22] for different choices of function ϕ. We consider an uncertainty set $\mathcal{D}^{\phi,i}_k$ defined as follows

$$\mathcal{D}^{\phi,i}_k = \left\{ F^i_k \in \mathcal{M}+ \,|\, I_\phi(F^i_k, \nu^i_k) \leq \theta^i_k \right\}, \tag{29}$$

where $\mathcal{M}^i+$ is the set of all probability measures on $\mathbb{R}^{r_i}$, with $r_1 = m$, $r_2 = n$, and $\theta^i_k > 0$. This uncertainty set is considered in [23]. We assume that for each $i = 1, 2$ and $k \in \mathcal{I}_i$, the true distribution of a^i_k belongs to the uncertainty set $\mathcal{D}^{\phi,i}_k\left(\mu^i_k, \Sigma^i_k\right)$.

Definition 2. *The conjugate of the function ϕ is a function $\phi^* : \mathbb{R} \to \mathbb{R} \cup +\infty$ such that*

$$\phi^*(s) = \sup_{t \geq 0} \left\{ st - \phi(t) \right\}.$$

We study some special cases of $\phi-$divergences, which are summarized in Table I. The data of Table I are taken from [21]. The following lemma provides the first reformulation of the constraints (4) and (5).

TABLE I
LIST OF SELECTED $\phi-$DIVERGENCES WITH THEIR CONJUGATE RESPECTIVELY

Divergence	$\phi(t), t \geq 0$	$\phi^*(s)$
Kullback-Leibler	$t\log(t) - t + 1.$	$e^s - 1$
Variation distance	$\|t-1\|.$	$-1, \quad s \leq -1,$ $s, \quad -1 \leq s \leq 1,$ $+\infty, \quad s > 1.$
Modified χ^2 - distance	$(t-1)^2.$	$-1, \quad s \leq -2,$ $s + \frac{s^2}{4}, \quad s > -2.$
Hellinger distance	$(\sqrt{t}-1)^2.$	$\frac{s}{1-s}, \quad s < 1,$ $+\infty, \quad s \geq 1.$

Lemma 6. *The constraint* (4) *is equivalent to*

$$\sup_{\lambda>0, \beta\in\mathbb{R}} \left\{ f_k^1(\lambda,\beta) \right\} \geq \alpha_k^1, \tag{30}$$

where $f_k^1(\lambda,\beta) = \beta - \lambda\theta_k^1 - \lambda\phi^*\left(\frac{-1+\beta}{\lambda}\right)\mathbb{P}_{\nu_k^1}(M_k^1) - \lambda\phi^*\left(\frac{\beta}{\lambda}\right)\left[1 - \mathbb{P}_{\nu_k^1}(M_k^1)\right]$, *and* $M_k^1 = \left\{\xi \in \mathbb{R}^m \mid \xi^T x \leq b_k^1\right\}$. *The constraint* (5) *is equivalent to*

$$\sup_{\lambda>0, \beta\in\mathbb{R}} \left\{ f_k^2(\lambda,\beta) \right\} \geq \alpha_k^2,$$

where $f_k^2(\lambda,\beta) = \beta - \lambda\theta_k^2 - \lambda\phi^*\left(\frac{-1+\beta}{\lambda}\right)\mathbb{P}_{\nu_k^2}(M_k^2) - \lambda\phi^*\left(\frac{\beta}{\lambda}\right)\left[1 - \mathbb{P}_{\nu_k^2}(M_k^2)\right]$, *and* $M_k^2 = \left\{\xi \in \mathbb{R}^n \mid \xi^T x \leq b_k^2\right\}$.

Proof. For $k \in \mathcal{I}_1$, consider the following optimization problem

$$v_k^{\mathrm{P}} = \inf_{F_k^1 \in \mathcal{D}_k^{\phi,1}} \mathbb{P}\left((a_k^1)^{\mathrm{T}} x \leq b_k^1\right).$$

The above problem is rewritten as

$$\begin{aligned} v_{\mathrm{P}}^k = \inf_{F\geq 0} & \int_{\mathbb{R}^m} \mathbf{1}_{M_k^1}(\xi) F(\xi) \mathrm{d}\xi \\ \text{s.t.} \quad (i) \quad & \int_{\mathbb{R}^m} f_{\nu_k^1}(\xi)\phi\left(\frac{F(\xi)}{f_{\nu_k^1}(\xi)}\right)\mathrm{d}\xi \leq \theta_k^1, \\ (ii) \quad & \int_{\mathbb{R}^m} F(\xi)\mathrm{d}\xi = 1, \end{aligned} \tag{31}$$

where the infimum value is taken over all positive measures on $\mathbb{R}^m$. The Lagrangian dual of (31) can be written as follows

$$v_{\mathrm{D}}^k = \sup_{\lambda \geq 0, \beta\in\mathbb{R}} \left\{ \beta - \lambda\theta_k^1 + \inf_{F(\xi)\geq 0} \int_{\mathbb{R}^m} g_k^1(\lambda,\beta) \right\},$$

where $g_k^1(\lambda,\beta) = \mathbf{1}_{M_k^1}(\xi)F(\xi) - \beta F(\xi) + \lambda f_{\nu_k^1}(\xi)\phi\left(\frac{F(\xi)}{f_{\nu_k^1}(\xi)}\right)\mathrm{d}\xi$, λ is the dual variable of the constraint (i) and β is the dual variable of the constraint (ii). Since $\theta_k^1 > 0$, the Slater's condition holds, then the strong duality holds, i.e., $v_{\mathrm{P}}^k = v_{\mathrm{D}}^k$. The rest of the proof follows from Theorem 1 [23]. ☐

We present an SOC reformulation of the constraints (4) and (5) by the following lemma.

Lemma 7. *The constraints* (4) *and* (5) *are equivalent to* (32) *and* (33), *respectively, given by:*

$$(\mu_k^1)^T x + \Phi^{(-1)}\left[H(\theta_k^1, 1-\alpha_k^1)\right]\left\|\left(\Sigma_k^1\right)^{\frac{1}{2}} x\right\|_2 \leq b_k^1, \quad \forall\, k \in \mathcal{I}_1, \tag{32}$$

$$-(\mu_k^2)^T y + \Phi^{(-1)}\left[H(\theta_k^2, 1-\alpha_k^2)\right]\left\|\left(\Sigma_k^2\right)^{\frac{1}{2}} y\right\|_2 \leq -b_k^2, \quad \forall\, k \in \mathcal{I}_2. \tag{33}$$

where $\Phi^{(-1)}$ *is the quantile of the standard Normal distribution and* H *is a function whose value is given in Table II.*

TABLE II
LIST OF SELECTED $\phi-$DIVERGENCES WITH THE FUNCTION f RESPECTIVELY

Divergence	$H(\theta,\epsilon) =$	θ, ϵ
Kullback-Leibler	$\inf_{x\in(0,1)} \frac{e^{-\theta}x^{1-\epsilon}-1}{x-1}$	$\theta > 0$ $0 < \epsilon < 1$
Variation distance	$1 - \epsilon + \frac{\theta}{2}$	$\theta > 0$ $0 < \epsilon < 1$
Modified χ^2 - distance	$1 - \epsilon + \frac{\sqrt{\theta^2+4\theta(\epsilon-\epsilon^2)}-(1-2\epsilon)\theta}{2\theta+2},$	$\theta > 0$ $0 < \epsilon < \frac{1}{2}$
Hellinger distance	$\frac{-B+\sqrt{\Delta}}{2},$ where $B = -(2-(2-\theta)^2)\epsilon - \frac{(2-\theta)^2}{2},$ $C = \left(\frac{(2-\theta)^2}{4} - \epsilon\right)^2,$ $\Delta = B^2 - 4C = (2-\theta)^2\left[4-(2-\theta)^2\right]\epsilon(1-\epsilon),$	$0 < \theta < 2-\sqrt{2}$ $0 < \epsilon < 1$

Proof. Using Lemma 6, we prove that the constraint (4) is equivalent to

$$\mathbb{P}_{\nu_k^1}(M_k^1) \geq H(\theta_k^1, 1-\alpha_k^1). \tag{34}$$

Since ν_k^1 follows a Normal distribution with mean vector μ_k^1 and covariance matrix Σ_k^1, it is well known that (34) is equivalent to the SOC constraint (32). We refer to Propositions 2, 3, and 4, [23] for the proof of the cases Kullback-Leibler, Variation distance and Modified χ^2 - distance. The proof of the case Hellinger distance is given in Appendix A. ☐

C. Second Order Cone Reformulation

In this section, we summarize our SOC reformulation results from Lemmas 1, 2, 3, 4, 5, and 7. They show that in all cases of uncertainty sets defined in Sections IV-A and IV-B, the feasible strategy sets (6) and (7) can be written as

$$S_{\alpha^1}^1 = \Big\{ x \in X \mid (\mu_{kj}^1)^{\mathrm{T}} x + \kappa_{\alpha_k^1} ||(\Sigma_{kw}^1)^{\frac{1}{2}} x||_2 \leq b_k^1, \quad \forall\, j = 1,2,\ldots,N_1,\ w = 1,2,\ldots,P_1,\ k \in \mathcal{I}_1 \Big\}, \tag{35}$$

and

$$S_{\alpha^2}^2 = \Big\{ y \in Y \mid -(\mu_{lj}^2)^{\mathrm{T}} y + \kappa_{\alpha_l^2} ||(\Sigma_{lw}^2)^{\frac{1}{2}} y||_2 \leq -b_l^2, \quad \forall\, j = 1,2,\ldots,N_2,\ w = 1,2,\ldots,P_2,\ l \in \mathcal{I}_2. \Big\}. \tag{36}$$

- If the uncertainty set is defined by (9), then $\kappa_{\alpha_k^i} = \sqrt{\frac{\alpha_k^i}{1-\alpha_k^i}}$ and $N_1 = P_1 = N_2 = P_2 = 1$, for all $i = 1, 2$, $k \in \mathcal{I}_i$.

- If the uncertainty set is defined by (12), then $\kappa_{\alpha_k^i} = \sqrt{\frac{\alpha_k^i}{1-\alpha_k^i}}\sqrt{\gamma_k^i}$ and $N_1 = P_1 = N_2 = P_2 = 1$, for all $i = 1, 2, \quad k \in \mathcal{I}_i$.
- If the uncertainty set is defined by (16), then $\kappa_{\alpha_k^i} = \left(\sqrt{\frac{\alpha_k^i}{1-\alpha_k^i}}\sqrt{\gamma_{k2}^i} + \sqrt{\gamma_{k1}^i}\right)$ and $N_1 = P_1 = N_2 = P_2 = 1$, for all $i = 1, 2, \quad k \in \mathcal{I}_i$.
- If the uncertainty set is defined by (21), then $\kappa_{\alpha_k^i} = \sqrt{\frac{\alpha_k^i}{1-\alpha_k^i}}$ and $N_1 = P_1 = N_2 = P_2 = M$, for all $i = 1, 2$, $k \in \mathcal{I}_i$.
- If the uncertainty set is defined by (25), then $\kappa_{\alpha_k^i} = \sqrt{\frac{\alpha_k^i}{1-\alpha_k^i}}$ and $N_1 = 2^m; P_1 = 2^{(m^2)}, N_2 = 2^n, P_2 = 2^{(n^2)}$, for all $i = 1, 2, \quad k \in \mathcal{I}_i$.
- If the uncertainty set is defined by (29), then $\kappa_{\alpha_k^i} = \Phi^{(-1)}\left[H(\theta_k^i, 1-\alpha_k^i)\right]$ and $N_1 = P_1 = N_2 = P_2 = 1$, where H and $\Phi^{(-1)}$ are defined in Lemma 7.

We assume that the strategy sets (35) and (36) satisfy the strict feasibility condition given by Assumption 1.

Assumption 1. 1) *There exists an $x \in S_{\alpha^1}^1$ such that the inequality constraints of $S_{\alpha^1}^1$ defined by* (35) *are strictly satisfied.*
2) *There exists an $y \in S_{\alpha^2}^2$ such that the inequality constraints of $S_{\alpha^2}^2$ defined by* (36) *are strictly satisfied.*

The conditions given in Assumption 1 are Slater's condition, which are sufficient for strong duality in a convex optimization problem. We use these conditions in order to derive equivalent SOCPs for the zero-sum game Z_α.

V. Existence and Characterization of Saddle Point Equilibrium

In this section, we show that there exists an SPE of the game Z_α if the distributions of the random constraint vectors of both the players belong to the uncertainty sets defined in Sections IV-A and IV-B. We further propose a primal-dual pair of SOCPs whose optimal solutions constitute an SPE of the game Z_α.

