Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P530-cd

Digital twins or equivalent infrastructure models? The role of modeling granularity in regional risk analysis of infrastructure

Fabrizio Nocera

Department of Civil, Environmental and Geomatic Engineering, University College London, UK.
E-mail: f.nocera@ucl.ac.uk

Paolo Gardoni

Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign, USA.
E-mail: gardoni@illinois.edu

Accurate regional risk analysis requires representative mathematical models of infrastructure. One of the main challenges in developing mathematical models of infrastructure is defining their modeling granularity, i.e., the level of detail in the topology of the model. Different modeling granularities affect our ability to capture the spatial variability of the impact arising from the changes in the capacities of infrastructure and service demands. A recent trend in infrastructure modeling is to develop detailed digital twins to mimic all aspects of the real infrastructure. However, detailed digital twins might require data not readily available, and their analyses often have prohibitive computational costs, making digital twins not always the most suitable option to model the performance of infrastructure. The goal of selecting the optimal modeling granularity is to allocate computational resources to the model that best delivers the desired information with the desired accuracy level. This paper presents a mathematical formulation to systematically select the appropriate modeling granularity of infrastructure. The formulation adaptively increases the granularity starting from a low-granularity infrastructure model until we reach the desired tradeoff among accuracy, simplicity, and computational efficiency. To define the tradeoff, we introduce metrics that measure the level of agreement between estimates of the quantities of interest computed using different levels of granularity. Such metrics include global measures that assess if a model is insufficiently detailed to capture the quantities of interest and local measures that identify specific regions of the model that may require further refinement. As an example, we apply the illustrated formulation to select the granularity of the potable water infrastructure model in Seaside, Oregon, to quantify its performance following a seismic event.

Keywords: Digital twin, Infrastructure, Model selection, Network granularity, Resilience

1. Introduction

Decision-making is often based on information about risks associated with possible courses of action or possible outcomes (Gardoni and Murphy 2014). Regional risk analysis requires defining all consequences relevant to the decision-making process and quantifying their probabilities (Gardoni et al. 2016). One of the first steps in regional risk analysis is developing representative mathematical models of infrastructure, i.e., a virtual representation of infrastructure for the intended analyses. One of the main challenges in developing a virtual representation of infrastructure is selecting the modeling resolution from multiple candidates for required analyses, each candidate having different computational costs, accuracy, and possibly the ability to provide information.

Past research (e.g., Guidotti et al. 2019; Sharma et al. 2021) has shown the importance of selecting the proper modeling resolution of infrastructure to adequately capture the impact of damaging events. Sharma et al. (2021) defined three types of resolution of infrastructure: the temporal, hierarchical, and spatial resolution. The temporal resolution refers to how often the model estimates the performance of infrastructure over time, while the hierarchical resolution determines the level of detail used to model infrastructure components. The spatial resolution, i.e., the *modeling granularity*, determines the level of detail in the topology of the model.

The selection of the temporal resolution should

be based on the frequency of changes in the infrastructure performance, such as hourly changes in water infrastructure performance. The hierarchical resolution should reflect the scope of the infrastructure performance assessment, such as the need for higher resolution in optimization problems targeted at minimizing the impact of damaging events on infrastructure performances by adopting mitigation and adaptation strategies. Finally, the spatial resolution (modeling granularity) should consider the scope of the infrastructure performance assessment, variability of hazard impact, and computational cost.

Infrastructure modeling has recently seen a shift towards using detailed digital twins that mimic all aspects of the infrastructure (e.g., Shirowzhan et al. 2020). However, the high computational costs and excessive level of detail often make digital twins unsuitable for modeling infrastructure performance. On the other hand, past research (e.g., Tomar and Burton 2021) defined simplified infrastructure models at an arbitrary granularity level without properly considering the appropriateness of the modeling granularity for the intended analyses. To ensure that computational resources are allocated optimally, the selection of modeling granularity should aim to provide the desired information at the desired accuracy level.

This paper illustrates a mathematical formulation to select the appropriate modeling granularity of infrastructure. First, we discuss a formulation to define equivalent simplified infrastructure models at different levels of granularity. Next, we formulate the selection of the appropriate modeling granularity as an iterative process. Starting from a coarser granularity, we obtain equivalent simplified infrastructure models at increasingly refined levels of granularity and estimate the quantities of interest using the simplified infrastructure models. We refine the modeling granularity until we achieve an acceptable balance between accuracy, simplicity, and computational efficiency. To assess the level of accuracy, we introduce metrics that measure the level of agreement between estimates of the quantities of interest computed using different levels of granularity. Such metrics include global measures that assess if a model is sufficiently detailed to capture the quantities of interest and local measures that identify specific regions of the model that may require further refinement. As an example, we apply the illustrated formulation to select the granularity of the potable water infrastructure model in Seaside, Oregon, to quantify its performance following a seismic event.

