

Spatial-Temporal Deep Learning for Hosting Capacity Analysis in Distribution Grids

Jiaqi Wu[✉], *Student Member, IEEE*, Jingyi Yuan[✉], *Student Member, IEEE*,
Yang Weng[✉], *Senior Member, IEEE*, and Raja Ayyanar[✉], *Fellow, IEEE*

Abstract—The widespread use of distributed energy sources (DERs) raises significant challenges for power system design, planning, and operation, leading to wide adaptation of tools on hosting capacity analysis (HCA). Traditional HCA methods conduct extensive power flow analysis. Due to the computation burden, these time-consuming methods fail to provide online hosting capacity (HC) in large distribution systems. To solve the problem, we first propose a deep learning-based problem formulation for HCA, which conducts offline training and determines HC in real time. The used learning model, long short-term memory (LSTM), implements historical time-series data to capture periodical patterns in distribution systems. However, directly applying LSTMs suffers from low accuracy due to the lack of consideration on spatial information, where location information like feeder topology is critical in nodal HCA. Therefore, we modify the forget gate function to dual forget gates, to capture the spatial correlation within the grid. Such a design turns the LSTM into the Spatial-Temporal LSTM (ST-LSTM). Moreover, as voltage violations are the most vital constraints in HCA, we design a voltage sensitivity gate to increase accuracy further. The results of LSTMs and ST-LSTMs on feeders, such as IEEE 34-, 123-bus feeders, and utility feeders, validate our designs.

Index Terms—Hosting capacity, deep learning, data-driven method, long short-term memory (LSTM), spatial-temporal correlation, distributed energy resource.

I. INTRODUCTION

NOWADAYS, a growing number of renewable energy-based distributed energy sources (DERs) have been used in the low-voltage distribution grid, e.g., photovoltaics (PVs). The widespread use of DERs brings many advantages, including voltage profile control, line loss reduction, and cost decrease. Meanwhile, challenges appear when the distribution grid is turned into an active grid. This is because DERs inevitably change load shapes, voltages, fault current profiles, etc., when the penetration level is substantial. For example, grid operators face over-voltages at solar-rich feeders instead of low voltage issues like the past. Specifically, solar energy

often peaks at valley load for communities with lots of roof-top PVs. The net load then turns negative and introduces reverse power flow and over-voltages at the lateral end, causing the malfunction of voltage regulators and protection coordination. Without proper awareness of potential capacity to accommodate PVs, grid planners and operators find it challenging to handle capacity planning and take appropriate control actions, which leads to hosting capacity analysis (HCA) to assess the distribution grid for further operation [1].

Specifically, HCA is to determine the value of the hosting capacity (HC). HC is denoted as the maximum active power that can be injected by DERs at a bus in an existing distribution grid without causing technical problems or requiring changes to power system facilities. [2], [3], [4], [5] provide systematic and comprehensive introductions to the research, development, evaluation, and enhancement of HC. Traditional HCA methods typically conduct power flow analysis for a baseline feeder model and modified scenarios with different load profiles, DER penetration, and other uncertainties caused by the environment, power equipment, and human activities [6]. Then, these methods implement different techniques to check the violation conditions of operation constraints and thus quantify the HC [7], [8].

Related works [2], [3], [4], [5] classified current HCA methods into the deterministic (worst-case), stochastic, streamlined, and iterative Integration Capacity Analysis (ICA) methods. The deterministic method [9], [10], [11] and the stochastic method [12], [13], [14], [15], [16], [17] focus on specific scenario(s) without considering time correlation. The deterministic method obtains an optimal solution on a single scenario, usually the worst-case scenario(s), while a stochastic method uses the probabilistic technique, e.g., Monte Carlo simulation, to model the uncertainties in the power system. However, they cannot capture the relation between the system variables over time, which is important in HCA [18], [19], [20], [21]. Differently, using historical time-series data, the streamlined method [22] and the iterative ICA [23], [24] method provide insight into how hosting capacity changes over time and the ability to derive a corresponding hosting capacity pattern subjected to time. To determine the HC, the streamlined method applies a set of equations and algorithms to evaluate power system criteria at each node, whereas the ICA method iteratively increases the DERs at each node until system violations occur. Though they can consider time-related scenarios and increase the calculation accuracy of the time-varying power system model, each separate step is a complex calculation, e.g., iterative optimal functions. Therefore, the

Manuscript received 8 November 2021; revised 17 February 2022 and 5 June 2022; accepted 5 July 2022. Date of publication 8 August 2022; date of current version 22 December 2022. This work was supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) through the Solar Energy Technologies Office under Award DE-EE0008773. Paper no. TSG-01794-2021. (*Corresponding author: Yang Weng.*)

The authors are with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: jiaqiwu1@asu.edu; jyuan46@asu.edu; yang.weng@asu.edu; rayyanar@asu.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TSG.2022.3196943>.

Digital Object Identifier 10.1109/TSG.2022.3196943

calculation burden is still the limit of traditional HCA for real-time tasks.

Since we aim to calculate HC in real time with high accuracy, the previous HCA methods cannot meet the requirements. Instead, we propose a machine learning-based problem formulation [25]. This formulation uses historical time-series data to conduct offline training and obtain the HC value based on real-time system conditions. Specifically, we model HCA as a supervised learning problem that uses data of power system features and operating conditions as input vector and HC data as target label. To achieve such a learning target, the mapping from historical data on different input features to the HC is highly nonlinear, and deep learning is a promising method to deal with the non-linearity. In order to capture the periodic pattern of the power system, e.g., hourly, daily, and yearly patterns, we use the recurrent neural network (RNN) as the basic learning framework [26], [27], [28]. In such a model, RNNs need some past context to predict the current output, but in practice, it can hardly capture the relationship in a long sequence. One of its variants, the long short-term memory (LSTM), has a much better performance of long-term learning [29]. While LSTM can improve our deep learning framework, the direct application on HCA still has some challenges.

Though the basic LSTM framework can capture the time-varying impacts in HCA, the impact of spatial information cannot be embedded directly. Such information includes the locations of nodes, current DERs, and potential DERs in the feeder. Ignoring these spatial relationships and simply relying on deep learning will limit the accuracy in our analysis [30], [31], [32]. Because of the flexibility of the deep neural network (DNN) and our motivation to consider both spatial and temporal correlation for HCA, we develop the basic LSTM to the Spatial-Temporal LSTM (ST-LSTM) [33], [34], [35], [36], [37]. We have two major contributions to achieve this goal. First, we modify the structure of LSTM cells. The most crucial LSTM design for capturing temporal correlation is the gate function. Reference [38] proposed the forget gate for the first time, and [29] emphasized that the forget gate is the most critical gate among all the gates. Therefore, we modified the forget gate to dual forget gates, which allows the model to transfer temporal and spatial memory in parallel. Second, to make the model perform well in our HC determination work, we design a sensitivity gate to use the voltage sensitivity data. The voltage sensitivity data is a particular data for HCA because it is related to the voltage violations constraint, which is one of the most critical limits in HCA [39]. Therefore, we design the voltage sensitivity gate to improve the accuracy further. This gate can work with the input gate, which is also an essential gate in the LSTM cell [40], to filter the new inputs.

We validate our new model using the IEEE 34-bus feeder, IEEE 123-bus feeder, and an Arizona generic utility feeder. The ST-LSTM has a much better performance for these feeders than the temporal or spatial LSTM models. Moreover, the designed sensitivity gate using the voltage sensitivity data in our LSTM model can improve the accuracy.

