Computer-Aided Civil and Infrastructure Engineering **00** (2017) 1–16

Efficient Infrastructure Restoration Strategies Using the Recovery Operator

Andrés D. González

Department of Civil & Environmental Engineering, Rice University, Houston, TX, USA and Department of Civil & Environmental Engineering, Department of Industrial Engineering, Universidad de los Andes, Bogotá, Colombia

Airlie Chapman

Department of Mechanical Engineering, University of Melbourne, Melbourne, VIC, Australia

Leonardo Dueñas-Osorio*

Department of Civil & Environmental Engineering, Rice University, Houston, TX, USA

Mehran Mesbahi

William E. Boeing Department of Aeronautics and Astronautics, University of Washington, Seattle, WA, USA

&

Raissa M. D'Souza

Complexity Sciences Center, Department of Mechanical and Aerospace Engineering, and Department of Computer Science, University of California, Davis, CA, USA and Santa Fe Institute, Santa Fe, NM 87501, USA

Abstract: *Infrastructure systems are critical for society's resilience, government operation, and overall defense. Thereby, it is imperative to develop informative and computationally efficient analysis methods for infrastructure systems, which reveal system vulnerabilities and recoverability. To capture practical constraints in systems analyses, various layers of complexity play a role, including limited element capacities, restoration resources, and the presence of interdependence among systems. High-fidelity modeling such as mixed integer programming and physics-based modeling can often be computationally expensive, making time-sensitive analyses challenging. Furthermore, the complexity of recovery solutions can reduce analysis transparency. An alternative, presented in this work, is a reduced-order representation, dubbed a recovery operator, of a high-fidelity time-dependent recovery model of a system of interdependent networks. The form of the operator is assumed to be a time-invariant linear dynamic model apt for infrastructure restoration. The recovery operator is generated by applying system identification techniques to numerous disaster and recovery scenarios. The proposed compact representation provides simple yet powerful information regarding systemic recovery dynamics, and enables generating fast suboptimal recovery policies in time-critical applications.*

*To whom correspondence should be addressed. E-mail: *leonardo.duenas-osorio@rice.edu*.

DOI: 10.1111/mice.12314

1 INTRODUCTION

Proper operation of critical infrastructure networks is vital to our society's well-being, as well as governance and safety. On the other hand, abnormal operation of such systems generates health and security issues, as

well as considerable economic loss. The current annual economic losses associated with adverse natural events such as earthquakes, landslides, cyclones, and floods, average between US$250 billion and US$300 billion, and by 2030 they are expected to increase up to US$415 billion (UNISDR, 2015). Thus, individuals, communities, governments, and other stakeholders must work toward a "prospective," "corrective," and "compensatory" management of risk, which is particularly amplified by the existing interdependencies between different utilities and infrastructure systems (UNISDR, 2015).

Considering the importance of interdependencies in infrastructure performance and resilience, despite being a relatively novel subject of study, the field has gained significant interest. Most of the early works on this subject focused on acknowledging and categorizing interconnectedness and interdependencies in systems of infrastructure networks (Rinaldi et al., 2001). Subsequently, some studies focused on the vulnerability of interdependent networks (Buldyrev et al., 2010; Vespignani, 2010; Gao et al., 2011; Hernandez-Fajardo and Dueñas-Osorio, 2013) and later, on reducing that vulnerability (Brummitt et al., 2012). Also, some studies started quantifying the recoverability and resilience of these interdependent networked systems (Vugrin et al., 2010; Pant et al., 2013; Ouyang, 2014; Shafieezadeh et al., 2014; Baroud et al., 2015; Ouyang and Wang, 2015), as well as developing recovery strategies for such systems (Lee et al., 2007; Cavdaroglu et al., 2011; González et al., 2016b). Nevertheless, these papers emphasize the complexity associated with generating recovery strategies and the difficulty of doing so via computationally efficient algorithms, thus the importance of developing new tools.

In this article, we propose a new methodology to efficiently develop recovery strategies for a damaged system of infrastructure networks, using system identification techniques. The proposed approach uses known recovery strategies associated with a set of different disaster scenarios—which may come from computer-generated or historical data—and applies system identification tools to extract the main recovery dynamics as a linear operator. This operator, defined as the recovery operator, in addition to containing relevant information about the recovery dynamics of the system, can be used to efficiently generate recovery strategies associated with any given damage scenario. Before detailing the new approach to deal with infrastructure recovery, we discuss in the following subsections two related fields of study that underpin this work. The first is associated with designing (or restoring) networked systems, particularly in relation to approaches that rely on network flows and mixed-integer programming (MIP). The second deals with system identification and linear time-invariant representations of nonlinear dynamics.

1.1 Literature review on the Network Design Problem (NDP) and optimization techniques applicable to interdependent infrastructure systems

The NDP consists of identifying the subgraph and flow configuration that satisfy a set of demands at minimum cost, for a given graph or network (Ahuja et al., 1993; Gendron et al., 1999). Johnson et al. (1978) presented a general formulation of the NDP, and proved its NP-completeness. Later, Dantzig et al. (1979) showed how the structure of the NDP can be used to develop a decomposition technique to enhance its speed and size capabilities. For instances when the focus is only on finding optimal flows, Assad (1978) highlighted the advantages of decomposition techniques for problems considering multiple commodities. Similarly, Kennington (1978) presented a survey of linear cost multicommodity network flows problems, emphasizing how realistic problems such as routing, assignment, urban transportation planning, and communication, among others, can be represented and studied using linear constraints. Even though these works acknowledge that some problems require nonlinear models for their parametrization, they emphasize that linear structures are more generalizable and easily adapted to many realistic network systems. Given that the NDP allows for a general and flexible description of networks under realistic constraints, the NDP and its derivatives are widely used in areas such as urban transportation, production planning, and scheduling, among others (Frangioni and Gendron, 2009; Pedersen et al., 2008; Crainic et al., 2006; Fortz and Poss, 2009; Poss, 2012; Lin, 2011). Considering the versatility shown by the NDP-like modeling, Lee et al. (2007) developed a network-flows-based approach to model multiple interdependent systems, and recover them after a damaging event. Nevertheless, Lee et al. (2007) emphasized the high complexity of the developed optimization model, and proposed a heuristic approach to allow solving it when computational power is limited. Later, Cavdaroglu et al. (2011) presented another NDP-like Mixed-Integer formulation to determine both the recovery strategy and the associated job scheduling for a damaged system of interdependent networks, but considering its complexity, proposed a partitioning heuristic approach. Exploiting the analytical structure of the problem, González et al. (2016a, b) proposed optimization models that extended the NDP to optimize the recovery process of a damaged system of networks with physical, logical, and geographical interdependencies, revealing the high computational cost associated with determining a guaranteed optimal

time-dependent recovery strategy. Even though NDP-based models have proved useful to model and optimize the recovery process of a system of infrastructure networks, they may not be ideal for time-critical applications due to their high computational complexity. Nevertheless, these models can be used to study prototypical disaster scenarios and to identify critical properties of the system dynamics. This information is key to study efficient approaches that allow identifying the main dynamics associated with the recovery process described by NDP-like models.

1.2 Literature review on the system identification and finite linear representations of nonlinear dynamics

System identification techniques have widespread use in multiple fields, such as physics, biology, and engineering. Such techniques have been embraced in civil and structural engineering for the identification of structural parameters, structural health monitoring, and structural response (Kijewski and Kareem, 2003; Jiang and Adeli, 2005; Beskhyroun et al., 2011; Sirca and Adeli, 2012), but to our knowledge they have not been used to identify the main recovery dynamics of systems of interdependent networks, such as infrastructure systems. The general area of system identification endeavors to form dynamic models of a process from system output data alone—in this article, the dynamics stem from interdependent system restoration. The system identification process varies depending on the assumptions placed on the dynamic model. On one hand, there is grey box modeling, which assumes a basic physical modeling based on preceding understanding of the system, and uses data to perform fitting on specific parameters of the model. On the other hand, there is black box modeling, which has access to limited knowledge of the studied system and assumes no prior model is available, thus leveraging heavily on the available data. Black box modeling usually relies on input–output information of the studied system, to infer the associated dynamics involved—thus, we will use this modeling approach, considering its general applicability. McGhee (1963) categorized these approaches as parametric space modeling, including differential equations of predetermined form and order, along with state models, and nonparametric or function space modeling, for example, impulse response, covariance function, spectral densities, and Volterra series. Comprehensive survey papers on the topic have been written by Åström and Eykhoff (1971) and Kerschen et al. (2006). For further reading, see the books by Ljung (1999), Söderström and Stoica (1989), and the International Federation of Automatic Control symposia series on system identification which has been ongoing since 1967.

System identification algorithms are typically applied to output data which is generated by experimental processes, but these can also be exercised on high-fidelity simulations to disseminate high-level features from the system. This novel application has become increasingly popular in the analysis of simulated fluid flow that is governed by nonlinear partial differential equations. Sequential time-series fluid simulation data, in the form of a "snapshot" sequence $\{\phi_0, \ldots, \phi_T\}$, are approximated as linear time-invariant dynamics $\phi_{t+1} = A\phi_t$ with an unknown linear operator A. The process of recovering the eigenvectors and eigenvalues of operator A is referred to as dynamic mode decomposition (DMD) (Schmid, 2010). DMD has shown to extract essential features of the underlying dynamics, reliably approximating models with low complexity (Rowley et al., 2009; Schmid et al., 2011). Rowley et al. (2009) connected the efficacy of DMD to its grounding in Koopman operator theory. The Koopman operator is an infinite dimensional *linear* operator which captures the *full* information of nonlinear systems (Koopman and von Neumann, 1932). DMD is a finite-dimensional approximation of Koopman operator theory over a special set of observables (Mezić, 2005). In addition to its success in fluid dynamics, linear operator theory has also been applied to bursting neuron models (Mauroy et al., 2014), large-scale power systems (Barocio et al., 2015), and multiresolution analysis of video streams (Kutz et al., 2015). Given such applications to complex systems, it is surmised to be applicable to restoration processes of infrastructure systems.