2. Modeling the time-varying performance of infrastructure

For completeness, this section briefly reviews the mathematical formulation for modeling the time-varying performance of infrastructure using graph theory. Following Sharma and Gardoni (2022), we model infrastructure as a collection of networks, each representing a specific function of the infrastructure (e.g., a network can describe the connectivity and physical damage of the infrastructure, and a flow network can describe its functionality). The collection of all networks is written as $\mathcal{G} = \{G^{[k]} = (V^{[k]}, E^{[k]}) : k = 1, \ldots, K\}$, where superscript $[k]$ denotes the function captured by the k^{th} network. The state of each network is characterized by a unique set of (i) capacity measures $\mathbf{C}^{[k]}(t)$, e.g., Gardoni et al. (2002); (ii) demand measures $\mathbf{D}^{[k]}(t)$, e.g., Gardoni et al. (2003); and (iii) supply measures $\mathbf{S}^{[k]}(t)$, e.g., Sharma and Gardoni (2022). In general, network measures are a function of dynamic state variables $\mathbf{x}^{[k]}(t)$, where the temporal dependence accounts for deterioration/aging processes (e.g., Jia and Gardoni 2018) and recovery activities (e.g., Sharma et al. 2020). Using the triplet $[\mathbf{C}^{[k]}(t), \mathbf{D}^{[k]}(t), \mathbf{S}^{[k]}(t)]$, we derive the general expression for the performance measures of the components of $G^{[k]}$ as $\mathbf{Q}^{[k]}(t) = \mathbf{Q}^{[k]}[\mathbf{C}^{[k]}(t), \mathbf{D}^{[k]}(t), \mathbf{S}^{[k]}(t)]$.

In addition, we model the interdependencies within and among infrastructure using network interfaces as proposed in Sharma and Gardoni (2022). An interface is defined as a boundary where networks interact. Therefore, interdependencies among networks can only exist at the interfaces. Following Sharma and Gardoni (2022), we can use interface functions to obtain the modified capacity, demand, and supply estimates $[\mathbf{C}'^{[k]}(t), \mathbf{D}'^{[k]}(t), \mathbf{S}'^{[k]}(t)]$. We can then obtain the modified derived performance $\mathbf{Q}'^{[k]}(t)$.

Next, we define an aggregate measure of the network performance $Q(t)$ derived from the component performances $\mathbf{Q}'^{[k]}(t)$. The modeled infrastructure and the type of analysis typically shape the definition of $Q(t)$. For example, in the case of transportation infrastructure, we can estimate $Q(t)$ from topology-based approaches (e.g., Nocera et al. 2019), whereas in the case of potable water infrastructure, we can estimate $Q(t)$ from flow-based approaches (e.g., Iannacone et al. 2022). Next, we divide the region of interest into n_a service areas and obtain $Q(t)$ by mapping $\mathbf{Q}'^{[k]}(t) \rightarrow Q_a^{[k]}(t),\ \forall a \in (1,\ldots,n_a)$. Lastly, we aggregate the $Q_a^{[k]}(t)$'s into a scalar $Q(t) = \sum_{a=1}^{n_a}[w_a Q_a^{[k]}(t)]$, where w_a denotes the weight for each area a.

Once we obtain $Q(t)$, we use descriptors, i.e., $\mathfrak{R}[Q(t)]$, that succinctly characterize the consequences (i.e., effects or impacts) of damaging events on the infrastructure. Examples of $\mathfrak{R}[Q(t)]$ include resilience metrics, or the percentage of the population impacted by the water pressure below a specified value. Likewise, if the time-varying performance of infrastructure is of interest $\mathfrak{R}[Q(t)]$ can be defined as $Q(t)$ itself.

3. Formulation to define equivalent simplified networks

This section presents the mathematical formulation to define equivalent simplified networks at different levels of granularity, building upon the mathematical formulation for modeling the time-varying infrastructure performance presented in Section 2. Defining equivalent simplified networks encompasses defining (i) the nodes of the equivalent network; (ii) the edges of the equivalent network; and (iii) equivalent state variables, capacity, and demand measures.

The definition of the nodes of equivalent networks (i.e., $V_{\text{eq}}^{[k]}$) is based on the topology of an initial detailed network (i.e., $G^{[k]}$). Starting from $G^{[k]}$, we rank the nodes $V^{[k]}$ based on a nodal measure of importance. Then, we define the nodes $V_{\text{eq}}^{[k]}$ (of the equivalent network $G_{\text{eq}}^{[k]}$) as the subset of nodes (of $G^{[k]}$) whose measure of importance is greater than a selected threshold. Mathematically, we write the nodal measure of importance $\mathcal{H}(v_i^{[k]}), \forall v_i^{[k]} \in V^{[k]}$ as

$$\mathcal{H}\left(v_i^{[k]}\right) = h\left(v_i^{[k]}\right) + \mathbf{1}_{\left\{v_i^{[k]} \in V_d^{[k]}\right\}} M \qquad (1)$$

