IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 November 2022, accepted 20 December 2022, date of publication 26 December 2022, date of current version 30 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3232278

RESEARCH ARTICLE

An Investigation of Contemporary Data-Driven Methods Applied to Complex Systems

SAUMITRA DWIVEDI[1,2], RICARDO DA SILVA TORRES[1,3], (Member, IEEE), IBRAHIM A. HAMEED[1], (Senior Member, IEEE), AND ANNIKEN SUSANNE T. KARLSEN[1]

[1]Department of ICT and Natural Sciences, Norwegian University of Science and Technology, 6025 Ålesund, Norway
[2]Department of Computer Science, Norwegian University of Science and Technology, 7491 Trondheim, Norway
[3]Wageningen Data Competence Center and Farm Technology Group, Wageningen University and Research, 6708 PB Wageningen, The Netherlands

Corresponding author: Saumitra Dwivedi (saumitra.dwivedi@ntnu.no)

This work was supported by the Norwegian University of Science and Technology (NTNU) (https://www.ntnu.no/).

ABSTRACT Understanding complex systems by the help of modeling, simulation, and control is a well-known challenge across several application domains. As such, these type of systems are not amenable to empirical models, typically due to their dynamic and non-linear nature. Data-driven methods, have in recent years gained traction regarding their application to such complex systems. This paper presents findings from a systematic literature review of data-driven methods applied to complex systems in recent years. The review is a result of a search strategy resulting in 1106 scientific publications, out of which forty one were identified as primary studies. It is our goal that the review findings will guide researchers with an evidence-base regarding data-driven methods, thereby equipping them with an arsenal of applied methods across various application domains. Also, the objective is to construct a knowledge-base based on these methods' contributions while applied to different types of problems. Additionally, the review findings may also guide researchers to realize potential gaps for future research.

INDEX TERMS Complex systems, data-driven methods, systematic map, systematic review.

I. INTRODUCTION

Several real-world systems are described by their component parts, which may exhibit collective behavior through self-organization and properties like emergence and non-linear evolution. Such systems can be found in several application domains, such as epidemiology [1], social sciences [2], and urban systems [3]. These systems are commonly referred to as complex systems [4], [5].

Understanding complex systems through modeling and simulation has been a well-known challenge across several scientific and technological domains. Several scientific methods have been developed, to predict or control such complex systems [6]. These methods are mainly based on expert knowledge and, more recently, measured data [7], [8]. With the advent of advanced computational systems, data-driven methods are gaining popularity due to the fact that the understanding of complex systems can be more robust if measured data is used [9].

According to Solomatine et al. [10], data-driven discovery or modeling is based on analyzing data about a system, in particular to find connections between the system state variables (input, internal, and output variables) without explicit knowledge of the physical behaviour. Data-driven discovery has been referred to as the fourth paradigm of scientific discovery, providing opportunities to engineers and scientist through vast quantities of data [11]. This fourth paradigm is a consequence of the first three paradigms: empirical experimentation, analytical derivation, and computational investigation [9]. Recent scientific publications, such as [12], [13], and [14], present efforts pertaining to data-driven modeling in the fields of Chaotic systems, Financial trading, Epidemiology, Video processing, Neuroscience, Control theory, and Fluid dynamics. According to Brunton and Kutz [9], such complex systems are not amenable to empirical models or derivations based on first principles (i.e., classic theoretical

The associate editor coordinating the review of this manuscript and approving it for publication was Hamed Azami.

work), due to their typically non-linear, dynamic, multi-scale (space and time) nature. As such, researchers are increasingly turning to data-driven approaches to model, predict, and control complex systems.

Motivated by the current rapid influx of new data-driven methods applied to several domains and a corresponding lack of recent systematic reviews having a broader scope of application, we have conducted a review with the following objectives:

1) To establish an evidence-base of state-of-the-art data-driven methods using a systematic literature review methodology,
2) To establish a broader knowledge-base of data-driven methods, by aiming the study scope to be across several application domains.

Based on these objectives we defined the following research questions:

- **RQ1:** What are major data-driven methods applied to complex systems described in literature?
- **RQ2:** What are major data-driven methods' main contributions?

In the next section we briefly discuss some recent secondary studies performed in relation to the application of data-driven methods. In section III, we detail the research methodology adopted for our study. Next, we show and discuss the results of the study in section IV and V respectively. In section VI, we present our conclusions.

II. RELATED WORKS

Applications of data-driven methods are being studied for various research and practice domains, in the form of comprehensive reviews. For instance, in [15] Wadoux et al. state soil as a complex system in which biological, chemical and physical interactions take place. The behaviour of these interactions changes in spatial scale from the atomic to the global, and in time. Wadoux et al. [15] discuss several issues and opportunities of knowledge discovery from soil data, with application of data-driven methods to soil science. Notably, [15] served as a starting point for further discussions on new epistemological challenges facing soil science research in the information era. In [16], Rajendra and Brahmajirao attempt to review application of deep learning methods to dynamical systems. A major focus of the review in [16], is to identify data-driven representations that make nonlinear systems amenable to linear analysis. How et al. [17] present a review of recently used state of charge (SOC) estimation methods highlighting the model-based and data-driven approaches for lithium-ion batteries. Woldaregay et al. [18] present a review identifying and discussing the current trends of machine learning applications within the context of medicine, specifically, pertaining to the controller of an artificial pancreas (closed-loop systems), modeling of personalized profiles, personalized decision support systems, and blood glucose alarm event applications. In [19], Peres et al. present a systematic review of current industrial artificial intelligence literature, focusing on its application in real manufacturing environments to identify the main enabling technologies and core design principles. The novelty of [19] is that they provide a clear definition and holistic view of Industrial Artificial Intelligence within Industry 4.0.

The above-mentioned studies comprehensively identify and discuss applications of data-driven methods focused on various individual research and practice domains pertaining to complex systems. Furthermore, studies like [19] are rare examples of comprehensive reviews with a cross-domain review scope, pertaining to applications of such data-driven methods. Indeed, deeper insights may be obtained, if such method applications are analyzed across types of problems and thereby reviewed for types of their contribution.

In this paper, we present findings from our cross-domain investigation of the use of such data-driven methods, while identifying and analyzing contributions of these methods' ability to tackle a variety of problems.

III. RESEARCH METHODOLOGY

In our study we have adopted a research methodology based on the guidelines presented by Kitchenham and Charters [20] and Petersen et al. [21], which cover strategies for conducting systematic review and mapping. Systematic review and mapping are becoming popular methodologies to establish an evidence-base and a knowledge-base about a research topic in a structured and comprehensive manner [21], [22].

The concepts of review used in our study is in line with the definitions provided by Kitchenham and Charters [20], Petersen et al. [21]. According to Kitchenham and Charters [20, p. vi], a systematic literature review is defined as "a form of secondary study that uses a well-defined methodology to identify, analyse and interpret all available evidence related to a specific research question in a way that is unbiased and to a degree repeatable." According to Petersen et al. [21], a systematic mapping study, on the other hand, is usually aimed at categorizing and summarizing the methods applied in several research publications related to a topic or research question. A systematic map helps to provide context for a more in-depth systematic review of a research topic/question by identifying available evidence [20], [23]. Despite the fact that systematic review and mapping studies have different objectives, they follow almost similar processes [24]. In our study, we used a combination of both methodologies, i.e., both systematic mapping and systematic review, to establish an evidence-base pertaining to the research questions. In particular, practices around the systematic mapping methodology were used to identify the methods applied across various application domains, and map their popularity among researchers. On the other hand, we used the systematic review methodology to identify and analyze the major contributions of such methods, when applied to a variety of problems. Specifically, in line with [20], the systematic mapping methodology is used to construct an evidence-base of use of methods across various studies. Whereas, the systematic review methodology is used to understand the contribution of the applied methods in the studies and hence to classify such methods based on a novel proposed taxonomy, culminating into a knowledge-base.

Systematic review and mapping studies follow a three phase process in line to [20].

- Planning: The first phase deals with specification of the research question and the review protocol.
- Conducting: The second phase forms a large portion of the review study where the studies are first identified for review through a systematic process. Next, the selected studies are extracted and synthesized for data related to the research questions of the review.
- Reporting: The third phase relates to reporting, evaluation and publication of the review.

For this study, we conducted the systematic mapping and review through an iterative four-stage process, as depicted in Figure 1. In the first stage, we conducted the literature search. In the second and third stages, we selected studies by first going through the title, abstracts, and keywords, and then going through the full text. In the final fourth stage, we extracted the data and performed a synthesis. Iterations were introduced to this four-stage process, in order to obtain a refined pool of publications being better capable of answering the research questions.

In the following subsections, we describe the four stages of the review and mapping process in detail.