The remainder of this paper is organized as follows. Section II proposes the problem formulation. Section III

illustrates the design of the dual forget gates and the sensitivity gate. Section IV provides the numerical validation using different models. Conclusion and discussions are in Section V.

II. PROBLEM FORMULATING

As our target is to provide real-time hosting capacity for each bus, we are not focusing on one single scenario. Different from the static snapshot hosting capacity, we first formulate a data-driven machine learning problem model for dynamic hosting capacity (DHC) analysis, which will help the distribution system operators enhance hosting capacity determination, facilitate optimal control, and dispatch of DERs in real time. Specifically, this can be achieved by integrating this function as a module in the data platform. The information of the HC values allows the other modules in the data platform, specifically the module on coordinated control of DERs, to determine which regions/buses are operating near their HC limit, how to control the DERs at these regions and around it, to mitigate HC violations such as overvoltage. This can be achieved by a cloud-based platform called end-to-end solar energy optimization platform (e-SEOP) that receives these data and performs multiple analytics, including dynamic HCA described in this paper and real-time control of DERs through the network of edge intelligent devices (EIDs) [41].

A. Per-Bus Hosting Capacity Analysis

The Per-bus Hosting Capacity Analysis determines the hosting capacity values for each bus, which are per-bus HC values. We use a two-bus example with current injection at both buses of $(p1, p2) = (0, 2)$ to introduce the per-bus HC values. To determine the HC value for bus 1, we gradually increase the injection by bus1 and keep all the other settings the same, including the injection by bus 2. If $(p1, p2) = (1, 2)$ is the maximum value without causing problem, the HC number of bus 1 is $1 - 0 = 1$. To determine the HC value for bus 2, we also keep all the other settings the same and only change the injection by bus 2. If $(p1, p2) = (0, 3.5)$ is the maximum value without causing problem, the HC number of bus 2 is $3.5 - 2 = 1.5$. Therefore, the HC numbers for both buses are $(hc1, hc2) = (1, 1.5)$. In this case, we have the per-bus hosting capacity under this scenario setting. Then, we can use the same metric for other scenarios and finally obtain time-series HC results for these time-series scenarios.

The per-bus hosting capacity values for each bus within a feeder are drastically different based on the impedances, short circuit ratios, and loads at the transformers. To determine per-bus HC values, traditional methods need to solve optimal functions or do iterative HCA to check the system violations for each bus respectively. Unlike these time-consuming methods, our learning-based HCA is to find a mapping rule between the power flow data and the hosting capacity data. With the offline trained model, the online prediction process can compute the per-bus HC values in real time.

B. Problem Formulation

With power flow data and per-bus hosting capacity data, we formulate our hosting capacity analysis (HCA) as a regression

problem. Our deep learning-base model has offline training and online prediction. In the training, the model uses the historical time-series data from system setups. Specifically, the power flow data from power flow analysis is the training input, including the voltage magnitude, voltage angle, load profiles, and PV profiles, etc. Moreover, the per-bus HC data is the desired output, which is also generated from simulation. Since the data from the simulation is required by our training, the detailed feeder settings for power flow analysis are also essential. Our learning model can capture the HC-related information among these data. Subsequently, we can apply the well-trained learning model to calculate the HC value according to the new inputs of the system in real time.

Our deep learning-based HCA will not calculate the per-bus HC values independently. Instead, we use two data sequences for better HC determination, considering power systems' temporal and spatial correlations. Specifically, the temporal sequence, which is the historical time-series data, can allow the model to learn the periodic pattern of the power system and obtain the HC changes over time [21]. The spatial sequences the topology of the distribution network. The per-bus HC values of the two adjacent buses could be different because of different load profiles, PV profiles, etc. Also, the constraints when determining the hosting capacity are highly related to the spatial relationships of buses. For example, the new PV may introduce the reverse current or the over-voltage violation at an upstream bus. Therefore, the spatial sequences provide a physical model embedding to improve the accuracy [42].

To capture the spatial configuration and temporal dynamics, and determine the HC with high accuracy, both spatial and temporal information needs to follow well-formed orders. Our proposed ST-LSTM model has two dimensions, which are the temporal and spatial sequences. The temporal sequence can be derived from the time-series models. However, for the spatial information, it is difficult to put all the buses in the complex distribution grid into a single sequence. A random traversal in the network will result in wrong spatial correlation. Thus, to convert the network into a spatial sequence, we propose traversing the network with paths. This method divides the distribution grid into different paths, from the slack bus or three-phase main trunk to each lateral end. These paths to the lateral ends will cover all the feeder buses, and we call it the longest paths method. Due to the fewer buses in one path, the dimensionality of the input data will also decrease, which will reduce the computation burden. In the following, we formally define the problem of learning-based hosting capacity analysis.

C. Mathematical Problem Definition

Mathematically, in the offline training, we define input tensor X as the values of all the time-series power flow data from time 0 to time T . Moreover, n is the number of input features of each bus, and m is the number of buses. X is a $\mathbb{R}^{T \times n \times m}$ tensor and $x_{s,i}(t)$ is the value of the i -th power system feature component of bus s at time-series t . $X(t) \in \mathbb{R}^{n \times m}$ denotes the feature values of all the buses at time t . Similarly, the output H is a $\mathbb{R}^{T \times 1 \times m}$ tensor, which represents the calculated HC

values. $H(t)$ will denote a slice of H at time t and $h_s(t)$ is the HC value of bus s .

With the historical data, we want to learn a regression model $f : X \rightarrow H$ to predict the future $H(\tau)$ based on $X(\tau)$. This mapping is achieved by updating the parameters of our neural network model. In our training, we decomposed the input tensor $X \in \mathbb{R}^{T \times n \times m}$ into slices. Each slice $X(t) \in \mathbb{R}^{n \times m}$ represents the input data at time series t . In our recurrent Spatial-Temporal LSTM model, at each time series step, the input is one slice of the data tensor.

The problem with using the deep learning-based model to calculate the HC value is defined as follows.

The offline training process:

- Given: a sequence of time-series historical power flow data $X(t)$, $t = 1, \dots, T$, where $x_{s,i}(t)$ is the value of the i -th power system feature component of bus s at time t .
- Target: learn the mapping $f : X \rightarrow H$ that outputs the time-series HC value $H(t)$, $t = 1, \dots, T$, where $h_s(t)$ is the HC value of bus s at time t .

The online testing and forecasting process:

- Given: new historical or online time-series power flow data $X(\tau)$, where $x_{s,i}(\tau)$ is the value of the i -th power system feature component of bus s at time τ .
- Find: use the trained model to find the time-series HC value $H(\tau)$, where $h_s(\tau)$ is the HC value of bus s at time τ .
- Online update: use the new time-series historical or online power flow data $X(\tau)$ to update the model $f : X \rightarrow H$ and have a better prediction performance.

III. SPATIAL-TEMPORAL LONG SHORT-TERM MEMORY

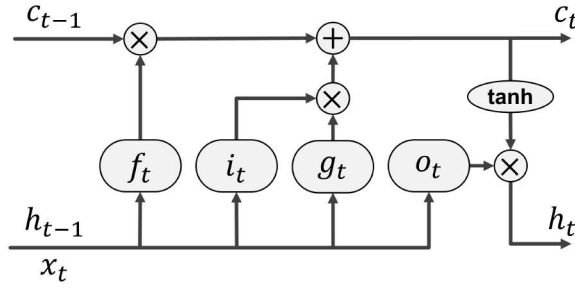
With the formulated problem, we will illustrate the proposed deep learning method, the Spatial-Temporal LSTM, for data-driven HCA. The basic LSTM model is first introduced to consider temporal correlation. While the basic LSTM model cannot consider spatial and temporal correlations simultaneously in the power system, we modified the forget gate in the LSTM cell to dual forget gates to process the temporal and spatial information in parallel. Moreover, another sensitivity gate is designed to boost the HCA accuracy with critical factors: the voltage violations constrain in the HCA.