where $h(v_i^{[k]})$ is the baseline importance of a node $v_i^{[k]} \in V^{[k]}$; $\mathbf{1}_{\{v_i^{[k]} \in V_d^{[k]}\}}$ is an indicator function equal to 1 if $v_i^{[k]} \in V_d^{[k]}$, and 0 otherwise; and M is a number large enough, which can be defined, for example, as the maximum value among $h(v_i^{[k]}), \forall v_i^{[k]} \in V^{[k]}$. In Equation (1), $h(v_i^{[k]})$ is defined as a node centrality measure, like diameter (Latora and Marchiori 2001), betweenness (Freeman 1977), or PageRank (Brin and Page 1998). Also, $V_d^{[k]}$ is the subset of nodes $v_i^{[k]} \in V^{[k]}$ having a relevant role when assessing the performance of the modeled infrastructure. For example, considering a water network, nodes representing water tanks or reservoirs belong to $V_d^{[k]}$ because these network components are essential for modeling purposes. Once we quantify $\mathcal{H}(v_i^{[k]})$, we define an importance threshold, ε_{eq}, to select the nodes in $V_{\text{eq}}^{[k]}$, such that $V_{\text{eq}}^{[k]} := \{v_i^{[k]} \in V^{[k]} | \mathcal{H}(v_i^{[k]}) \geq \varepsilon_{\text{eq}}\}$. Furthermore, a single value of ε_{eq} would imply a homogeneous granularity of $G_{\text{eq}}^{[k]}$ However, the formulation also allows defining $\boldsymbol{\varepsilon}_{\text{eq}}$ as a vector to obtain a heterogeneous granularity by assigning each component of $\boldsymbol{\varepsilon}_{\text{eq}}$ to different regions within the footprint of the infrastructure. For instance, we can define a higher level of granularity within the region of interest, whereas we can have a lower modeling granularity outside the region of interest.

Next, we define the edges of $G_{\text{eq}}^{[k]}$ by exploring the paths having as origin and destination the nodes in $V_{\text{eq}}^{[k]}$. Then, we add an edge to the set of edges of the equivalent network $E_{\text{eq}}^{[k]}$ when a path does not require passing through multiple nodes in $V_{\text{eq}}^{[k]}$. Lastly, we verify that the number of incoming/outgoing edges from a node in $G_{\text{eq}}^{[k]}$ does not exceed the number of incoming/outgoing edges in $G^{[k]}$. Otherwise, we remove the least important newly added edges. The selection of which outgoing edges from $v_i^{[k]}$ need to be removed is also based on a centrality measure (of the edges) in $G_{\text{eq}}^{[k]}$. However, if removing the least important edge results in a disconnected network, we add

the removed edge and remove the second to last important edge. If removing the second to last important edge results in a disconnected network, we keep adding the removed edge and remove an edge with less importance.

Once we obtain the topology of $G_{\text{eq}}^{[k]} = (V_{\text{eq}}^{[k]}, E_{\text{eq}}^{[k]})$, we define a mapping to appropriately approximate $G^{[k]}$ and obtain an equivalent $G_{\text{eq}}^{[k]}$. The equivalency depends on the modeled infrastructure, type of analysis, and the selected $\mathfrak{R}[Q(t)]$. The mapping translates state variables, capacity, and demand from $G^{[k]}$ into $G_{\text{eq}}^{[k]}$. For instance, in the case of hydraulic flow analysis for the water infrastructure, reducing $V^{[k]}$ into $V_{\text{eq}}^{[k]}$ requires assigning the water service demand to $V_{\text{eq}}^{[k]}$. For this case, we assign the water demand of $v_i^{[k]} \in V^{[k]}$ to the nearest $v_i^{[k]} \in V_{\text{eq}}^{[k]}$.

4. Formulation to select the modeling granularity of infrastructure

This section presents the mathematical formulation to select the appropriate modeling granularity. In the following subsections, first, we explain the use of an equivalent network for modeling the time-varying performance of infrastructure. Then, we explain the iterative formulation to select the appropriate modeling granularity. Finally, we introduce the metrics that facilitate the decision on the level of modeling granularity.

4.1. *Modeling the infrastructure performance using equivalent simplified networks*

Once we obtain an equivalent $G_{\text{eq}}^{[k]}$, we quantify $\mathbf{S}'^{[k]}(t, \boldsymbol{\varepsilon}_{\text{eq}})$ and $\mathbf{Q}'^{[k]}(t, \boldsymbol{\varepsilon}_{\text{eq}})$ as discussed in Section 2. Also, we map $\mathbf{Q}'^{[k]}(t, \boldsymbol{\varepsilon}_{\text{eq}}) \rightarrow Q_a^{[k]}(t, \boldsymbol{\varepsilon}_{\text{eq}}), \forall a \in [1, \ldots, n_a(\boldsymbol{\varepsilon}_{\text{eq}})]$, and we aggregate the $Q_a^{[k]}(t, \boldsymbol{\varepsilon}_{\text{eq}})$'s into a scalar $Q(t, \boldsymbol{\varepsilon}_{\text{eq}}) = \sum_{a=1}^{n_a(\boldsymbol{\varepsilon}_{\text{eq}})} [w_a(\boldsymbol{\varepsilon}_{\text{eq}}) Q_a^{[k]}(t, \boldsymbol{\varepsilon}_{\text{eq}})]$, where $w_a(\boldsymbol{\varepsilon}_{\text{eq}})$ is the weight for each a for a given level of granularity. The estimated $Q(t, \boldsymbol{\varepsilon}_{\text{eq}})$ (and the resulting $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}})]$) are a function of $\boldsymbol{\varepsilon}_{\text{eq}}$ because the ability to capture the spatial variability of $\mathbf{S}'^{[k]}(t, \boldsymbol{\varepsilon}_{\text{eq}})$ depend on the modeling granularity.