A. SEARCH STRATEGY

As guided by Kitchenham and Charters [20], we recognize that the quality of systematic review and mapping studies highly depends on the rigour of the search process. Therefore, we directed our search strategy to find studies relating to the research questions in a comprehensive manner. We achieved this by a careful choice of search venues, and a search string based on research questions, taking care of synonyms, abbreviations, and alternative spellings.

1) SEARCH VENUES

We used Scopus[1] since it indexes most of the studies found in other databases like IEEE Xplore,[2] ACM Digital library[3] etc., while not including grey literature (e.g., unpublished materials) [25]. This contributes to a comprehensive search of literature, while maintaining high research publication quality (through peer-reviewed sources), to be considered for mapping and review.

2) SEARCH KEYWORDS AND STRING

We crafted the search string used for literature search, while taking into account synonyms, abbreviations and alternative words. We achieved this by the use of Boolean operators "AND" and "OR," in conjunction with quotation marks ("") and "*." Table 1 presents the used search string.

TABLE 1. Search string.

String
(TITLE-ABS-KEY("nonlinear dynamic*") OR TITLE-ABS-KEY("non-linear dynamic*") OR TITLE-ABS-KEY("complex system*")) AND (TITLE-ABS-KEY("data-driven") OR TITLE-ABS-KEY("data-based")) AND (TITLE-ABS-KEY(simulation) OR TITLE-ABS-KEY(model*))

TABLE 2. Inclusion and exclusion criteria (Final iteration).

Inclusion Criteria (IC)	Exclusion Criteria (EC)
IC1: The study is focused on complex systems and/or non-linear dynamical systems	EC1: Papers published earlier than 2019
IC2: The study deals with the use of data-driven methods	EC2: The paper does not have an abstract, or is published just as an abstract
IC3: Subject are related to Computer science, Engineering or Mathematics	EC3: The paper is written in a language other than English
	EC4: Non-peer reviewed publication sources and publication sources with low Citescore (less than 6)
	EC5: Subject areas related to – Chemical Engineering – Neuroscience – Arts and humanities – Chemistry – Material science – Agriculture and biological sciences – Biochemistry – Genetics and molecular biology – Psychology
	EC6: Paper doesn't deal with primary studies focused on analyzing and modeling the complexity of the system.

B. SELECTION CRITERIA

We created the study selection strategy, to filter studies capable of providing context or information to answer the defined research questions. To obtain evidence on methods applied to complex systems, we included studies dealing with complex systems, in particular, non-linear dynamical systems. Also, to select studies representing the state of the art, we included articles published between 1^{st} January 2019 till 21^{st} April 2021 (date of search). Moreover, we selected studies fully published (not just as abstracts), in English language. To achieve this, we created an inclusion-exclusion criteria (IC-EC) as given in Table 2. We then applied the IC-EC successively to the paper abstracts and then to the full-text of the papers (stage 2 and 3 in Figure 1) to obtain the final pool of studies for data extraction and synthesis.

[1]Scopus – https://www.elsevier.com/solutions/scopus (As of Feb. 2022).
[2]IEEE Xplore – https://ieeexplore.ieee.org/Xplore/home.jsp (As of Feb. 2022).
[3]ACM Digital library – https://dl.acm.org/ (As of Feb. 2022).

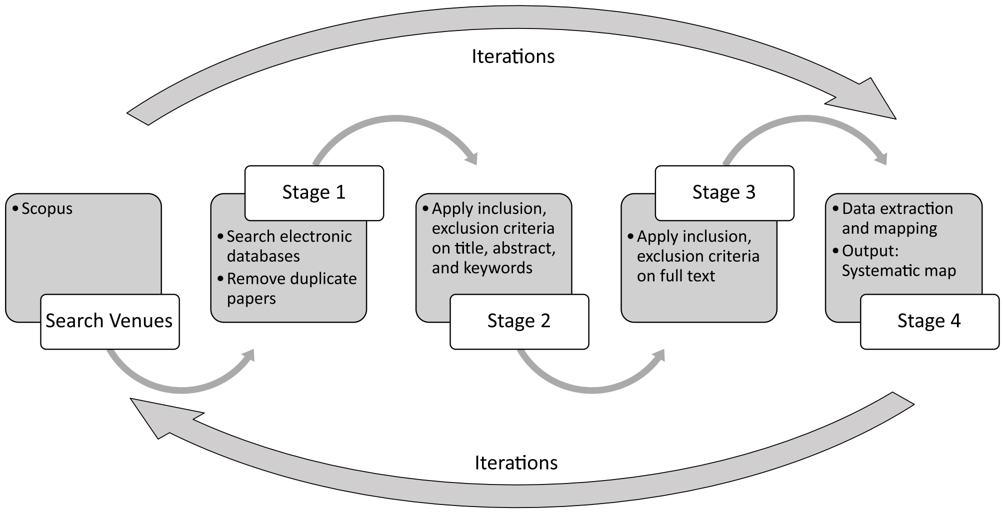


FIGURE 1. Systematic review and mapping process adopted in this study as an iterative four-stage process.

1) ITERATIONS

For our study, we designed the review and mapping process as an iterative process, in line with [26]. Iterations were introduced to refine the list of papers to be reviewed, in order to obtain a pool of publications being better capable of answering the research questions.

We employed refinements based on the following criteria through iterations:

- We deemed it important to consider only high quality papers for review and mapping. Hence EC4 was introduced, through which CiteScore[4] was used to filter out low impact journals and conference proceedings.
- We decided to limit the scope of the review to be related to computer science, engineering, and mathematics only. Thus IC3 and EC5 were introduced.
- We decided to exclude publications, with insufficient focus and discussion regarding its contribution to analyze or model the complexity of the system. Thus EC6 was introduced.

C. DATA EXTRACTION AND SYNTHESIS STRATEGY

To obtain data relevant to the research questions, we extracted the following information:

- Applied modeling method: Data regarding the applied method and its components.
- Problem type: Data about the type of problem addressed in the studies.
- Method contribution: Data about the main contribution of the applied method, usually indicating the novelty in the studies.

Furthermore, we cross-mapped the data extracted as listed above, to synthesize insights pertaining to the research questions, i.e.,

- to investigate RQ1, we analyzed the use of methods for various problem types, gaining insight on the popularity of methods as per their use for different problem types.
- to investigate RQ2, we created a taxonomy to classify the contribution type of the used methods and performed analysis of contribution types for each problem type. This provided us with insights in the form of a knowledge-base regarding the contextual use of methods, for different types of problems.

IV. RESULTS

Based on the methodology described in previous sections, we obtained search results, extracted data and synthesis:

A. SEARCH RESULTS

We identified 1106 records through search in the Scopus database. Next we screened records for eligibility through IC-EC. Figure 2 shows the result of the final iteration of the search process. Finally, we ended up with 41 publications, used for data extraction and synthesis (see Table 5 in Appendix).

B. EXTRACTED DATA

The review data consisted of highly inspirational state of the art methods applied to a variety of problems. For instance, [32], [33], [34] deal with control of non-affine/nonlinear discrete time systems. The methods applied in [32] show an insightful combination of data-driven dynamic linearization to the conventional control methods like active disturbance rejection control (ADRC). References [34] and [33] produce novel Iterative learning control (ILC) and Model free adaptive control (MFAC) frameworks, by utilizing full form dynamic linearization (FFDL) technique in the iteration and time domains. A similar attempt of using data-driven methods was evident in [35], to develop a data-driven adaptive learning consensus protocol for multi-agent systems (MAS) with a

[4]CiteScore https://www.elsevier.com/connect/editors-update/citescore-a-new-metric-to-help-you-choose-the-right-journal (As of Feb. 2022)