If the fundamental features of the dynamics can be represented via a linear operator, then numerous desirable directions follow. From a computational standpoint, evolving a linear time-invariant system $\phi_{t+1} = A\phi_t$ is preferable over a nonlinear system. Also, using approximative linear operators enables the application of linear systems algorithms and theory to nonlinear systems. Applying linear stability analysis to establish global stability is one example (Mauroy and Mezic, 2013). Sootla et al. (2016) leverage the Koopman operator in the design of temporal pulse control of bistable monotone systems. Brunton et al. (2016) demonstrated the potential use of these operators to design optimal control laws for fully nonlinear systems using techniques from linear optimal control. In this venue, linear operator theory has the potential to open the application of many techniques from the rich field of linear controller design (Anderson and Moore, 1990; Stengel, 1994), thus hinting its applicability to modeling the dynamics associated with subjects such as the efficient recovery of systems of interdependent infrastructure networks.

1.3 Article structure

This article is organized as follows: Section 2 details the use of system identification techniques to approximate optimal restoration dynamics. For this purpose, we propose the construction of a linear time-invariant operator defined as the recovery operator, and we show how to use it to efficiently construct recovery strategies for any given disaster scenario. Section 3 applies the proposed techniques to study the recovery dynamics of a realistic case study: the recovery process of the water, gas, and power networks in Shelby County, TN, United States, due to earthquakes in the New Madrid Seismic Zone (NSMZ). Using this illustrative example we showcase some of the capabilities of the recovery operator, and the recovery strategies generated with it. Then, Section 4 presents relevant conclusions and ideas for future research. Finally, Appendix A provides details of the high-fidelity MIP model, defined as the time-dependent Interdependent Network Design Problem (td-INDP) (González et al., 2016a, 2016b), which was used to create "snapshots" of the recovery process that served as training data to generate the recovery operator.

2 SYSTEM IDENTIFICATION TECHNIQUES APPLIED TO INFRASTRUCTURE RECOVERY

Providing good recovery strategies for a given networked system is a complex task. In particular, using time-dependent optimization models to determine efficient strategies (such as the MIP models proposed by Lee et al. (2007); Cavdaroglu et al. (2011); Nurre et al. (2012), or the one presented in Appendix A), may be computationally expensive, and for scenarios in which it is necessary to obtain results in real time, they may not be suitable. One option to overcome such a problem is to precalculate prototypical disaster configurations and their respective recovery strategies, to identify and assess the associated risks, vulnerabilities, and challenges, and to develop predisaster recovery plans (U.S. Federal Emergency Management Agency (FEMA), 2011). Unfortunately, given the large number of feasible disaster configurations, creating and maintaining plans for every possible disaster also becomes a prohibitive task. Thus, an alternative option introduced in this article is to create a system identification (system ID) framework that would take families of prototypical failure scenarios and their respective optimal recovery strategies as the input/output data to be studied, that later can be used to generate tailor-made recovery strategies. In particular, we propose constructing a linear operator, following a truncated Koopman operator approach, to approximate the recovery dynamics observed in a system of interconnected networks. The proposed system ID framework uses failure/recovery input/output "snap-shot" data to divulge key features of the system and its approximate recovery dynamics, and provides tools to analyze and generate good recovery strategies in an efficient manner.

2.1 Time-invariant recovery operator $\hat{A}$

Assume there is a set of elements E for which we want to study and model their failure and recovery dynamics. Without loss of generality, since the system of infrastructure networks is being modeled using graphs, the set of elements may refer to the nodes in the system (i.e., $E = \mathcal{N}$), the arcs (i.e., $E = \mathcal{A}$), or both (i.e., $E = \mathcal{N} \cup \mathcal{A}$). Now, define $\phi(t)$ as the vector that describes the damage state information for all elements $e \in E$ at time t. Each entry $\phi_e(t)$ of such a vector is either 1 if the element e is damaged at time t, or 0 otherwise.

Let us also assume that given a disaster scenario at a time t (where the damage state of each element in the system is known), there is an optimal recovery process. Then, mirroring field conditions, such a recovery process could be seen as a process where the optimal recovery jobs (and by extension the damage states) for time $t+1$ could be determined based only on the damage states at time t. Then, there exists a mapping f such that:

$$\phi(t+1) = f(\phi(t)) \tag{1}$$

Such a mapping would imply a fully deterministic system, where any given action, under the same conditions, would always result in a fixed outcome. What we propose in this article is approximating $f(\phi(t))$ with a time-invariant operator $\hat{A}$, such that

$$\phi(t+1) \approx \hat{A}\phi(t) \tag{2}$$

The operator $\hat{A}$ is selected to minimize the approximation error over all time periods, namely,

$$\hat{A} = \arg \min_{A \in \mathbb{R}^{|E|\times|E|}} \|\Phi_0 - A\Phi_1\| \tag{3}$$

where (Φ_0, Φ_1) is the input–output history pair

$$\Phi_0 = [\phi(0), \phi(1), \ldots, \phi(|\mathcal{T}|-1)] \tag{4}$$

$$\Phi_1 = [\phi(1), \phi(2), \ldots, \phi(|\mathcal{T}|)] \tag{5}$$

This operator is associated with a given disaster scenario. To construct an analog operator for a set of n disaster scenarios, we can construct $\tilde{\Phi}_0$ and $\tilde{\Phi}_1$ as

$$\tilde{\Phi}_0 = \left[\Phi_0^1, \Phi_0^2, \ldots, \Phi_0^n\right] \tag{6}$$

$$\tilde{\Phi}_1 = \left[\Phi_1^1, \Phi_1^2, \ldots, \Phi_1^n\right] \tag{7}$$

where (Φ_0^i, Φ_1^i) is the input–output pair of disaster scenario i.

Then, it can be seen that in this case

$$\hat{A} = \arg\min_{A \in \mathbb{R}^{|E|\times|E|}} \|\tilde{\Phi}_1 - A\tilde{\Phi}_0\| \tag{8}$$

Assuming that the norm used in (8) to quantify the error is the commonly used square root of the sum of the squares of each entry (also known as the Frobenius norm), and that the set of damage states studied can produce a set of $|E|$ linearly independent vectors, then the optimization problem associated with finding $\hat{A}$ is known, and the minimizer of this problem assumes the closed form

$$\hat{A} = \tilde{\Phi}_1\tilde{\Phi}_0^T\left(\tilde{\Phi}_0\tilde{\Phi}_0^T\right)^{-1} \tag{9}$$

where $\tilde{\Phi}_1\tilde{\Phi}_0^T$ denotes the one-period temporal cross-correlation, and $(\tilde{\Phi}_0\tilde{\Phi}_0^T)^{-1}$ is associated with the auto-correlation of the damage states. Note that the generation of the proposed recovery operator is highly efficient, and can handle large-scale systems, because it is based on least-squares optimization and linear operations. Thus, in general, the main source of computational complexity of the proposed method comes from assembling the input set of benchmark damage scenarios (and their respective recovery strategies), which can be done well before the occurrence of the damaging event. Also note that the set of damage states should include damage and recovery information of all elements to obtain nonsingular solutions. We propose using the Frobenius norm to penalize large deviations more, since later on, when using the operator to generate recovery strategies, large deviations would result in proposed recovery times too far from the optimal, which would reduce the quality of the overall recovery strategies. Nevertheless, note that the reader is free to select the type of norm to penalize deviations, but Equation (9) would not return the optimal $\hat{A}$, and other optimization methods would be needed.

The operator $\hat{A}$ can be viewed as a matrix of dimensions $|E| \times |E|$ composed of real values. In general, each position $\hat{A}_{\hat{i}\hat{j}}$ indicates the influence that the damage state of element $\hat{i}$ at any given time t has over the damage state of element $\hat{j}$ at time $t+1$. This operator contains important information about the recovery dynamics of the system. In particular, diverse relevant analyses could be performed using the recovery operator, such as finding the recovery modes of the system and their rates of convergence via DMD, based on the eigendecomposition of $\hat{A}$.

Note that the proposed method to construct the recovery operator assumes that each individual element has the same relative importance, thus errors in their individual recovery times are weighted identically. The method could be expanded to consider prior information about the relative importance of each element or each time period modeled. Assuming there are positive diagonal matrices W_E and W_T that describe the relative importance (weights) associated with the studied elements and recovery periods, respectively, the general optimization problem to construct the recovery operator would then be

$$\hat{A} = \arg\min_{A \in \mathbb{R}^{|E|\times|E|}} \|W_E(\tilde{\Phi}_1 - A\tilde{\Phi}_0)W_T\|. \tag{10}$$

The more important a given element (or period) is, the higher its respective values in W_E (or W_T), thus resulting in a larger penalization for its errors.

2.2 Efficient generation of recovery strategies

The simplicity of the linear dynamics (2) provides a computationally efficient method to generate approximate recovery strategies given a set of initial damage conditions.