4.2. *Iterative formulation to select the granularity of networks*

We formulate the selection of the modeling granularity as an iterative process. First, we obtain samples of $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}})]$, which can be used to estimate a measure of the error when searching the appropriate granularity of $G^{[k]}$. The estimate of $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}})]$ can be computed either using $G^{[k]}$ (denoted as $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}}^{+})]$) or an initial low-granularity $G_{\text{eq}}^{[k]}$ (i.e., $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq},0}^{-})]$) and then incrementally higher granularity networks (i.e., $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq},u}^{-})]$), where subscript u denotes the u^{th} iteration in the search of the appropriate granularity of $G^{[k]}$.

After we estimate $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}}^{+})]$ (or $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq},0}^{-})]$), the search of the appropriate granularity of $G^{[k]}$ begins from a lower level of granularity by choosing an initial value of $\boldsymbol{\varepsilon}_{\text{eq}}$ (e.g., selected based on a predetermined value of the cardinality of $V_{\text{eq}}^{[k]}$.) For instance, $\boldsymbol{\varepsilon}_{\text{eq}}$ could be selected so that $|V_{\text{eq}}^{[k]}| = \lfloor \boldsymbol{\alpha}_1 |V^{[k]}| \rfloor$, where $\alpha_1 \in (0, 1)$. Next, we iteratively refine the model granularity until we achieve the desired tradeoff among accuracy, simplicity, and computational efficiency. At each iteration, we increase the model granularity by updating the cardinality of $V_{\text{eq}}^{[k]}$, i.e., by updating the value of $\boldsymbol{\alpha}_u$ at the $u+1$ iteration.

4.3. *Global and local metrics to estimate the accuracy of networks at different levels of granularity*

In this section, we introduce the metrics that measure the accuracy of $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}})]$ and facilitate the decision on the refinement of the granularity of $G_{\text{eq}}^{[k]}$. Such metrics include global measures that assess if $G_{\text{eq}}^{[k]}$ is insufficiently detailed to capture $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}})]$ and local measures that identify specific regions of $G_{\text{eq}}^{[k]}$ that may require further refinement.

4.3.1. *Global metrics*

Let us consider $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}}^{-})]$ and $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}}^{+})]$ as a general time-invariant $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}})]$ estimated using two levels of granularity, where $\mathfrak{R}[Q(t, \boldsymbol{\varepsilon}_{\text{eq}}^{-})]$ is obtained using a lower granularity of $G^{[k]}$.

Then, let $\mathfrak{D}$ be defined as

$$\mathfrak{D} = \left(\mathbf{I} - \mathbf{I}\right) \cdot \begin{pmatrix} \mathfrak{R}\left[Q\left(t, \varepsilon_{\text{eq}}^{+}\right)\right] \\ \mathfrak{R}\left[Q\left(t, \varepsilon_{\text{eq}}^{-}\right)\right] \end{pmatrix} \tag{2}$$

where $\mathfrak{D}$ is the difference between $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}}^{+})]$ and $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}}^{-})]$; $(\mathbf{I} - \mathbf{I})$ is the matrix composed of the two submatrices $\mathbf{I}$ and $-\mathbf{I}$; and $\mathbf{I}$ is the identity matrix of size equal to the number of components in $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}})]$. Similarly, let $\mathfrak{M}$ be defined as

$$\mathfrak{M} = \frac{1}{2}\left(\mathbf{I}\ \mathbf{I}\right) \cdot \begin{pmatrix} \mathfrak{R}\left[Q\left(t, \varepsilon_{\text{eq}}^{+}\right)\right] \\ \mathfrak{R}\left[Q\left(t, \varepsilon_{\text{eq}}^{-}\right)\right] \end{pmatrix} \tag{3}$$