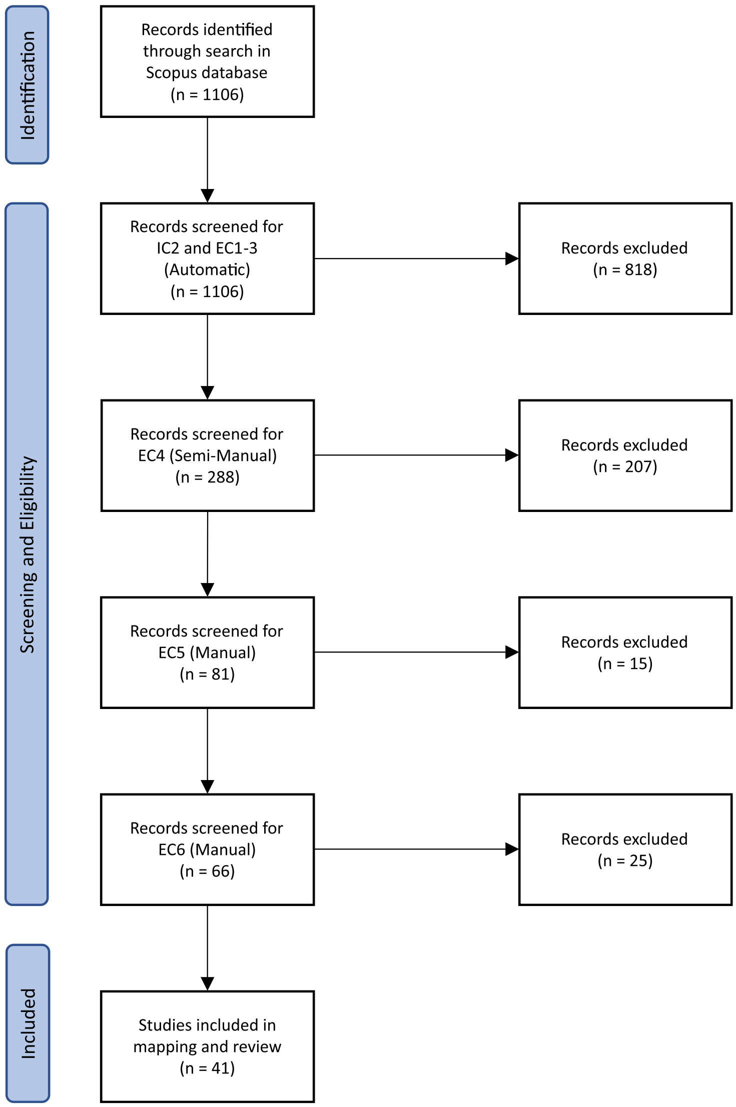


FIGURE 2. Prisma flow diagram for the final iteration of the review and mapping process.

strong learning ability to improve consensus performance by learning from both time dynamics and spatial dynamics.

Many studies have also used Reinforcement learning (RL) for data-driven control. For instance, in [36] Gros and Zanon utilized non-linear model predictive control (NMPC) as a function approximator in the RL based control framework. [37] in turn, used RL to solve the globally robust optimal output regulation problem (GROORP) of partially linear systems. The system studied in [37] was challenging due to unknown inherent dynamics with both static and non-linear dynamic uncertainties. Other examples of solving highly challenging control based problems using RL are [38] and [39], where RL based control methods like Internal Reinforce Q-learning (IrQ-L) is proposed to handle tracking and formation control of nonlinear multi-agent systems. As an augmentation to standard RL based control frameworks, [40] propose model-based predictive control (MPC) as a safety filter, that can turn a highly non-linear and safety critical dynamical systems into inherently safe systems, to which any RL-based algorithm can be applied without any safety certification. Additionally, [41] illustrate the use of deep neural networks in a learning-based predictive control of cooling system of a large business centre. In their paper, Terzi et al. present an application of LSTM Neural network, as a dynamical model of a real complex system (cooling system), using data sampled from its routine operations. Reference [42] propose a Multi Objective Optimization based predictive control framework to enable smooth and accurate tracking in repetitive tasks.

All studies mentioned above propose state-of-the-art learning-based methods for control of complex systems. However, in addition to the well-known data-driven methods like Dynamic linearization, Reinforcement Learning, and Artificial Neural Networks, as discussed above, some atypical data-driven control methods are proposed in [43], [44], [45], and [46] which presented novel methods like hierarchical predictive learning, decoupled data based control (D2C), and event-triggered communication schemes.

A variety of data-driven methods are found in the review data, to predict system dynamical evolution. For instance, [47], [48], [49] use Neural networks in conjunction with Newmark-β and inhibitor methods to accurately predict evolution of system behavior. Additionally, [50] provide a unique example of network-based modeling of complex material flow based logistics and manufacturing systems, where Funke and Becker use Stochastic block models to predict evolution of system behavior. Moreover, [51] illustrate a unique application of a data informed agent-based modeling (ABM) method to simulate autonomous shared taxi/urban transport and perform sensitivity analysis on the number of cars and the number of charging points.

The above mentioned methods, like dynamic linearization, reinforcement learning, and LSTM neural networks, mostly focus on temporal evolution of system dynamics. Other studies suggest more robust data-driven analysis and identification of features of the dynamical system. For example, [52] apply a Koopman operator for data-driven control. The use of Koopman operators in their paper exploits their ability to produce higher or lower dimensional representation of the original state space as an embedded dynamic model. Hence, the latent variable describes the non-linearities of the dynamical system. A similar method is seen in the work of Korda and Mezic [53], who propose a new algorithm for construction of Koopman eigenfunctions from data, thereby stating that the embedding mapping must consist of the (generalized) eigenfunctions of the Koopman operator (or linear combinations thereof). Reference [54] also propose a novel method of extracting dynamics of nonlinear systems using their time-series data. The proposed method integrates the Koopman operator and linear systems theory to construct a linear model that approximately represents the dynamics of a nonlinear system on a linear space of reduced dimension, based on the available time series data. In particular, the study uses the temporal trajectories generated by this low dimensional linear system as features in machine learning to classify time-series data, in terms of the 'distinction' of system dynamics.

Other studies also suggest various methods for robust system identification. For example, [55], [56], and [57] use statistical methods to produce mathematical models of dynamical systems. Reference [55] propose a generalized power-law formulation to model stiffness and damping in an oscillator, by fitting parameters of the differential equation on experimental data. Reference [56] illustrate a systematic approach to identify dynamics of Covid-19 spread through different regions using data to identify generalized logistic and Susceptible-Exposed-Infectious-Quarantined-Recovered (SEIQR) with distributed time delay models. A similar approach is taken by Volkening et al. [57], to forecast elections using compartmental models of infection.

More recently popular methods, like Sparse Identification of nonlinear dynamics (SINDy) and Dynamic Mode Decomposition (DMD) are used by [58] and [59]. Reference [58] incorporates the concept of SINDy based methods, and knowledge in the field of classical mechanics to identify interpretable and sparse expressions of total energy and the Lagrangian of the system that shelters the hidden dynamics. Reference [59] prove the minimization problem associated with SINDy and DMD methods as specific cases of a more generalized objective function, which yields more robust recovery of inherent dynamics in presence of sparse and noisy measurements.

Additionally, we find few studies employing model order reduction and machine learning in conjunction with system identification methods. For example, [60] proposes to model input-output relationships in a real Tokamak machine, by using a combination of autoencoders, capable of compressing input features, and the Hammerstein-weiner method, for identifying the system dynamics. Similarly, [61] uses a Data knowledge based fuzzy neural network to identify of nonlinear system dynamics. The method proposed by Wu et al. [61] cannot only take advantage of the current data-driven model but also the existing knowledge from a reference model, hence overcoming the problem of incomplete datasets for model training. Also, [62] present a unique methodology to test a dynamical system for its non-linearity based on data. Thus, method of [62] can be used to determine whether or not the data motivates nonlinear modeling. Furthermore, it also can determine whether or not a non linear dynamical model has captured the predictability for the data. Reference [63] propose a novel non-intrusive model reduction method method to learn low-dimensional models of dynamical systems with analytically given non-polynomial nonlinear terms. The rest of linear and polynomially nonlinear dynamics are learnt as a least squares problem. Reference [64] propose an artificial neural network based partial differential equation (PDE) integrator in conjunction with proper orthogonal decomposition (POD) based model order reduction. Reference [65] propose a similar approach, to construct a time-stepping predictor on the basis of a non-black-box dimensionality reduction approach using kernel based Principal component analysis (kernel-based-PCA).

Other studies like [66], [67], and [68] form examples of novel data-driven modeling methods applied to complex problems like weather predictions, blast furnace industrial process modeling. Reference [66] use data mining techniques to establish prior knowledge from historical data, to integrate with support vector machine modeling of the blast furnace ironmaking industrial process. Reference [68] apply evolutionary algorithm in the same blast furnace problem to construct explainable features based on the domain knowledge, to handle the characteristics of data such as nonlinear, dynamic and time lag. Reference [67] propose a novel hybrid data assimilation (DA) with neural network (MLP) based framework to make accurate numerical weather predictions. Their paper formalizes the optimization of single

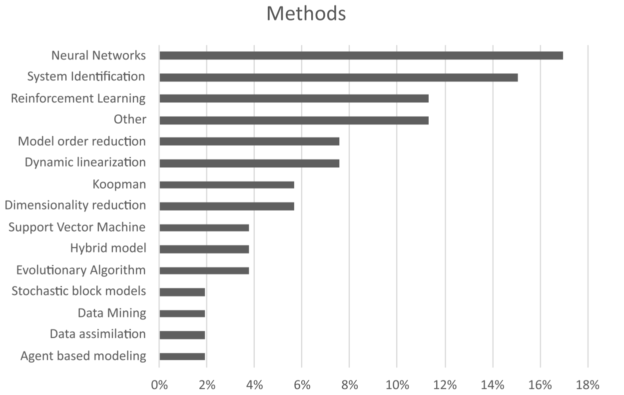


FIGURE 3. This figure illustrates the overall popularity of data-driven methods, applied across all types of problems. Specifically, the chart represents the use of a method as percentage of total use of all methods.