Theorem 1. *Consider the game Z_α, where the distributions of the random constraint vectors a_k^i, $k \in \mathcal{I}_i$, $i = 1, 2$, belong to the uncertainty sets described in Sections IV-A and IV-B. Then, there exists an SPE of the game for all $\alpha \in (0,1)^p \times (0,1)^q$.*

Proof. Let $\alpha \in (0,1)^p \times (0,1)^q$. For uncertainty sets described in Sections IV-A and IV-B, the strategy sets $S_{\alpha^1}^1$ and $S_{\alpha^2}^2$ are given by (35) and (36), respectively. It is easy to see that $S_{\alpha^1}^1$ and $S_{\alpha^2}^2$ are convex and compact sets. The function $u(x, y)$ is a bilinear and continuous function. Hence, there exists an SPE from the minimax theorem [4]. □

A. Equivalent Primal-Dual Pair of Second-Order Cone Programs

From the minimax theorem [4], (x^*, y^*) is an SPE for the game Z_α if and only if

$$x^* \in \arg\max_{x \in S_{\alpha^1}^1} \min_{y \in S_{\alpha^2}^2} u(x, y), \tag{37}$$

$$y^* \in \arg\min_{y \in S_{\alpha^2}^2} \max_{x \in S_{\alpha^1}^1} u(x, y). \tag{38}$$

We start with the optimization problem

$$\min_{y \in S_{\alpha^2}^2} \max_{x \in S_{\alpha^1}^1} u(x, y).$$

By introducing auxiliary variables t_{kjw}^1, the inner optimization problem $\max_{x \in S_{\alpha^1}^1} u(x, y)$ can be equivalently written as

$$\begin{aligned}
&\max_{x,\, t_{kjw}^1} \; x^{\mathrm{T}} G y + g^{\mathrm{T}} x + h^{\mathrm{T}} y \\
&\text{s.t.} \\
&(i) \quad -x^{\mathrm{T}} \mu_{kj}^1 - \kappa_{\alpha_k^1} \left\| t_{kjw}^1 \right\|_2 + b_k^1 \geq 0, \\
&\forall\ j = 1, 2 \ldots, N_1,\ w = 1, 2 \ldots, P_1,\ k \in \mathcal{I}_1, \\
&(ii) \quad t_{kjw}^1 - \left(\Sigma_{kw}^1\right)^{\frac{1}{2}} x = 0, \\
&\forall\ j = 1, 2 \ldots, N_1,\ w = 1, 2 \ldots, P_1,\ k \in \mathcal{I}_1, \\
&(iii) \quad C^1 x = d^1,\ x_r \geq 0,\ \forall\ r = 1, 2, \ldots, m.
\end{aligned} \tag{39}$$

Let $\lambda^1 = \left(\lambda_{kjw}^1\right)$, δ_{kjw}^1, and ν^1 be the Lagrange multipliers of constraints $(i), (ii)$, and equality constraints given in the constraint (iii) of (39), respectively. Here, for any $j = 1, \ldots, N_1, w = 1, \ldots, P_1, k \in \mathcal{I}_1$, λ_{kjw}^1 is a real number, δ_{kjw}^1 is an $m \times 1$ real vector, and ν^1 is a $K_1 \times 1$ real vector. Then, the Lagrangian dual problem of the SOCP (39) can be written as

$$\begin{aligned}
&\min_{\lambda_1 \geq 0,\ \delta_{kjw}^1,\ \nu^1} \max_{x \geq 0,\ t_{kjw}^1} \Big\{ x^{\mathrm{T}} G y + g^{\mathrm{T}} x + h^{\mathrm{T}} y \\
&+ \sum_{k \in \mathcal{I}_1} \sum_{j=1}^{N_1} \sum_{w=1}^{P_1} \Big[\lambda_{kjw}^1 \big(-x^{\mathrm{T}} \mu_{kj}^1 - \kappa_{\alpha_k^1} \left\| t_{kjw}^1 \right\|_2 + b_k^1 \big) \\
&+ (\delta_{kjw}^1)^{\mathrm{T}} \Big(t_{kjw}^1 - \left(\Sigma_{kw}^1\right)^{\frac{1}{2}} x \Big) \Big] + (\nu^1)^{\mathrm{T}} (d^1 - C^1 x) \Big\}.
\end{aligned}$$

By reformulating the objective function of the above optimization problem as the sum of two functions such that one depends on x and other depends on t_{kjw}^1, we have

$$\begin{aligned}
&\min_{\lambda_1 \geq 0, \delta_{kjw}^1, \nu^1} \max_{x \geq 0} \Big\langle \; x^{\mathrm{T}} \Big[G y - (C^1)^{\mathrm{T}} \nu^1 + g \\
&- \sum_{k \in \mathcal{I}_1} \sum_{j=1}^{N_1} \sum_{w=1}^{P_1} \big(\lambda_{kjw}^1 \mu_{kj}^1 + \left(\Sigma_{kw}^1\right)^{\frac{1}{2}} \delta_{kjw}^1 \big) \Big] \\
&+ \max_{t_{kjw}^1} \sum_{k \in \mathcal{I}_1} \sum_{j=1}^{N_1} \sum_{w=1}^{P_1} \Big[(\delta_{kjw}^1)^{\mathrm{T}} t_{kjw}^1 - \kappa_{\alpha_k^1} \lambda_{kjw}^1 \left\| t_{kjw}^1 \right\|_2 \Big] \\
&+ h^{\mathrm{T}} y + (\nu^1)^{\mathrm{T}} d^1 + \sum_{k \in \mathcal{I}_1} \sum_{j=1}^{N_1} \sum_{w=1}^{P_1} \lambda_{kjw}^1 b_k^1 \Big\rangle.
\end{aligned}$$

The first term of the objective function is a function of x

$$x^{\mathrm{T}}\Big[Gy-(C^1)^{\mathrm{T}}\nu^1+g - \sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\big(\lambda^1_{kjw}\mu^1_{kj}+\big(\Sigma^1_{kw}\big)^{\frac{1}{2}}\delta^1_{kjw}\big)\Big]. \quad (40)$$

The above term is unbounded on the domaine $x \geq 0$, unless the following condition holds

$$Gy-\sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\big(\lambda^1_{kjw}\mu^1_{kj}+\big(\Sigma^1_{kw}\big)^{\frac{1}{2}}\delta^1_{kjw}\big) -(C^1)^{\mathrm{T}}\nu^1+g\leq 0.$$

When the above condition holds, it is clear that the maximum value of (40) is zero and it holds at $x = 0$. The second term of the objective function is a function of t^1_{kjw}

$$\sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\Big[(\delta^1_{kjw})^{\mathrm{T}}t^1_{kjw}-\kappa_{\alpha^1_k}\lambda^1_{kjw}\left\|t^1_{kjw}\right\|_2\Big]. \quad (41)$$

The above term is unbounded on the domaine $t^1_{kjw} \in \mathbb{R}^m$, unless the following condition holds

$$||\delta^1_{kjw}|| \leq \kappa_{\alpha^1_k}\lambda^1_{kjw}, \quad \forall\ k\in\mathcal{I}_1, j=1,2\ldots,N_1, w=1,2\ldots,P_1.$$

When the above condition holds, it is clear that the maximum value of (41) is zero and it holds at $t^1_{kjw} = 0$. Then, the Lagrangian dual problem of the SOCP (39) can be written as

$$\begin{aligned}
&\min_{\lambda_1\geq 0,\delta^1_{kjw},\nu^1}\big(h^{\mathrm{T}}y+(\nu^1)^{\mathrm{T}}d^1+\sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\lambda^1_{kjw}b^1_k\big)\\
&\text{s.t. } (i)\quad Gy-\sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\big[\lambda^1_{kjw}\mu^1_{kj}+\big(\Sigma^1_{kw}\big)^{\frac{1}{2}}\delta^1_{kjw}\big]\\
&\quad -(C^1)^{\mathrm{T}}\nu^1+g\leq 0,\\
&\quad (ii)\quad ||\delta^1_{kjw}||\leq\kappa_{\alpha^1_k}\lambda^1_{kjw},\\
&\quad \forall\ k\in\mathcal{I}_1, j=1,2\ldots,N_1, w=1,2\ldots,P_1.
\end{aligned}$$

Under Assumption 1, the Lagrangian dual problem of (39) has zero duality gap [24], which implies that the above optimization problem is equivalent to the problem $\max_{x\in S^1_{\alpha^1}} u(x,y)$. Hence, the problem $\min_{y\in S^2_{\alpha^2}}\max_{x\in S^1_{\alpha^1}} u(x,y)$ is equivalent to the following SOCP

$$\begin{aligned}
&\min_{y,\ \nu^1,\ \delta^1_{kjw},\ \lambda^1_{kjw}\geq 0}\ h^{\mathrm{T}}y+(\nu^1)^{\mathrm{T}}d^1+\sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\lambda^1_{kjw}b^1_k\\
&\text{s.t.}\\
&(i)\quad Gy-\sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\big[\lambda^1_{kjw}\mu^1_{kj}+\big(\Sigma^1_{kw}\big)^{\frac{1}{2}}\delta^1_{kjw}\big]\\
&-(C^1)^{\mathrm{T}}\nu^1+g\leq 0,\\
&(ii)\quad ||\delta^1_{kjw}||\leq\kappa_{\alpha^1_k}\lambda^1_{kjw},\\
&\forall\ k\in\mathcal{I}_1,\ j=1,2\ldots,N_1,\ w=1,2\ldots,P_1,\\
&(iii)\quad -(\mu^2_{lj})^{\mathrm{T}}y+\kappa_{\alpha^2_l}||(\Sigma^2_{lw})^{\frac{1}{2}}y||\leq -b^2_l,\\
&\forall\ j=1,2,\ldots,N_2,\ w=1,2,\ldots,P_2,\ l\in\mathcal{I}_2,\\
&(iv)\quad C^2y=d^2,\ y_s\geq 0,\ \forall\ s=1,2,\ldots,n,
\end{aligned}\quad (42)$$

where the constraints (iii) and (iv) are due to the fact that $y \in S^2_{\alpha^2}$ and the representation of $S^2_{\alpha^2}$ in (36). Similarly, problem $\max_{x\in S^1_{\alpha^1}}\min_{y\in S^2_{\alpha^2}} u(x,y)$ is equivalent to the following SOCP

$$\begin{aligned}
&\max_{x,\ \nu^2,\ \delta^2_{ljw},\ \lambda^2_{ljw}\geq 0}\ g^{\mathrm{T}}x+(\nu^2)^{\mathrm{T}}d^2-\sum_{l\in\mathcal{I}_2}\sum_{j=1}^{N_2}\sum_{w=1}^{P_2}\lambda^2_{ljw}b^2_l\\
&\text{s.t.}\\
&(i)\quad G^{\mathrm{T}}x-\sum_{l\in\mathcal{I}_2}\sum_{j=1}^{N_2}\sum_{w=1}^{P_2}\big[-\lambda^2_{ljw}\mu^2_{lj}+\big(\Sigma^2_{lw}\big)^{\frac{1}{2}}\delta^2_{ljw}\big]\\
&-(C^2)^{\mathrm{T}}\nu^2+h\geq 0,\\
&(ii)\quad ||\delta^2_{ljw}||\leq\kappa_{\alpha^2_l}\lambda^2_{ljw},\ \lambda^2_{ljw}\geq 0,\\
&\forall\ \ l\in\mathcal{I}_2,\ j=1,2,\ldots,N_2,\ w=1,2,\ldots,P_2,\\
&(iii)\quad (\mu^1_{kj})^{\mathrm{T}}x+\kappa_{\alpha^1_k}||(\Sigma^1_{kw})^{\frac{1}{2}}x||\leq b^1_k,\\
&\forall\ j=1,2,\ldots,N_2,\ w=1,2,\ldots,P_2,\ k\in\mathcal{I}_1,\\
&(iv)\quad C^1x=d^1,\ x_r\geq 0,\ \forall\ r=1,2,\ldots,m.
\end{aligned}\quad (43)$$

It follows from the duality theory of SOCPs that (42) and (43) form a primal-dual pair of SOCPs [24]. Next, we show that the equivalence between the optimal solutions of (42)-(43) and an SPE of the game Z_α.

Theorem 2. *Consider the zero-sum game Z_α, where the feasible strategy sets of player 1 and player 2 are given by* (35) *and* (36)*, respectively. Let Assumption 1 holds. Then, for a given $\alpha \in (0,1)^p \times (0,1)^q$, (x^*, y^*) is an SPE of the game Z_α if and only if there exists $(\nu^{1*}, \delta^{1*}_{kjw}, \lambda^{1*}_{kjw} \geq 0)$ and $(\nu^{2*}, \delta^{2*}_{ljw}, \lambda^{2*}_{ljw} \geq 0)$ such that $(y^*, \nu^{1*}, \delta^{1*}_{kjw}, \lambda^{1*}_{kjw})$ and $(x^*, \nu^{2*}, \delta^{2*}_{ljw}, \lambda^{2*}_{ljw})$ are optimal solutions of* (42) *and* (43)*, respectively.*

Proof. Let (x^*, y^*) be an SPE of the game Z_α. Then, x^* and y^* are the solutions of (37) and (38), respectively. Therefore, there exists $(\nu^{1*}, \delta^{1*}_{kjw}, \lambda^{1*}_{kjw} \geq 0)$ and $(\nu^{2*}, \delta^{2*}_{ljw}, \lambda^{2*}_{ljw} \geq 0)$ such that $(y^*, \nu^{1*}, \delta^{1*}_{kjw}, \lambda^{1*}_{kjw})$ and $(x^*, \nu^{2*}, \delta^{2*}_{ljw}, \lambda^{2*}_{ljw})$ are optimal solutions of (42) and (43) respectively. On the other hand, let $(y^*, \nu^{1*}, \delta^{1*}_{kjw}, \lambda^{1*}_{kjw})$ and $(x^*, \nu^{2*}, \delta^{2*}_{ljw}, \lambda^{2*}_{ljw})$

be optimal solutions of (42) and (43), respectively. Under Assumption 1, (42) and (43) are strictly feasible. Therefore, strong duality holds for primal-dual pair (42)-(43). Then, we have

$$g^{\mathrm{T}}x^* + (\nu^{2*})^{\mathrm{T}}d^2 - \sum_{l\in\mathcal{I}_2}\sum_{j=1}^{N_2}\sum_{w=1}^{P_2}\lambda_{ljw}^{2*}b_l^2 = h^{\mathrm{T}}y^* + (\nu^{1*})^{\mathrm{T}}d^1 + \sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\lambda_{kjw}^{1*}b_k^1. \quad (44)$$

Consider the constraint (i) of (42) at optimal solution $(y^*, \nu^{1*}, \delta_{kjw}^{1*}, \lambda_{kjw}^{1*})$ and multiply it by x^{T}, for any $x \in S_{\alpha^1}^1$, we have

$$x^{\mathrm{T}}Gy^* + g^{\mathrm{T}}x \leq x^{\mathrm{T}}(C^1)^{\mathrm{T}}\nu^{1*} + \sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\left[x^{\mathrm{T}}\mu_{kj}^1\lambda_{kjw}^{1*} + x^{\mathrm{T}}(\Sigma_{kw}^1)^{\frac{1}{2}}\delta_{kjw}^{1*}\right]. \quad (45)$$

By using the Cauchy-Schwartz inequality, for any $k \in \mathcal{I}_1,\ j = 1, 2\ldots, N_1,\ w = 1, 2\ldots, P_1$, we have

$$x^{\mathrm{T}}(\Sigma_{kw}^1)^{\frac{1}{2}}\delta_{kjw}^{1*} \leq \|(\Sigma_{kw}^1)^{\frac{1}{2}}x\|_2\|\delta_{kjw}^{1*}\|_2.$$

Using the constraint (ii) of (43), the above constraint implies that

$$(x^*)^{\mathrm{T}}(\Sigma_{kw}^1)^{\frac{1}{2}}\delta_{kjw}^{1*} \leq \|(\Sigma_{kw}^1)^{\frac{1}{2}}x\|_2\kappa_{\alpha_k^1}\lambda_{kjw}^{1*}.$$

Since $x \in S_{\alpha^1}^1$, we have

$$C^1x = d^1.$$

Then, the constraint (45) implies that

$$x^{\mathrm{T}}Gy^* + g^{\mathrm{T}}x \leq (\nu^{1*})^{\mathrm{T}}d^1 + \sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\left[x\mathrm{T}\mu_{kj}^1\lambda_{kjw}^{1*} + (\Sigma_{kw}^1)^{\frac{1}{2}}x\|_2\kappa_{\alpha_k^1}\lambda_{kjw}^{1*}\right],$$

which in turn implies by using the constraint (iii) of (43) that

$$x^{\mathrm{T}}Gy^* + g^{\mathrm{T}}x \leq (\nu^{1*})^{\mathrm{T}}d^1 + \sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\lambda_{kjw}^{1*}b_k^1.$$