Assume that for a given time t, there is a known damage state $\phi(t)$, and we want to generate a recovery strategy starting from it along with the damage states associated with such strategy, namely, $\{\breve{\phi}(t), \breve{\phi}(t+1), ..., \breve{\phi}(|\mathcal{T}|)\}$, with $\breve{\phi}(t) = \phi(t)$. Assuming that $\hat{A}$ has already been calculated (which in practice can be done during non-time-critical periods), then we know that for the period $t+1$ the approximate damage state can be calculated by $\hat{A}\breve{\phi}(t)$. As the approximate dynamics are continuous, then $\hat{A}\breve{\phi}(t)$ will require a projection step to describe a binary damage state. Thus, we can define a threshold value $0 < \bar{a} < 1$, that will determine the ranges in which a noninteger damage state will be rounded to a binary state. Then, the damage state for period $t+1$ would be

$$\breve{\phi}(t+1) = \lceil(\hat{A}\breve{\phi}(t) - \mathbf{1}\bar{a})\rceil \tag{11}$$

Generalizing for any positive integer of m periods in the future, we have that

$$\breve{\phi}(t+m) = \lceil(\hat{A}^m\breve{\phi}(t) - \mathbf{1}\bar{a})\rceil \tag{12}$$

The proposed method uses the recovery operator to replicate the recovery dynamics from the generating data, without enforcing any constraint on the predicted recovery strategy. If there are constraints that need to be enforced, such as a strict maximum resource utilization per recovery period, it is possible to use the recovery operator to formulate a general iterative approach as follows:

Assume that our current iteration starts at period t and its related initial damage state is $\breve{\phi}(t)$. Then, $\hat{A}\breve{\phi}(t)$ can be interpreted as the relative ranking or importance

of recovering each node during period t. In particular, this ranking is associated with the order in which the elements should be recovered such that the proposed recovery fits better the behavior described by the input recovery strategies. Thus, this ranking could be used to indicate the set of elements in which the resources should be used first (while ensuring that all desired constraints are satisfied). For example, if there is a strict maximum resource utilization, one could select the set of elements with higher importance and an associated total resource utilization less than the maximum allowed for that period. Then, considering the elements selected to be recovered at time t, calculate $\breve{\phi}(t+1)$, update the vector of relative importances (now $\hat{A}\breve{\phi}(t+1)$), and repeat.

Note that in the case of a recovery process, the spectral radius of the operator (denoted as $\rho(A)$) is expected to be strictly less than one, to guarantee that when using any disaster configuration as the initial damage state $\breve{\phi}(0)$, the damage states of the system will converge to a full recovery as t increases (since $\lim_{m\to\infty} A^m = \mathbf{0}$ if and only if $\rho(A) < 1$).

2.3 Measuring the quality of the generated recovery strategies

To use the strategies that are generated using the recovery operator, it is important to quantify their quality. In general, it is desirable that the recovery operator strategies resemble as close as possible strategies considered to be optimal. Using the strategies generated by a high-fidelity optimization model as a benchmark, it is possible to evaluate how much the recovery times deviate from one strategy to the other. In particular, the ideal is that the recovery times proposed for each individual component by the recovery operator approach deviate as little as possible compared to the optimal strategy.

In some instances, however, focusing only on comparing the proposed approximate recovery sequences with the optimal ones may not be enough. In particular, even though the recovery times of an approximate recovery strategy may deviate little compared to the optimal one, the resultant performance recovery and costs associated with implementing them may differ greatly. Moreover, the approximate recovery strategy may be an unfeasible one. In particular, if the elements of the studied system are highly interdependent, or if the available resources are very limited, the feasibility of the solutions will be sensitive to changes in the recovery times of each element. Thus, in addition to studying the errors in the estimated recovery times, it is also important to compare the cumulative percentage of repaired elements associated with the estimated and benchmark recovery strategies. If this cumulative function describes different behaviors for the two strategies, this would indicate that the performance recovery, the resource utilization, and the costs associated with each strategy are inconsistent.

Whenever there are inconsistencies in the predicted times of recovery or the cumulative percentage of repaired elements, it is important to evaluate the impact of these inconsistencies on the quality of the predicted recovery strategy. To estimate the impact of these recovery time deviations in an efficient manner, we could construct an optimization model to both evaluate the performance of a given recovery strategy, and detect any infeasibility associated with it, by modifying the high-fidelity optimization model originally used to determine the benchmark optimal recovery strategies (González et al., 2016a). Assume that such high-fidelity model is defined based on two arrays of variables, X and Y, where the first indicates in which period each element should be recovered (i.e., the recovery strategy), and the second encompasses all remaining operational variables, such as flow of commodities, undersupply, etc. Thus, without loss of generality, assume that the high-fidelity optimization model can be written as $\min_{X,Y}\{F(X,Y) : C(X,Y) \geq 0\}$, where $F(X,Y)$ is the objective function (such as recovery cost, total recovery time, etc.) to be minimized, and $C(X,Y)$ represents the set of constraints that determine which solutions are feasible. Now, let χ be a fixed recovery strategy generated using the recovery operator (in which case χ at period t would be described by $\breve{\phi}(t-1) - \breve{\phi}(t)$). Then, the modified optimization problem

$$\min_{Y} \quad F(\chi, Y) \tag{13}$$

$$\text{s.t.} \quad C(\chi, Y) \geq 0 \tag{14}$$

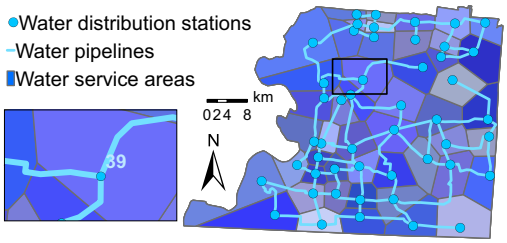
could be used to evaluate the performance and feasibility of strategy χ. In particular, Equation (13) could be used to evaluate if there is a performance detriment, whereas inequalities (14) would determine if the predicted strategies are not feasible, indicating the exact periods and elements associated with such an infeasibility. In general, the optimization problem described by Equations (13) and (14) will be more efficient than the original high-fidelity optimization problem, since its number of decision variables is highly reduced. For example, the optimization problem proposed in Appendix A (used to find the optimal recovery strategy) is NP-hard, but its modified version (used to evaluate given recovery strategies) can be solved in polynomial time (if partial functionality is allowed) (González et al., 2016a).

3 ILLUSTRATIVE EXAMPLE—SHELBY COUNTY, TN

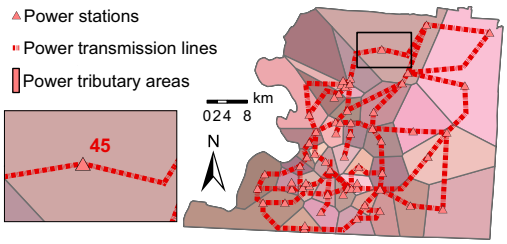
To illustrate some of the capabilities of the proposed recovery operator when studying multiple interdependent infrastructure networks and designing recovery strategies for them simultaneously, this case study examines the power, water, and gas networks in Shelby County, TN, United States. The site includes the city of Memphis, and is known to be in the area of influence of the NMSZ. These test networks were adapted from Hernandez-Fajardo and Dueñas-Osorio (2013), Dueñas-Osorio et al. (2007), and Song and Ok (2010). Note that the proposed method could be used for a broad range of systems beyond those previously mentioned, such as transportation systems and telecommunications, among others. Figure 1 shows the gas, water, and power networks, as well as the intersection areas that constitute the set $\mathcal{S}$.

In this illustrative example, we modeled two types of interdependencies, the physical interdependence between the power network and the gas and water networks, and the geographical interdependence between the gas and the water networks. The first is of relevance given that pumping stations require sufficient power supply to work adequately, whereas power stations require water for cooling down purposes. The second is associated with the geographical colocation between the gas and water networks, which enables potential reductions in the geographical preparation costs by scheduling simultaneous repair jobs of components from each network that share the same corridors.

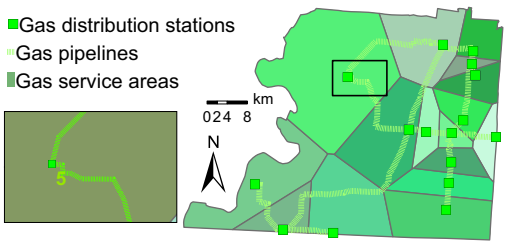
For this case study, we used randomly generated disaster realizations consistent with earthquakes of magnitudes $M_w \in \{6, 7, 8, 9\}$ and epicenter at the 35.3°N and 90.3°W site. We simulated 1,000 damage realizations associated with each magnitude, and used a high-fidelity MIP model to optimize each of their recovery strategies. In general, the proposed recovery operator is constructed such that it can be applied to diverse initial disaster scenarios, making the presented methodology not depend directly on the causes of failure or their likelihood. Nevertheless, the quality of the obtained recovery operator will depend on how accurate and realistic the input disaster realizations are. In this study, the authors assume hazard propagation curves that relate each magnitude with PGA and PGV values in space. With these seismic hazard models, and the subsequent use of fragility curves associated with each element in the system of systems (from HAZUS), we obtain the probabilities of failure that are then used to randomly generate the initial damage states. For further details regarding the disaster scenarios used, please refer to González et al. (2016b). The MIP model used to optimize the



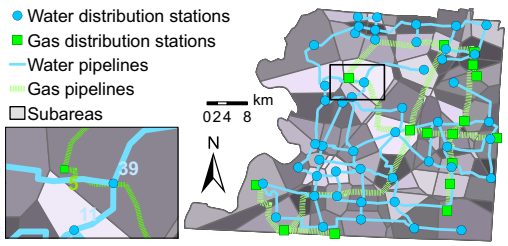
(a) Water network



(b) Power network



(c) Gas network



(d) Superposition of the gas and water networks

Fig. 1. Critical infrastructure networks in Shelby County, TN (adapted from González et al., 2014b).

recovery strategies, denominated the td-INDP, is detailed in Appendix A (each MIP instance was solved using Xpress-MP, in a PC with Windows 7 64-bit OS, 32 GB of RAM, and a processor Intel Core i7 4770 CPU @3.40 GHz). The available resources—modeled using Equation (A12)—were set such that a maximum of six elements could be recovered per period (thus, at least a node and an arc from each network could be recovered during each period), allowing the resultant recovery process to better showcase the effects of interdependencies. For each disaster realization studied, whenever the optimizer had spent more than 120 seconds solving the td-INDP, the optimizer was set to return the best