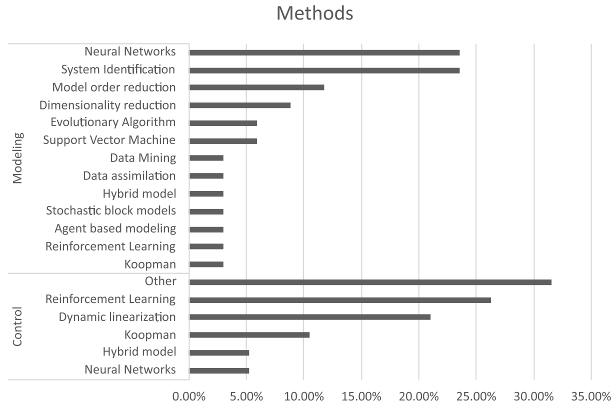


FIGURE 4. This figure illustrates the popularity of data-driven methods, applied for individual types of problems (i.e. modeling and control). Specifically, the chart represents the use of a method as percentage of total use of all methods, for a type of problem.

DA methods (EnKF and 3DVar) through MLP and construct new datasets, adding future short term predictions to correct the current assimilation results, which significantly improves the quality of the results.

Studies like [69], [70], [71], and [72] use data-driven modeling and control methods exhibiting both robust identification of system state and dynamical evolution. For example, [69] propose a hybrid method for complex tendon driven robots, by merging the latest mechanics-based rigid multi-body dynamic models with neural network learning-based models for the difficult-to-capture friction and tendon modeling. A similar approach was proposed by [72], who proposes a combination of physics and data-driven partial process models for hybrid modeling of Electrohydrodynamic jet printing. Reference [70] proposes a novel spatiotemporal neural network taking into account a complex nonlinear distributed parameter system's relationship with space as well as time. The proposed method uses full information from all sensors without the use of model reduction techniques. Reference [71] propose a deep learning network called Semi-Supervised Dynamic Feature Extracting (SSDFE) network to extract nonlinear dynamic features of semi-supervised process data for output quality prediction. In this network, the encoder–decoder with long short-term memory (LSTM) cells consists of the dynamic feature extractor. Simultaneously, an efficient integer differential evolution (IDE) algorithm is utilized to search the optimal Variable Time delay (VTD) values in the training process of the SSDFE network, where VTDs are also regarded as model parameters. In this proposed IDE-SSDFE modeling algorithm, the nonlinear dynamic features and VTD values are cooperatively obtained, which significantly improves the prediction performance of the quality predictor.

C. SYNTHESIS

As described in section 2.3, we classify our extracted data based on the following criteria:

1) Problem type: Modeling or Control

TABLE 3. Data-driven methods.

Data-driven Methods
Agent based modeling (ABM)
Data assimilation (DA)
Data Mining
Dimensionality reduction
Dynamic linearization (DL)
Evolutionary Algorithm
Hierarchical Learning
Hybrid model (based on Differential Equations)
Koopman
Model order reduction (MOR)
Multi-Objective Optimization (MOO)
Neural Networks
Reinforcement Learning (RL)
Stochastic block models (SBM)
Support Vector Machine (SVM)
System Identification (including SINDy, DMD)

a) Modeling: Studies dealing with modeling of complex systems with an objective of system behavior prediction or classification.
b) Control: Studies dealing with system control based objectives or problems, i.e., studies in which the tackled problem lies in the scope of system control.

2) Method contribution: C1 or C2
The concept behind the proposed taxonomy for method contribution is inspired by several studies and books like [9], [12], [27], [28], [29], and [30]. Such literature details fundamental concepts and knowledge regarding data-driven methods, thereby, also providing a fundamental distinction among such methods. In our study, we observe that a method, either learns the temporal characteristics of the system evolution, or provides a more relevant description of the state of a system. We consequently, define a novel taxonomy to classify methods based on their main contributions, as follows:

a) C1: Studies in which applied methods contribute mostly to model temporal evolution of system behaviour.

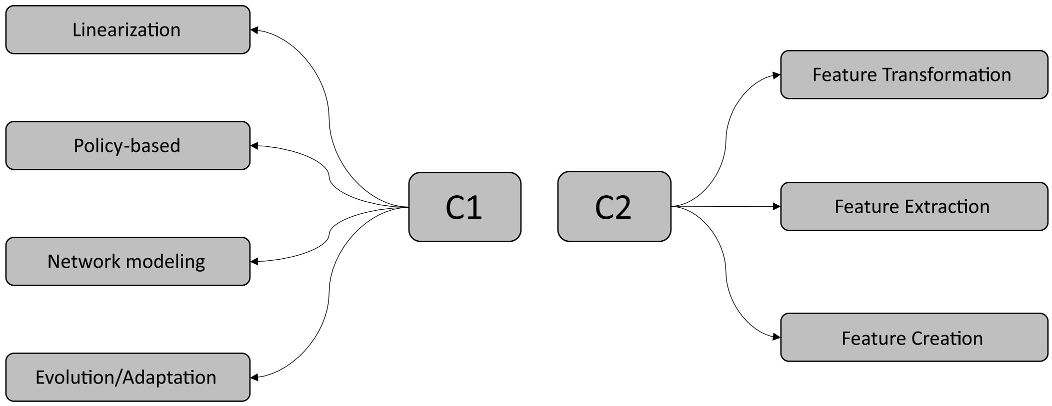


FIGURE 5. **This figure illustrates a conceptual map of a taxonomy to classify method contribution based on method characteristics.**

b) C2: Studies in which applied methods attempt to structure and/or identify the inherent features associated to dynamics of the complex system.

We also map the proposed taxonomy to the fundamental concepts, usually used in literature, involved in a data-driven method as depicted in Figure 5.

V. DISCUSSIONS

In this section we discuss our findings, as the result of the systematic mapping and review, while also answering the research questions.

***RQ1**: What are major applied data-driven methods described in the literature?*

Based on extracted data as described in the previous section, we are able to identify several state-of-the-art data-driven methods, applied in literature. We have enumerated those methods in Table 3. These methods are usually used in combinations, as per system complexity and system prediction/control scope. In Figure 6 we have also graphed the co-occurrence of the applied methods in 41 studies. In order to gain some insight on the popularity of the methods, we recorded the number of times a method is used in 41 studies. Figure 3 provides an overview of the methods' popularity. In particular, Figure 4 illustrates the popularity of methods across problem types.

***RQ2**: What are major data-driven methods' main contributions?*

To outline main contributions of several applied methods found in 41 studies, we created a taxonomy, defined in the previous section. Figure 7 illustrates the methods as members of the classification for evident method contribution. In order to enumerate several applied methods with regards to their use for types of problems and their types of contribution, we cross-mapped the two classification schemes, defined in the previous section, as given by Table 4.

A. TRENDS IN USE OF CONTEMPORARY METHODS

As evident from the method popularity graphs (i.e., Figure 3 and Figure 4), a wide variety of data-driven methods are employed in literature, with a prevalence of studies focusing on the use of Neural Networks, System identification, and Reinforcement learning. This is true for both modeling- and control-based problems. Moreover, machine learning related methods like Neural Networks, Reinforcement learning, Support Vector Machine, among others, are often applied in combination, as illustrated by the co-occurrence graph in Figure 6, based on the problem scope. For instance, in problems related to modeling of system behavior, the machine learning methods are commonly used in combination with Dimensionality reduction, Model order reduction, and System identification based methods. Similarly, for system control-based problems, Reinforcement learning and Dynamic linearization are often combined with other methods.

In addition to the above popular methods, there are few upcoming methods, which are gaining traction recently. For example, many studies are using Koopman operators for modeling as well as control based problems. System Identification methods based on SINDy and DMD methods are also good examples of such upcoming methods. The methods have been devised very recently, for instance SINDy [12], uses sparse data to identify the governing dynamics of the dynamical system while producing parsimonious representative models. In a similar way, DMD methods are also trending as they provide a better way for dimensionality reduction. In contrast to traditional methods, like Principal component analysis (PCA), DMD based methods provide a predetermined temporal behavior (oscillatory, damped, growth) associated to the decomposed modes, which can make those modes more physically meaningful [13].