Then, for any $x \in S_{\alpha^1}^1$, we have

$$x^{\mathrm{T}}Gy^* + g^{\mathrm{T}}x + h^{\mathrm{T}}y^* \leq h^{\mathrm{T}}y^* + (\nu^{1*})^{\mathrm{T}}d^1 + \sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\lambda_{kjw}^{1*}b_k^1. \quad (46)$$

Similarly, for any $y \in S_{\alpha^2}^2$, we have

$$(x^*)^{\mathrm{T}}Gy + g^{\mathrm{T}}x^* + h^{\mathrm{T}}y \geq g^{\mathrm{T}}x^* + (\nu^{2*})^{\mathrm{T}}d^2 + \sum_{l\in\mathcal{I}_2}\sum_{j=1}^{N_2}\sum_{w=1}^{P_2}\lambda_{ljw}^{2*}b_l^2. \quad (47)$$

Take $x = x^*$ and $y = y^*$ in (46) and (47), then from (44), we get

$$u(x^*, y^*) = h^{\mathrm{T}}y^* + (\nu^{1*})^{\mathrm{T}}d^1 + \sum_{k\in\mathcal{I}_1}\sum_{j=1}^{N_1}\sum_{w=1}^{P_1}\lambda_{kjw}^{1*}b_k^1 = g^{\mathrm{T}}x^* + (\nu^{2*})^{\mathrm{T}}d^2 + \sum_{l\in\mathcal{I}_2}\sum_{j=1}^{N_2}\sum_{w=1}^{P_2}\lambda_{ljw}^{2*}b_l^2. \quad (48)$$

It follows from (46), (47), and (48) that

$$u(x, y^*) \leq u(x^*, y^*) \leq u(x^*, y),\ \forall\, x \in S_{\alpha^1}^1,\ y \in S_{\alpha^2}^2,$$

which in turn implies that (x^*, y^*) is an SPE of the game Z_α. □

VI. Numerical results

A. Competition in Financial Market

In this section, we consider a competition of two firms in financial market. They invest in the same set of portfolios. Let $P = \{1, 2, \ldots, N_P\}$ be the set of portfolios. Let $\mathcal{A}_j$ be the set of assets in the portfolio j. Assume that the sets $\mathcal{A}_j$ and $\mathcal{A}_k$ are disjoint, for any $j \neq k$. Let $x_k = (x_{kj})_{j\in\mathcal{A}_k}$ be the investment vector of firm 1 in portfolio k and $y_k = (y_{kj})_{j\in\mathcal{A}_k}$ be the investment vector of firm 2 in portfolio k. Let $x = (x_k)_{k\in P}$ and $y = (y_k)_{k\in P}$ be the investment vector of firm 1 (resp. firm 2). The set of investments X of firm 1 is defined as follows

$$X = \left\{x \,\middle|\, \sum_{j\in\mathcal{A}_k} x_{kj} = W_k^1,\ \forall j \in \mathcal{A}_k,\ k \in P\right\},$$

and the set of investments Y of firm 2 is defined as follows

$$Y = \left\{y \,\middle|\, \sum_{j\in\mathcal{A}_k} y_{kj} = W_k^2,\ \forall j \in \mathcal{A}_k,\ k \in P\right\},$$

where W_k^i is the total investment of firm i in portfolio k, for any $i = 1, 2$ and $k \in P$. Let $L_k^i = (L_{kj}^i)_{j\in\mathcal{A}_k}$ be a random loss vector of firm i from portfolio k. Then, for a given investment vector x_k and y_k, the total loss of firm 1 (resp. firm 2) caused by portfolio k is defined as $(L_k^1)^{\mathrm{T}}x_k$ (resp. $(L_k^1)^{\mathrm{T}}y_k$). Each firm wants to make sure that their random loss is below a maximal allowable loss level with high probability. This condition is modeled by the following inequality

$$\mathbb{P}\left\{(L_k^1)^{\mathrm{T}}x_k \leq b_k^1\right\} \geq \alpha_k^1, \quad (49)$$

and

$$\mathbb{P}\left\{(L_l^2)^{\mathrm{T}}y_l \leq b_l^2\right\} \geq \alpha_l^2, \quad (50)$$

where b_k^i are deterministic vectors and α_k^i are confidence levels, $i = 1, 2,\ k \in P$. We assume that the true distribution of random loss vectors is unknown, but only known to belong to some uncertainty set $\mathcal{D}_k^i$ defined in Section IV. Then, the feasible strategy sets of two firms are given by

$$\inf_{F_k^1\in\mathcal{D}_k^1}\mathbb{P}\left\{(L_k^1)^{\mathrm{T}}x_k \leq b_k^1\right\} \geq \alpha_k^1,\ \forall\, k \in P,$$

and

$$\inf_{F_l^2 \in \mathcal{D}_l^2} \mathbb{P}\left\{(L_l^2)^{\mathrm{T}} y_l \leq b_l^2\right\} \geq \alpha_l^2, \ \forall\, l \in P.$$

We assume that the total profit of both firm is zero, i.e., for each profile of strategies $(x, y) \in X \times Y$, if firm 1 gains a profit $u(x, y)$, then firm 2 gains a profit $-u(x, y)$. Firm 1 wants to maximize u w.r.t x, for $y \in S^2_{\alpha^2}$ and firm 2 wants to minimize u w.r.t y, for $x \in S^1_{\alpha^1}$. We assume that u has the form (1), i.e., $u(x, y) = x^{\mathrm{T}} G y + g^{\mathrm{T}} x + h^{\mathrm{T}} y$.

In order to find an SPE of (8), we solve the two SOCP problems (42) and (43) using **coneqp** solver in CVXOPT. We compare the uncertainty sets defined in Section (IV) with the true model, in which we assume that the true distribution of random loss vectors is known and follows Gaussian distribution. In this case, it is well known that the constraints (49) and (50) are equivalent to SOC constraints [25]. An SPE in true model can be computed by solving an SOCP reformulation [8].

B. Case Study

All the numerical results below are performed using Python 3.8.8 on an Intel Core i5-1135G7, Processor 2.4 GHz (8M Cache, up to 4.2 GHz), RAM 16G, 512G SSD. We consider two firms investing in a portfolio consists of four assets, i.e., $P = \{1\}$ and $\mathcal{A}_1 = \{1, 2, 3, 4\}$. We generate randomly the vectors g and h in (1) in $[-3, 3]^4$ by the command *"numpy.random.uniform(-3,3,size=(4,1))"*. The matrix G in (1) is randomly generated by the command *"numpy.random.uniform(-3,3,size=(4,4))"*. We take the confidence levels of two firms as $\alpha^1 = \alpha^2 = 0.9$, the total investment of two firms in the portfolio W_1^1 and W_1^2 are randomly generated on $[20, 80]$ by the command *"numpy.random.uniform(20,80)"*. The maximal allowable loss levels of two firms b_1^1 and b_1^2 are randomly generated on $[100, 500]$ by the command *"numpy.random.uniform(100,500)"*. The probability distribution of the loss of two firms L_1^1 and L_1^2 are assumed to be Normal distributions with mean vector μ_1^1 (resp. μ_1^2) and covariance matrix Σ_1^1 (resp. Σ_1^2). The mean vectors are randomly generated on $[8, 12]^4$ using the command *"numpy.random.uniform(8,12, size=(4,1))"*. The covariance matrix are defined as follows

$$\Sigma_1^i = \frac{AA^{\mathrm{T}}}{4} + \mathbf{I}_4, \forall\, i = 1, 2,$$

where A is a 4×4 random matrix whose all entries belong to $[0, 1]$ generated by the command *"A=numpy.random.random(size=(4,4))"* and $\mathbf{I}_4$ denotes 4×4 identity matrix. For any $i = 1, 2$, we define sample mean vector μ^i_{sample} and Σ^i_{sample} by generating randomly a sample of 100 observations $\xi_1^i, \ldots, \xi_{100}^i$, which follow Normal distribution with mean vector μ_1^i and covariance matrix Σ_1^i. To do that, we generate a standard Gaussian vector by the command "*x=numpy.random.normal(0,1,4)*". We generate a Gaussian vector with mean vector μ_1^i and Σ_1^i by taking $\xi_j^i = Bx + \mu_1^i$, where B is the Cholesky factorization of Σ_1^i. To get the Cholesky factorization of a matrix, we use the command "*numpy.linalg.cholesky*". The sample mean vector μ^i_{sample} and the covariance matrix Σ^i_{sample} are defined as follows

$$\mu^i_{sample} = \frac{1}{100} \sum_{j=1}^{100} \xi_j^i,$$

$$\Sigma^i_{sample} = \frac{1}{99} \sum_{j=1}^{100} (\xi_j^i - \mu^i_{sample})(\xi_j^i - \mu^i_{sample})^{\mathrm{T}}.$$

Now, we define other parameters for each model. For the uncertainty set (12), we take $\gamma_1^i = 1.1$, for any $i = 1, 2$. For the uncertainty set (16), we take $\gamma_{11}^i = \gamma_{12}^i = 1$, for any $i = 1, 2$. We take the uncertainty set (21) similarly as the uncertainty set (9) by choosing $M = 1$. For the uncertainty set (25), we take the radius vector $\epsilon^i_{\mu,1} = (0.1, 0.1, 0.1, 0.1)^4$ and the radius matrix $\epsilon^i_{\Sigma,1} = 0.1 \times \mathbf{I}_4$, for any $i = 1, 2$, where $\mathbf{I}_4$ is 4×4 identity matrix. For the uncertainty set (29), we take $\theta_1^i = 0.05$, for any $i = 1, 2$.

For the above instance, we compute an SPE of the true model, where the true distribution of random loss vectors L_1^1 and L_1^2 follow Gaussian distributions with mean vector μ_1^1 (resp. μ_1^2) and covariance matrix Σ_1^1 (resp. Σ_1^2). We obtain an SPE (x^*, y^*) given by

$$x^* = (18.91, 19.45, 19.45, 20.22)^{\mathrm{T}},$$
$$y^* = (19.01, 20.15, 20.45, 18.71)^{\mathrm{T}}.$$

The profit of firm 1 for this instance is $u(x^*, y^*) = -275.52$. Now, we calculate an SPE of the models defined in Section (IV). For the uncertainty sets (9), (12), (16), (21) and (25), we take $\mu_1^i = \mu^i_{sample}$ and $\Sigma_1^i = \Sigma^i_{sample}$, for any $i = 1, 2$. For the uncertainty set (29), we assume that the nominal distribution ν_1^i follows a Gaussian distribution with mean vector μ^i_{sample} and covariance matrix Σ^i_{sample}. We compare the optimal profit value of firm 1 in above models with the optimal profit value of firm 1 in the true model. The results are given in Table III. We can see that for this instance, the models defined by $\phi-$divergence give better solution than the models defined by moments since the optimal profit value in $\phi-$divergence uncertainty sets approximates well the optimal profit value in true model. We also present the time analysis

TABLE III
LIST OF OPTIMAL PROFIT VALUES $u(x^*, y^*)$

True model	Known Mean Known Covariance	Known Mean Unknown Covariance	Unknown Mean Unknown Covariance	Polytopic
-257.52	-221.11	-222.5	-224.8	-221.11
Componentwise Bounds	Kullback Leibler	Variation Distance	Modified χ^2 - distance	Hellinger Distance
-223.3	-255.1	-256.23	-255.8	-253.9

for a large numbers of assets size model by considering the number of assets between 100 and 1000. For each case of number of assets, we randomly generate 10 instances of the known mean known covariance model, where the parameters are defined similarly as above and we calculate the average running time (in seconds) to solve the two optimization problems (42) and (43). The results are given in Figure 1.

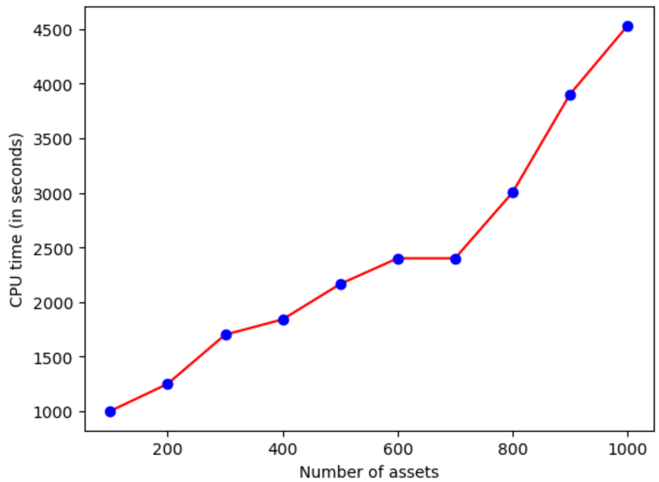


Fig. 1. CPU time (in seconds) to solve (42) and (43) in known mean known covariance cases with different number of assets.

It is clear from Figure 1 that our optimization problems can be solved efficiently in high dimension up to 1000 assets.

VII. Conclusion and Future Work

We study a more general two player zero-sum game than the model considered in [1] under various moments based and statistical based uncertainty sets. We propose a reformulation of the chance constraints using distributionally chance-constrained optimization framework and show that there exists a mixed strategy SPE of the game. Under Slater's condition, the SPE of the game can be obtained from the optimal solutions of a primal-dual pair of SOCPs. We present a competition of two firms in financial market as an application to figure out out theoretical results. The numerical experiments are performed using randomly generated data on the game up to 1000 assets and it is clear from our time analysis that the SOCPs problems can be computed efficiently. For our future works, we will study tractable reformulation of the zero-sum game problem with different payoff structure in a different game model and apply the game problem in a different application to the competition in financial market considered in this paper.

Acknowledgement

This research was supported by DST/CEFIPRA Project No. IFC/4117/DST-CNRS-5th call/2017-18/2 and CNRS Project No. AR/SB:2018-07-440.