where $\mathfrak{M}$ is the average between $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}}^{+})]$ and $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}}^{-})]$. In the case of time-invariant $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}})]$, we combine the information from $\boldsymbol{\mu}_{\mathfrak{D}}$, i.e., the mean vector of $\mathfrak{D}$ in Equation (2), and $\|\text{diag}(\boldsymbol{\Sigma}_{\mathfrak{D}\mathfrak{M}})\|$, i.e., the Euclidean norm of the diagonal matrix of the cross-covariance matrix $\boldsymbol{\Sigma}_{\mathfrak{D}\mathfrak{M}} = E[(\mathfrak{D} - \mu_{\mathfrak{D}})(\mathfrak{M} - \mu_{\mathfrak{M}})^{\mathrm{T}}]$. In the case the granularity of $G_{\text{eq}}^{[k]}$ is sufficiently detailed, then $\boldsymbol{\mu}_{\mathfrak{D}} \rightarrow \mathbf{0}$ and $\|\text{diag}(\boldsymbol{\Sigma}_{\mathfrak{D}\mathfrak{M}})\| \rightarrow \mathbf{0}$. We write the *global Network-Resolution-Root-Mean-Square-Error*, $\mathbf{GN}$, as

$$\mathbf{GN} = \sqrt{\left(\boldsymbol{\mu}_{\mathfrak{D}}\right)^2 + \|\text{diag}\left(\boldsymbol{\Sigma}_{\mathfrak{D}\mathfrak{M}}\right)\|} \tag{4}$$

It follows that $\mathbf{GN} \rightarrow \mathbf{0}$ as the granularity of $G_{\text{eq}}^{[k]}$ becomes appropriate to capture the changes in $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}})]$.

Similarly, in the case of a time-varying $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}})]$, we define $\mathfrak{D}_t$ with mean $\boldsymbol{\mu}_{\mathfrak{D}_t}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathfrak{D}_t\mathfrak{D}_t}$ in a similar fashion as $\mathfrak{D}$ in Equation (2). Likewise, we also define the $\mathfrak{M}_t$ with mean $\boldsymbol{\mu}_{\mathfrak{M}_t}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathfrak{M}_t\mathfrak{M}_t}$. As in the time-invariant case, if $G_{\text{eq}}^{[k]}$ is sufficiently detailed to capture the changes in $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}})]$, then $\sum_{t=1}^{\tau} \boldsymbol{\mu}_{\mathfrak{D}_t} \rightarrow \mathbf{0}$ (where τ is the time-span of the performance assessment of infrastructure) and $\sum_{t=1}^{\tau} \|\text{diag}(\boldsymbol{\Sigma}_{\mathfrak{D}_t\mathfrak{M}_t})\| \rightarrow \mathbf{0}$. In the case of time-varying $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}})]$, we combine the information from $\mu_{\mathfrak{D}_t}$ and $\|\text{diag}(\boldsymbol{\Sigma}_{\mathfrak{D}_t\mathfrak{M}_t})\|$ to assess the accuracy of $G_{\text{eq}}^{[k]}$ at different levels of granularity. We write the *temporal global Network-Resolution-Root-Mean-Square-Error*, $\mathbf{TGN}$ as

$$\mathbf{TGN} = \sqrt{\sum_{t=1}^{\tau}\left[\left(\mu_{\mathfrak{D}_t}\right)^2 + \|\text{diag}\left(\boldsymbol{\Sigma}_{\mathfrak{D}_t\mathfrak{M}_t}\right)\|\right]} \tag{5}$$

4.3.2. *Local metrics*

The use of only global metrics to estimate the accuracy of networks at different levels of granularity does not guarantee the selection of the appropriate model granularity because of the local spatial variability in $\mathfrak{R}[Q(t, \varepsilon_{\text{eq}})]$. Similarly, global metrics alone would not suffice if decision-makers desire increased granularity in selected spatial portions of the network. We also introduce local metrics to estimate the accuracy of networks at different levels of granularity. The use of local metrics allows controlling the accuracy of networks at different levels of granularity at the service area level. The principle behind the local metrics of accuracy is the same as the one illustrated for $\mathbf{GN}$ and $\mathbf{TGN}$. However, because multiple nodes in $G^{[k]}$ correspond to the same node in $G_{\text{eq}}^{[k]}$, we define the local metrics of accuracy at the level of the service area by defining the quantities $\mathfrak{D}_a$ and $\mathfrak{M}_a$, $\forall a \in [1, \ldots, n_a(\varepsilon_{\text{eq}}^{+})]$ in a similar fashion as $\mathfrak{D}$ and $\mathfrak{M}$ in Equations (2) and (3). Using the defined quantities $\mathfrak{D}_a$ and $\mathfrak{M}_a$, we then estimate the *local Network-Resolution-Root-Mean-Square-Error*, $\mathbf{LN}$ and the *local temporal Network-Resolution-Root-Mean-Square-Error*, $\mathbf{TLN}$ that can be written like $\mathbf{GN}$ and $\mathbf{TGN}$ in Equations (4) and (5).