B. CONTEXTUAL USE OF METHODS

The classification of studies based on type of method contributions provides us with applicable insight into contextual usage of methods. The contexts for the use of data-driven methods are found to be, in overall, to understand the inherent features associated with system dynamics, and/or for modeling temporal evolution of system behavior. Figure 7 provides us with an applicable knowledge-base as regards to

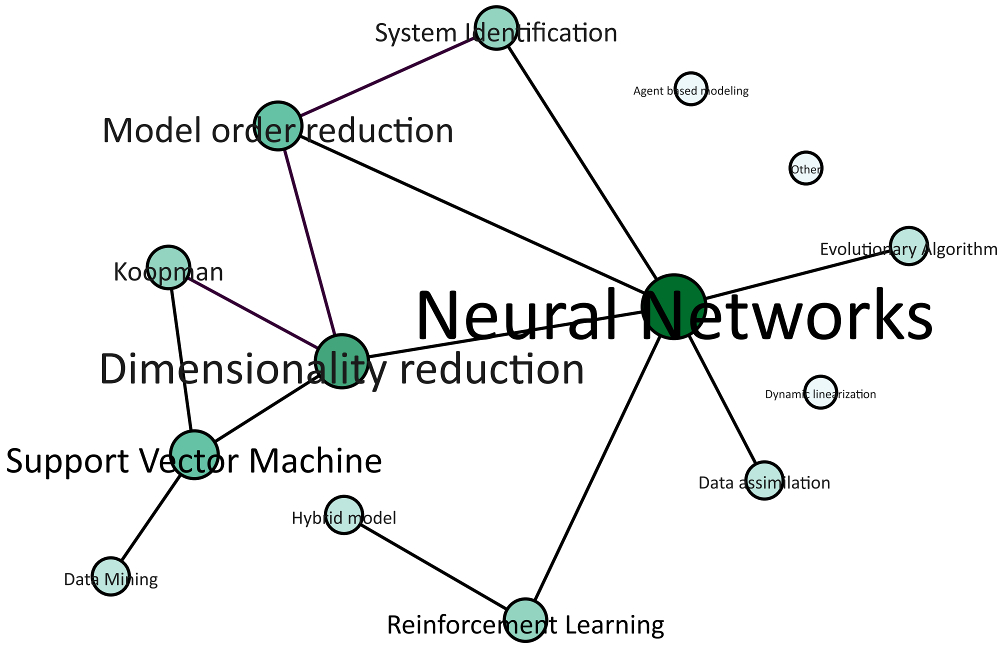


FIGURE 6. This figure shows Co-occurrence graph of applied methods in 41 studies. The color and size of the nodes represent the degree of co-occurrence, i.e. more number of times a method is used in conjunction with other methods, bigger is the size and darker is the color of the representative node. The edges represent the use of method-pairs.

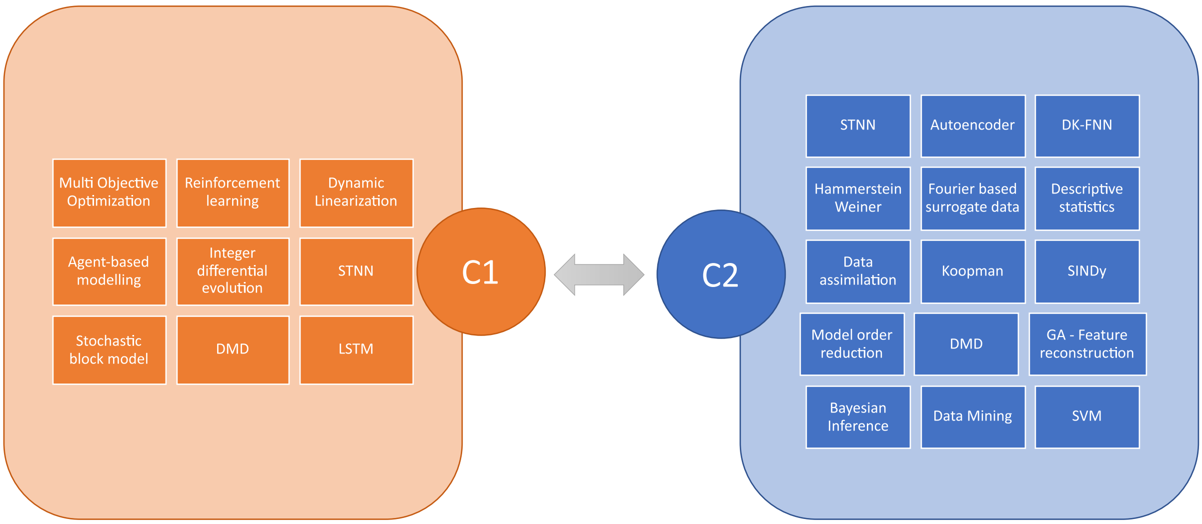


FIGURE 7. This figure enumerates several data-driven methods as nuts and bolts of the two major contribution types, based on the classification scheme defined in Section 3.3.

the utility of various methods found in the 41 studies we have investigated.

Also, the use of methods in conjunction with each other, also depends on the type of problem (modeling or control). The cross-mapping as given by Table 4 contributes to the knowledge-base, i.e. in regards to the contextual use of methods for different problem types.

C. IMPLICATIONS FOR PRACTICE AND RESEARCH

This paper provides an evidence-base and knowledge-base of contemporary data-driven methods applied to cross-domain complex systems. This provides important implications in research as well as practice. Based on our findings related to the most popular data-driven methods and their contextual use on different problem types, we provide applicable insights to IT managers and other practitioners on how to experiment with such methods according to the scope of problems. To give an example, practitioners dealing with projects associated with control-based problems/challenges can refer to the evidence-base and knowledge-base, to plan experiments using methods like Reinforcement learning, Dynamic linearization and Koopman operators, while also considering

TABLE 4. Problem type and method contribution cross-map.

	C1	C2
Modeling	9 papers RL, IDE, LSTM, STNN, DMD, SBM, MOR, ABM Nb : 9,10,11,16,20,25,33,36,38	18 papers MOR, SINDy, Differential equation, Bayesian Inference, DA, STNN, Hammerstein Wiener, Fourier-based surrogate, SVM, Feature Reconstruction,DK-FNN,SSDFE, Koopman, DMD Nb : 2,3,7,8,10,12,14,18,19,20,24,25,26,29,31,34,36,39
Control	16 papers DL, RL, MOO, LSTM, Hierarchical Learning Nb : 4,5,6,13,17,21,22,23,27,28,30,32,35,37,40,41	3 papers Koopman, Differential equation Nb : 1,6,15

the contribution type needed from such methods to tackle their problem/challenges. Such use of evidence-base and knowledge-base, pertaining to planning experiments, can contribute to more informed decisions for practitioners and assist managers to properly direct future investments. At the same time, our findings also guide researchers with an evidence-base of the use of data-driven methods, for a variety of problems, equipping them with an arsenal of applied methods across various application domains. The findings of this review also makes it easier for researchers to realize potential research gaps. For instance, the application of data-driven methods as frameworks based on agent-based modelling (ABM) has not been very evident among the selected 41 studies. Moreover, ABM's limited application is mostly oriented towards modelling of complex systems, and thus may prove to be a potential candidate to explore gaps for future research. A similar potential research gap is indicated by the cross-mapping summarized by Table 4. We observe that there are limited studies with methods used in the context of feature structuring and identification, with regards to control-based problems. This might suggest potential future research on data-driven methods for control-based problems, with an orientation to understand features associated with complex systems.

D. EFFICIENCY OF MAPPING AND REVIEW

The proposed methodology presented in the paper involves elements of systematic literature review, which results from a comprehensive study. Usually, such studies may prove to be highly time consuming, based on a given set of resources.

In our study, due to limited resources, we aimed at automating several parts of the methodology. As given in Figure 2, the search, and screening parts of the review and mapping process were fully or partially automated, distilling the final pool of publications to a manageable number for further manual process components. For instance, tools given by Scopus were used to apply several exclusion criteria. Moreover, the Citescore database provided by Elsevier was used to remove low impact publication venues.

These types of automated tools and measures helped us to increase the efficiency of the entire study.

E. THREATS TO VALIDITY AND LIMITATIONS OF THE STUDY

Based on [31], we accounted for threats to validity in our study, regarding (1) descriptive validity, (2) generalizability, (3) interpretive validity and (4) repeatability, as follows:

1) *Descriptive validity* relates to the accuracy of the data collection. The data collection was performed according to the methodology described in Section 2. We mitigated this threat by carefully directing the search strategy to find studies relating to the research questions, while including only high quality publications. We achieved this by a careful choice of search venues, a search string based on research questions, taking care of synonyms, abbreviations and alternative spellings, and study selection using a robust set of IC-EC.
   Nevertheless, we must express a potential threat to validity to the study, originating from using Scopus as the only search venue. Even though Scopus does cover a large part of relevant literature as compared to Web of Science and Google Scholar, as studied by [25] in the general fields of computer science and engineering, using just Scopus as search venue does not guarantee a complete coverage of literature.
2) *Generalizability* relates to the extent to which the study can be generalized to other areas outside the context of this study. We aimed our study to be cross-domain in nature, hence contributing to a generalized applicability of the findings. Nevertheless, one of the greatest threat to generalizability of this study is associated to the volatile nature of what is state-of-the-art as regards to data-driven methods. Due to increasing popularity of using such methods for a variety of problems, new methods are getting recognized rapidly and being included in literature. With such fast arrival of new methods, the representative quality of data, sampled

TABLE 5. Systematic review data.