Appendices

Appendix A: Proof of Lemma 7 - Case Hellinger distance

For $i = 1, 2$ and $k \in \mathcal{I}_i$, it suffices to calculate the value of $\sup_{\lambda>0,\beta\in\mathbb{R}} \{f_k^i(\lambda,\beta)\}$ with Hellinger distance divergence. We consider two cases as follows

- **Case 1:** $\frac{\beta}{\lambda} < 1 \Leftrightarrow \beta < \lambda$. We have

$$\phi^*\left(\frac{\beta}{\lambda}\right) = \frac{\beta}{\lambda-\beta},$$
$$\phi^*\left(\frac{\beta-1}{\lambda}\right) = \frac{\beta-1}{\lambda+1-\beta}.$$

Therefore,

$$\sup_{\lambda>0,\beta\in\mathbb{R}} \{f_k^i(\lambda,\beta)\} =$$
$$\sup_{\lambda>0,\beta\in\mathbb{R}} \mathbb{P}_{\nu_k^i}(M_k^i)\frac{\lambda^2}{(\lambda-\beta)(\lambda-\beta+1)} + \frac{\beta^2}{\beta-\lambda} - \lambda\theta_k^i.$$

Since $\lambda > 0$ and $\beta < \lambda$, let $\gamma = \lambda - \beta$, we deduce that

$$\sup_{\lambda>0,\beta\in\mathbb{R}} \{f_k^i(\lambda,\beta)\}$$
$$= \sup_{\lambda>0,\gamma>0} \left\{\lambda^2\left(\frac{\mathbb{P}_{\nu_k^i}(M_k^i)}{\gamma(\gamma+1)} - \frac{1}{\gamma}\right) + \lambda(2-\theta_k^i) - \gamma\right\}.$$

Let $Q(\lambda,\gamma) = \lambda^2\left(\frac{\mathbb{P}_{\nu_k^i}(M_k^i)}{\gamma(\gamma+1)} - \frac{1}{\gamma}\right) + \lambda(2-\theta_k^i) - \gamma$. Note that $0 \le \mathbb{P}_{\nu_k^i}(M_k^i) \le 1$ and $\gamma > 0$. Therefore, $Q(\lambda,\gamma)$ is a second-order polynomial of λ and the coefficient of λ^2 is negative. It is well known that the maximum value of a second order function $f(x) = ax^2 + bx + c$ with $a < 0$ is $c - \frac{b^2}{4a}$ and it holds at $x = \frac{-b}{2a}$. Hence, the maximum value of $Q(\lambda,\gamma)$ holds at $\lambda^* = \frac{\gamma(\gamma+1)(2-\theta_k^i)}{2(1+\gamma-\mathbb{P}_{\nu_k^i}(M_k^i))}$. Since $\theta_k^i < 2$, it is clear that $\lambda^* > 0$. Then, the optimal value of $\sup_{\lambda>0,\beta\in\mathbb{R}} \{f_k^i(\lambda,\beta)\}$ holds when $\lambda = \lambda^*$ and we have

$$\sup_{\lambda>0,\beta\in\mathbb{R}} \{f_k^i(\lambda,\beta)\}$$
$$= \sup_{\gamma>0} \left\{-\gamma + \frac{(2-\theta_k^i)^2\gamma(\gamma+1)}{4(\gamma+1-\mathbb{P}_{\nu_k^i}(M_k^i))}\right\}. \quad (51)$$

Let $u = \gamma + 1 - \mathbb{P}_{\nu_k^i}(M_k^i)$, then $u > 1 - \mathbb{P}_{\nu_k^i}(M_k^i)$. Rewriting (51) as a function of u, we have:

$$\sup_{\lambda>0,\beta\in\mathbb{R}} \{f_k^i(\lambda,\beta)\} = \sup_{u>1-\mathbb{P}_{\nu_k^i}(M_k^i)} F(u),$$
$$= \sup_{u>1-\mathbb{P}_{\nu_k^i}(M_k^i)} \left\{au + \frac{b}{u} + c\right\},$$

where $a = \left(\frac{(2-\theta_k^i)^2}{4} - 1\right)$,
$b = \frac{(2-\theta_k^i)^2\mathbb{P}_{\nu_k^i}(M_k^i)(\mathbb{P}_{\nu_k^i}(M_k^i)-1)}{4}$,
$c = 1 - \mathbb{P}_{\nu_k^i}(M_k^i) + \frac{(2-\theta_k^i)^2(2\mathbb{P}_{\nu_k^i}(M_k^i)-1)}{4}$. Note that $a < 0$ and $b \le 0$. We have: $F'(u) = a - \frac{b}{u^2}$. Hence, it can be shown that F is decreasing on $(u^*, +\infty)$, increasing on $(-u^*, u^*)$ and decreasing on $(-\infty, -u^*)$, where $u^* = \sqrt{\frac{b}{a}}$. Or,

$$u^* = \sqrt{\frac{(2-\theta_k^i)^2}{4-(2-\theta_k^i)^2}\mathbb{P}_{\nu_k^i}(M_k^i)(1-\mathbb{P}_{\nu_k^i}(M_k^i))}. \quad (52)$$

We have $F(u^*) = -2\sqrt{ab} + c$. We consider 2 cases as follows

1: $u^* \leq 1 - \mathbb{P}_{\nu_k^i}(M_k^i)$. Since F is decreasing on $(u^*, +\infty)$, it is also decreasing on $(1 - \mathbb{P}_{\nu_k^i}(M_k^i), +\infty)$. Hence, $\sup_{u>1-\mathbb{P}_{\nu_k^i}(M_k^i)} F(u) = 0$, where the optimal value holds when $u \to 1 - \mathbb{P}_{\nu_k^i}(M_k^i) \Leftrightarrow \gamma \to 0$, which violates (30).

2: $u^* > 1 - \mathbb{P}_{\nu_k^i}(M_k^i) > 0$. Then, the optimal value of $\sup_{u>1-\mathbb{P}_{\nu_k^i}(M_k^i)} F(u)$ holds when $u = u^*$. Therefore,

$$\sup_{\lambda>0,\beta\in\mathbb{R}} \left\{ f_k^i(\lambda, \beta) \right\} = F(u^*) = -2\sqrt{ab} + c.$$

Then, (30) is equivalent to

$$\begin{aligned} &-2\sqrt{\frac{(2-\theta_k^i)^2}{4}\left(1 - \frac{(2-\theta_k^i)^2}{4}\right)\mathbb{P}_{\nu_k^i}(M_k^i)(1-\mathbb{P}_{\nu_k^i}(M_k^i))} \\ &\geq \left(1 - \frac{(2-\theta_k^i)^2}{2}\right)\mathbb{P}_{\nu_k^i}(M_k^i) + \frac{(2-\theta_k^i)^2}{4} - (1-\alpha_k^i). \end{aligned} \tag{53}$$

By taking square on both side of (53), we obtain a second order inequality of $\mathbb{P}_\nu(K)$ as follows

$$\mathbb{P}_{\nu_k^i}(M_k^i)^2 + B\mathbb{P}_{\nu_k^i}(M_k^i) + C \geq 0,$$

where B, C are defined in Table II. By solving the equality $x^2 + Bx + C = 0$, we have two solutions $x_{\min} < x_{\max}$ where $x_{\min} = \frac{-B-\sqrt{\Delta}}{2}$, $x_{\max} = \frac{-B+\sqrt{\Delta}}{2}$. It is clear that (53) is equivalent to either $\mathbb{P}_{\nu_k^i}(M_k^i) \geq x_{\max}$ or $\mathbb{P}_{\nu_k^i}(M_k^i) \leq x_{\min}$. Since $\theta_k^i < 2 - \sqrt{2}$, we deduce that $1 - \frac{(2-\theta_k^i)^2}{2} < 0$. Therefore, we have

$$\begin{aligned} &\left(1 - \frac{(2-\theta_k^i)^2}{2}\right)x_{\min} + \frac{(2-\theta_k^i)^2}{4} - (1-\alpha_k^i) \\ &> \left(1 - \frac{(2-\theta_k^i)^2}{2}\right)x_{\max} + \frac{(2-\theta_k^i)^2}{4} - (1-\alpha_k^i). \end{aligned} \tag{54}$$

On the other hand, we have

$$\begin{aligned} &-2\sqrt{\frac{(2-\theta_k^i)^2}{4}\left(1 - \frac{(2-\theta_k^i)^2}{4}\right)x(1-x)} \\ &= \pm\left[\left(1 - \frac{(2-\theta_k^i)^2}{2}\right)x + \frac{(2-\theta_k^i)^2}{4} - (1-\alpha_k^i)\right], \end{aligned}$$

where $x = x_{\min}$ or $x = x_{\max}$. Note that $-2\sqrt{\frac{(2-\theta_k^i)^2}{4}\left(1 - \frac{(2-\theta_k^i)^2}{4}\right)x(1-x)} < 0$. Using (54), we deduce that

$$\begin{aligned} &-2\sqrt{\frac{(2-\theta_k^i)^2}{4}\left(1 - \frac{(2-\theta_k^i)^2}{4}\right)x_{\max}(1-x_{\max})} \\ &= \left[\left(1 - \frac{(2-\theta_k^i)^2}{2}\right)x_{\max} + \frac{(2-\theta_k^i)^2}{4} - (1-\alpha_k^i)\right], \end{aligned}$$

and

$$\begin{aligned} &-2\sqrt{\frac{(2-\theta_k^i)^2}{4}\left(1 - \frac{(2-\theta_k^i)^2}{4}\right)x_{\min}(1-x_{\min})} \\ &= -\left[\left(1 - \frac{(2-\theta_k^i)^2}{2}\right)x_{\min} + \frac{(2-\theta_k^i)^2}{4} - (1-\alpha_k^i)\right]. \end{aligned}$$

or $x_{\max}$ satisfies (53) while $x_{\min}$ does not satisfy (53). Then, (53) is equivalent to $\mathbb{P}_{\nu_k^i}(M_k^i) \geq x_{\max}$.

- **Case 2:** $1 \leq \frac{\beta}{\lambda} \Leftrightarrow \lambda \leq \beta$. We have

$$\phi^*\left(\frac{\beta}{\lambda}\right) = +\infty,$$

which implies that $\sup_{\lambda>0,\beta\in\mathbb{R}} \left\{ f_k^i(\lambda, \beta) \right\} = -\infty$, which violates (30).

REFERENCES

[1] H. N. Nguyen, A. Lisser, V. V. Singh, and M. Arora, "Zero-sum games with distributionally robust chance constraints," in *17th International Conference on Internet and Web Applications and Services (ICIW)*, pp. 7–12, IARIA, 2022.

[2] A. A. Cournot, *Researches into the Mathematical Principles of the Theory of Wealth*. Macmillan Company, New York, 1897.

[3] J. Nash, "Non-cooperative games," *Annals of Mathematics*, pp. 286–295, 1951.

[4] J. von Neumann, "On the theory of games," *Math. Annalen*, vol. 100, no. 1, pp. 295–320, 1928.

[5] I. Adler, "The equivalence of linear programs and zero-sum games," *International Journal of Game Theory*, vol. 42, no. 1, pp. 165–177, 2013.

[6] G. B. Dantzig, "A proof of the equivalence of the programming problem and the game problem," in *Activity analysis of production and allocation* (T. Koopmans, ed.), pp. 330–335, John Wiley Sons, New York, 1951.

[7] A. Charnes, "Constrained games and linear programming," *Proceedings of National Academy of Sciences of the USA*, vol. 39, pp. 639–641, 1953.

[8] V. V. Singh and A. Lisser, "A second-order cone programming formulation for zero sum game with chance constraints," *European Journal of Operational Research*, vol. 275, pp. 839–845, 2019.

[9] J. F. Nash Jr, "Equilibrium points in n-person games," *Proceedings of the National Academy of Sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[10] G. Debreu, "A social equilibrium existence theorem," *Proceedings of National Academy of Sciences*, vol. 38, pp. 886–893, 1952.

[11] U. Ravat and U. V. Shanbhag, "On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games," *SIAM Journal of Optimization*, vol. 21, no. 3, pp. 1168–1199, 2011.

[12] V. V. Singh, O. Jouini, and A. Lisser, "Existence of Nash equilibrium for chance-constrained games," *Operations Research Letters*, vol. 44, no. 5, pp. 640–644, 2016.

[13] V. V. Singh, O. Jouini, and A. Lisser, "Distributionally robust chance-constrained games: Existence and characterization of Nash equilibrium," *Optimization Letters*, vol. 11, no. 7, pp. 1385–1405, 2017.

[14] P. Shen, A. Lisser, V. V. Singh, N. Gupta, and E. Balachandar, "Games with distributionally robust joint chance constraints," *Optimization Letters*, vol. 15, pp. 1931–1953, 2021.

[15] V. V. Singh, A. Lisser, and M. Arora, "An equivalent mathematical program for games with random constraints," *Statistics and Probability Letters*, vol. 174, p. 109092, 2021.

[16] R. Ji and M. A. Lejeune, "Risk-budgeting multi-portfolio optimization with portfolio and marginal risk constraints," *Annals of Operations Research*, vol. 262, pp. 547–578, 2018.

[17] L. El-Ghaoui, M. Oks, and F. Oustry, "Worst-case value-at-risk and robust portfolio optimization: A conic programming approach," *Operations Research*, vol. 51, no. 4, pp. 543–556, 2003.

[18] N. Rujeerapaiboon, D. Kuhn, and W. Wiesemann, "Chebyshev inequalities for products of random variables," *Mathematics of Operations Research*, vol. 43, no. 3, pp. 887–918, 2018.

[19] J. Cheng, E. Delage, and A. Lisser, “Distributionally robust stochastic knapsack problem,” SIAM *Journal of Optimization*, vol. 24, no. 3, pp. 1485–1506, 2014.
[20] E. Delage and Y. Ye, “Distributionally robust optimization under moment uncertainty with application to data-driven problems,” *Operations Research*, vol. 58, no. 3, pp. 595–612, 2010.
[21] A. Ben-Tal, D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen, “Robust solutions of optimization problems affected by uncertain probabilities,” *Management Science*, vol. 59, no. 2, pp. 341–357, 2013.
[22] L. Pardo, *Statistical Inference Based on Divergence Measures*. Chapman and Hall/CRC Press, New York, 2018.
[23] R. Jiang and Y. Guan, “Data-driven chance constrained stochastic program,” *Mathematical Programming*, vol. 158, no. 1, pp. 291–327, 2016.
[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, New York, 2004.
[25] R. Henrion, “Structural properties of linear probabilistic constraints,” *Optimization*, vol. 56, no. 4, pp. 425–440, 2007.