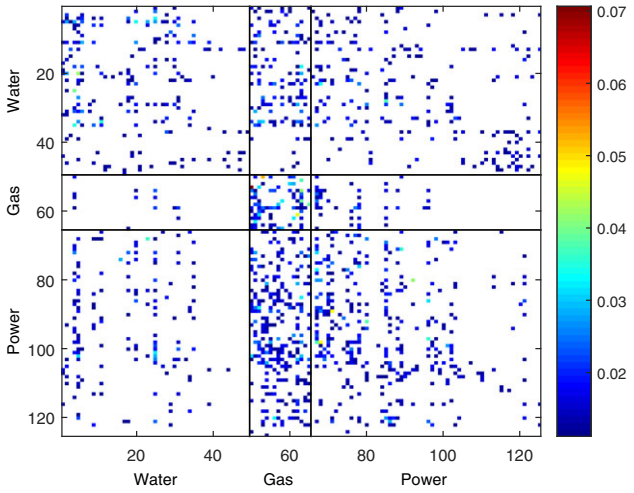


Fig. 2. Representation of the recovery operator $\hat{A}$ considering nodes only. Matrix entries with absolute value below 0.011 are not displayed. The rows and columns correspond to the assigned labels for each element.

recovery strategy found and its respective optimality gap (to give analytical guarantees of the quality of the returned solution). Each individual optimality gap was calculated by comparing the objective functions associated with the best recovery strategy found and solutions of constrained linear relaxations of the problem. The average optimality gaps were 0%, 0%, 1.8%, and 18.8%, for earthquake magnitudes 6, 7, 8, and 9, respectively.

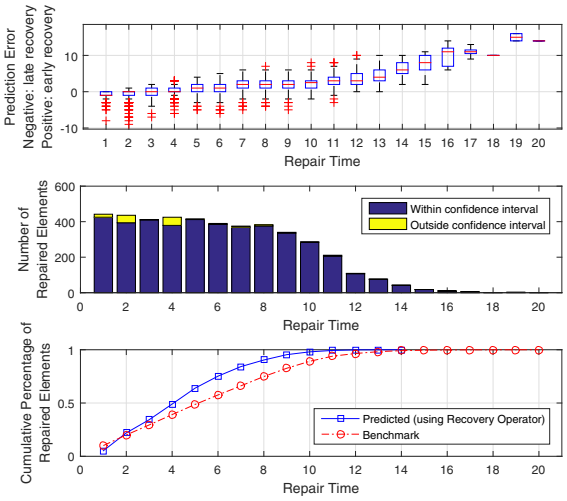
To showcase the generality and versatility of the proposed methodologies, the following subsections present two different cases, associated with identifying and approximating the recovery dynamics of the nodes only, and of both nodes and arcs.

3.1 Recovery operator for nodes

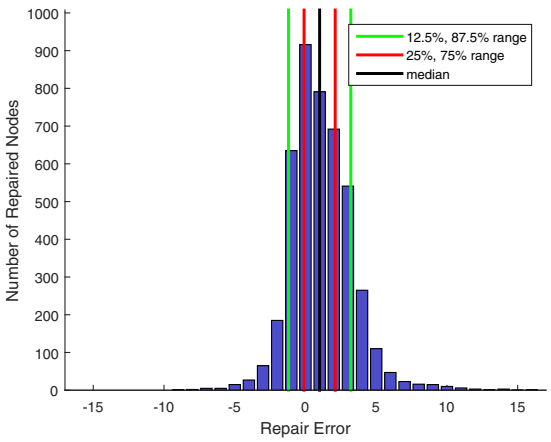
Figure 2 represents the recovery operator $\hat{A}$ found as proposed in Section 2, based on the full set of randomly generated disaster scenarios and their respective td-INDP recovery strategies. Each entry $\hat{A}_{e\tilde{e}}$ (value of $\hat{A}$ at row e and column $\tilde{e}$) shows the strength in which the damage state of element $\tilde{e}$ at any given time period influences the damage state of element e in the next period. The coefficient of determination associated with $\hat{A}$ was $R^2 = 0.7729$, and its root-mean-square (RMS) error was 0.0099, which are good indicators of the accuracy of the found operator and its predictability power. The spectral radius of this operator is 0.9250, which indicates that the system will tend to full recovery, independently of the initial damage state. Note that Figure 2 does not show the lowest values in the operator or its diagonal entries, to highlight the nontrivial stronger effects. To select the minimum value to display, first we took all the entries in $\hat{A}$ (except for the diagonal) and sorted their absolute values (in increasing order). Then, we calculated the knee (point of maximum curvature) of the curve associated with this list. Since this point indicates the value for which the rate of growth in the list changes the most, we used it as the minimum to display. In particular, this value was 0.011 for Figure 2. This recovery operator uses $E = \mathcal{N}$, thus it is showing the relation between the damage states of all pairs of nodes, with $|\mathcal{N}| = 125$. By drawing horizontal and vertical lines to separate the nodes in each axis into the three infrastructure networks studied, we obtain nine different sections that depict the intra- and internetwork recovery dependencies. For simplicity, let us name each block as $\hat{A}_{k\tilde{k}} : k, \tilde{k} \in \mathcal{K}$, which shows the strength in which the damage state of nodes in network $\tilde{k}$ affects the states of network k. The diagonal blocks (i.e., $\hat{A}_{kk} : k \in \mathcal{K}$) depict the strength in which the current damage state of the nodes in each given network influences the recovery of the nodes of the same network, whereas the off-diagonal blocks (i.e., $\hat{A}_{k\tilde{k}} : k, \tilde{k} \in \mathcal{K} | k \neq \tilde{k}$) depict how the state of the nodes in each network impacts the recovery of the other networks. Since each row and column of the recovery operator corresponds to a specific element in each network, the structure and values exhibited by the operator can provide valuable insights about the properties of the system of systems. For example, as expected, Figure 2 shows that the damage states of the nodes in each network have a strong influence within their network, but also that some networks have a strong influence on the recovery of the others. In particular, it can be seen that $\hat{A}_{\text{water, gas}}$ and $\hat{A}_{\text{power, gas}}$ are dense compared to the other blocks, showing that the states of the gas nodes have a notable influence over the other networks. This can also be interpreted as a priority to repair the nodes in the gas network before recovering the nodes of other networks—consistent with the td-INDP training data, which does not include safety constraints. Such trends can be explained considering the relative topologies of each network, since the studied gas network has fewer elements and much less redundancies than the other networks, which forces initial repairs in the gas networks to have a larger influence on the overall performance of the system of networks (assuming that M^+_{iklt} and M^-_{iklt} from Equation (A1) are equal for all networks). On the other side, blocks $\hat{A}_{\text{gas, water}}$ and $\hat{A}_{\text{gas, power}}$ are sparse compared to the others, which indicates that states of the gas network do not strongly depend on the states of the others. This also supports the interpretation that the gas network had a priority to be recovered. Note that these influence relationships are not trivially observable without relying on

the recovery operator, since as mentioned, the gas was not considered to be physically dependent on the power or the water in the training td-INDP data. Note that the element ordering used in Figure 2 was adequate to easily separate the elements from different networks, but other element sorting may also be useful. For example, one could sort the elements such that their closeness in the operator reflects its geographical colocation, providing a graphical depiction that would facilitate finding the influence that each geographical region has over others.

3.1.1 Generated recovery strategies for nodes. To evaluate the power and accuracy of the proposed methodology to efficiently generate recovery strategies using the recovery operator, we separated all disaster scenarios into 10 different groups (of 100 scenarios each) and applied cross-validation. In particular, for each group we generated recovery strategies, using a recovery operator created using only the benchmark (td-INDP-generated) recovery strategies of the other nine groups. For each disaster scenario in each group, seeded with the initial disaster state $\phi(0) = \check{\phi}(0)$, we used its associated recovery operator to efficiently generate a recovery strategy as described in Section 2.2. The generated strategies were compared with their associated td-INDP-based recovery strategies, to study the quality of approximation for each time period. For this case study, assuming no prior information was provided, we used a threshold value (number used in Equation (12) to round a noninteger damage state to a binary state) $\bar{a} = 0.5$. The average cross-validation RMS error was 0.03165, and the average R^2 was 0.7681. The average error in the predicted recovery times was 0.92195, and its standard deviation was 2.16407. These values show that the proposed operator represents a good fit for the recovery dynamics, and can be used to adequately propose quasi-optimal recovery strategies. Figure 3a shows the box-plot of the errors between the predicted strategies with respect to their benchmark, the number of elements with repair times accurately predicted for each period, and the cumulative percentage of recovery for both the predicted and the benchmark strategies. Figure 3b shows the associated histogram of the errors between the two strategies. It can be observed that the predicted recovery strategies can replicate the td-INDP recovery strategies accurately, given that the median error is less than one period, and in general the estimated recovery times have an error below three periods. In fact, the overall distribution of errors shows that more than 50% of the errors are within only one period from the median, and more than 75% within two periods of the median, indicating that the generated strategies are also precise. Note that even though the accuracy is better for the starting recovery periods, a good level of accuracy is achieved even when performing a long-range prediction (i.e., when estimating recovery times that are far from the current time).



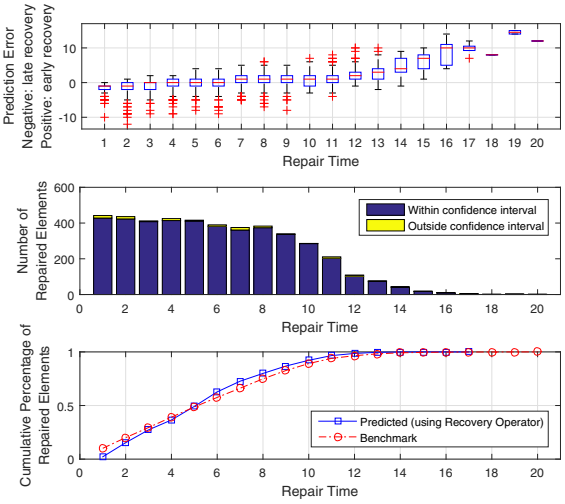
(a) Box-plot of the errors, number of repaired elements, and cumulative percentage of recovery, respectively, for each recovery period.