A. Illustration of Basic LSTM

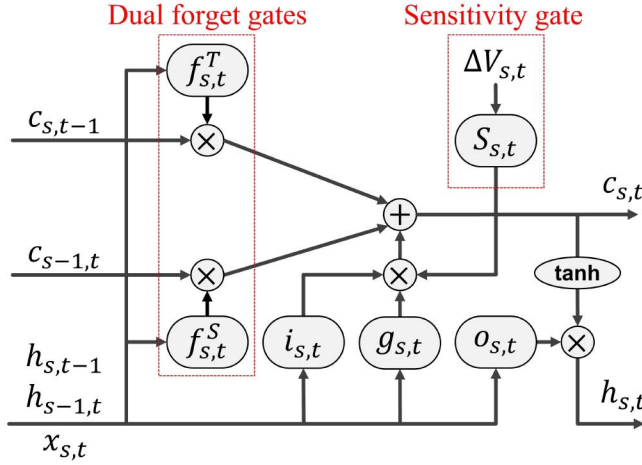
To consider the correlation over time in learning, we use the Recurrent Neural Network (RNN), which is designed for sequential information. An RNN is a series of identical feed-forward neural networks, one unit for each time or step, which are known as 'RNN cells' [43], [44]. The input of an RNN cell has two parts. One is the output from the last cell, which is hidden state h_{t-1} , and the other is the input vector for this cell, which is x_t . In this way, an RNN can use its internal state (memory) to process sequences of inputs. By transferring the information through these repeating cells, the network can be trained recurrently. Each RNN cell will have an output h_t as

$$h_t = \phi(W([h_{t-1}, x_t]) + b), \quad (1)$$

where $\phi(\cdot)$ is the activation function to extract non-linearity.



(a) The basic LSTM cell.



(b) The ST-LSTM cell.

Fig. 1. The comparison between cell structures of the basic LSTM and the proposed ST-LSTM. In our ST-LSTM cell design, the red rectangle boxes are the dual forget gates and the sensitivity gate.

Though RNNs are designed to process sequential data, it has drawbacks for HCA. One is that RNNs cannot transfer information correctly between two cells with relatively long distances. To solve this problem, LSTMs improve basic RNNs via gate functions, which are capable of learning short-term and long-term dependencies [45]. Such gate design also avoids gradient exploding and vanishing, which are another significant drawback of RNNs [45]. Therefore, LSTMs have become an effective model for sequential prediction problems.

Different from the basic recurrent unit, each LSTM unit, which is shown in Fig. 1(a), use the cell state c_t to memorize the long-term information and the hidden state h_t to capture the short-term information, where the RNNs only use the hidden state h_t . To control the information stream in the model, the designed three gates recurrently update the c_t and the h_t , which is the most important design of the LSTM. The functions of a single cell are

$$i_t = \sigma(W_i([h_{t-1}, x_t]) + b_i), \quad (2)$$

$$f_t = \sigma(W_f([h_{t-1}, x_t]) + b_f), \quad (3)$$

$$g_t = \tanh(W_g([h_{t-1}, x_t]) + b_g), \quad (4)$$

$$o_t = \sigma(W_o([h_{t-1}, x_t]) + b_o), \quad (5)$$

where i_t , f_t and o_t represent the input, forget, and output gates of the t -th cell, respectively. In order to handle long-term

dependencies in data, the input gate works together with the candidate g_t to decide what information should be input and the forget gate decides what information should be forgotten. Additionally, the output gate decides what information should be output. These three gates use the sigmoid function, which is represented as $\sigma(\cdot)$, as the activation function. This nonlinear function $\sigma(\cdot)$ can map the input values between 0 to 1, where 1 means to keep the information while 0 means to abandon the information. Each of the gates and the candidate is a small neural network to learn functional mapping. h_{t-1} is the hidden output vector of the last cell and x_t represents the input vector. Specifically, x_t is the power flow data in our problem formulation. Besides, W_i , W_f , W_o and W_c are the weight matrices and b_i , b_f , b_o and b_c are corresponding biases to parameterize the functional gates and candidate.

Since we have the output of all the gates, the next step is to use them to update the cell state c_t and the hidden state h_t . The c_t updating function is

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t. \quad (6)$$

The former term is the previous cell state c_{t-1} controlled by the forget gate f_t , which is to decide how much long-term memory should be forgotten. The latter term is the mapped new input candidate value controlled by the input gate, which decides how much new input should be added. The two terms of this function will work together to update the cell state c_t using the element-wise (Hadamard) product \odot . Subsequently, the h_t is denoted as

$$h_t = o_t \odot \tanh(c_t), \quad (7)$$

where the h_t is the hidden state that needs to be transferred to the next cell. In this way, the LSTM model can follow a sequence to conduct training. The output h_t of each cell is the result we want to obtain, which is the hosting capacity value in our problem.

B. Temporal and Spatial Sequence in HCA

A basic LSTM is assertive when processing temporal sequential data, e.g., the historical time-series power flow data. Thus, this model can learn the HC changes over time with respect to different system conditions like loads, inverter set-points, etc. However, a power system is a complex network where buses will have mutual effects with each other. We expect the model to be coupled or coordinated because we are looking for violations not only at the bus under consideration but anywhere in the feeder due to the power injection at this bus. An overvoltage violation may occur at an upstream bus even though the bus where the power is injected does not see overvoltage. Therefore, the connectivity relationship is crucial in HCA, and ignoring the spatial correlation will cause low accuracy.

Since an LSTM is trained as a simple chain, and one cell can only receive information from the last cell, this recurrent network calculates both the c_t and the h_t according to one sequential data series. Thus, the LSTM network is unable to process two data sequences of different dimensions simultaneously. In order to consider both temporal and spatial

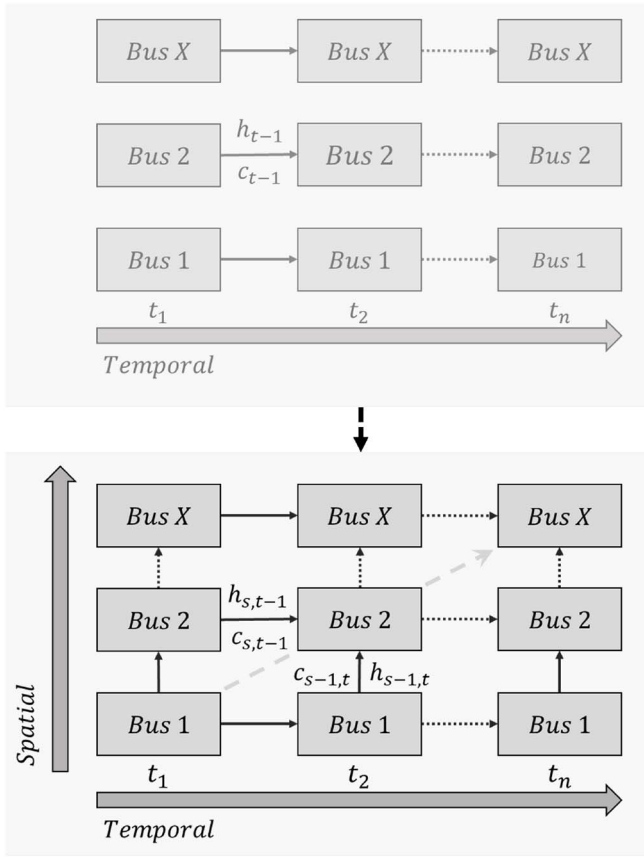


Fig. 2. The comparison between the structures of the basic LSTM and the ST-LSTM. The top sub-figure is the structure of the basic LSTM model, and the bottom sub-figure is the structure of the ST-LSTM model. Compared with the basic LSTM model, the ST-LSTM model can capture the spatial correlation among buses.

correlation for HC, we embed the topology information of the power system by modifying the LSTM to the Spatial-Temporal LSTM. The spatial sequences we used are the paths in the network, and each of the paths is from the feeder head to the lateral end.