Semantic Patterns to Structure Timeframes for Event Ordering Enhancement

Nour Matta
IT and Digital Society Laboratory
University of Technology of Troyes
Troyes, France
email: Nour.matta@utt.fr

Nada Matta
IT and Digital Society Laboratory
University of Technology of Troyes
Troyes, France
email: Nada.matta@utt.fr

Nicolas Declercq
Namkin
Troyes, France
email: n.declercq@namkin.fr

Agata Marcante
Research & Development
Namkin
Troyes, France
email: a.marcante@namkin.fr

***Abstract*— Event ordering is a field in Event Extraction that deals with the temporality aspect and order of occurrences of events mentioned in a text. Event Ordering is essential because any analysis of causalities and consequences of specific actions or changes of state requires a time evaluation. Standard approaches using machine learning models, with or without inferences, start by identifying events in text and then generate the temporal relationships between them individually. With no consideration of flashbacks, flash-forward, and direct speech temporal aspect, available models lack performance. In this paper, we introduce a novel approach to group events in temporal frames that we refer to as Timeframes. Three types of timeframes will be presented: Publication, Narrative, and Spoken. The purpose of this paper is to highlight the need of this approach, define the different timeframes, introduce their extraction process, evaluate the extraction and compare the event ordering with and without the timeframe approach.**

Keywords-Timeframe; Event Extraction; Event Ordering; Natural Language Processing.

I. INTRODUCTION

Event extraction is one of the most important tasks of Information Extraction through Natural Language Processing [1], [2]. It enables the extraction of events in text and aims to identify the different participants and attributes of the extracted events. Some examples of the extracted information can be the cause, place, time, means, or goal, which can be identified through dependency analysis [3]. Moreover, evaluating the influence of a particular event or a specific action requires an account of temporality [4]. In traditional event extraction, available approaches are very performant when it comes to the analysis of single sentences. Some approaches can support complex sentences. But even though models aim to extract events from a "text" and create temporal relations between them, the performance lacks, and soon the extracted information easily becomes unreadable or inaccurate from a temporal relation point of view [5]. Furthermore, when focusing on temporality event extraction, many approaches focused on ordering the multiple events mentioned one by one and creating relationships [6], [7] without considering the fact that multiple 'processes' can be part of a preparatory stage of a single event. In this paper, we introduce the use of timeframes, an approach used for time analysis in different domains, to improve the temporal relation made between events in a text.

It is important to note that within the same text, multiple timeframes can be identified and multiple time references can be used. A small example would be a news report about a company announcing the launching of a new product. We have the time when the announcement was made, the timeframe within the announcement (such as the date of the launching), and the time of the publication of the news. Another example would be in a narrative text in which the author talks about multiple events while going back and forth in time. Our main goal is to identify the events in a text, create temporal relations between them and identify the different timeframes if there are multiple ones. We aim to assign each event to its timeframe enabling improved readability of the extracted event, their temporal relation, and finally their interpretation.

In Section 2, we will start by defining what timeframes are and how they are used for time analysis. We will also present the different conceptualizations of events in linguistics. In Section 3, the related work, we will go through the different event extraction approaches. In Section 3, we will be presenting the timeframe approach along with part of the semantic pattern identified. In Section 4, we will provide an example of application of the proposed algorithm. In Section 5, the result and evaluation done on 120 news article from multiple source and multiple topic. Lastly, Section 5 provides the future work.

II. TIMEFRAMES AND EVENTS

This section is divided into two main parts: the timeframes and the events. For the timeframe, we will go through the analysis of temporality in fields other than text mining and show how those conceptualizations can be helpful in the analysis of temporality in text. As for the events, we will go through their definition in the event extraction field and how it is viewed from a linguistic point of view.

A. Timeframes

A timeframe is a certain period of time in which an event should happen or has already happened [8]. This leads us to question the meaning of time. In philosophy [9], the Platonist understanding of time is segregated from the Relationist definition. Platonists picture time as an "empty container" of events that exist regardless of whether anything is placed in it. In this perspective, Platonists consider that it is possible that changes in the universe can cease to exist for a certain period. On the other hand, Relationists view time as a set of events and the temporal relationship between these events. While dealing with event extraction and their temporal relationship, the Relationist understanding is used.

The study of the temporal relation between actions, events, states, and their influences is applicable in different domains other than event extraction. In their study on temporality in video games, Zagal et al. [4] distinguished multiple types of games: the ones with game time being equivalent to the real-world time, the ones in which action can speed up or skip time, the ones where specific action triggers events of a specific duration, and finally the ones where certain events occur without affecting the game time as if time had stopped. To analyze the temporality for each game type, Zagal et al. [4] defined timeframes, creating relations between those timeframes and between events within the timeframe and coordinating them. Reflecting on that approach, from a textual perspective, the authors also set the duration to specific events as shown in "1)", which can make flashbacks "2)" and flash-forwards "3)". They can also skip time "4)" and even focus on a specific event or describe elements making the time indirectly stop "5)".

1) *John ran for an hour.*
2) *Henry was looking at the photo. He took it a few years back when he was in New York.*
3) *John is preparing his luggage; he will be leaving in the morning.*
4) *Five years later, Henry went back to New York.*
5) *John looked through the window for a few seconds. It was a rainy day; people were walking while holding their umbrellas. He went to his desk.*

Distinguishing the different timeframes and specifying the events that happened in each frame enables the focus on specific events based on their occurrence time and aims to improve coordination between multiple timeframes in the text. However, in the event extraction field, events are ordered one by one without having a more global representation. Some of the concepts that must be considered while dealing with temporality are the duration, the time point, calendar, narrative time, timeline, countdown, and temporal relation [4], [8]. Each of these concepts plays a specific role in the pattern and the extracted knowledge. The use of timeframe also enables the consideration of the release date of the text as a timeframe on its own in order to improve topic tracking and event follow-up.

B. Events

In the event extraction field and the event-based decision systems, events are usually defined as happenings or changes that occurred in a specific interval of time. They can be associated with the change of states (canceled, ongoing, recently done, past or future plans) and can have multiple occurrences [2], [10].

Other than Natural Language Processing, linguists also worked on defining what an event is, distinguishing it from a state, and partitioning it onto atomic and extended events. Using the tense of the verb, the duration, and time reference along with temporal connectors, some set of rules and patterns are proposed. Vendler [11] was one of the first to work on defining the concept of event in linguistics, while working on verbs and tenses, he first identified the tense as the location of a happening in the time (past, present, or future) and its aspect which refers to the state of an event (completed, ongoing or interrupted). Later on, he defined "Eventualities" [11] as a concept that groups the state and non-state. Some particularities of each group were identified:

6) *Jack was ill on Sunday*.
7) *Jack wrote a letter on Sunday*.

"6)" is an example of a state, in which we cannot determine if the state "ill" started before or during the "Sunday" and if is stopped during "Sunday" or after. While the non-state "wrote" started on ended Sunday. And comparing the duration of "was ill" and wrote, we can presume that "wrote" has a shorter duration than "was ill".

Their conceptualization of eventualities goes as follows: non-state was divided into Activities and Events; activity refers to actions that had a duration but with no endpoint or consequent state while events have a quantification or an ending result:

8) *Alex ran.*
9) *Alex ran to the store.*
10) *Alex ran a mile.*

"8)" is considered an activity while "9)" and "10)" are events. Events are then distinguished [12]. Where an accomplishment is considered to have a duration and accept the progressive (continuous tense) while achievement is strange in progressiveness. It is important to note that Kamp highlighted the ambiguity between those concepts, starting with the very first division between distinguishing a state and a none state.

Using the conceptualization made by Vendler, Moens et al. [10] defined another conceptualization. Eventualities are divided into States and Events. And they considered two dimensions for distinguishing events: the duration and the consequence. For the duration, they considered events as Atomic or Extended. Extended Events have a notion of duration. For the consequence, they started by defining the term "culmination" as an event that has a consequence, a change of state. A "nucleus", as shown in Figure 1, is the combination of a preparatory process of a culmination, the culmination, and the consequent state. If we consider the example "9)", "Alex ran to the store", we can regard it as a culmination, in which running is a preparatory process to the culmination of arriving at the store. The consequent state is "being in the store".

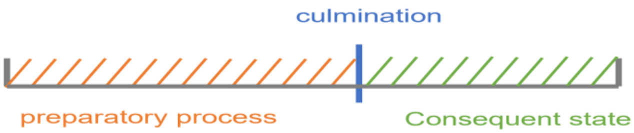


Figure 1. Moen et al. nulcreus definition [10].

Figure 2 presents the 4 subcategories of events. An Atomic Event with no consequence is considered as a point, for instance, "He hiccupped". A culmination is an atomic event with a consequence [10]. An extended event is a process and is considered a culminated process if it has a consequence. It is important to highlight that many elements were used to distinguish the different categories of events. This work and the pattern identified in it are essential for our approach especially the use of a nucleus. When using the timeframe approach, identifying the culminations in a text and all the processes, the preparatory stage, and the consequent state are one of our goals. The slice difference in our approach is that we intend to associate different events on the nucleus timeline.

	Atomic Event	Extended Event
- Consequence	Point	Process
+ Consequence	Culmination	Culminated Process

Figure 2. Moens' event conceptualization [10].

The pattern identified by Moens will be used and associated with other patterns to enable the representation of events extracted for each timeframe. When trying to identify events and states, the use of adverbs, the tense of the verb, and the use of semantic dependencies, such as the verbs' objects, were used for identifying the categories. The same verb can be considered a point, a process, a culmination of a culminated process depending on its use.

It is important to note that in linguistics, verbs tend to be classified as states and events. Adjectives are considered states of the elements they describe. In the event extraction field, nouns are also identified as events depending on their context, for example:

11) Two years after his graduation, John moved to New York.

In this sentence, the noun "graduation" is considered as an event. In order to manage these types of events, a pattern concerning nouns was added. If a noun is a temporal reference, in this case, "after graduation", then this noun is an event. Several studies define principles to identify and extract events along with their arguments; we summarized them in the related work.

III. RELATED WORK

This section will be partitioned as follows: we will start by going through event extraction techniques and more importantly Event Ordering. Then we will go through different research works that addressed temporality aspects in the temporality recognition techniques.

A. *Event Extraction*

Event Extraction is one of the most crucial branches of Information Extraction from textual data. In this field, an event is presented as an occurrence of a happening [5], [10]. Events may be related to a specific date, actors, places, objects, or other events. Event Extraction models provide effective benefits to build decision support systems or question-answering models [13]. The first work on event extraction started in the health care field in order to analyze bacterial behavior [14], and thanks to the progressive advancement of data science and big data, this field gained popularity to enable insights extraction from the published or shared textual data. Some of those data sources are social media posts, news messages, web pages, articles, or research papers. Knowledge discovery strategies that are event-oriented start with event identifications along with the related entities that play a role in those events and enable decision models based on the extracted information [15], [16].

Event extraction is divided into two types [7]: Open-domain event extraction and Closed-domain event extraction. Open-domain event extraction is a non-specific event detection and extraction. This approach is independent of annotated data. The detection of events and their arguments is based on their syntactic and semantic role. The arguments types are not specific to the event type but rather standard arguments such as actor or agent, location, object, time, cause, etc. This method is used for topic detection, story segmentation, and first story detection. Open-domain event extraction is also referred to as a data-driven approach [5] where researchers aim to convert data into knowledge using statistical methods and data mining. On the other hand, closed-domain event extraction is the extraction of previously defined events of interest. Those events are relative to the application domain. For instance, in the finance sector, business intelligence decision support models require economic event analysis. Event extraction enables the detection of business events along with the entities involved in each event [17]. An example could be a merger of two companies, the models aim to distinguish the old companies that had emerged from the newly created company. Another scenario can be seen in police reports analysis. Homicide or accident incidents may be detected along with specific arguments for each type of event that are precisely named and pertinent to certain occurrences including the location, the period, the victims, the offenders, etc. This type of event extraction can be used for event trigger detection, event mention detection, event argument extraction, and causal relation event analysis. There exist multiple methods that can be used for closed-domain event extraction. The pipeline approach [18] is a popular method that divides the extraction into multiple classification stages. The process is as follows: first, it uses keyword matching to identify the event trigger. Second, arguments are classified based on semantic and syntactic dependencies. Arguments can be simple words, noun phrase segments, or verb phrase segments. Lastly, a role is assigned for each argument. Another popular approach is the join-based approach in which the events type and the arguments are simultaneously classified. The purpose of this simultaneous classification is to avoid errors generated by the trigger classification set. To improve argument role identification, the reinforcement learning approach and the incremental learning model [6] can be applied. These methods distinguish different arguments related to an event within the same sentence. Deep-learning methods [19] train their models based on annotated datasets and use transformers such as BERT. Their performances depend on the quantity and quality of the annotations taken as input.

Knowledge Graphs based on event extraction are one of the potential representations for the events and their related entity [20]. Two types of representation are popular: event-centric and entity-centric knowledge graphs. Event-centric uses only events as nodes and the entities connected to the events are added inside the nodes as attributes for better visualization of all the events and how events are related to one another. On the other hand, entity centric considers all entities as nodes. This view enables a detailed representation of each event. In our work, we focus on events and only the temporal relations between them. In other words, we will be using the event-centric representation to represent the events and their relations in the provided figures.

B. Event Ordering

In order to study causal, consequential, or impart effects of events and happenings, time analysis of events occurring is essential [21]. Temporal relations analysis is a key factor for narrative processing, storyline construction, causality, and impact evaluation. In the question-answering field, temporal analysis was used to evaluate the correctness of the answer based on the temporal aspect. Two main categories of questions were identified. The first set of questions had permanent answers or answers for a long period of time such as “Who was the US president in 1996”. The second set had answers that were short-term and change a lot over time such as: “Who is the US president today” [13].

Two main temporal event relation extraction approaches are available: the data-driven approach and the hybrid approach which uses both data-based and constraint-driven approaches [22]. Data-driven models are trained using annotated datasets, and the most method for event ordering is the join-based approach. The constraint-based approach [23] enables applying inferences while processing the data to generate more information. An example of inferences is the transitivity of the event order or the reverse relation generation. For example, if event A happened before event B, the relation event B happened before event A can be created. Another example would be considering the three events A, B, and C, if A happened before B and B happened before C, then A happened before C.

There exist multiple datasets for temporal relation extraction, different versions of them, and those datasets differ in the relationships available, the type of entities linked, and when the relationship is considered or ignored. An example of the dataset would be TCR [24] which considers only five types of temporal relationships: “before”, “after”, “is included”, “includes”, and “simultaneously”. TimeBank [25] on the other hand also considers “ended by”, “During”, “Begun by”, “Begins”, “IBefore”, “IAfter”, “During”, and “Ends”. While TimeBank Dense [26] has a relation named “vague” that relates events with unclear temporal relations. Some datasets have annotation features of the event aspects label. In TimeML [27] for example, “Progressive”, “Perfective”, “Perfective Progressive” or none are assigned to each event.