5. Selecting the modeling granularity of an example water infrastructure

This section illustrates the mathematical formulation to select the network granularity for modeling the water infrastructure in Seaside, Oregon. We use a set of 100 hypothetical earthquake scenarios originating from the Cascadia Subduction Zone. Then, for each earthquake scenario, we use Ground Motion Prediction Equations (Boore and Atkinson 2008) to obtain values of the Peak Ground Acceleration (PGA) and Peak Ground Velocity (PGV) at the location of the vulnerable components of the water infrastructure (i.e.,

pumping stations, tanks, and pipelines). We model the water infrastructure as a collection of two networks, i.e., the structural network $G^{[1]}$ (describing the connectivity and the physical damage to the water infrastructure) and the flow network $G^{[2]}$ (describing the functionality in terms of transmission of water.) Following the mathematical formulation presented in Section 2, $G^{[1]}$ has state variables $\mathbf{x}^{[1]}(t)$ that include material and geometric properties of pipelines, water tanks, and water pumps. Also, $G^{[1]}$ has structural capacity $\mathbf{C}^{[1]}(t)$, and demand $\mathbf{D}^{[1]}(t)$, where the elements of $\mathbf{C}^{[1]}(t)$ and $\mathbf{D}^{[1]}(t)$ are the capacity and demand models associate with the network components. We define the supply $\mathbf{S}^{[1]}(t)$ as $\mathbf{S}^{[1]}(t) = \mathbf{D}^{[1]}(t) \odot \mathbf{1}_{\{\mathbf{D}^{[1]}(t) \preceq \mathbf{C}^{[1]}(t)\}}$, where $\odot$ is the elementwise product of the elements of $\mathbf{D}^{[1]}(t)$ and $\mathbf{1}_{\{\mathbf{D}^{[1]}(t) \preceq \mathbf{C}^{[1]}(t)\}}$, $\mathbf{1}_{\{\mathbf{D}^{[1]}(t) \preceq \mathbf{C}^{[1]}(t)\}}$ is an indicator function equal to 1 when $\mathbf{D}^{[1]}(t) \preceq \mathbf{C}^{[1]}(t)$, and 0 otherwise, and $\preceq$ is the elementwise inequality. Next, we write $\mathbf{Q}^{[1]}(t)$ as the instantaneous reliability of each of the components (Sharma et al. 2021). Similarly, the water flow network $G^{[2]}$ has state variables $\mathbf{x}^{[2]}(t)$ that include flow pressure, the length and diameter of the pipelines, their roughness coefficient, as well as pump curves, and tank characteristics. Also, $G^{[2]}$ has capacity $\mathbf{C}^{[2]}(t)$, and demand $\mathbf{D}^{[2]}(t)$, where the elements of $\mathbf{C}^{[2]}(t)$ and $\mathbf{D}^{[2]}(t)$ are defined in terms of the volumetric water flow. In the case of $G^{[2]}$, we define the supply $\mathbf{S}^{[2]}(t)$ as the solution of a pressure-driven hydraulic flow-based analysis (obtained using the Python package *WNTR* (Klise et al. 2017)). Next, we write $\mathbf{Q}^{[2]}(t)$ as $\mathbf{Q}^{[2]}(t) = [\mathbf{S}^{[2]}(t) \oslash \mathbf{D}^{[2]}(t)] \odot \mathbf{1}_{\{\mathbf{D}^{[2]}(t) \succ 0\}}$, in which $\oslash$ is the elementwise division and $\succ$ is the elementwise comparison operator. We obtain the overall performance measure $Q(t)$ by dividing the region into n_a areas. Lastly, we use $\mathfrak{R}[Q(t)] = Q(t_{0^+})$ where t_{0^+} is the time in the immediate aftermath of the earthquake. The selected $\mathfrak{R}[Q(t)]$ quantifies the direct impact of the hazard on the ability of the infrastructure to fulfill the water service demand.

Next, we define equivalent simplified networks at different levels of granularity. We use the PageRank centrality as $h(v_i^{[2]})$, measuring the likelihood of passing through a node in a random walk over $G^{[2]}$. Then, $V_d^{[2]}$ includes the nodes representing reservoirs, water tanks, and water pumps. Using $M = 4.81 \times 10^{-4}$, we compute $\mathcal{H}(v_i^{[2]})$ in Equation (1). We obtain the topology of $G_{\text{eq}}^{[2]}$ with the lowest level of granularity using the value of ε_{eq} corresponding to $|V_{\text{eq}}^{[2]}| = \lfloor \alpha_1 |V^{[2]}| \rfloor$, where $\alpha_1 = 0.10$. Once we obtain the topology of $G_{\text{eq}}^{[2]}$, we estimate the equivalent $\mathbf{x}^{[2]}(t)$. Following Walski et al. (2003), we obtain equivalent length, diameter, and roughness coefficient for the pipelines in $G_{\text{eq}}^{[2]}$. Similarly, we aggregate the nodal values of the water service demand from $v_i^{[2]} \in V^{[2]}$ to the nearest (equivalent) $v_i^{[2]} \in V_{\text{eq}}^{[2]}$, $\forall v_i^{[2]} \in V^{[2]}$.

Then, for every earthquake scenario, we estimate $Q(t_{0^+}, \varepsilon_{\text{eq}})$ by translating the physical damage into the changes in $\mathbf{C}^{[2]}(t, \varepsilon_{\text{eq}})$ and $\mathbf{D}^{[2]}(t, \varepsilon_{\text{eq}})$ to estimate $\mathbf{S}^{[2]}(t, \varepsilon_{\text{eq}})$ (details in Iannacone et al. 2022).