In Fig. 2 and Fig. 3, we use the comparisons of the structures and information flow between the basic LSTM and ST-LSTM to show how cells transfer c_t and the h_t in the ST-LSTM model. The cell design to embed temporal and spatial correlation is detailed in the next subsection.

C. Design of the Dual Forget Gates

Specifically, to process temporal and spatial sequences simultaneously, we reconstruct the structure of the cells as shown in Fig. 2. Though the forget gate is critical in the basic LSTM, the single gate cannot process information from two dimensions separately. Therefore, we modified this gate to dual forget gates. The dual forget gates are designed to decide what information the cell should memorize from these two dimensions. Consequently, one new cell can receive the cell states and hidden states from the last temporal cell and the last spatial cell separately. Mathematically, we design the dual forget gates and modify the other functional gates from (2), (3), (4),

and (5) accordingly as follows,

$$i_{s,t} = \sigma(W_i([h_{s,t-1}, h_{s-1,t}, x_t]) + b_i), \quad (8)$$

$$f_{s,t}^T = \sigma(W_{f_t}([h_{s,t-1}, h_{s-1,t}, x_t]) + b_{f_t}), \quad (9)$$

$$f_{s,t}^S = \sigma(W_{f_s}([h_{s,t-1}, h_{s-1,t}, x_t]) + b_{f_s}), \quad (10)$$

$$g_{s,t} = \tanh(W_c([h_{s,t-1}, h_{s-1,t}, x_t]) + b_c), \quad (11)$$

$$o_{s,t} = \sigma(W_o([h_{s,t-1}, h_{s-1,t}, x_t]) + b_o), \quad (12)$$

where $f_{s,t}^T$ and $f_{s,t}^S$ are the dual forget gates. $f_{s,t}^T$ represents the temporal forget gate based on temporal information, while $f_{s,t}^S$ is the spatial forget gate based on spatial information. Besides, $h_{s,t-1}$ is the hidden state from the last time step $t-1$ while the $h_{s-1,t}$ is the hidden state from the last space step $s-1$. The designed dual forget gates are highlighted in Fig. 1(b).

To decide which temporal information needs to be memorised, we multiply the cell state $c_{s,t-1}$ element-wise by our temporal-based forget gate $f_{s,t}^T$. Similarly, the element-wise multiplication of the cell state $c_{s-1,t}$ and spatial-based forget gate $f_{s,t}^S$ is to determine which spatial information to keep/leave. Thus, the new cell state is

$$c_{s,t} = f_{s,t}^T \odot c_{s,t-1} + f_{s,t}^S \odot c_{s-1,t} + i_{s,t} \odot g_{s,t}. \quad (13)$$

The design of the dual forget gates causes the cell state c_t and the hidden state h_t of this cell to be impacted by both the temporal forget gate and the spatial forget gate. In this case, the ST-LSTM model can capture the temporal and spatial correlation in parallel.

D. Design of the Sensitivity Gate

Previous input features come from power flow analysis and lack specific HC-related information. However, this information cannot be directly added because it covers information over different locations. To improve the performance of the ST-LSTM, we design a new gate, namely the sensitivity gate, to decide which information should be input. The sensitivity gate, collaborating with the input gate, use the voltage sensitivity information to control the input when adding new information into the cell state.

During hosting capacity analysis, we notice that the limit of voltage violations constraint plays a crucial role in HCA, which is similarly concluded by [5]. It is to check the over-voltage on the other buses after adding PV generators at one selected bus. Specifically, [39] illustrates that voltage increase at load bus bars is a severely limiting factor when installing DERs. Based on this, we use this voltage sensitivity data as the input of the sensitivity gate. The voltage sensitivity data is the voltage changes of the other buses when adding a unit number of generation at one bus.

The voltage sensitivity is considered over buses, where the locations of nodes lead to different impacts on the other buses. To embed the impact based on locations, we use the average sensitivity data. We define the average voltage sensitivity of bus s as the number of the average voltage change of the other buses on the path it belongs to when adding a generator at bus s . Our simulation found that the maximum bus voltage change of one path always happens at the lateral end. Additionally, the voltage change of the neighbor nodes has a

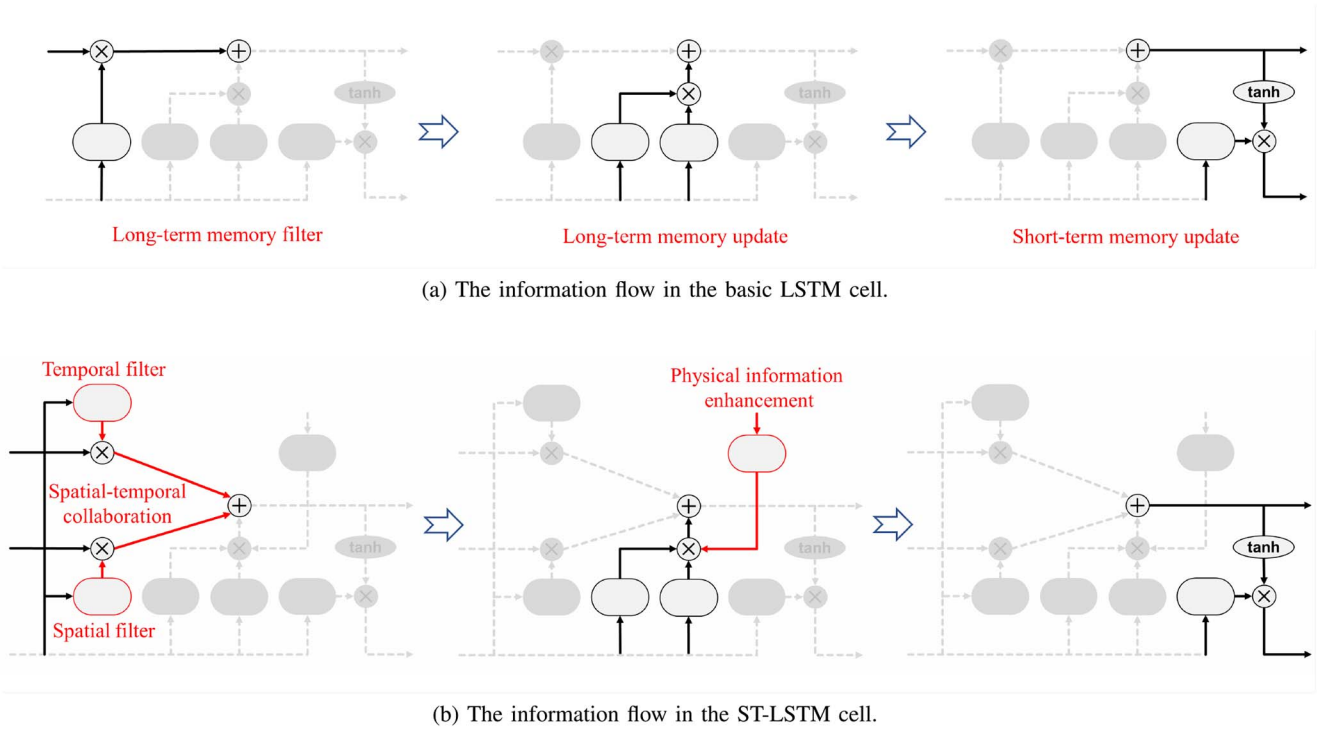


Fig. 3. The comparison between the information flow of the basic LSTM and the ST-LSTM.

higher error than the average bus voltage change. Thus, we use the average bus voltage change to reflect the impact brought by each individual bus on the entire path.