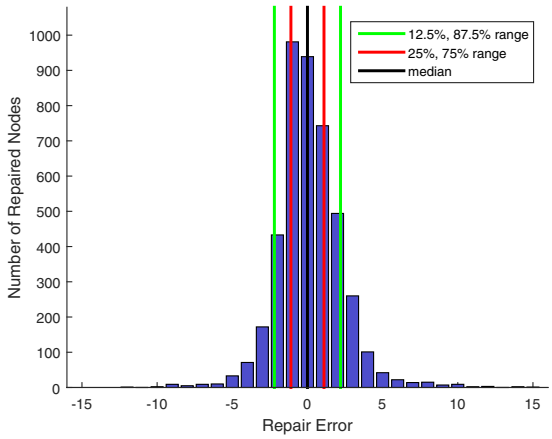
(b) Histogram of the errors

Fig. 3. Comparison between predicted recovery strategy for nodes (using the recovery operator) and the benchmark strategies (based on the td-INDP), using $\bar{a} = 0.5$.

Note that the quality of the estimated solutions depends directly on the threshold value $\bar{a}$ used. For higher accuracy, the parameter $\bar{a}$ could be tuned to minimize the difference between the benchmark and the predicted recovery, that is, by solving an additional optimization problem of the form $\arg\min_{\bar{a}} \|\check{\phi}(t+k) - (\hat{A}^k \check{\phi}(t) - \mathbf{1}\bar{a})\|$. To exemplify the effects of choosing a "tuned-up" $\bar{a}$, Figures 4a and b show a comparison between the benchmark and the predicted recovery times



(a) Box-plot of the errors, number of repaired elements, and cumulative percentage of recovery, respectively, for each recovery period.



(b) Histogram of the errors

Fig. 4. Comparison between predicted recovery strategy for nodes (using the recovery operator) and the benchmark strategies (based on the td-INDP), using $\bar{a} = 0.4$.

using $\bar{a} = 0.4$. The average error in the predicted recovery times using this value of $\bar{a}$ changed to 0.20954, and its standard deviation to 2.05361. These values show that the proposed recovery strategies generated using the optimized threshold resemble more accurately the optimal recovery strategies. From Figure 4a, it can be seen that the relative number of outliers observed for each recovery time is less than the one observed using $\bar{a} = 0.5$. Similarly, the predicted cumulative fraction of elements recovered depicts the benchmark much more accurately than if using $\bar{a} = 0.5$. In particular, it is important to ensure that the predicted cumulative fraction of elements recovered is not extensively far above the

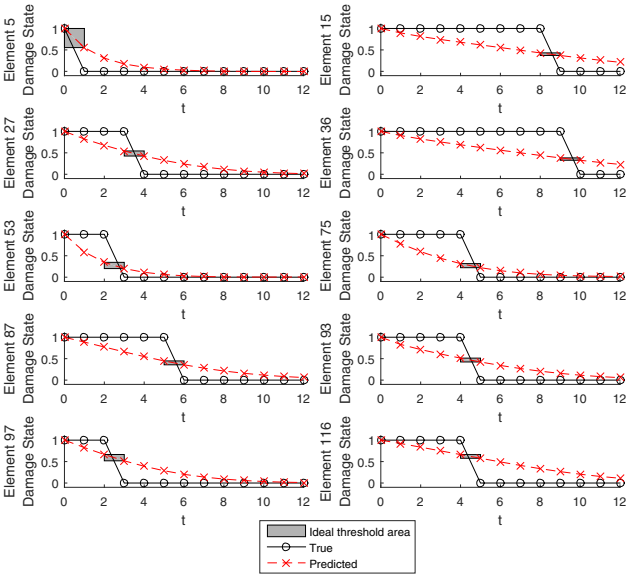


Fig. 5. Recovery state estimation for 10 randomly selected components.

benchmark, as it would be suggesting likely infeasibilities in the recovery strategies, due to the extended use of more resources than available. Moreover, Figure 4b shows that the histogram of the observed errors is better centered around zero, relative to its analogous in Figure 3b.

To provide insight on the quality of the operator strategies, and on how Equation (12) is used to generate such recovery strategies, Figure 5 shows the evolution of the failure states for a set of 10 nodes, comparing their td-INDP recovery strategy and their associated values of $\hat{A}^k\check{\phi}(0)$. For each component, a value of 1 describes a fully failed state, whereas 0 describes a fully recovered state. For each element in Figure 5, there is a gray box that highlights in the horizontal axis the time period in which the element was recovered according to the td-INDP strategy, and in the vertical axis the range of values that $\bar{a}$ can take such that the generated recovery strategy perfectly replicates the td-INDP strategy. As expected, the faster an element is recovered according to the td-INDP strategy, the larger is the range of values that $\bar{a}$ can take to return an accurate recovery strategy based on the recovery operator, since the slope of $\hat{A}^k\check{\phi}(0)$ for that particular element would be higher. For example, it can be observed that the range of threshold values that would generate an accurate recovery time for element 36 is much narrower than the same range associated with element 5. Note that another approach to fine-tune $\bar{a}$ would be finding the threshold value that better fits these accuracy ranges across all elements and scenarios. In this case, $\bar{a}$ would maximize the number

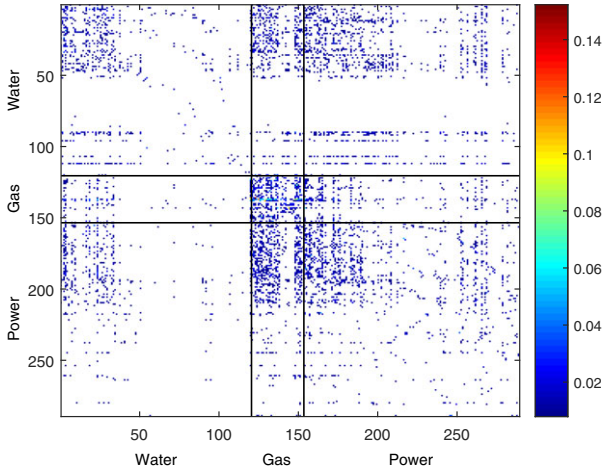


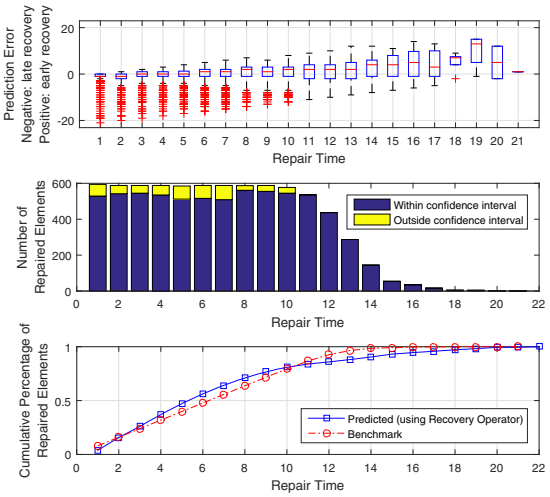
Fig. 6. Representation of the recovery operator $\hat{A}$ considering both nodes and arcs. Matrix entries with absolute value below 0.0079 are not displayed. The rows and columns correspond to the assigned labels for each element.

of times in which the recovery periods were accurately predicted, instead of minimizing the overall error.

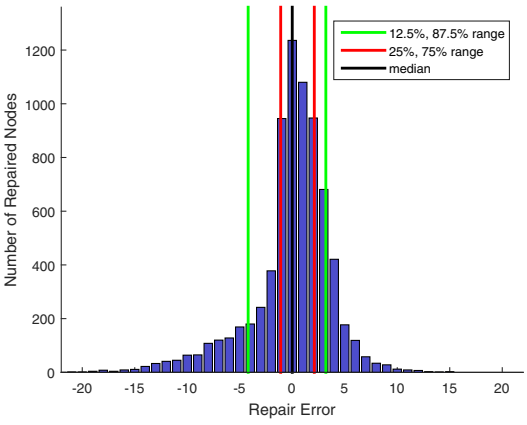
3.2 Recovery operator for nodes and arcs

Now, to model the recovery process of all the elements in the studied system of networks, assume $E = \mathcal{N} \cup \mathcal{A}$. In this case, the total number of elements increases from 125 to 289. Figure 6 shows the calculated recovery operator $\hat{A}$ for this new set of elements, using a display threshold of 0.0079. The coefficient of determination associated with this operator was $R^2 = 0.8857$ and the RMS error was 0.0117, which show an increase in accuracy and predictability power due to the inclusion of additional significant elements, with respect to the case of nodes only. The spectral radius of this operator is 0.9944, which again guarantees a predicted full recovery of the system. As observed in the recovery operator associated with nodes only, the blocks $\hat{A}_{\text{water, gas}}$ and $\hat{A}_{\text{power, gas}}$ are dense relative to other blocks, whereas $\hat{A}_{\text{gas, water}}$ and $\hat{A}_{\text{gas, power}}$ are the most sparse. This shows that the recovery operator provides consistent results regarding the relationships between networks. Nevertheless, it can be seen that for each block in this operator, the upper-left corners (associated with the node elements, as opposed to the arc elements, located at the lower-right corners) are densest. This indicates that in general the recovery relationships between nodes tend to be stronger than the ones observed for the arcs.