Next, we quantify the changes in $Q(t_{0^+}, \varepsilon_{\text{eq}})$ when ε_{eq} correspond to $|V_{\text{eq}}^{[2]}| = \lfloor \alpha_2 |V^{[2]}| \rfloor$, where $\alpha_2 = 2\alpha_1$. After quantifying $Q(t_{0^+}, \varepsilon_{\text{eq}})$ with $G_{\text{eq}}^{[2]}$ corresponding to the values of α_1 and α_2, we obtain the value of α_3 as $\alpha_3 = \alpha_2 + \lambda_1 \cdot E[\partial \Xi(\alpha)/\partial \alpha]$, where $\lambda_1 = 4.24$, and $\Xi(\alpha)$ is the measure of the error. We estimate $E[\partial \Xi(\alpha)/\partial \alpha]$ as

$$E\left[\frac{\partial \Xi(\alpha)}{\partial \alpha}\right] = \frac{E\left[Q\left(t_{0^+}, \varepsilon_{\text{eq},u+1}\right) - Q\left(t_{0^+}, \varepsilon_{\text{eq}}^{+}\right)\right]}{\alpha_{u+1} - \alpha_u} - \frac{E\left[Q\left(t_{0^+}, \varepsilon_{\text{eq},u}\right) - Q\left(t_{0^+}, \varepsilon_{\text{eq}}^{+}\right)\right]}{\alpha_{u+1} - \alpha_u} \tag{6}$$

The value of λ_1 is selected as $1/10$ of the value of λ_1 that would result in $|V_{\text{eq}}^{[2]}| = |V^{[2]}|$ in one step. Based on λ_1 and $E[\partial \Xi(\alpha)/\partial \alpha]$, we obtain the value of $\alpha_3 = 0.29$. At each iteration, we update the value of λ_u as $1/10$ of the maximum value of λ_u at the u^{th} iteration. We estimate $\mathfrak{D} = Q(t_{0^+}, \varepsilon_{\text{eq}}^{+}) - Q(t_{0^+}, \varepsilon_{\text{eq},u})$ and $\mathfrak{M} = [Q(t_{0^+}, \varepsilon_{\text{eq}}^{+}) + Q(t_{0^+}, \varepsilon_{\text{eq},u})]/2$ using Equations (2) and (3), and the value of GN using Equation (4). At the first iteration, $GN = 0.05$. For illustration purposes, we assume the desired level of accuracy corresponds to $GN = 0.02$, and we repeat this process until the value of $GN \leq 0.02$. Figure 1 shows the selected modeling granularity

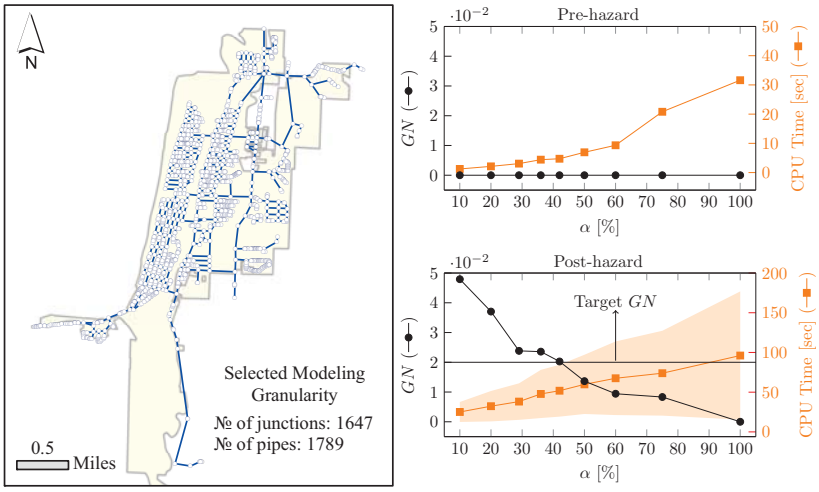

Fig. 1. Selected topology of $G_{\text{eq}}^{[2]}$, changes in the global metric of accuracy and CPU time at each iteration (adapted from Nocera and Gardoni (2022))

of $G_{\text{eq}}^{[2]}$ (corresponding to $\alpha = 0.42$) and the values of GN at each iteration of the iterative search. Also, we show the changes in GN at each u^{th} iteration and the mean CPU time per run (with its confidence band) for the pre-hazard scenario and the 100 simulated earthquake scenarios. In Figure 1, we show the values of GN and the mean CPU time per run also for values of $\alpha > 0.42$. As the modeling granularity increases, the results show a relatively small increase in accuracy at a high computational cost. The estimates of the mean time per run indicate the savings in the computational cost if $G_{\text{eq}}^{[2]}$ would be used in fully probabilistic analyses. The results show that using the selected modeling granularity would save (on average) 54% of CPU time per run in fully probabilistic analyses.