Moreover, we can observe that, when adding a generator at bus s , the voltage change of the buses from the same path will be higher than the buses from the other paths, which shows the independence of paths. Thus, this weaker interactive impact between two paths supports our design of the longest paths method.

Mathematically, the sensitivity gate function is

$$S_{s,t} = \sigma(W_d \Delta V_{s,t} + b_d), \quad (14)$$

where $\Delta V_{s,t}$ is the vector of voltage sensitivity data and it will be mapped using the weight matrix. In this gate, we also use the sigmoid function as our activation function, which can extract non-linear impact of voltage sensitivity on HC. The designed sensitivity gate is highlighted in Fig. 1(b).

Consequently, the cell state function is updated from (13) to

$$c_{s,t} = f_{s,t}^T \odot c_{s,t-1} + f_{s,t}^S \odot c_{s-1,t} + i_{s,t} \odot S_{s,t} \odot g_{s,t}. \quad (15)$$

Since we do not modify the output gate, the function of $h_{s,t}$ will be remained as

$$h_{s,t} = o_{s,t} \odot \tanh(c_{s,t}). \quad (16)$$

In summary, to capture the temporal and spatial correlation, the designed dual forget gates can allow the ST-LSTM model to process information in these two dimensions in parallel. To improve the model's accuracy, we embed the physical information by designing the sensitivity gate, which can consider the voltage violations constraint in HCA. Furthermore, to achieve the goal of online prediction, we keep using new time-series data to update the model.

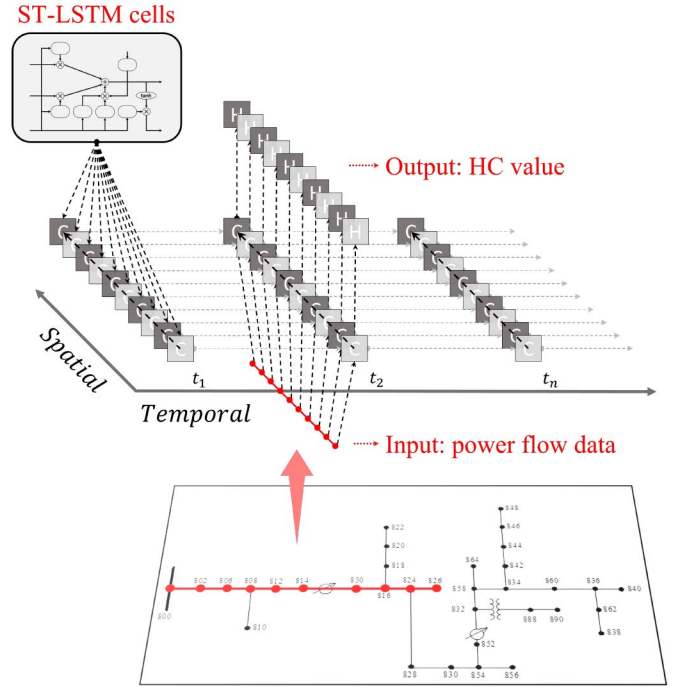


Fig. 4. The detailed illustration of the ST-LSTM structure using one path (the red line in this figure) of the IEEE 34-bus feeder at t_2 as an example. This path is from the substation to the lateral end. Each box with the character 'C' represents the ST-LSTM cell for bus s at time t_n . Each box with the character 'H' represents the calculated HC value for bus s at time t_n .

IV. NUMERICAL RESULTS

The proposed Spatial-Temporal LSTM model for hosting capacity analysis is validated extensively on various test cases.

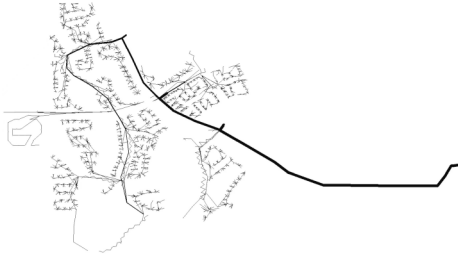


Fig. 5. The Arizona generic utility feeder.

This section illustrates the results on test distribution grids, including the IEEE 123-bus feeder and a utility high penetration feeder. This utility feeder, shown in Fig. 5, is an actual 12.47 kV, 9 km-long Arizona utility feeder with 3.8 MW of residential roof-top PV installed, leading to a penetration level of more than 200% (3.8 MW/1.6 MW) as compared to the feeder total gross load during peak PV production hours [46]. To simplify the calculation, we assume that all equipped PVs at residential customers are integrated into the secondary side of each distribution transformer. The simplified model of this feeder contains 2100 bus nodes and 371 distribution transformers [47].

We will validate the performance of the designed dual forget gates and the sensitivity gates via the comparison with baseline models, which are the temporal sequence LSTM and the spatial sequence LSTM. These comparison results will support our design.

A. Dataset Preparation

To prepare training and testing data for validation, we use CYME to generate time-series data by conducting power flow analysis and hosting capacity analysis. CYME is a power engineering simulation software by EATON which can be used to analyze power systems [48]. In CYME, the Load Flow module can provide power flow analysis, and the Integration Capacity Analysis module can calculate the HC. Specifically, we use the Integration Capacity Analysis module in CYME software to calculate the hosting capacity as part of the training data set [49], [50]. This module can set different constraints, including voltage violation constraints, thermal loading constraints, etc. Then, the software will progressively add PV and run power flow simulations until violations of one or more operation standards appear. Our learning-based model is to find a mapping rule from the power flow data to the HC values for each bus. Since these different constraints already limit the simulated HC datasets we used in our model, our learning-based model can naturally consider the constraints. We have time-series models of the IEEE 123-bus feeder and the Arizona utility high penetration feeder in CYME.

We use the result of Load Flow analysis as our input vector. Specifically, we use the voltage magnitude, voltage angles, load profiles, and PV profiles, etc. The h_t of each cell is the HC value. We use the min-max normalization method to preprocess the input vector. For each feature in the input vector, we convert the minimum and maximum values to 0 and 1, respectively. All other values are converted to a decimal number

between 0 and 1. For the HC data, we directly use a selected constant to scale the value. This linearity scale method allows us to retrieve the output to the expected HC value easily.

To use the voltage sensitivity data on one selected path, first, we add a generator at one bus. Then, we calculate the average voltage change at the other buses on this path. This average voltage change vector is the input of the sensitivity gate in the ST-LSTM cell. In this way, the sensitivity gate can capture the overall impact of each bus on this path.