3.2.1 Generated recovery strategies for both nodes and arcs. Following the procedure detailed in Section 3.1.1, for each of the disaster scenarios, we generated recovery strategies for both nodes and arcs simultaneously. For this case, the threshold value $\bar{a}$ used was also 0.5. The average cross-validation RMS error was 0.03748, and the average R^2 was 0.88184. The average error in the predicted recovery times was 0.15993, and its standard deviation was 3.29823. These values show that the recovery operator can be used to adequately generate quasi-optimal recovery strategies, with an average error even smaller than the case with only nodes. Nevertheless, note that the average deviation was slightly higher, which can be expected since there are two different types of elements being analyzed simultaneously now. As in the previous examples, Figure 7 overviews different aspects associated with the errors of the



(a) Box-plot of the errors, number of repaired elements, and cumulative percentage of recovery, respectively, for each recovery period.



(b) Histogram of the errors

Fig. 7. Comparison between predicted recovery strategy for nodes and arcs (using the recovery operator) and the benchmark strategies (based on the td-INDP), using $\bar{a} = 0.5$.

generated recovery strategies compared to the td-INDP strategies. Figures 7a and b show that the generated strategies are also accurate, having a median error close to 0, but with a slightly increased variability. However, Figure 7a shows that the predicted cumulative recovery function offers a close fit to the benchmark recovery at all periods, despite having to model two different types of elements simultaneously. In particular, Figure 7b shows that the error distribution has a reduced symmetry compared to the one associated with nodes only, favoring outliers in the left tail. This may be explained by the weaker recovery relationships observed in the arcs compared to the ones observed between the nodes, reducing the predictability power for these types of elements. Nevertheless, despite this increased variability, more than 50% of the proposed recovery times are within two periods with respect to the td-INDP strategies, and more than 75% within five periods.

4 CONCLUSIONS

In this article, we presented a novel approach to efficiently generate approximate recovery strategies for interdependent infrastructure systems, using a linear time-invariant recovery operator. The recovery operator is generated from a set of high-fidelity recovery scenarios, to identify and approximate their main recovery dynamics, and then to develop tailored recovery strategies after specific disaster scenarios. For our illustrative example, we used the td-INDP to generate these benchmark recovery strategies associated with a set of predetermined disaster scenarios. The recovery strategies generated using the proposed recovery operator show good agreement with the high-fidelity recovery scenario even for long-term planning, both in the time of recovery and the resources used in each period. Also note that as our proposed method is a data-driven process that operates on "snap-shots" of input/output data, the recovery operator approach is not restricted to any particular recovery strategy algorithm.

As the linear operator approach is a compact representation of the recovery dynamics, it provides a useful tool with varied perspectives for decision makers and stakeholders. First, the linear generation of the approximate recovery strategy can be exploited for its computational savings when compared to the computationally expensive high-fidelity algorithms. This provides an efficiently computable coarse recovery strategy as well as a good initial strategy to warm-start high-fidelity optimization models. Second, the operator describes time-correlations within the dynamics through the zero structure of $\hat{A}$, exposing sometimes hidden interdependencies between network elements that can be difficult to discern from the high-fidelity modeling framework. Analysis of these interdependencies can help isolate particular significant sections of the networks during the recovery process and help inform recovery modifications. Finally, as the approximate dynamics are linear, they can be examined using a plethora of techniques from linear systems theory. The linear operator representation means that linear time-invariant system analysis tools can be applied to examine features like stability, convergence, and modal features of the dynamics. The area of linear controller design is also open for exploration, addressing such issues as optimal placements of resources in the network and designing inputs into the system for efficient recovery.

Currently, the construction of the proposed recovery operator assumes that the system to be recovered is deterministic, and that the recovery process can be seen as a time-invariant process, where the optimal recovery strategy depends only on the current damage states of each element (and not their recovery/damage history). For future work, we will explore and extend the capabilities of the recovery operator to consider hysteresis, uncertainty in the recovery process, and time-variant recovery dynamics. This would allow, among others, studying the impact of resource and behavioral inertia and the occurrence of multiple subsequent damaging events. Note that if the available resources change in time, the underlying recovery dynamics may be affected; thus, to improve the resultant accuracy it may be ideal to construct different recovery operators for different time domains. Also, applying tools from linear controller analysis and design is of particular interest, as notions of controllability and observability can now be applied to recovery dynamics. For example, Equation (2) could be extended to consider the effects of an input matrix B and a control vector u, such that $\phi(t+1) \approx \hat{A}\phi(t) + Bu(t)$. Connections between control-theoretic measures and infrastructure-centric measures such as resilience would present a powerful link between the two fields, enabling stakeholders to determine optimal control configurations to maximize recoverability and resilience of their systems across multiple time scales. Moreover, as hinted in Section 2.2, the recovery operator offers relevant information about the relative importance of each element in the recovery process, that could be exploited in determining possible element clustering configurations and the development of importance rankings.

ACKNOWLEDGMENTS

This publication was partially funded by the "Research Program 2012" Grant from the Office of the Vice

President for Research - Universidad de los Andes (Bogotá, Colombia). The authors gratefully acknowledge the support by Centro de Estudios Interdisciplinarios Básicos y Aplicados (CEIBA), the U.S. National Science Foundation (Grant CMMI-1436845), the U.S. Department of Defense (Grant W911NF-13-1-0340), and FICO for providing the Xpress-MP licenses used in the computational experiments.

REFERENCES

Ahuja, R. K., Magnanti, T. L. & Orlin, J. B. (1993), *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, Upper Saddle River, NJ.

Anderson, B. D. O. & Moore, J. B. (1990), *Optimal Control: Linear Quadratic Methods*, Prentice Hall, Upper Saddle River, NJ.

Assad, A. A. (1978), Multicommodity network flows—a survey, *Networks*, **8**(1), 37–91.

Åström, K. & Eykhoff, P. (1971), System identification—a survey, *Automatica*, **7**(2), 123–62.

Barocio, E., Pal, B. C., Thornhill, N. F. & Messina, A. R. (2015), A dynamic mode decomposition framework for global power system oscillation analysis, *IEEE Transactions on Power Systems*, **30**(6), 2902–12.

Baroud, H., Barker, K., Ramirez-Marquez, J. E. & Rocco, C. M. (2015), Inherent costs and interdependent impacts of infrastructure network resilience, *Risk Analysis*, **35**(4), 642–62.

Beskhyroun, S., Wegner, L. D. & Sparling, B. F. (2011), New methodology for the application of vibration-based damage detection techniques, *Structural Control and Health Monitoring*, **19**(1), 88–106.

Brummitt, C. D., D'Souza, R. M. & Leicht, E. A. (2012), Suppressing cascades of load in interdependent networks, *Proceedings of the National Academy of Sciences of the United States of America*, **109**(12), E680–89.

Brunton, S. L., Brunton, B. W., Proctor, J. L. & Kutz, J. N. (2016), Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control, *PLOS ONE*, **11**(2), e0150171.

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. (2010), Catastrophic cascade of failures in interdependent networks, *Nature*, **464**(7291), 1025–28.

Cavdaroglu, B., Hammel, E., Mitchell, J. E., Sharkey, T. C. & Wallace, W. A. (2011), Integrating restoration and scheduling decisions for disrupted interdependent infrastructure systems, *Annals of Operations Research*, **203**(1), 279–94.

Crainic, T. G., Li, Y. & Toulouse, M. (2006), A first multilevel cooperative algorithm for capacitated multicommodity network design, *Computers & Operations Research*, **33**(9), 2602–22.

Dantzig, G. B., Harvey, R. P., Lansdowne, Z. F., Robinson, D. W. & Maier, S. F. (1979), Formulating and solving the network design problem by decomposition, *Transportation Research Part B: Methodological*, **13**(1), 5–17.

Dueñas-Osorio, L., Craig, J. I., & Goodno, B. J. (2007), Seismic response of critical interdependent networks, *Earthquake Engineering & Structural Dynamics*, **36**(2), 285–306, https://doi.org/10.1002/eqe.626.

Fortz, B. & Poss, M. (2009), An improved Benders decomposition applied to a multi-layer network design problem, *Operations Research Letters*, **37**(5), 359–64.

Frangioni, A. & Gendron, B. (2009), Reformulations of the multicommodity capacitated network design problem, *Discrete Applied Mathematics*, **157**(6), 1229–41.

Gao, J., Buldyrev, S. V., Havlin, S. & Stanley, H. E. (2011), Robustness of a network of networks, *Physical Review Letters*, **107**(19), 195701-1–195701-5.

Gendron, B., Crainic, T. G. & Frangioni, A. (1999), Multicommodity capacitated network design, in B. Sansò and P. Soriano (eds.), *Telecommunications Network Planning*, Springer US, pp. 1–19.

González, A. D., Dueñas-Osorio, L., Medaglia, A. L. & Sánchez-Silva, M. (2014a), Resource allocation for infrastructure networks within the context of disaster management, in G. Deodatis, B. Ellingwood, and D. Frangopol (eds.), *Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures*, CRC Press, New York, pp. 639–46.

González, A. D., Dueñas-Osorio, L., Medaglia, A. L. & Sánchez-Silva, M. (2016a), The time-dependent interdependent network design problem (td-INDP) and the evaluation of multi-system recovery strategies in polynomial time, in H. Huang, J. Li, J. Zhang, and J. Chen (eds.), *The 6th Asian-Pacific Symposium on Structural Reliability and Its Applications*, Shanghai, China, 544–50.

González, A. D., Dueñas-Osorio, L., Sánchez-Silva, M. & Medaglia, A. L. (2016b), The interdependent network design problem for optimal infrastructure system restoration, *Computer-Aided Civil and Infrastructure Engineering*, **31**(5), 334–50.

González, A. D., Sánchez-Silva, M., Dueñas-Osorio, L. & Medaglia, A. L. (2014b), Mitigation strategies for lifeline systems based on the interdependent network design problem, in M. Beer, S.-K. Au, and J. W. Hail (eds.), *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, American Society of Civil Engineers (ASCE), Reston, VA, pp. 762–71.