There are three type of temporal relations available [28]: (1) relations between two events, (2) relations between events and time expressions, and (3) finally relations between time expressions and time expressions. Time expressions can be labeled as duration, date, time, or set. To compute the time value, TimeN provides a time normalization system that converts time expressions to actual Date Time relations [29]. It’s important to note that not all models consider all three types of relationships. Events can be nouns and verbs, the models match the tense of the verbs but don’t consider the culmination, point, process, and culmination process aspects. Events are assigned polarities such as negative and positive and modularity such as “would”, “may”, or “could”.

Multiple models are presented such as NavyTime [28], ClearTK [27], UTTime [30], CAVEO [26], Sequential Models based on LSTM [31], TEKMN [13], or structured learning approach [21]. Temporal label dependencies and constraints are used to improve relations between events [27]. Some worked on the linguistic and syntactic rules, such as Leeuwenberg et al. [23] or Laokulrat et al. [30]. The models vary in the search process, especially in terms of the distance between the events and time expression. Some models search for temporal relation intra-sentences; others go beyond single sentences. For instance, NavyTime searches for temporal relations of events are not only limited to the same sentence, but also in adjacent sentences, and between paragraphs. Another example would be UTTime which considers the relations between all events and the document creation time, events and time expression, and events mentioned in the same sentence and in consecutive sentences. The performance of the models depends on the datasets used along with the temporal relation types, and the overall search process, some models are built using multiple datasets.

The models that uses inference may go not only be limited to temporal relationship in their search. Since a cause occurs before its consequences, causal analysis enables temporal inferences [24]. So other than the previously mentioned transitivity and reverse inferences, the causal relation generates an ordering inference between the event and its cause. And vice versa, some models work on both temporal and causal relations and add constraints to learn more about the causal relations based on event ordering. Some of the datasets that used causal relations are Causal-TimeBank [32], Event Storyline [33], and FinReason [34].

Lastly, TIMERS [35] is considered a document-level approach that integrates causal prerequisite links, chain reasoning, and future events to improve temporal relationship extraction.

In this paper, we introduced timeframes as frames of time that will contain multiple events that happened in a specific period of time. This approach aims to group events together by detecting relationships between the different timeframes identified even though they are not mentioned in adjacent sentences. After detecting the relationship, we will apply one of the related work approaches for the event order. In this paper, we will be using the CAEVO model based on the TimeBank-Dense dataset for its availability. In the next section, we will present the different types of timeframes identified, their use, and how to extract them.

C. Temporality Recognition Techniques

In the Question Answering field, temporal analysis is a must for determining if an answer to a question will change throughout time or not. In their work, Pal et al. [36] identified multiple classes of information temporality: short duration, medium duration, long duration, and permanent. They tried classifying the question/answer under those categories but ended up grouping the short term and medium term together and long term and permanent together. It enabled distinguishing between “Who won the competition X in 2022?” and “Who won the last competition?”. One of the questions will have permanent information and the other will change throughout the years. It is important to consider these types of classifications to identify information that is true regardless of the timeframe of the text and relations that are relative to the timeframe of the text. Recent work focuses on identifying the attention in complex questions and the use of multiple sentences that contain the answer [13]. Note that in their work Kwiatkowski et al. [37] mentioned descriptive sentences or informative sentences, in which information is given without a particular event being mentioned.

Temporality plays a very important part in social science and social discourse analysis [38]. Coordination between different events from multiple resources is also used when clustering news and following up on events. Sources vary between news and social media posts, such as tweets [39]. It is also essential to consider time relations when analyzing the influence of social media and the media in general on social events, such as protests and violence and study the sentiments behind it [40].

The question answering field provided a very important aspect to consider when extracting events and information. Completed Events and states with a specific date tend to be permanent information while unfinished events and events with reference to the text temporality tend to be true in a specific timeframe. Coordination of events between multiple texts will be considered in our approach and will be based on the timeframe concept. Our approach introduces the use of multiple types of timeframes and how to extract them. We will be using several models and patterns already provided in order to optimize the model’s performance.

IV. TIMEFRAME APPROACH

For the extraction of timeframes that will be used to improve the temporal relation analysis between events, we identified three types of timeframes, (1) the Publication Timeframe, (2) the Narrative Timeframe, and (3) the Spoken Timeframe. Those timeframes were inspired by the identified timeframes for temporal analysis in video games with adaptation to the text constraints [4]. The Publication Timeframe reflects the publication date or year of the analyzed text. The Narrative Timeframe is the timeframe of the events happening in the text; we may find multiple Narrative Timeframes in a single document. Finally, the Spoken Timeframe is a particular type of timeframe that may not always appear in a text. It is used when an announcement, a speech, or a dialogue is present. The events and information that are mentioned in that context will be analyzed in their own timeframe in order to reduce event relationship complexity. The timeframe will consist of two main parts: (1) the text belonging to the timeframe, and (2) the extracted information related to it.

A. Publication Timeframe Extraction

All text documents have by default a Publication Timeframe and a Narrative Timeframe. To identify the publication date, the type of text affects the extraction. If a post on social media is being analyzed then, the date is usually available as metadata to the text. When dealing with online news, most publishers put the date at the beginning of the text. Considering the presence of the title, we will check the first three sentences, for the presence of dates using Named Entity Recognition. If no dates were found, the last sentence will be checked. In case a sentence was identified as the publication date, it will be extracted from the document in order to avoid confusion with the rest of the text. The date will be set in the information field of the timeframe. Figure 3 provides the Publication Timeframe extraction function. It takes two elements as input: a text, and the patterns that identify the publication date. The returned list contains two elements: The Publication Timeframe and the text. The text is returned since it is modified in case the pattern was found in a sentence.

```
Extraction_Publication(Single_Text, pub_pattern):
    sent = split_sentences(single_Text)
    tf_pub = [[],[]]
    to_check = [sent[0],sent[1],sent[2],sent[-1]]
    for element in to_check:
        identified = check_pattern(pub_pattern,element)
        if identified:
            timeframe_pub = [[date],[element]]
            remove element from Single_Text
            return [timeframe_pub , Single_Text]
    return [[], Single_Text]
```

Figure 3. Publication Timeframe Extraction.

Table 1 provides some of the patterns used to identify the Publication Timeframe. Please note that for the first two patterns, their presence in the sentence is enough while for the last two, they must be alone in the sentence to be considered a sign of Publication Timeframe.

TABLE I. SOME OF THE PATTERN USED TO DETECT PUBLICATION TIMEFRAMES

pub_patterns
‘Updated’ + <date>
‘Published’ + <date>
<number> + [hours, days, months, years] + ‘ago’
<date>

When identifying temporal relations between timeframes, the publication timeframe will have relations with the narrative timeframe. This will enable possible relation of the events mentioned in the narrative timeframes and the publication date.

B. Spoken Timeframe Extraction

```
Extraction_Spoken(Single_Text, say_pattern):
    list_sentence = split_in_sentences(single_Text)
    tf_speech_element = [[],[]]
    tfs_speech =[]
    id = 0
    before = false
    for sentence in list_sentence:
        identified = check say_pattern in sentence
        if identified and not before:
            before = true
            tf_speech_element = [id,[sentence]]
            replace(sentence,Single_Text,"tf_speech_"+id)
        else:
            if identified and before:
                add sentence to tf_speech_element[2]
                remove sentence from Single_Text
            else:
                if before:
                    before = false
                    add tf_speech_element to tfs_speech
                    id = id + 1
                    tfs_speech = [[],[]]
    return [tfs_speech , Single_Text]
```

Figure 4. Spoken Timeframe Extraction.

This timeframe will be treated before the Narrative Time. The search for verbs that reflect speaking and punctuation that are proper to dialogue will be the main task. If nothing is identified, we skip to the next stage, else a spoken frame will be created. If the “spoken” elements are all available in successive sentences, they will all be extracted and set in a single Spoken Timeframe. If multiple sentences have ‘spoken’ elements but are not successive, a Spoken Timeframe should be created for each nonconsecutive part. But in order to enable relations between the Narrative Timeframe and the Spoken Timeframe, identification will be assigned to each extracted Spoken Timeframe, and the extracted sentences will be replaced by the Spoken Timeframe Identification. If any dates are mentioned, they can be added to the information field of the timeframe. Note the tense in the Spoken Timeframe reflects a relationship between the Spoken and Narrative Timeframe it belongs to, so if a unique tense is identified, a relation between the Spoken and the Narrative Timeframe will be identified. For example, if future tense is identified in the “spoken” element, then the relationship will most probably be “after”. To keep track of this relationship, the relation if available will be added with the timeframe identification.

Figure 4 provides the Spoken Timeframe extraction function. It takes two elements as input: a text, and the patterns that identify the speaking patterns. The returned list contains two elements: The Spoken Timeframe list and the text. The Spoken Timeframe list contains all the Spoken Timeframes identified in the text. Each one contains an identification that distinguishes different segments in which the patterns were identified along with the sentences. Note that if consecutive sentences contain the patterns, then they will be grouped in the same timeframe. The return text is the remaining text with the identifications of the Spoken Timeframes.

We used a single pattern to identify the presence of a direct speech in a sentence. First, some direct speech may contain multiple sentences between quotation mark which reduced the accuracy of the dependency parsing. This is why, during the pattern matching phase, any text between quotation was replaced by “” and we analyzed the dependencies of the quotations. If the quotation mark is the object of the verb in the sentence, then we consider that the current sentence belongs to a Spoken Timeframe.

A small modification was brought to our algorithm to enable the next step of this approach that differs from Matta et al. [1]. In order to evaluate the temporal relation between the different timeframes, we should keep track of the tense of the verbs available in the direct speech. Therefore, we applied the algorithm of verb tense extraction. The tense extraction algorithm is presented in section C. The tense in the direct speech reflects the relations that the spoken timeframe has with the narrative timeframe it belongs to. For instance, if the tense is in the past such as “‘He worked on the program yesterday’, Simon said.”, then the spoke timeframe reflects on events that occurred before the narrative timeframe they were mentioned in.

C. Narrative Timeframe Extraction

The starting point of the Narrative Timeframe is having an empty information field and the whole text inside of it. The purpose of using multiple timeframes is to distinguish between current time in a text and in case a flashback is mentioned or flash-forward is mentioned, the information should be treated accordingly. Using the VerbNet parser [41], we detect any temporal relation. We associate a change in the timeframe when the relationship is not related to a specific event. For example, “before going to bed” is related to the event “go to bed” while “a few years ago” is a temporal relation with the current timeframe. We also consider “later that day” or “later that year” elements within the same timeframe.

In this section, we will present the elements that trigger the creation of a new Narrative Timeframe. The temporal relationship elements that will create a new Narrative Timeframe are: “a few years later”, “(number) years later”. The same goes for “months” and “days” instead of “years” and “ago” instead of “later”. Dates are relatively important; if a date is mentioned, it will be assigned as information about the timeframe. If no dates are mentioned, temporal relations that start with 'this' for example, 'this year', 'this month', 'today', will be considered as time information of the timeframe. If multiple dates are separately mentioned, each will be assigned a timeframe.

```
Extraction_Narrative(Single_Text, narrative_pattern, tense_pattens):
    list_sentence = split_in_sentences(single_Text)
    tf_nar_element = ["tf_nar_0",[],[]]
    tfs_narrative = []
    id = 0
    tocheck = 0
    for sentence in list_sentence:
        identified = check narrative_pattern in sentence
        if tocheck == 0:
            if not identified :
                add sentence to tf_nar_element[1]
            else:
                dominant_tense = check_tense(tf_nar_element)
                id=id+1
                add "tf_nar_"+id to tf_nar_element[1]
                add tf_nar_element to tfs_narrative
                sentence_tense = check_tense(sentence,tense_pattens)
                tf_nar_element = ["tf_nar_"+id,[sentence],[sentence_tense]]
                if (dominant_tense != sentence_tense):
                    tocheck = 1
        else:
            if not identified :
                sentence_tense = check_tense(sentence)
                if sentence_tense == tf_nar_element[2]:
                    add sentence to tf_nar_element[1]
                else:
                    add tf_nar_element to tfs_narrative
                    id=id+1
                    sentence_tense = check_tense(sentence)
                    tf_nar_element = ["tf_nar_"+id,[sentence],[sentence_tense]]
                    tocheck = 0
    add tf_nar_element to tfs_narrative
    return tfs_narrative
```

Figure 5. Narrative Timeframe Extraction.

Figure 5 provides the algorithm used for the Narrative Timeframes extraction. It takes as input the text, the patterns that identify the existence of a new timeframe, and the patterns that check the tense of a verb. The patterns that check the tense of the verbs are based on the part-of-speech tagging, dependencies, and the lemmatization of the verb. The lemmatization is the original form of a word without conjugation. We use it only to detect the verbs 'be' and "have". Those elements are provided by Natural Language Processing tools such as Spacy [42]. For the part-of-speech tags of interest, we used:

- "VB" is assigned to the verbs base form
- "VBD" is assigned to verbs in the past tense
- "VBG" is the gerund (a verb that ends with 'ing')
- "VBN" assigned to the verb in past participle form
- "VBP" is assigned to the verbs in non-third person singular present form
- "VBZ" is assigned to the verbs in the third person singular present form

We considered the 12 principal tenses, and Table 2 provides some of the tenses and their respective patterns. We grouped the 12 verb tenses in the respective 5 tense categories: past anterior, past, present, future, future anterior [11]. For example, present continuous and present simple will both be present while present perfect, past simple, and past continuous will be considered as past. Based on the verb tense, the function check_tense will return the category of the verb tense identified.

As for the patterns that identify the presence of a new Narrative Timeframe, Table 3 presents some of them.