B. Experiment Setups

1) *Baseline Methods*: As shown in Fig. 2, the basic LSTM model can only flow one type of sequential information, either temporal or spatial sequence. Thus, we use the basic LSTM as the baseline model to test the performance using either temporal or spatial sequences.

For temporal sequence LSTM, we have time-series models. In each training step, we input the power flow data from power flow analysis of all the time-series at one bus and can have the output as the HC data vector for all the time series at this bus. For spatial sequence LSTM, we use different paths in the feeder, all from the slack bus to the lateral ends. Then, we collect the data of one path for each time and put them into the ST-LSTM model. In this way, we can obtain the HC data vector of all the buses at this time. For our ST-LSTM model, we make a similar setting to the spatial sequence LSTM. The difference is that we use the outputs from the last time as part of the input each time. In the IEEE 123-bus feeder, we have 12 time-series models, 8 of which are the training set and 4 of which are the testing set. The utility feeder has 24 time-series models where 16 are training sets and 8 are testing sets.

2) *Proposed ST-LSTM (Longest Paths Method for Spatial Correlation)*: For the IEEE example feeders, we will show the result of the IEEE 123-bus feeder. The longest path in the IEEE 123-bus feeder is the red path shown in Fig. 6. This path contains 24 buses and two critical regulators. Node 67 on this path is one of HCA's most critical nodes. Besides the longest path, different paths will compose the whole network. Because the IEEE 123-bus feeder does not have long lateral branches, we can add PV generators on the three-phase main trunk, we have paths from the slack bus to each lateral end. Since we can obtain the HC value of all the buses on the trained path, theoretically, we can find a finite number of paths to cover the whole feeder. In this feeder, we have 39 paths and calculate the HC value of all buses in the feeder. The computed results may be slightly different on the buses from the overlapped paths. To reduce this overfitting risk, we use a boosted method which takes the average of the computed value from different paths. Compared to the IEEE feeder, the utility feeder has fewer overlapped paths, and the paths are from the three-phase main trunk to each lateral end. We divide the whole network into 5 different zones and use a depth-first search (DFS) based script to find about 6 paths in each zone. In this way, we can obtain the HC value of all the buses.

3) *Proposed ST-LSTM (Online Update for Better Performance)*: The proposed ST-LSTM is a regression model to find a mapping rule from historical power flow data

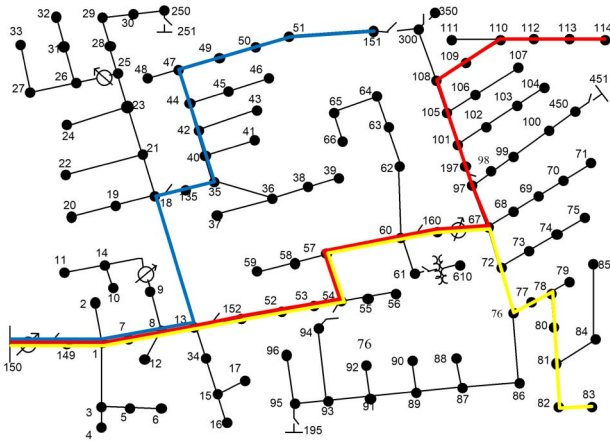


Fig. 6. Three examples of paths in the IEEE 123-bus feeder. The red path is the longest path.

TABLE I
COMPARISON AMONG THE TWO BASIC LSTM MODELS AND THE ST-LSTM MODEL ON THE LONGEST PATH OF THE IEEE 123-BUS FEEDER

Test Scenario	MSE	Percentage Error
LSTM (Temporal sequence)	0.124	252%
LSTM (Spatial sequence)	0.29	22%
ST-LSTM (Without sensitivity gate)	0.0025	5%

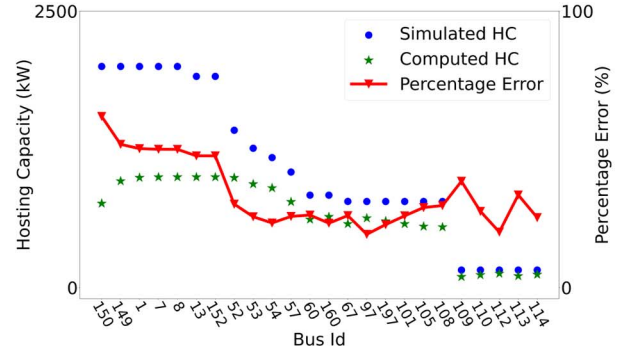
to the hosting capacity data. After we conduct the offline training and get a well-trained model, we use the computed mapping rule to determine the HC value according to new power flow data. However, along with the continuous changes of the load profiles, PV profiles, climate, facility settings, etc., the trained model may not have a good prediction after a long time. To solve this problem for better performance, we continue to update our model using the new power flow data and corresponding HC data.

4) *Performance Evaluation Metrics*: We use two criteria to evaluate the performance of the models. The first is the Mean square error (MSE) criteria, and the second is the percentage accuracy. The model uses the MSE to conduct back-propagation, and the percentage accuracy can give an intuitive comparison between the simulated HC and the calculated HC. Specifically, simulated HC is the simulation result from CYME, and the calculated HC is the output of the deep learning models.

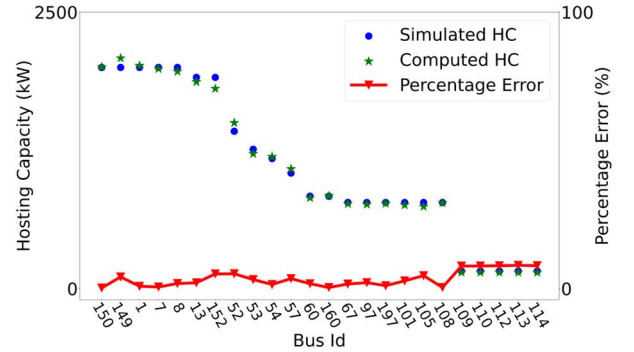
C. Accuracy of ST-LSTM With Multi-Dimension Information

To validate the performance of our ST-LSTM with the designed dual forget gates, this section shows the results of the longest path in the IEEE 123-bus feeder using different models, which are the temporal sequence LSTM, the spatial sequence LSTM, and the ST-LSTM models.

As shown in Table I, the temporal sequence LSTM has the worst performance. HC is highly correlated to spatial information, while the temporal sequence LSTM lacks this crucial information. Compared with the spatial sequence LSTM, the ST-LSTM model has lower MSE and percentage error. Fig. 7 shows the comparisons between these two models at time t , which is time-series 11.



(a) Results of the longest path of the IEEE 123-bus feeder at time-series 11 using spatial LSTM model.



(b) Results of the longest path of the IEEE 123-bus feeder at time-series 11 using ST-LSTM model.

Fig. 7. Computation results of the longest path in the IEEE 123-bus feeder at time-series 11 using different models. The simulated HC values are from software simulation, and the computed HC values are from the learning-based models. The ST-LSTM model has a better performance compared to the basic spatial LSTM model in the IEEE 123-bus feeder.