Nb	Title	Contribution Type	Problem Type
1	Active Learning of Dynamics for Data-Driven Control Using Koopman Operators [52]	C2	Control
2	Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms [63]	C2	Modeling
3	Linear Priors Mined and Integrated for Transparency of Blast Furnace Black-Box SVM Model [66]	C2	Modeling
4	Active Disturbance Rejection Control for Nonaffined Globally Lipschitz Nonlinear Discrete-time Systems [32]	C1	Control
5	Data-Driven Adaptive Consensus Learning From Network Topologies [35]	C1	Control
6	A hybrid dynamic model for the AMBIDEX tendon-driven manipulator [69]	Both	Control
7	Discovering interpretable dynamics by sparsity promotion on energy and the lagrangian [58]	C2	Modeling
8	A generalised power-law formulation for the modelling of damping and stiffness nonlinearities [55]	C2	Modeling
9	Combining analytics and simulation methods to assess the impact of shared, autonomous electric vehicles on sustainable urban mobility [51]	C1	Modeling
10	Learning time-stepping by nonlinear dimensionality reduction to predict magnetization dynamics [65]	C2	Modeling
11	Complex networks of material flow in manufacturing and logistics: Modeling, analysis, and prediction using stochastic block models [50]	C1	Modeling
12	Bayesian system ID: optimal management of parameter, model, and measurement uncertainty [59]	C2	Modeling
13	Data-driven economic NMPC using reinforcement learning [36]	C1	Control
14	A Data-Driven Method for Hybrid Data Assimilation with Multilayer Perceptron [67]	C2	Modeling
15	Optimal Construction of Koopman Eigenfunctions for Prediction and Control [53]	C2	Control
16	A novel long short-term memory neural-network-based self-excited force model of limit cycle oscillations of nonlinear flutter for various aerodynamic configurations [47]	C1	Modeling
17	Data-Driven Multiobjective Controller Optimization for a Magnetically Levitated Nanopositioning System [42]	C1	Control
18	Domain knowledge based explainable feature construction method and its application in ironmaking process [68]	C2	Modeling
19	COVID-19: data-driven dynamics, statistical and distributed delay models, and observations [56]	C2	Modeling
20	A Spatiotemporal Neural Network Modeling Method for Nonlinear Distributed Parameter Systems [70]	Both	Modeling
21	Data-driven global robust optimal output regulation of uncertain partially linear systems [37]	C1	Control
22	Consensus of multiagent systems with nonlinear dynamics using an integrated sampled-data-based event-triggered communication scheme [43]	C1	Control
23	Optimal Tracking Control of Nonlinear Multiagent Systems Using Internal Reinforce Q-Learning [38]	C1	Control
24	An artificial neural network framework for reduced order modeling of transient flows [64]	C2	Modeling
25	Hybrid Modeling of Electrohydrodynamic Jet Printing [72]	Both	Modeling
26	Data-driven order reduction in Hammerstein–Wiener models of plasma dynamics [60]	C2	Modeling
27	Learning-based predictive control of the cooling system of a large business centre [41]	C1	Control
28	Data-Driven Hierarchical Predictive Learning in Unknown Environments [44]	C1	Control
29	Forecasting elections using compartmental models of infection [57]	C2	Modeling
30	A predictive safety filter for learning-based control of constrained nonlinear dynamical systems [40]	C1	Control
31	Data-Based Testing for Nonlinearity in Dynamical Systems: The Use of Surrogate Data [62]	C2	Modeling
32	Model and Data Based Approaches to the Control of Tensegrity Robots [45]	C1	Control
33	Neural machine-based forecasting of chaotic dynamics [48]	C1	Modeling
34	Data-Knowledge-Based Fuzzy Neural Network for Nonlinear System Identification [61]	C2	Modeling
35	Model-Free Control for Unknown MIMO Nonaffine Nonlinear Discrete-Time Systems With Experimental Validation [33]	C1	Control
36	Cooperative Deep Dynamic Feature Extraction and Variable Time-Delay Estimation for Industrial Quality Prediction [71]	Both	Modeling
37	A Data-Driven ILC Framework for a Class of Nonlinear Discrete-Time Systems [34]	C1	Control
38	Data-Based Reinforcement Learning for Nonzero-Sum Games with Unknown Drift Dynamics [49]	C1	Modeling
39	Dynamics reconstruction and classification via Koopman features [54]	C2	Modeling
40	A Switching Control Scheme With Increment Estimate of Unmodelled Dynamics [46]	C1	Control
41	Data-Driven Optimal Formation Control for Quadrotor Team With Unknown Dynamics [39]	C1	Control

for state-of-the-art in our study, may get outdated very soon.

3) *Interpretive validity* relates to the measure of reasonability of conclusions drawn from the extracted data. While the risk of researcher bias cannot be completely mitigated, the data extraction and synthesis was thoroughly reviewed by all authors, to minimize this threat.
4) *Repeatability* relates to the measure of detail while reporting the research process, in order to be able to replicate the results of the study. One of the main reasons we adopted this methodology, to be based on systematic review and mapping, was due to its repeatable nature [20]. This way, we mitigated threat to repeatability.

VI. CONCLUSION

Data-driven methods, have recently gained popularity regarding their application to understand complex systems in terms of modeling, simulation and control of such systems. In our paper, we present findings of a systematic mapping and review of application of data-driven methods to such complex systems.

Our systematic mapping and review process resulted in a selection of 41 studies. To obtain data relevant to the research questions, these studies were subjected to data extraction in context with the application of data-driven methods and their major contributions to these studies. To investigate RQ1, data was then mapped and classified to find popular and upcoming data-driven methods used in recent research studies. To evaluate RQ2, a classification scheme or taxonomy was created to analyze contributions of data-driven methods applied to different types of problems.

As illustrated by Figure 3 and Figure 4, the results of this study demonstrate current popular data-driven methods being applied to various problem types in the form of an evidence-base. Often these methods are applied in conjunction with each other, whereby co-occurrences are illustrated in Figure 6. Depending on problem type, the use of methods in combination contribute to the evidence-base. Additionally, based on the kind of method contributions and the problem type the methods were used for, a knowledge-base could be constructed. Table 4 and Figure 7 compile this knowledge-base.

Since what is state-of-the art regarding data-driven methods is volatile in nature, our detailed description of the review process aims at making it easy for readers to run future review updates. At present, section V-C. highlights existent research gaps in the application of data-driven methods for complex systems.

DECLARATIONS OF INTEREST

None

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Saumitra Dwivedi: Conceptualization, Methodology, Validation, Writing - Original Draft, Formal analysis, Visualization. **Ricardo da Silva Torres:** Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision, Project administration. **Ibrahim A. Hameed:** Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision. **Anniken Susanne T. Karlsen:** Conceptualization, Methodology, Validation, Writing - Review & Editing, Supervision, Project administration.

APPENDIX

See Table 5.

ACKNOWLEDGMENT

The authors would like to express their special thanks to Prof. Letizia Jaccheri (NTNU) and Prof. Gunnar Tufte (NTNU) for contributing to the work.

REFERENCES

[1] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys. Rev. Lett.*, vol. 86, p. 3200, May 2001, doi: 10.1103/PhysRevLett.86.3200.

[2] E. M. Joshua, *Generative Social Science: Studies in Agent-Based Computational Modeling* (Princeton Studies in Complexity). Princeton, NJ, USA: Princeton Univ. Press, 2006. [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=421475&site=ehost-live

[3] M. Namazi-Rad, L. Padgham, P. Perez, K. Nagel, and A. Bazzan, *Agent Based Modelling of Urban Systems: First International Workshop, ABMUS 2016, Held in Conjunction with AAMAS, Singapore, Singapore, May 10, 2016, Revised, Selected, and Invited Papers* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2017. [Online]. Available: https://books.google.no/books?id=TYjgDQAAQBAJ

[4] H. Sayama, *Introduction to the Modeling and Analysis of Complex Systems*. New York, NY, USA: Open SUNY Textbooks, 2015. [Online]. Available: https://knightscholar.geneseo.edu/oer-ost/14

[5] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, 2nd ed. Boulder, CO, USA: Westview Press, 2015.

[6] T. Deisboeck and J. Y. Kresh, *Complex Systems Science in Biomedicine*. Boston, MA, USA: Springer, 2006, doi: 10.1007/978-0-387-33532-2.

[7] J. Y. Kresh, "Integrative systems view of life: Perspectives from general systems thinking," in *Complex Systems Science in Biomedicine*, T. S. Deisboeck and J. Y. Kresh, Eds. Boston, MA, USA: Springer, 2006, pp. 3–29, doi: 10.1007/978-0-387-33532-2_1.

[8] C. R. Shalizi, "Methods and techniques of complex systems science: An overview," in *Complex Systems Science in Biomedicine*, T. S. Deisboeck and J. Y. Kresh, Eds. Boston, MA, USA: Springer, 2006, pp. 33–114, doi: 10.1007/978-0-387-33532-2_2.