Hernandez-Fajardo, I. & Dueñas-Osorio, L. (2013), Probabilistic study of cascading failures in complex interdependent lifeline systems, *Reliability Engineering & System Safety*, **111**, 260–72.

Jiang, X. & Adeli, H. (2005), Dynamic wavelet neural network for nonlinear identification of highrise buildings, *Computer-Aided Civil and Infrastructure Engineering*, **20**(5), 316–30.

Johnson, D., Lenstra, J. & Kan, A. (1978), The complexity of the network design problem, *Networks*, **8**, 279–85.

Kennington, J. L. (1978), A survey of linear cost multicommodity network flows, *Operations Research*, **26**(2), 209–36.

Kerschen, G., Worden, K., Vakakis, A. F. & Golinval, J.-C. (2006), Past, present and future of nonlinear system identification in structural dynamics, *Mechanical Systems and Signal Processing*, **20**(3), 505–92.

Kijewski, T. & Kareem, A. (2003), Wavelet transforms for system identification in civil engineering, *Computer-Aided Civil and Infrastructure Engineering*, **18**(5), 339–55.

Koopman, B. O. & von Neumann, J. (1932), Dynamical systems of continuous spectra, *Proceedings of the National Academy of Sciences of the United States of America*, **18**(3), 255–63.

Kutz, J. N., Fu, X., Brunton, S. L. & Erichson, N. B. (2015), Multi-resolution dynamic mode decomposition for foreground/background separation and object tracking, in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, vol. 2016, IEEE, 921–29.

Lee II, E. E., Mitchell, J. E. & Wallace, W. A. (2007), Restoration of services in interdependent infrastructure systems: a network flows approach, *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, **37**(6), 1303–17.
Lin, D.-Y. (2011), A dual variable approximation-based descent method for a bi-level continuous dynamic network design problem, *Computer-Aided Civil and Infrastructure Engineering*, **26**(8), 581–94.
Ljung, L. (1999), *System Identification (2nd Ed.)—Theory for the User*, Prentice Hall PTR, Upper Saddle River, NJ.
McGhee, R. B. (1963), Identification of nonlinear dynamic systems by regression analysis methods, Ph.D. thesis.
Mauroy, A. & Mezic, I. (2013), A spectral operator-theoretic framework for global stability, in *52nd IEEE Conference on Decision and Control*, no. 3, IEEE, 5234–39.
Mauroy, A., Rhoads, B., Moehlis, J. & Mezić, I. (2014), Global isochrons and phase sensitivity of bursting neurons, *SIAM Journal on Applied Dynamical Systems*, **13**(1), 306–38.
Mezić, I. (2005), Spectral properties of dynamical systems, model reduction and decompositions, *Nonlinear Dynamics*, **41**(1–3), 309–25.
Nurre, S. G., Cavdaroglu, B., Mitchell, J. E., Sharkey, T. C. & Wallace, W. A. (2012), Restoring infrastructure systems: an integrated network design and scheduling (INDS) problem, *European Journal of Operational Research*, **223**(3), 794–806.
Ouyang, M. (2014), Review on modeling and simulation of interdependent critical infrastructure systems, *Reliability Engineering & System Safety*, **121**, 43–60.
Ouyang, M. & Wang, Z. (2015), Resilience assessment of interdependent infrastructure systems: with a focus on joint restoration modeling and analysis, *Reliability Engineering & System Safety*, **141**(September), 74–82.
Pant, R., Barker, K. & Zobel, C. W. (2013), Static and dynamic metrics of economic resilience for interdependent infrastructure and industry sectors, *Reliability Engineering & System Safety*, **125**, 1–11.
Pedersen, M. B., Crainic, T. G. & Madsen, O. B. G. (2008), Models and tabu search metaheuristics for service network design with asset-balance requirements, *Transportation Science*, **43**(2), 158–77.
Poss, M. (2012), Models and algorithms for network design problems, *4OR*, **10**(2), 215–16, https://doi.org/10.1007/s10288-011-0174-8.
Rinaldi, S. M., Peerenboom, J. P. & Kelly, T. K. (2001), Identifying, understanding, and analyzing critical infrastructure interdependencies, *IEEE Control Systems Magazine*, **21**(6), 11–25.
Rowley, C. W., Mezić, I., Bagheri, S., Schlatter, P. & Henningson, D. S. (2009), Spectral analysis of nonlinear flows, *Journal of Fluid Mechanics*, **641**, 115–27.
Schmid, P. J. (2010), Dynamic mode decomposition of numerical and experimental data, *Journal of Fluid Mechanics*, **656**(4), 5–28.
Schmid, P. J., Li, L., Juniper, M. P. & Pust, O. (2011), Applications of the dynamic mode decomposition, *Theoretical and Computational Fluid Dynamics*, **25**(1–4), 249–259.
Shafieezadeh, A., Leelardcharoen, K. & Dueñas-Osorio, L. (2014), A framework for assessing the effectiveness of resilience enhancement strategies for interdependent infrastructure systems, in *Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures*, CRC Press/Balkema, Leiden, The Netherlands, pp. 573–80.
Sirca, G. & Adeli, H. (2012), System identification in structural engineering, *Scientia Iranica*, **19**(6), 1355–64.
Söderström, T. & Stoica, P. (1989), *System Identification*, Prentice-Hall, Upper Saddle River, NJ.
Song, J. & Ok, S.-Y. (2010), Multi-scale system reliability analysis of lifeline networks under earthquake hazards, *Earthquake Engineering & Structural Dynamics*, **39**, 259–79.
Sootla, A., Mauroy, A. & Goncalves, J. (2016), *Shaping Pulses to Control Bistable Monotone Systems Using Koopman Operator*, in *IFAC-PapersOnLine*, **49**(18), 698–703.
Stengel, R. F. (1994), *Optimal Control and Estimation*, Dover, New York.
UNISDR (2015), *Making Development Sustainable: The Future of Disaster Risk Management*, Global Assessment Report on Disaster Risk Reduction.
U.S. Federal Emergency Management Agency (FEMA), (2011), *National Disaster Recovery Framework*, Technical report, U.S. Department of Homeland Security.
Vespignani, A. (2010), Complex networks: the fragility of interdependency, *Nature*, **464**(7291), 984–5.
Vugrin, E. D., Warren, D. E., Ehlen, M. A. & Camphouse, R. C. (2010), A framework for assessing the resilience of infrastructure and economic systems, in *Sustainable and Resilient Critical Infrastructure Systems: Simulation, Modeling, and Intelligent Engineering*, Springer Berlin Heidelberg, Germany, pp. 77–116.

APPENDIX A: THE TIME-DEPENDENT INTERDEPENDENT NETWORK DESIGN PROBLEM

Let us define the time-dependent Interdependent Network Design Problem (td-INDP) as the problem of finding the least-cost time-dependent recovery strategy of a partially destroyed system of infrastructure networks, subject to budget, resources, and operational constraints, while considering interdependencies between the networks (González et al., 2016a, 2016b). Under such a definition, the reconstruction costs can include the costs of the resources used, labor costs, the costs of preparing the geographical locations for the reconstruction process (González et al., 2014a, 2016a, 2016b), and costs related to the system not being able to perform adequately. Similarly, there are different types of interdependencies that the problem should be able to consider, such as physical, cyber, geographical, and logical (Rinaldi et al., 2001).

A.1 Formulation

The following td-INDP formulation (González et al., 2016a), uses as an input the information associated with a partially destroyed system of interconnected infrastructure networks (costs, capacities, damaged components, resources, etc.), and returns the optimal recovery strategy (maximizing the performance and minimizing recovery costs). Such recovery strategies become the

data that will be used to generate the recovery operator. The sets, parameters, and decision variables used in the td-INDP formulation are described as follows:

Sets

- $\mathcal{N}$ Set of nodes before a destructive event
- $\mathcal{A}$ Set of arcs before a destructive event
- $\mathcal{T}$ Set of periods for the recovery process (time horizon)
- $\mathcal{S}$ Set of geographical spaces (spatial distribution of the area that contains the infrastructure networks)
- $\mathcal{L}$ Set of commodities flowing in the system
- $\mathcal{R}$ Set of limited resources to be used in the reconstruction process
- $\mathcal{K}$ Set of infrastructure networks
- $\mathcal{N}_k^*$ Set of nodes in network $k \in \mathcal{K}$ that require their demands to be fully satisfied to be functional
- $\mathcal{N}_k$ Set of nodes in network $k \in \mathcal{K}$ before a destructive event
- $\mathcal{N}_k'$ Set of destroyed nodes in network $k \in \mathcal{K}$ after the event
- $\mathcal{A}_k$ Set of arcs in network $k \in \mathcal{K}$ before a destructive event
- $\mathcal{A}_k'$ Set of destroyed arcs in network $k \in \mathcal{K}$ after the event
- $\mathcal{L}_k$ Set of commodities flowing in network $k \in \mathcal{K}$

Parameters

- v_{rt} Availability of resource r at time t
- h_{ijkrt} Usage of resource r related to recovering arc (i, j) in network k at time t
- p_{ikrt} Usage of resource r related to recovering node i in network k at time t
- M_{iklt}^+ Costs of excess of supply of commodity l in node i in network k at time t
- M_{iklt}^- Costs of unsatisfied demand of commodity l in node i in network k at time t
- α_{ijkst} Indicates if repairing arc (i, j) in network k at time t requires preparing space s
- β_{ikst} Indicates if repairing node i in network k at time t requires preparing space s
- $\gamma_{ijk\tilde{k}t}$ Indicates if at time t node i in network k depends on node j in network $\tilde{k} \in \mathcal{K}$
- g_{st} Cost of preparing geographical space s at time t
- f_{ijkt} Cost of recovering arc (i, j) in network k at time t
- q_{ikt} Cost of recovering node i in network k at time t
- c_{ijklt} Commodity l unitary flow cost through arc (i, j) in network k at time t
- u_{ijkt} Total flow capacity of arc (i, j) in network k at time t
- b_{iklt} Demand/supply of commodity l in node i in network k at time t