TABLE II
COMPARISON BETWEEN THE BASIC ST-LSTM MODEL AND THE GCN-BASED ST-LSTM MODEL IN THE IEEE 123-BUS FEEDER

	Basic ST-LSTM	GCN-based ST-LSTM
Percentage Error	25.2%	66.1%

Remark 1: For the ST-LSTM model, we need to find a method to build a spatial sequence and capture the structural relationship for simulation. In computer science, graph neural network (GNN) is a popular way to embed the topology information [51]. Therefore, we implement the GNN in the ST-LSTM to compare with the proposed longest path method for spatial correlation. Specifically, we add a graph convolutional network (GCN), which is a representative type of the GNNs, layer before the input of the ST-LSTM cell [51]. This GCN layer is built based on the adjacency matrix, mapping the original input power flow data to output. However, the direct model combination did not improve the learning performance in the IEEE 123-bus feeder. In the simulation, we use the order of bus numbers as the spatial sequence. The results of the basic ST-LSTM model and GCN-based ST-LSTM model are shown in Table II. Neither of the models has a sensitivity gate to consider spatial correlation as proposed. Thus, we proposed the longest paths method to

TABLE III

COMPARISON BETWEEN THE SPATIAL SEQUENCE LSTM MODEL AND THE ST-LSTM MODEL ON OTHER PATHS OF THE IEEE 123-BUS FEEDER

Test Scenario	Path 2	Path 3
LSTM (Spatial sequence)	28.5%	45.9%
ST-LSTM (Without sensitivity gate)	6.5%	6.8%

TABLE IV

COMPARISON OF PERCENTAGE ERROR BETWEEN THE SPATIAL LSTM AND ST-LSTM IN THE UTILITY FEEDER

Test Scenario	Zone 1	Zone 2	Zone 3
LSTM (Spatial sequence)	9.3%	12.5%	18.3%
ST-LSTM (Without sensitivity gate)	7.2%	12%	12.2%

divide the feeder network into paths for a better learning performance.

D. Accuracy of the ST-LSTM on Different Paths

To prove that the high accuracy of the ST-LSTM is not an accidental experimental result, this section shows the results of other paths in the IEEE 123-bus feeder using the spatial sequence LSTM and the ST-LSTM models.

We choose another two paths in the IEEE 123-bus feeder. In Fig. 6, the yellow and blue paths represent Path 2 and Path 3 in the IEEE 123-bus feeder, respectively.

The numerical results comparison between the spatial sequence LSTM and the ST-LSTM of path 2 and path 3 are shown in Table III. We can see, compared with the spatial sequence LSTM, the ST-LSTM model has a significantly smaller percentage error on both of the two paths. That is to say, the ST-LSTM model will increase the accuracy of HC calculation for different paths in the IEEE 123-bus feeder.

E. Practicality of the ST-LSTM

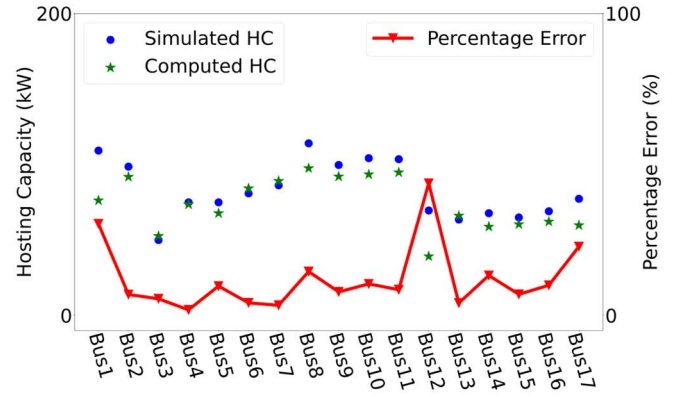
This section will show the accuracy of the ST-LSTM in a utility feeder, which will further verify the practicality of the ST-LSTM in HCA.

The network we used is a generic Arizona utility high penetration feeder. We divide this utility feeder into different zones. The division of the zones was primarily based on the location of the buses and where those buses are connected into the three-phase main trunk. In each zone, we also use a script to find all the paths. In this way, we can obtain the HC value for all the buses. The comparison is shown in Table IV. The accuracy of the ST-LSTM is higher compared with the basic LSTM as well. The detailed results of time t , which is time-series 16, of one of the paths using the spatial sequence LSTM and the ST-LSTM models are shown in Fig. 8(a) and Fig. 8(b) respectively.

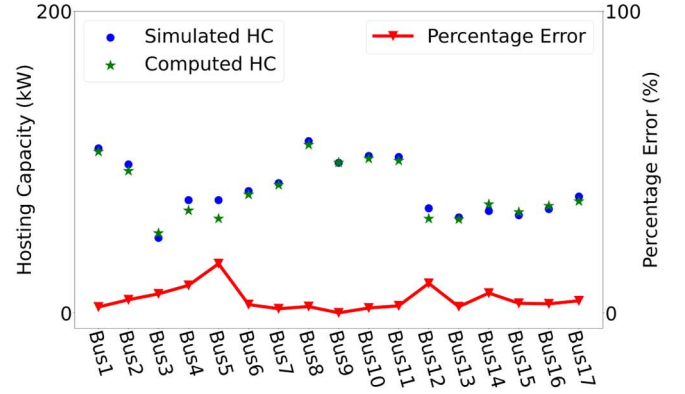
F. Accuracy of the ST-LSTM With the Sensitivity Gate

Since we already verified the performance of the ST-LSTM with the designed dual forget gates, this section will show the accuracy of the ST-LSTM with and without the sensitivity gate, which will further validate the performance of the sensitivity gate of ST-LSTM.

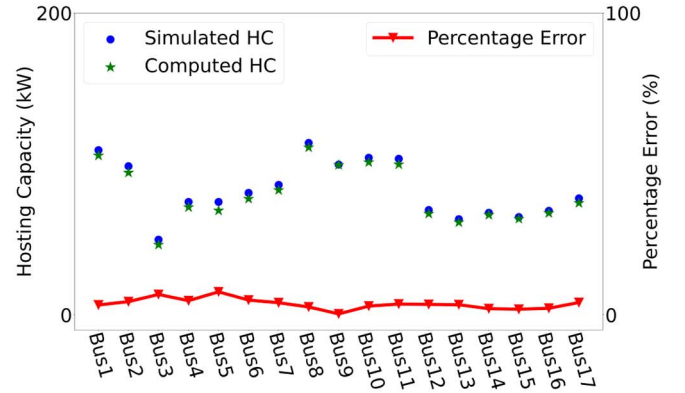
In our review and simulation, we found that the voltage sensitivity data correlates with the HC. To make use of



(a) Results of one path in the utility feeder at time-series 16 using spatial LSTM.



(b) Results of one path in the utility feeder at time-series 16 using ST-LSTM without the sensitivity gate.



(c) Results of one path in the utility feeder at time-series 16 using ST-LSTM with the sensitivity gate.

Fig. 8. Computation results of one path in the utility feeder at time-series 16 using different models. The simulated HC values are from software simulation, and the computed HC values are from the learning-based models. The ST-LSTM model with the sensitivity gate has the best performance in the utility feeder.

this correlation, we use the voltage sensitivity data as the input of the sensitivity gate. The comparison between the ST-LSTM without and with the sensitivity gate is shown in Table V and the detailed results of one time t , which is time-series 16, of one of the paths using the ST-LSTM with and without sensitivity gate models are shown in Fig. 8(b) and Fig. 8(c) respectively. Notice, the sensitivity gate can significantly decrease the percentage error.