6. Conclusions

Accurate risk analysis requires representative mathematical models of infrastructure. One of the main challenges in developing mathematical models of infrastructure is selecting the modeling granularity. Different modeling granularities may affect the ability to capture the spatial variability of the impact arising from the changes in the ability of critical infrastructure to fulfill service demands. This paper presented a mathematical formulation to select the granularity for the modeling of infrastructure to obtain the quantities of interest with the desired level of accuracy. First, we obtained equivalent simplified infrastructure models using estimates of topological characteristics of a detailed infrastructure model. Then, we formulated the selection of the model granularity as an iterative process. The formulation adaptively increases the granularity starting from a low-granularity infrastructure model until we reach the desired tradeoff among accuracy, simplicity, and computational efficiency. To define the tradeoff, we introduced metrics that measure the level of agreement between estimates of the quantities of interest computed using different levels of granularity. As an example, we applied the formulation to select the granularity of the potable water infrastructure model in Seaside, Oregon, to quantify its performance following a seismic event. The example showed that using the presented formulation would save significant CPU time per run in fully probabilistic analyses with a prescribed level of accuracy.

References

Boore, D. M. and G. M. Atkinson (2008). Ground-motion prediction equations for the average horizontal component of pga, pgv, and 5%-

damped psa at spectral periods between 0.01 s and 10.0 s. *Earthquake Spectra 24*(1), 99–138.

Brin, S. and L. Page (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems 30*(1-7), 107–117.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35–41.

Gardoni, P., A. Der Kiureghian, and K. M. Mosalam (2002). Probabilistic capacity models and fragility estimates for reinforced concrete columns based on experimental observations. *Journal of Engineering Mechanics 128*(10), 1024–1038.

Gardoni, P., K. M. Mosalam, and A. Der Kiureghian (2003). Probabilistic seismic demand models and fragility estimates for rc bridges. *Journal of Earthquake Engineering 7*(spec01), 79–106.

Gardoni, P. and C. Murphy (2014). A scale of risk. *Risk analysis 34*(7), 1208–1227.

Gardoni, P., C. Murphy, and A. Rowell (2016). *Risk analysis of natural hazards: Interdisciplinary challenges and integrated solutions*. Springer.

Guidotti, R., P. Gardoni, and N. Rosenheim (2019). Integration of physical infrastructure and social systems in communities' reliability and resilience analysis. *Reliability Engineering & System Safety 185*, 476–492.

Iannacone, L., N. Sharma, A. Tabandeh, and P. Gardoni (2022). Modeling time-varying reliability and resilience of deteriorating infrastructure. *Reliability Engineering & System Safety 217*, 108074.

Jia, G. and P. Gardoni (2018). State-dependent stochastic models: A general stochastic framework for modeling deteriorating engineering systems considering multiple deterioration processes and their interactions. *Structural Safety 72*, 99–110.

Klise, K. A., M. Bynum, D. Moriarty, and R. Murray (2017). A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study. *Environmental Modelling & Software 95*, 420–431.

Latora, V. and M. Marchiori (2001). Efficient behavior of small-world networks. *Physical Review Letters 87*(19), 198701–1–198701–4.

Nocera, F. and P. Gardoni (2022). Selection of the modeling resolution of infrastructure. *Computer-Aided Civil and Infrastructure Engineering 37*(11), 1352–1367.

Nocera, F., A. Tabandeh, R. Guidotti, J. Boakye, and P. Gardoni (2019). Physics-based fragility functions: Their mathematical formulation and use in the reliability and resilience analysis of transportation infrastructure. In P. Gardoni (Ed.), *Routledge handbook of sustainable and resilient infrastructure*, pp. 239–262. Routledge.

Sharma, N. and P. Gardoni (2022). Mathematical modeling of interdependent infrastructure: An object-oriented approach for generalized network-system analysis. *Reliability Engineering & System Safety 217*, 108042.

Sharma, N., F. Nocera, and P. Gardoni (2021). Classification and mathematical modeling of infrastructure interdependencies. *Sustainable and Resilient Infrastructure 6*(1-2), 4–25.

Sharma, N., A. Tabandeh, and P. Gardoni (2020). Regional resilience analysis: A multiscale approach to optimize the resilience of interdependent infrastructure. *Computer-Aided Civil and Infrastructure Engineering 35*(12), 1315–1330.

Shirowzhan, S., W. Tan, and S. M. Sepasgozar (2020). Digital twin and CyberGIS for improving connectivity and measuring the impact of infrastructure construction planning in smart cities. *ISPRS International Journal of Geo-Information 9*(4).

Tomar, A. and H. V. Burton (2021). Active learning method for risk assessment of distributed infrastructure systems. *Computer-Aided Civil and Infrastructure Engineering 36*(4), 438–452.

Walski, T. M., D. V. Chase, D. A. Savic, W. Grayman, S. Beckwith, and E. Koelle (2003). *Advanced water distribution modeling and management*. Haestad press.