[9] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge, U.K.: Cambridge Univ. Press, 2019, doi: 10.1017/9781108380690.

[10] D. Solomatine, L. See, and R. Abrahart, "Data-driven modelling: Concepts, approaches and experiences," in *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications*, R. J. Abrahart, L. M. See, and D. P. Solomatine, Eds. Berlin, Germany: Springer, 2008, pp. 17–30, doi: 10.1007/978-3-540-79881-1_2.

[11] T. Hey, S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA, USA: Microsoft Research, Oct. 2009. [Online]. Available: https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/

[12] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 15, pp. 3932–3937, Apr. 2016, doi: 10.1073/pnas.1517384113.

[13] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition*. Philadelphia, PA, USA: SIAM, 2016, doi: 10.1137/1.9781611974508.

[14] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, "Machine learning for fluid mechanics," *Annu. Rev. Fluid Mech.*, vol. 52, no. 1, pp. 477–508, Jan. 2020, doi: 10.1146/annurev-fluid-010719-060214.

[15] A. M. J.-C. Wadoux, M. Román-Dobarco, and A. B. McBratney, "Perspectives on data-driven soil research," *Eur. J. Soil Sci.*, vol. 72, no. 4, pp. 1675–1689, 2021, doi: 10.1111/ejss.13071.

[16] P. Rajendra and V. Brahmajirao, "Modeling of dynamical systems through deep learning," *Biophys. Rev.*, vol. 12, no. 6, pp. 1311–1320, Dec. 2020, doi: 10.1007/s12551-020-00776-4.

[17] D. N. How, M. A. Hannan, M. H. Lipu, and P. J. Ker, "State of charge estimation for lithium-ion batteries using model-based and data-driven methods: A review," *IEEE Access*, vol. 7, pp. 136116–136136, 2019, doi: 10.1109/ACCESS.2019.2942213.

[18] A. Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, and G. Hartvigsen, "Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes," *Artif. Intell. Med.*, vol. 98, pp. 109–134, Jul. 2019, doi: 10.1016/j.artmed.2019.07.007.

[19] R. S. Peres, X. Jia, J. Lee, K. Sun, A. W. Colombo, and J. Barata, "Industrial artificial intelligence in industry 4.0—Systematic review, challenges and outlook," *IEEE Access*, vol. 8, pp. 220121–220139, 2020, doi: 10.1109/ACCESS.2020.3042874.

[20] B. A. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ. Durham Univ. Joint Rep., Durham, U.K., Tech. Rep. EBSE 2007-001, Jul. 2007. [Online]. Available: https://www.elsevier.com/_data/promis_misc/525444systematicreviewsguide.pdf

[21] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *Proc. 12th Int. Conf. Eval. Assessment Softw. Eng.* Swindon, U.K.: BCS Learning & Development, 2008, pp. 68–77.

[22] A. Fink, *Conducting Research Literature Reviews: From the Internet to Paper*, 4th ed. Thousand Oaks, CA, USA: Sage, 2014.

[23] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016412120600197X

[24] T. E. Idrissi, A. Idri, and Z. Bakkoury, "Systematic map and review of predictive techniques in diabetes self-management," *Int. J. Inf. Manage.*, vol. 46, pp. 263–277, Jun. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268401218304705

[25] A. Martín-Martín, E. Orduna-Malea, M. Thelwall, and E. D. López-Cózar, "Google scholar, web of science, and Scopus: A systematic comparison of citations in 252 subject categories," *J. Informetrics*, vol. 12, no. 4, pp. 1160–1177, Nov. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1751157718303249

[26] P. Brereton, M. Turner, and R. Kaur, "Pair programming as a teaching tool: A student review of empirical studies," in *Proc. 22nd Conf. Softw. Eng. Educ. Training*, 2009, pp. 240–247, doi: 10.1109/CSEET.2009.11.

[27] S. Hochreiter and J. J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 37–45, 1997, doi: 10.1162/neco.1997.9.8.1735.

[28] N. Buduma and N. Locascio, *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. Sebastopol, CA, USA: O'Reilly Media, 2017. [Online]. Available: https://books.google.no/books?id=n0IlDwAAQBAJ

[29] A. Zheng and A. Casari, *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. Sebastopol, CA, USA: O'Reilly Media, 2018. [Online]. Available: https://books.google.no/books?id=sthSDwAAQBAJ

[30] A. Burkov, *The Hundred-Page Machine Learning Book*. Quebec City, QC, Canada: Andriy Burkov, 2019. [Online]. Available: https://books.google.no/books?id=ZF3KwQEACAAJ

[31] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, pp. 1–18, Aug. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584915000646

[32] R. Chi, Y. Hui, B. Huang, and Z. Hou, "Active disturbance rejection control for nonaffined globally Lipschitz nonlinear discrete-time systems," *IEEE Trans. Autom. Control*, vol. 66, no. 12, pp. 5955–5967, Dec. 2021, doi: 10.1109/TAC.2021.3051353.

[33] S. Xiong and Z. Hou, "Model-free adaptive control for unknown MIMO nonaffine nonlinear discrete-time systems with experimental validation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 4, pp. 1727–1739, Apr. 2022, doi: 10.1109/TNNLS.2020.3043711.

[34] X. Yu, Z. Hou, and M. M. Polycarpou, "A data-driven ILC framework for a class of nonlinear discrete-time systems," *IEEE Trans. Cybern.*, vol. 52, no. 7, pp. 6143–6157, Jul. 2022, doi: 10.1109/TCYB.2020.3029596.

[35] R. Chi, Y. Hui, B. Huang, Z. Hou, and X. Bu, "Data-driven adaptive consensus learning from network topologies," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3487–3497, Aug. 2022, doi: 10.1109/TNNLS.2021.3053186.

[36] S. Gros and M. Zanon, "Data-driven economic NMPC using reinforcement learning," *IEEE Trans. Autom. Control*, vol. 65, no. 2, pp. 636–648, Feb. 2020, doi: 10.1109/TAC.2019.2913768.

[37] A. Odekunle, W. Gao, and Y. Wang, "Data-driven global robust optimal output regulation of uncertain partially linear systems," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 5, pp. 1108–1115, Sep. 2019, doi: 10.1109/JAS.2019.19117678.

[38] Z. Peng, R. Luo, J. Hu, K. Shi, S. K. Nguang, and B. K. Ghosh, "Optimal tracking control of nonlinear multiagent systems using internal reinforce Q-learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 4043–4055, Aug. 2022, doi: 10.1109/TNNLS.2021.3055761.

[39] W. Zhao, H. Liu, F. L. Lewis, and X. Wang, "Data-driven optimal formation control for quadrotor team with unknown dynamics," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7889–7898, Aug. 2022, doi: 10.1109/TCYB.2021.3049486.

[40] K. P. Wabersich and M. N. Zeilinger, "A predictive safety filter for learning-based control of constrained nonlinear dynamical systems," *Automatica*, vol. 129, Jul. 2021, Art. no. 109597, doi: 10.1016/j.automatica.2021.109597.

[41] E. Terzi, T. Bonetti, D. Saccani, M. Farina, L. Fagiano, and R. Scattolini, "Learning-based predictive control of the cooling system of a large business centre," *Control Eng. Pract.*, vol. 97, Apr. 2020, Art. no. 104348, doi: 10.1016/j.conengprac.2020.104348.

[42] X. Li, H. Zhu, J. Ma, T. J. Teo, and C. S. Teo, "Data-driven multiobjective controller optimization for a magnetically levitated nanopositioning system," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 4, pp. 1961–1970, Feb. 2020, doi: 10.1109/TMECH.2020.2999401.

[43] C. Peng, J. Zhang, and Q.-L. Han, "Consensus of multiagent systems with nonlinear dynamics using an integrated sampled-data-based event-triggered communication scheme," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 3, pp. 589–599, Mar. 2019, doi: 10.1109/TSMC.2018.2814572.

[44] C. Vallon and F. Borrelli, "Data-driven hierarchical predictive learning in unknown environments," in *Proc. IEEE 16th Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2020, pp. 104–109, doi: 10.1109/CASE48305.2020.9216872.

[45] R. Wang, R. Goyal, S. Chakravorty, and R. E. Skelton, "Model and data based approaches to the control of tensegrity robots," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 3846–3853, Jul. 2020, doi: 10.1109/LRA.2020.2979891.

[46] Y. Zhang, H. Niu, X. Chen, J. Tao, and X. Li, "A switching control scheme with increment estimate of unmodeled dynamics," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6054–6061, Sep. 2021, doi: 10.1109/TII.2020.3038672.

[47] W. Li, S. Laima, X. Jin, W. Yuan, and H. Li, "A novel long short-term memory neural-network-based self-excited force model of limit cycle oscillations of nonlinear flutter for various aerodynamic configurations," *Nonlinear Dyn.*, vol. 100, no. 3, pp. 2071–2087, 2020, doi: 10.1007/s11071-020-05631-5.