TABLE II. SOME OF THE PATTERN USED TO DISTINGUISH VERB TENSE

Verb Tense	**tense_patterns**
Present Simple	pos = "VBZ" or pos = "VBP"
Present Continuous	verb with pos="VBG" and has_child = {dep= "AUX", pos = "VBZ" or "VBP", lemma="be"}
Past Simple	verb with pos="VBD"
Past Continuous	verb with pos="VBG" and has_child = {dep= "AUX", pos = "VBD", lemma="be"}
Future Simple	verb with pos="VB" and has_child = {dep= "AUX", pos = "MD", lemma="be"}

TABLE III. SOME OF THE PATTERNS THAT IDENTIFY NARRATIVE TIMEFRAMES

Narrative_Patterns
A few ['years', 'months', 'days'] ['later', 'ago', 'back']
['earlier', 'later'] ['this', 'that'] ['years', 'months', 'days']
In <date>
['starting', 'from', 'starting from'] <date>
<number> ['years', 'months', 'days'] ['later', 'ago', 'back']

The algorithm goes as follows: an empty Narrative Timeframe is initialized. We go through all the sentences and we check the presence of a pattern. If no pattern is identified, we add the sentence to the timeframe. If patterns that trigger the creation of a new Narrative Timeframe are identified, we generate an identification to the new timeframe and we add to the previous timeframe to keep track of their connection. We then check the tense of the previous timeframe and the tense of the new one and we save the current timeframe element in the list of Narrative Timeframes. If the tenses are similar, we just add the sentence to the new timeframe. If the tenses are similar, there is a high risk that the author switches back to the previous timeframe. In that case, for the upcoming sentences, we keep track of any changes in the tenses. This is the only case in which the change of tense will trigger a change in the Narrative Timeframe. In future work, just like the event ordering approaches, a change in tense will trigger relations between events. Examples (12) and (13) clarify the need for this process:

12) *John is thinking about his life in New York. A few years ago, he had to move out because of his parents' job. He misses his friends dearly.*

13) *Alice graduated with a master's degree. A few months later, she found a job in an international company. She was finally able to move out.*

In 12), the change of the tense use can simulate a go back to the previous timeframe or just a need to change timeframes. In our current algorithm, we will just separate the three timeframes and handle the relationships between different timeframes in future works. As for 13), the continuity in the tense simulates just a skip in time with no need for further tense monitoring of verb tenses. In the next section, we will be evaluating our approach.

V. EXAMPLE OF APPLICATIONS

In this section, we will present an output for each of the three algorithms provided above. In this section, we will present the output for each of the three algorithms provided above. We used news articles to evaluate the performance of the algorithms. The articles differ in themes, sources, and sizes. The topics and the data sources used are:

- Sports from www.nba.com
- Automotive from www.autoweek.com
- Aeronautics from www.aerotime.aero
- Healthcare from www.healthcarenews.com
- Energy from www.euronews.com
- Politics from www.bbc.com, www.cnn.com, and www.glogalnews.ca

We aimed to have multiple styles in writing, and multiple sources. In politics, the main goal of the multiple sources was the evaluation of the publication timeframe extraction. We tool 20 news articles for each topic so in total we got 120 articles to evaluate. The results will be presented as follow: we will start by providing an example of the output of each of the algorithm alone then we will evaluate the overall output of all the data. Please note that the evaluation of the algorithms is manually done since this approach is still in its early stages and no automatic or pre-annotated data is available.

A. Publication Timeframe Extraction Results

For each site, we started by testing the performance of the Publication Timeframe since the scraper used is not customized for each website. We were able to identify the Publication Timeframe.

Figure 6. Output of Publication Timeframe Extraction Algorithm.

Please note that a cleaning phase is necessary before applying the approach. Figure 6 presents one of the outputs of the algorithms. The figure provides part of the extracted text from a news site with the sentence with the pattern of interest highlighted in the input of the algorithm. We can notice the Publication Timeframe along with the rest of the text from which we removed the sentence with the pattern.

B. Spoken Timeframe Extraction Results

Figure 7 provides the output of one of the provided data. Please note that for better visualization, long paragraphs with no pattern were replaced by '...' in the figure.

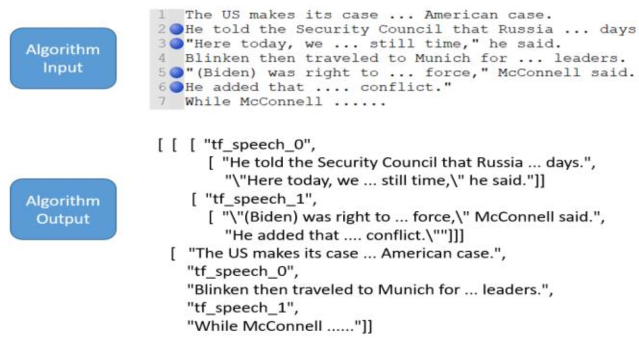


Figure 7. Output of Spoken Timeframe Extraction Algorithm.

In the example provided in Figure 7, we identified 4 sentences with patterns. They were distributed into 2 groups of consecutive sentences with patterns. The sentences in the example input were marked by blue dots next to them. In the output, we can notice that the algorithm provided a list with two elements in it, a Spoken Timeframe list and a list of the remaining sentences. The Spoken Timeframe list had 2 Spoken Timeframes in it, each having identification and consecutive sentences with patterns. As for the remaining text list, we notice that the extracted sentences were indeed replaced by their respective timeframe identification.

C. Narrative Timeframe Extraction Results

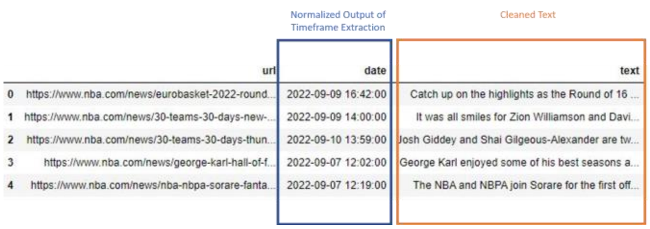
Finally, for the Narrative Timeframe, for the representation we used a text that had 2 Spoken Timeframes already identified in it. This enables an explicit viewing of how the identification and the linking between Narrative and Spoken is provided. The Spoken identifications in this example are "tf_speech_0" and "tf_speech_1". We also added blue dots next to the sentences with the patterns identified.

Figure 8. Output of Spoken Timeframe Extraction Algorithm.

We can notice in the output the Narrative Timeframe list returned by our algorithm in Figure 8. It contains the three expected timeframes having their respective identification, the sentences ordered that belong to the timeframe, and the tense of the last sentence to enable comparisons.

D. Results Analysis

The evaluation of the results will go as follow: we will tackle each timeframe type by order, Publication, Spoken, and Narrative. We started by separating the Publication Timeframe and we normalized the date available.



	url	date	text
0	https://www.nba.com/news/eurobasket-2022-round...	2022-09-09 16:42:00	Catch up on the highlights as the Round of 16 ...
1	https://www.nba.com/news/30-teams-30-days-new-...	2022-09-09 14:00:00	It was all smiles for Zion Williamson and Davi...
2	https://www.nba.com/news/30-teams-30-days-thun...	2022-09-10 13:59:00	Josh Giddey and Shai Gilgeous-Alexander are tw...
3	https://www.nba.com/news/george-karl-hall-of-f...	2022-09-07 12:02:00	George Karl enjoyed some of his best seasons a...
4	https://www.nba.com/news/nba-nbpa-sorare-fanta...	2022-09-07 12:19:00	The NBA and NBPA join Sorare for the first off...

Figure 9. Output of Publication Timeframe Extraction Algorithm Normalized.

Figure 9 shows a segment of the output of the Publication timeframe after a Normalization of the available date. When the hour is not available it is by default assigned "00:00:00". Once the data is cleaned we ran the Spoken Timeframe algorithm and finally the Narrative Timeframe algorithm. We started by evaluating the repartition of the Spoken Timeframes.

Figure 10. Pie Chart representing the repartition of the spoken timeframes.

Figure 10 provides the different repartitions of the of the spoken timeframes per domain. Sport and aeronautics both having around 25% with spoken timeframe, Automotive around 50% and energy, health and politics two third of the articles had Spoken Timeframes. The amount of data used is not sufficient to provide a deduction. It's important to highlight the fact that the longer text with Spoken Timeframes are the more likely we find multiple Spoken Timeframes. On the other hand, when the shorter the text were the more likely we found no or a single Spoken Timeframe. We also took this chance to evaluate manually the performance of our algorithm. In order to do that we started for each text identifying to sentences and how must they be grouped. We were able to determine that most Spoken Timeframe within the same Narrative Timeframe are usually completing the same idea or belong to the same speech. Nothing can be concluded though, since the data is not enough and even though we did mention large news articles, the longest is around 2 pages, we need to apply our approach on books and narratives to be more representative. We are pleased to get the same output, but we got one specific article that had a direct speech inside a different direct speech. The punctuation and the overall structure of the text was complex for us while manually extracting the timeframe and the algorithm had a two missing elements. This issue is not frequent so it will be handled later on. For the repartition of the narrative, almost 43% had only a single timeframe and the rest had varying amount the maximum being 6 Narrative timeframe per text.

It is important to mention that some dissimilarities were noticed when comparing the manually extracted narrative timeframe and our algorithm's output. To be more specific, we have two patterns that provoke the creation of a new Narrative Timeframe. The first one occurs when a temporal expression is identified in the beginning of a sentence. When that's the case no errors were identified. On the other hand, the other pattern is more complex. We must have a previously detected pattern where the tenses were different from the previous one. The purpose is to identify an end of a flashback or flash forward. While the use of a time expression with no change in the tense is considered a time skip or time jump in which we don't go back initial timeframe that occurred before the flashback or flash forward. The latter pattern generated new Narrative Timeframe when there was no need to generate a new timeframe. To be more specific, out of the 68 news articles that showed multiple timeframe 13 showed this pattern. Out of the 13 papers, only 5 showed an unnecessary Narrative Timeframe. This issue is not considered relevant for now since this will be solved in the timeframe temporal relation extraction in our future work. We will than relate the new timeframe to the previous timeframe or the one prior to it.

In order to evaluate the usefulness of our approach from an event ordering point of view, we decided to apply one of the related work event ordering algorithm to compare the performance with and without the timeframe approach. We selected the CAEVO event ordering [26] which is a CAscading EVent Ordering architecture, that using inferences to generate more temporal relationship. The cascade name reflects the effect inferences have on the relations. The model is built on the TimeBank Dense dataset and have the "Vague" relation along with "Before", "After", "Includes", "Is Included", and "Simultaneous". Just to put into perspective the importance of Event Ordering and Timeframe Event Ordering, we made a representation that highlight their use.

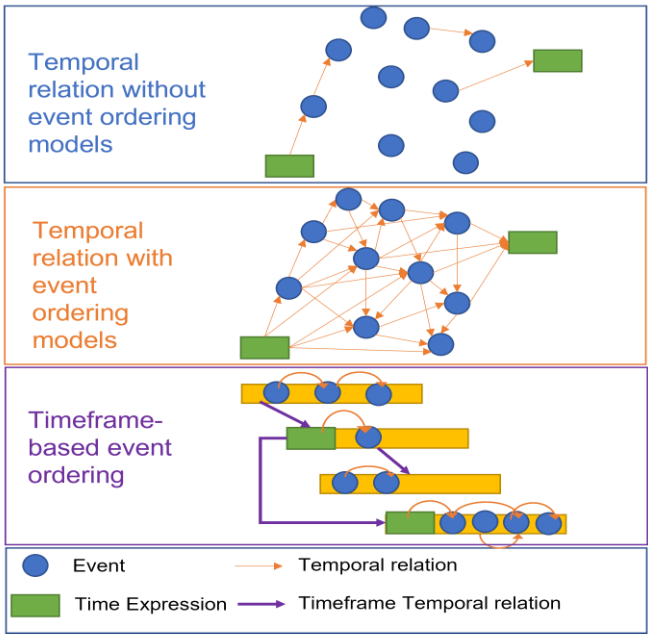


Figure 11. Temporal relations representation of multiple approaches.

Figure 11 shows the difference in the representation of temporal relations extract using multiple approaches. The first output shows a representation of an output without using an event ordering model. We can notice that hardly 4 events are connected and only 2 events are related to a time expression. The second presents the output of the same event extraction but after applying an event ordering model. This time most events are connected but the interpretation and the usability are complex. Finally, the third provides our desired representation using the temporal timeframes. This approach grouped events that occurred in the same period of time and limited the relation extraction between events from different timeframe. In this article we proposed the different type of timeframes and how to extract them without extracting the relationships between the timeframes. So in the current state of our approach, the Timeframe-based Event ordering is missing the temporal relation between the timeframes but using CAEVO, we evaluate the performance of the event ordering model intra-timeframe compared to not using our approach.

To elaborate on the event ordering evaluation, the first element done was evaluate the number of event which is identical in both, since the timeframe approach keep all the sentences in the text but segments them onto multiple timeframes. The temporal relation extracted highlighted the importance of our approach. Without the timeframe-based approach, on average more than 70% of the temporal relations were labeled "Vague" per text. The other most frequent relations were "After" and "Before". The rest of the temporal relations were really rare, 36% of the text only had the dominant relations, and the percentage of the three remaining relations were less than 4% in the text. Having an average of 73 temporal relations per document, after using timeframes, the average number of temporal relations dropped to 29. The most frequent temporal relations were the "Before" and "After" relations 61% of the text.

We had notice 7 articles where no change was detected, this is due to the fact that these texts had no Spoken Timeframes and a single Narrative Timeframe. The rest of the articles that had a single Narrative Timeframe showed a difference in the results due to the presence of Spoken Timeframe. When multiple narrative timeframes were identified we also noticed I a small decrease in the "non-vague" relations. This decrease is due to the fact that the relation between the time expression that led to the Narrative Timeframe split and the previous sentence were broken. These temporal relations are not considered lost since they will be restored once the timeframe to timeframe relations are added to the approach. But the best element is the "Vague" relations between two timeframes are no longer present. In the following section, we will present possible ways to generate the relation between the timeframes.

VI. FUTURE WORK

After distributing the text onto the different timeframe the main target becomes in identifying the temporal relationship between the different timeframes available. Our approach adds a new type of temporal relation which is the relation between two timeframes. It is important to note that we identified three main type of timeframe to timeframe relations:

- Publication timeframe to Narrative Timeframe
- Spoken Timeframe to Narrative Timeframe
- Narrative Timeframe to Narrative Timeframe

In this perspective, the publication timeframe will have a direct relation with the first narrative timeframe in the text. The tense of the verbs, along with the time expression when mentioned at the beginning of the narrative timeframe will be used to identify if the narrative timeframe is about event that happened 'Before', 'After', or 'During' the Publication Timeframe. The same goes for the Spoken Timeframe and the Narrative Timeframe. The Spoken Timeframe will only be related to the Narrative Timeframe it occurred in. The tense and temporal expression used will enable the value assigned to the relation. The last relation which aims to join two Narrative Timeframes is more complexed. We can have relations between consecutive Narrative timeframe and none consecutive. In the current state of our Narrative Timeframe generation, two triggers can generate the creation of a Narrative Timeframe. The first one is the presence of a time expression at the beginning of a sentence. The time expression will be used in this case for identifying the relation with the previous timeframe. When a specific date is mention a relation with the Publication timeframe can be added. The most complexed part is when a new timeframe is generated due to verb tense changes, in the specific pattern. For now, with no time expression and only verb tense to compare, the relation will be associated to the previous Narrative Timeframe and the first timeframe that has common verb tenses. For now, the relation generated will be labeled "Vague" since the tenses are not enough to deduct the relations. The proposed approach and criteria to extract

Timeframe relation is more developed and presented in Matta et al. [43].