Decision variables

- δ_{iklt}^+ Excess of supply of commodity l in node i in network k at time t
- δ_{iklt}^- Unmet demand of commodity l in node i in network k at time t
- x_{ijklt} Flow of commodity l through arc (i, j) in network k at time t
- w_{ikt} Binary variable that indicates if node i in network k is functional at time t
- y_{ijkt} Binary variable that indicates if arc (i, j) in network k is functional at time t
- Δw_{ikt} Binary variable that indicates if node i in network k should be recovered at time t
- Δy_{ijkt} Binary variable that indicates if arc (i, j) in network k should be recovered at time t
- Δz_{st} Binary variable that indicates if space s has to be prepared at time t

Based on the previous description of the variables and parameters involved, the proposed model for the td-INDP is represented as follows:

minimize

$$\sum_{t\in\mathcal{T}|t>0}\left(\sum_{s\in\mathcal{S}} g_{st}\Delta z_{st} + \sum_{k\in\mathcal{K}}\left(\sum_{(i,j)\in\mathcal{A}_k'} f_{ijkt}\Delta y_{ijkt} + \sum_{i\in\mathcal{N}_k'} q_{ikt}\Delta w_{ikt}\right)\right) + \sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{K}}\left(\sum_{l\in\mathcal{L}_k}\sum_{i\in\mathcal{N}_k}\left(M_{iklt}^+\delta_{iklt}^+ + M_{iklt}^-\delta_{iklt}^-\right) + \sum_{l\in\mathcal{L}_k}\sum_{(i,j)\in\mathcal{A}_k} c_{ijklt}x_{ijklt}\right) \tag{A1}$$

subject to

$$\sum_{j:(i,j)\in\mathcal{A}_k} x_{ijklt} - \sum_{j:(j,i)\in\mathcal{A}_k} x_{jiklt} = b_{iklt} - \delta_{iklt}^+ + \delta_{iklt}^- \quad \forall k\in\mathcal{K}, \forall i\in\mathcal{N}_k, \forall l\in\mathcal{L}_k, \forall t\in\mathcal{T} \tag{A2}$$

$$\sum_{l\in\mathcal{L}_k} x_{ijklt} \le u_{ijkt}y_{ijkt}, \quad \forall k\in\mathcal{K}, \forall (i,j)\in\mathcal{A}_k, \forall t\in\mathcal{T} \tag{A3}$$

$$\sum_{l\in\mathcal{L}_k} x_{ijklt} \le u_{ijkt}w_{ikt}, \quad \forall k\in\mathcal{K}, \forall (i,j)\in\mathcal{A}_k, \forall t\in\mathcal{T} \tag{A4}$$

$$\sum_{l\in\mathcal{L}_k} x_{ijklt} \le u_{ijkt}w_{jkt}, \quad \forall k\in\mathcal{K}, \forall (i,j)\in\mathcal{A}_k, \forall t\in\mathcal{T} \tag{A5}$$

$$w_{ikt}|b_{iklt}| \le |b_{iklt}| - \delta_{iklt}^-, \quad \forall k\in\mathcal{K}, \forall i\in\mathcal{N}_k^*, \forall l\in\mathcal{L}_k, \forall t\in\mathcal{T} \tag{A6}$$

$$\sum_{i\in\mathcal{N}_k} w_{ikt}\gamma_{ijk\tilde{k}t} \ge w_{j\tilde{k}t}, \quad \forall k,\tilde{k}\in\mathcal{K}, \forall j\in\mathcal{N}_{\tilde{k}}, \forall t\in\mathcal{T} \tag{A7}$$

$$w_{ik0} = 0, \quad \forall k \in \mathcal{K}, \forall i \in \mathcal{N}'_k \tag{A8}$$

$$y_{ijk0} = 0, \quad \forall k \in \mathcal{K}, \forall (i,j) \in \mathcal{A}'_k \tag{A9}$$

$$w_{ikt} \leq \sum_{\tilde{t}=1}^{t} \Delta w_{ik\tilde{t}}, \quad \forall k \in \mathcal{K}, \forall i \in \mathcal{N}'_k, \forall t \in \mathcal{T} \mid t > 0 \tag{A10}$$

$$y_{ijkt} \leq \sum_{\tilde{t}=1}^{t} \Delta y_{ijk\tilde{t}}, \quad \forall k \in \mathcal{K}, \forall (i,j) \in \mathcal{A}'_k, \forall t \in \mathcal{T} \mid t > 0 \tag{A11}$$

$$\sum_{k \in \mathcal{K}} \left(\sum_{(i,j) \in \mathcal{A}'_k} h_{ijkrt} \Delta y_{ijkt} + \sum_{i \in \mathcal{N}'_k} p_{ikrt} \Delta w_{ikt} \right) \leq v_{rt}, \quad \forall r \in \mathcal{R}, \forall t \in \mathcal{T} \mid t > 0 \tag{A12}$$

$$\Delta w_{ikt} \alpha_{ikst} \leq \Delta z_{st}, \quad \forall k \in \mathcal{K}, \forall i \in \mathcal{N}'_k, \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \mid t > 0 \tag{A13}$$

$$\Delta y_{ijkt} \beta_{ijkst} \leq \Delta z_{st}, \quad \forall k \in \mathcal{K}, \forall (i,j) \in \mathcal{A}'_k, \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \mid t > 0 \tag{A14}$$

$$\delta^{+}_{iklt} \geq 0, \quad \forall k \in \mathcal{K}, \forall i \in \mathcal{N}_k, \forall l \in \mathcal{L}_k, \forall t \in \mathcal{T} \tag{A15}$$

$$\delta^{-}_{iklt} \geq 0, \quad \forall k \in \mathcal{K}, \forall i \in \mathcal{N}_k, \forall l \in \mathcal{L}_k, \forall t \in \mathcal{T} \tag{A16}$$

$$x_{ijklt} \geq 0, \quad \forall k \in \mathcal{K}, \forall (i,j) \in \mathcal{A}_k, \forall l \in \mathcal{L}_k, \forall t \in \mathcal{T} \tag{A17}$$

$$w_{ikt} \in \{0,1\}, \quad \forall k \in \mathcal{K}, \forall i \in \mathcal{N}_k, \forall t \in \mathcal{T} \tag{A18}$$

$$y_{ijkt} \in \{0,1\}, \quad \forall k \in \mathcal{K}, \forall (i,j) \in \mathcal{A}'_k, \forall t \in \mathcal{T} \tag{A19}$$

$$\Delta w_{ikt} \in \{0,1\}, \quad \forall k \in \mathcal{K}, \forall i \in \mathcal{N}'_k, \forall t \in \mathcal{T} \mid t > 0 \tag{A20}$$

$$\Delta y_{ijkt} \in \{0,1\}, \quad \forall k \in \mathcal{K}, \forall (i,j) \in \mathcal{A}'_k, \forall t \in \mathcal{T} \mid t > 0 \tag{A21}$$

$$\Delta z_{st} \in \{0,1\}, \quad \forall s \in \mathcal{S}, \forall t \in \mathcal{T} \mid t > 0 \tag{A22}$$

Equation (A1) details the five different costs that constitute the objective function. The first is the total cost of preparing the required geographical spaces for the recovery of nodes and arcs—related to perforation and excavations processes, constructing foundations, etc. The second and third costs are associated with the individual reconstruction of arcs and nodes, respectively. The fourth is the cost caused by having excess or deficit of commodities to properly satisfy the demands. Finally, the fifth represents the cost associated with the flow of commodities through each network. This objective function assumes that each term involved is quantified in similar units (such as monetary units), thus having a similar weight on the recovery process. Nevertheless, the objective function can be easily generalized such that it can include different weights for each term, by adding a weight/importance coefficient associated with each of them, effectively accounting for the possible heterogeneity of each type of cost. Constraints (A2) enforce adequate flow balance of commodities for each node, relating the total amount of commodities entering and leaving the node with its respective demands. Constraints (A3), (A4), and (A5) relate the functionality of each link, its starting and ending nodes, respectively, with the amount of commodities that could flow through it. Note that, on one hand, these serve as capacity constraints that limit the flow capacities when the corresponding nodes and arc are functional. On the other hand, when an arc or a node is not functioning (y or w binary variables are 0), the corresponding flows (variable x) are forced to be 0 as well. Constraints (A6) ensure that nodes in $\mathcal{N}^*_k$ are functional only if their demands are fully satisfied. Constraints (A7) guarantee that nodes need to be functional for their dependent nodes to become functional as well. Constraints (A8) and (A9) enforce that initially damaged nodes and arcs, respectively, are initially not functional either. Constraints (A10) and (A11) ensure that initially damaged nodes and arcs, respectively, can become functional only after they have been recovered. Note that if an initially damaged element has not been recovered by period t, the cumulative sum of the respective recovery decision variable (Δw or Δy) will be 0, thus forcing the functionality variable (w or y) to be 0 as well. Constraints (A12) ensure that for each period, the resources used to perform the planned recovery jobs are at most their availability. Constraints (A13) and (A14) guarantee that if a node or an arc, respectively, need to be repaired, the geographical spaces that accommodate them need to be prepared. In this case, if for a given element the recovery variable (Δw or Δy) is 1 at a period t, indicating that such an element should be recovered at that period, then the geographical preparation variable (Δz) is forced to be 1 for all associated spaces, thus indicating that these need to be prepared. Last, constraints (A15)–(A22) show the nature of each decision variable, indicating if they are continuous nonnegative or binary. For further details regarding the assumptions and other properties of the td-INDP, refer to González et al. (2016a, 2016b).