[48] R. Wang, E. Kalnay, and B. Balachandran, "Neural machine-based forecasting of chaotic dynamics," *Nonlinear Dyn.*, vol. 98, no. 4, pp. 2903–2917, Dec. 2019, doi: 10.1007/s11071-019-05127-x.

[49] Q. Zhang and D. Zhao, "Data-based reinforcement learning for nonzero-sum games with unknown drift dynamics," *IEEE Trans. Cybern.*, vol. 49, no. 8, pp. 2874–2885, Aug. 2019, doi: 10.1109/TCYB.2018.2830820.

[50] T. Funke and T. Becker, "Complex networks of material flow in manufacturing and logistics: Modeling, analysis, and prediction using stochastic block models," *J. Manuf. Syst.*, vol. 56, pp. 296–311, Jul. 2020, doi: 10.1016/j.jmsy.2020.06.015.

[51] O. Dlugosch, T. Brandt, and D. Neumann, "Combining analytics and simulation methods to assess the impact of shared, autonomous electric vehicles on sustainable urban mobility," *Inf. Manage.*, vol. 59, no. 5, Jul. 2022, Art. no. 103285, doi: 10.1016/j.im.2020.103285.

[52] I. Abraham and T. D. Murphey, "Active learning of dynamics for data-driven control using Koopman operators," *IEEE Trans. Robot.*, vol. 35, no. 5, pp. 1071–1083, Oct. 2019, doi: 10.1109/TRO.2019.2923880.

[53] M. Korda and I. Mezic, "Optimal construction of Koopman eigenfunctions for prediction and control," *IEEE Trans. Autom. Control*, vol. 65, no. 12, pp. 5114–5129, Dec. 2020, doi: 10.1109/TAC.2020.2978039.

[54] W. Zhang, Y.-C. Yu, and J.-S. Li, "Dynamics reconstruction and classification via Koopman features," *Data Mining Knowl. Discovery*, vol. 33, no. 6, pp. 1710–1735, Nov. 2019, doi: 10.1007/s10618-019-00639-x.

[55] M. Civera, S. Grivet-Talocia, C. Surace, and L. Z. Fragonara, "A generalised power-law formulation for the modelling of damping and stiffness nonlinearities," *Mech. Syst. Signal Process.*, vol. 153, May 2021, Art. no. 107531, doi: 10.1016/j.ymssp.2020.107531.

[56] X. Liu, X. Zheng, and B. Balachandran, "COVID-19: Data-driven dynamics, statistical and distributed delay models, and observations," *Nonlinear Dyn.*, vol. 101, no. 3, pp. 1527–1543, Aug. 2020, doi: 10.1007/s11071-020-05863-5.

[57] A. Volkening, D. F. Linder, M. A. Porter, and G. A. Rempala, "Forecasting elections using compartmental models of infection," *SIAM Rev.*, vol. 62, no. 4, pp. 837–865, Jan. 2020, doi: 10.1137/19M1306658.

[58] H. K. Chu and M. Hayashibe, "Discovering interpretable dynamics by sparsity promotion on energy and the Lagrangian," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2154–2160, Apr. 2020, doi: 10.1109/LRA.2020.2970626.

[59] N. Galioto and A. A. Gorodetsky, "Bayesian system ID: Optimal management of parameter, model, and measurement uncertainty," *Nonlinear Dyn.*, vol. 102, no. 1, pp. 241–267, Sep. 2020, doi: 10.1007/s11071-020-05925-8.

[60] A. G. Spinosa, A. Buscarino, L. Fortuna, M. Iafrati, and G. Mazzitelli, "Data-driven order reduction in Hammerstein–Wiener models of plasma dynamics," *Eng. Appl. Artif. Intell.*, vol. 100, Apr. 2021, Art. no. 104180, doi: 10.1016/j.engappai.2021.104180.

[61] X. Wu, H. Han, Z. Liu, and J. Qiao, "Data-knowledge-based fuzzy neural network for nonlinear system identification," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 2209–2221, Sep. 2020, doi: 10.1109/TFUZZ.2019.2931870.

[62] M. Waller, "Data-based testing for nonlinearity in dynamical systems: The use of surrogate data," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 2, pp. 679–688, Mar. 2019, doi: 10.1109/TCST.2017.2771431.

[63] P. Benner, P. Goyal, B. Kramer, B. Peherstorfer, and K. Willcox, "Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms," *Comput. Methods Appl. Mech. Eng.*, vol. 372, Dec. 2020, Art. no. 113433, doi: 10.1016/j.cma.2020.113433.

[64] O. San, R. Maulik, and M. Ahmed, "An artificial neural network framework for reduced order modeling of transient flows," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 77, pp. 271–287, Oct. 2019, doi: 10.1016/j.cnsns.2019.04.025.

[65] L. Exl, N. J. Mauser, T. Schrefl, and D. Suess, "Learning time-stepping by nonlinear dimensionality reduction to predict magnetization dynamics," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 84, May 2020, Art. no. 105205, doi: 10.1016/j.cnsns.2020.105205.

[66] S. Chen and C. Gao, "Linear priors mined and integrated for transparency of blast furnace black-box SVM model," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3862–3870, Jun. 2020, doi: 10.1109/TII.2019.2940475.

[67] L. Huang, H. Leng, X. Li, K. Ren, J. Song, and D. Wang, "A data-driven method for hybrid data assimilation with multilayer perceptron," *Big Data Res.*, vol. 23, Feb. 2021, Art. no. 100179, doi: 10.1016/j.bdr.2020.100179.

[68] Y. Li and C. Yang, "Domain knowledge based explainable feature construction method and its application in ironmaking process," *Eng. Appl. Artif. Intell.*, vol. 100, Apr. 2021, Art. no. 104197, doi: 10.1016/j.engappai.2021.104197.

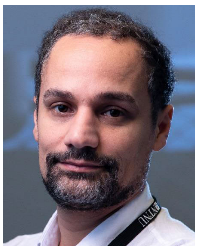
[69] K. Choi, J. Kwon, T. Lee, C. Park, J. Pyo, C. Lee, S. Lee, I. Kim, S. Seok, Y.-J. Kim, and F. C. Park, "A hybrid dynamic model for the AMBIDEX tendon-driven manipulator," *Mechatronics*, vol. 69, Aug. 2020, Art. no. 102398, doi: 10.1016/j.mechatronics.2020.102398.

[70] X. Lu and X. Cui, "A spatiotemporal neural network modeling method for nonlinear distributed parameter systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1916–1926, Mar. 2021, doi: 10.1109/TII.2020.2996996.

[71] L. Yao and Z. Ge, "Cooperative deep dynamic feature extraction and variable time-delay estimation for industrial quality prediction," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 3782–3792, Jun. 2021, doi: 10.1109/TII.2020.3021047.

[72] I. A. Spiegel, P. Sammons, and K. Barton, "Hybrid modeling of electrohydrodynamic jet printing," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 6, pp. 2322–2335, Nov. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8846571/

SAUMITRA DWIVEDI received the B.Tech. degree in petroleum engineering from the Indian Institute of Technology (IIT-ISM) Dhanbad, India, and the M.Sc. degree in simulation and visualization from the Norwegian University of Science and Technology (NTNU), Ålesund, Norway, where he is currently pursuing the Ph.D. degree with the Department of ICT and Natural Sciences, Faculty of Information Technology and Electrical Engineering.



RICARDO DA SILVA TORRES (Member, IEEE) received the B.Sc. degree in computer engineering and the Ph.D. degree in computer science from the University of Campinas, Brazil, in 2000 and 2004, respectively. He was a Professor with the University of Campinas, from 2005 to 2019. He has been a Professor of visual computing with the Norwegian University of Science and Technology (NTNU), since 2019. He is currently a Professor of data science and artificial intelligence with Wageningen University and Research. He has been developing multidisciplinary e-science research projects involving multimedia analysis, multimedia retrieval, machine learning, databases, information visualization, and digital libraries.

IBRAHIM A. HAMEED (Senior Member, IEEE) received the Ph.D. degree in AI from Korea University, South Korea, in 2010, and the Ph.D. degree in field robotics from Aarhus University, Denmark, in 2012. He is currently a Professor and the Deputy Head of the Research and Innovation, Department of ICT and Natural Sciences, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology (NTNU), Ålesund, Norway. His research interests include control systems, robotics, AI, and machine learning.

ANNIKEN SUSANNE T. KARLSEN received the M.Sc. degree in information technology from the University of Aalborg, Denmark, and the Ph.D. degree in information science from the University of Bergen, Norway. She is currently an Associate Professor with the Department of ICT and Natural Sciences, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology (NTNU). She teaches and researches within technology management and digital transformation. She is also the Head of the Sustainable Digital Transformation Research Group. Her research interests include maritime, offshore, health, and banking.