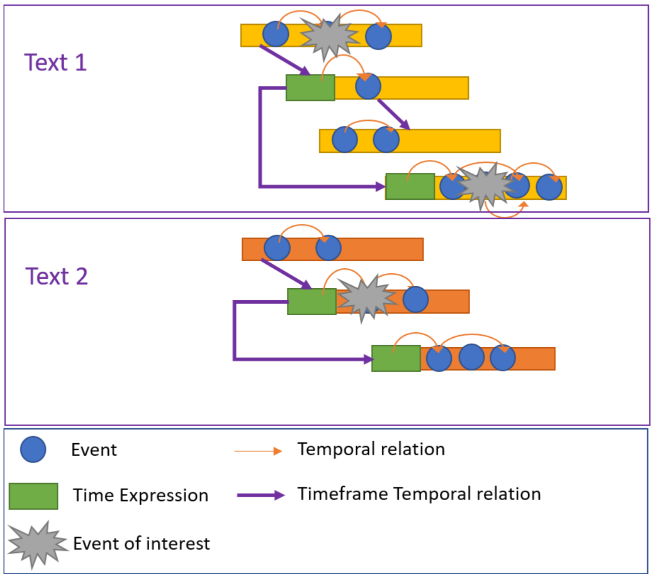


Figure 12. Intra-text and extra-text event desambiguation.

The last step in our work will be event disambiguation. We can notice in Figure 12, the event of interest that is mentioned in two different text, and mentioned twice in the same text. We intent to use Event Coreference Resolution, to identify the mention of the same event in the same text and use the timeframes to identify the change in state and entities of these events through text. And later on, identify the mention of specific event in multiple text, and using publication timeframes, narrative timeframes, and the relations between them, identify how the event of interest is progressing, or even analyses how object, entities are evolving through time. In that case, we would consider a person, an object, an organization, and view the events that are related to them.

VII. Conclusion

Finally, event extraction is an essential task in the Natural Language Processing field. It enables the use of text data in order to build decision-making systems and for event monitoring. It also enables story follow up, first story detection, event extraction along with the entities participating in the events. Those event can also be represented onto graphs, event centric or entity centric in order to have a clear visualization of the identified information. Event ordering is a branch in Event Extraction that focus on the identification of the temporal relation between the different event extracted. In this paper, we highlighted the need for timeframes to improve event ordering in the event extraction field. Three types of timeframes were presented: The Publication, the Narrative, and the Spoken Timeframe. Publication Timeframes will be used for multiple text analysis as a temporal indicator of the text. Narrative Timeframes enable the distinguishing of multiple periods of time used in a text, notably when a flashback or a flash-forward occurs. Finally, the Spoken Timeframe enables the distinction between the Narrative Timeframes and the timeframe of "spoken" elements in a text, such as announcements or dialogs. We set a few semantic patterns for the identification and extraction of the different timeframes. In future work section, we provided possible relation extraction methods for the timeframes and the events of each timeframe. Some of the process and criteria had been introduced but are still in their early stages. We intend to evaluate the performance on longer texts and a larger number. We will also be distinguishing the multiple classes of events: point, process, culmination, and culminated process in order to identify states available in timeframes. This work will complete our study on detection and representation context from text [44].

References

[1] N. Matta, N. Matta, N. Declercq, and A. Marcante, 'Semantic Patterns to Structure TimeFrames in Text', INTELLI 2022, May 2022, pp. 16–23. Accessed: Sep. 13, 2022. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=intelli_2022_1_40_60016

[2] X. Wu, T. Wang, Y. Fan, and F. Yu, 'Chinese Event Extraction via Graph Attention Network', ACM Trans. Asian Low-Resour. Lang. Inf. Process., vol. 21, no. 4, p. 71:1-71:12, Jan. 2022, doi: 10.1145/3494533.

[3] W. Zhao, J. Zhang, J. Yang, T. He, H. Ma, and Z. Li, 'A novel joint biomedical event extraction framework via two-level modeling of documents', Inf. Sci., vol. 550, pp. 27–40, Mar. 2021, doi: 10.1016/j.ins.2020.10.047.

[4] J. Zagal and M. Mateas, 'Time in Video Games: A Survey and Analysis', Simul. Gaming - Simulat Gaming, vol. 41, Jan. 2010, doi: 10.1177/1046878110375594.

[5] F. Hogenboom, F. Frasincar, U. Kaymak, F. de Jong, and E. Caron, 'A Survey of event extraction methods from text for decision support systems', Decis. Support Syst., vol. 85, pp. 12–22, May 2016, doi: 10.1016/j.dss.2016.02.006.

[6] Q. Li et al., 'Reinforcement Learning-Based Dialogue Guided Event Extraction to Exploit Argument Relations', IEEEACM Trans. Audio Speech Lang. Process., vol. 30, pp. 520–533, 2022, doi: 10.1109/TASLP.2021.3138670.

[7] W. Xiang and B. Wang, 'A Survey of Event Extraction From Text', IEEE Access, vol. 7, pp. 173111–173137, 2019, doi: 10.1109/ACCESS.2019.2956831.

[8] M. J. Buehner and J. May, 'Knowledge mediates the timeframe of covariation assessment in human causal induction', Think. Reason., vol. 8, no. 4, pp. 269–295, Nov. 2002, doi: 10.1080/13546780244000060.

[9] W. H. Newton-Smith, 'The Structure of Time', Routledge & CRC Press, Jun. 01, 2020. https://www.routledge.com/The-Structure-of-Time/Newton-Smith/p/book/9781138394063 (accessed Feb. 09, 2022).

[10] M. Moens and M. Steedman, 'Temporal Ontology and Temporal Reference', Comput. Linguist., vol. 14, no. 2, pp. 15–28, 1988.

[11] Z. Vendler, 'Verbs and Times', Philos. Rev., vol. 66, no. 2, pp. 143–160, 1957, doi: 10.2307/2182371.

[12] Z. Vendler, Linguistics in Philosophy. Cornell University Press, 2019. doi: 10.7591/9781501743726.

[13] X. Duan et al., 'Temporality-enhanced knowledgememory network for factoid question answering', Front. Inf. Technol. Electron. Eng., vol. 19, no. 1, pp. 104–115, Jan. 2018, doi: 10.1631/FITEE.1700788.

[14] Y. Chen, Z. Ding, Q. Zheng, Y. Qin, R. Huang, and N. Shah, ‘A History and Theory of Textual Event Detection and Recognition’, IEEE Access, vol. 8, pp. 201371–201392, 2020, doi: 10.1109/ACCESS.2020.3034907.

[15] M. Mejri and J. Akaichi, ‘A Survey of Textual Event Extraction from Social Networks’, 2017.

[16] H. Zhou, H. Yin, H. Zheng, and Y. Li, ‘A survey on multi-modal social event detection’, Knowl.-Based Syst., vol. 195, p. 105695, May 2020, doi: 10.1016/j.knosys.2020.105695.

[17] G. Jacobs and V. Hoste, ‘SENTiVENT: enabling supervised information extraction of company-specific events in economic and financial news’, Lang. Resour. Eval., vol. 56, no. 1, pp. 225–257, Mar. 2022, doi: 10.1007/s10579-021-09562-4.

[18] J. Liu, L. Min, and X. Huang, ‘An overview of event extraction and its applications’, ArXiv211103212 Cs, Nov. 2021, Accessed: Jul. 18, 2022. [Online]. Available: http://arxiv.org/abs/2111.03212

[19] S. Li, H. Ji, and J. Han, ‘Document-Level Event Argument Extraction by Conditional Generation’, ArXiv210405919 Cs, Apr. 2021, Accessed: Aug. 05, 2021. [Online]. Available: http://arxiv.org/abs/2104.05919

[20] F. Li et al., ‘Time event ontology (TEO): to support semantic representation and reasoning of complex temporal relations of clinical events’, J. Am. Med. Inform. Assoc., vol. 27, no. 7, pp. 1046–1056, Jul. 2020, doi: 10.1093/jamia/ocaa058.

[21] Q. Ning, Z. Feng, and D. Roth, ‘A Structured Learning Approach to Temporal Relation Extraction’, ArXiv190604943 Cs, Jun. 2019, Accessed: Mar. 22, 2022. [Online]. Available: http://arxiv.org/abs/1906.04943

[22] C.-G. Lim, Y.-S. Jeong, and H.-J. Choi, ‘Survey of Temporal Information Extraction’, J. Inf. Process. Syst., vol. 15, no. 4, pp. 931–956, 2019, doi: 10.3745/JIPS.04.0129.

[23] A. Leeuwenberg and M.-F. Moens, ‘Structured Learning for Temporal Relation Extraction from Clinical Records’, in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, Valencia, Spain, Apr. 2017, pp. 1150–1158. Accessed: Feb. 09, 2022. [Online]. Available: https://aclanthology.org/E17-1108

[24] Q. Ning, Z. Feng, H. Wu, and D. Roth, ‘Joint Reasoning for Temporal and Causal Relations’, ArXiv190604941 Cs, Jun. 2019, Accessed: Mar. 22, 2022. [Online]. Available: http://arxiv.org/abs/1906.04941

[25] J. Pustejovsky et al., ‘The TimeBank corpus’, Proc. Corpus Linguist., Jan. 2003.

[26] N. Chambers, T. Cassidy, B. McDowell, and S. Bethard, ‘Dense Event Ordering with a Multi-Pass Architecture’, Trans. Assoc. Comput. Linguist., vol. 2, pp. 273–284, Oct. 2014, doi: 10.1162/tacl_a_00182.

[27] S. Bethard, ‘ClearTK-TimeML: A minimalist approach to TempEval 2013’, in Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, USA, Jun. 2013, pp. 10–14. Accessed: Feb. 07, 2022. [Online]. Available: https://aclanthology.org/S13-2002

[28] N. Chambers, ‘NavyTime: Event and Time Ordering from Raw Text’, in Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, USA, Jun. 2013, pp. 73–77. Accessed: Jun. 07, 2022. [Online]. Available: https://aclanthology.org/S13-2012

[29] H. Llorens, L. Derczynski, R. Gaizauskas, and E. Saquete, ‘TIMEN: An Open Temporal Expression Normalisation Resource’, in Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12), Istanbul, Turkey, May 2012, pp. 3044–3051. Accessed: Apr. 01, 2022. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2012/pdf/128_Paper.pdf

[30] N. Laokulrat, M. Miwa, Y. Tsuruoka, and T. Chikayama, ‘UTTime: Temporal Relation Classification using Deep Syntactic Features’, in Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, USA, Jun. 2013, pp. 88–92. Accessed: Aug. 07, 2022. [Online]. Available: https://aclanthology.org/S13-2015

[31] P. K. Choubey and R. Huang, ‘A Sequential Model for Classifying Temporal Relations between Intra-Sentence Events’, ArXiv170707343 Cs, Jul. 2017, Accessed: Mar. 22, 2022. [Online]. Available: http://arxiv.org/abs/1707.07343

[32] P. Mirza and S. Tonelli, ‘CATENA: CAusal and TEmporal relation extraction from NAtural language texts’, p. 12, 2016.

[33] T. Caselli and P. Vossen, ‘The Event StoryLine Corpus: A New Benchmark for Causal and Temporal Relation Extraction’, in Proceedings of the Events and Stories in the News Workshop, Vancouver, Canada, Aug. 2017, pp. 77–86. doi: 10.18653/v1/W17-2711.

[34] P. Chen, K. Liu, Y. Chen, T. Wang, and J. Zhao, ‘Probing into the Root: A Dataset for Reason Extraction of Structural Events from Financial Documents’, in Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Online, Apr. 2021, pp. 2042–2048. doi: 10.18653/v1/2021.eacl-main.175.

[35] P. Mathur, R. Jain, F. Dernoncourt, V. Morariu, Q. H. Tran, and D. Manocha, ‘TIMERS: Document-level Temporal Relation Extraction’, in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Online, Aug. 2021, pp. 524–533. doi: 10.18653/v1/2021.acl-short.67.

[36] A. Pal, J. Margatan, and J. Konstan, ‘Question temporality: identification and uses’, in Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, New York, NY, USA, Feb. 2012, pp. 257–260. doi: 10.1145/2145204.2145246.

[37] T. Kwiatkowski et al., ‘Natural Questions: A Benchmark for Question Answering Research’, Trans. Assoc. Comput. Linguist., vol. 7, pp. 453–466, Aug. 2019, doi: 10.1162/tacl_a_00276.

[38] J. Hamann and L. Suckert, ‘Temporality in Discourse: Methodological Challenges and a Suggestion for a Quantified Qualitative Approach’, Forum Qual. Sozialforschung Forum Qual. Soc. Res., vol. Vol 19, p. No 2 (2018), Mar. 2018, doi: 10.17169/FQS-19.2.2954.

[39] V. Moutinho, João Cordeiro, and P. Brazdil, ‘Association and Temporality between News and Tweets’, pp. 500–507, Sep. 2019, doi: 10.5220/0008362105000507.

[40] T. Poell, ‘Social media, temporality, and the legitimacy of protest’, Soc. Mov. Stud., vol. 19, no. 5–6, pp. 609–624, Nov. 2020, doi: 10.1080/14742837.2019.1605287.

[41] S. W. Brown, J. Bonn, J. Gung, A. Zaenen, J. Pustejovsky, and M. Palmer, ‘VerbNet Representations: Subevent Semantics for Transfer Verbs’, in Proceedings of the First International Workshop on Designing Meaning Representations, Florence, Italy, Aug. 2019, pp. 154–163. doi: 10.18653/v1/W19-3318.

[42] E. Partalidou, E. Spyromitros-Xioufis, S. Doropoulos, S. Vologiannidis, and K. I. Diamantaras, ‘Design and implementation of an open source Greek POS Tagger and Entity Recognizer using spaCy’, in 2019 IEEE/WIC/ACM

International Conference on Web Intelligence (WI), Oct. 2019, pp. 337–341.

[43] N. Matta, N. Matta, N. Declercq, and A. Marcante, Evolution Discovery in Textual Data. 18th Annual International Conference on Information Technology & Computer Science (ATINER 2022); in press.

[44] N. Matta, N. Matta, E. Giret, and N. Declercq, 'Enhancing Textual Knowledge Discovery using a Context-Awareness Approach', in 2021 International Conference on Computational Science and Computational Intelligence (CSCI), Dec. 2021, pp. 233–237. doi: 10.1109/CSCI54926.2021.00071.