TABLE V

COMPARISON OF PERCENTAGE ERROR FOR ST-LSTM WITHOUT AND WITH SENSITIVITY GATE ON UTILITY FEEDER

Test Scenario	Zone 1	Zone 2	Zone 3
ST-LSTM (Without sensitivity gate)	7.2%	12.1%	12.2%
ST-LSTM (With sensitivity gate)	2.1%	6.9%	7.9%

V. CONCLUSION

This paper uses a developed deep learning-based method, namely Spatial-Temporal LSTM, to calculate the distribution grids' hosting capacity (HC). There are three main contributions of this paper. First, the paper proposed a deep learning-based method to calculate the HC in real time. Second, the paper modifies the basic LSTM to the ST-LSTM. This paper updates the forget gate to dual forget gates to build a correlation between temporal and spatial sequences because the forget gate is exceptionally significant in the basic LSTM model. Third, this paper increases the accuracy of the deep learning-based HC calculation method using the designed sensitivity gate. These gates use the result of the voltage sensitivity analysis. Comparing basic LSTM and the ST-LSTM on IEEE example feeders confirms that the new ST-LSTM has a considerable increase in the average and the respective HC accuracy at the buses. To further validate the design of the ST-LSTM deep learning framework, we test the performance on a generic Arizona utility high penetration feeder, which also reinforces the superiority of our new designs.

In our future work, we plan to theoretically investigate how voltage sensitivity data can improve the performance of deep learning-based HCA. Furthermore, we will use a more efficient way to embed spatial information, for example, a tree-structured LSTM network or a combination of the GNN and LSTM network may be feasible research directions to extend the ST-LSTM design.

REFERENCES

- [1] J. Yuan, Y. Weng, and C.-W. Tan, "Quantifying hosting capacity for rooftop PV system in LV distribution grids," 2019, *arXiv:1909.00864*.
- [2] S. M. Ismael, S. H. A. Aleem, A. Y. Abdelaziz, and A. F. Zobaa, "State-of-the-art of hosting capacity in modern power systems with distributed generation," *Renew. Energy*, vol. 130, pp. 1002–1020, Jan. 2019.
- [3] M. Rylander, J. Smith, and L. Rogers, *Impact Factors, Methods, and Considerations for Calculating and Applying Hosting Capacity*, Electr. Power Res. Inst., Washington, DC, USA, 2018.
- [4] E. Mulenga, M. H. Bollen, and N. Etherden, "A review of hosting capacity quantification methods for photovoltaics in low-voltage distribution grids," *Int. J. Electr. Power Energy Syst.*, vol. 115, Feb. 2020, Art. no. 105445.
- [5] M. Z. Ul Abideen, O. Ellabban, and L. Al-Fagih, "A review of the tools and methods for distribution networks' hosting capacity calculation," *Energies*, vol. 13, no. 11, p. 2758, 2020.
- [6] Q. Cui, Y. Weng, and C.-W. Tan, "Electric vehicle charging station placement method for urban areas," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6552–6565, Nov. 2019.
- [7] J. Yuan, Q. Cui, Z. Ma, and Y. Weng, "Learning EV placement factors with social welfare and economic variation modeling," in *Proc. North Amer. Power Symp.*, 2019, pp. 1–6.
- [8] J. Yuan, Y. Weng, and C.-W. Tan, "Determining maximum hosting capacity for PV systems in distribution grids," *Int. J. Electr. Power Energy Syst.*, vol. 135, Feb. 2022, Art. no. 107342.
- [9] J. Smith, R. Dugan, M. Rylander, and T. Key, "Advanced distribution planning tools for high penetration PV deployment," in *Proc. IEEE Power Energy Soc. General Meeting*, 2012, pp. 1–7.
- [10] S. Elsaiah, M. Benidris, and J. Mitra, "Analytical approach for placement and sizing of distributed generation on distribution systems," *IET Gener. Transm. Distrib.*, vol. 8, no. 6, pp. 1039–1049, 2014.
- [11] H. Hu, Q. Shi, Z. He, J. He, and S. Gao, "Potential harmonic resonance impacts of PV inverter filters on distribution systems," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 151–161, Jan. 2015.
- [12] F. Ding and B. Mather, "On distributed PV hosting capacity estimation, sensitivity study, and improvement," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1010–1020, Jul. 2017.
- [13] M. H. Bollen and S. K. Rönnerberg, "Hosting capacity of the power grid for renewable electricity production and new large consumption equipment," *Energies*, vol. 10, no. 9, p. 1325, 2017.
- [14] Z. Ren, W. Yan, X. Zhao, Y. Li, and J. Yu, "Probabilistic power flow for distribution networks with photovoltaic generators," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2013, pp. 1–5.
- [15] C. Wan, Z. Xu, Z. Y. Dong, and K. P. Wong, "Probabilistic load flow computation using first-order second-moment method," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2012, pp. 1–6.
- [16] Y. Weng, J. Yu, and R. Rajagopal, "Probabilistic baseline estimation based on load patterns for better residential customer rewards," *Int. J. Electr. Power Energy Syst.*, vol. 100, pp. 508–516, Sep. 2018.
- [17] X. Geng, L. Tong, A. Bhattacharya, B. Mallick, and L. Xie, "Probabilistic hosting capacity analysis via Bayesian optimization," 2020, *arXiv:2011.05193*.
- [18] H. Samet and M. Khorshidsavar, "Analytic time series load flow," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 3886–3899, Feb. 2018.
- [19] J. E. Quiroz, M. J. Reno, and R. J. Broderick, "Time series simulation of voltage regulation device control modes," in *Proc. IEEE Photovoltaic Specialists Conf.*, 2013, pp. 1700–1705.
- [20] A. K. Jain, K. Horowitz, F. Ding, N. Gensollen, B. Mather, and B. Palmintier, "Quasi-static time-series PV hosting capacity methodology and metrics," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf.*, 2019, pp. 1–5.
- [21] M. J. Reno, J. Deboever, and B. Mather, "Motivation and requirements for quasi-static time series (QSTS) for distribution system analysis," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2017, pp. 1–5.
- [22] M. Rylander, J. Smith, and W. Sunderman, "Streamlined method for determining distribution system hosting capacity," in *Proc. IEEE Rural Electric Power Conf.*, 2015, pp. 3–9.
- [23] M. Alturki, A. Khodaei, A. Paaso, and S. Bahramirad, "Optimization-based distribution grid hosting capacity calculations," *Appl. Energy*, vol. 219, pp. 350–360, Jun. 2018.
- [24] K. Joshi and R. R. Gokaraju, "An iterative approach to improve PV hosting capacity for a remote community," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2018, pp. 1–5.
- [25] H. Li, Y. Weng, E. Farantatos, and M. Patel, "A hybrid machine learning framework for enhancing PMU-based event identification with limited labels," in *Proc. Int. Conf. Smart Grid Synchronized Meas. Anal.*, 2019, pp. 1–8.
- [26] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," 2015, *arXiv:1506.00019*.
- [27] A. R. S. Parmezan, V. M. Souza, and G. E. Batista, "Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model," *Inf. Sci.*, vol. 484, pp. 302–337, May 2019.
- [28] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.
- [29] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
- [30] V. Quintero-Molina, M. Romero-L., and A. Pavas, "Assessment of the hosting capacity in distribution networks with different DG location," in *Proc. IEEE Manchester PowerTech*, 2017, pp. 1–6.
- [31] F. Ding, B. Mather, and P. Gotseff, "Technologies to increase PV hosting capacity in distribution feeders," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2016, pp. 1–5.
- [32] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed energy resources topology identification via graphical modeling," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2682–2694, Jul. 2017.
- [33] P. Zhao, H. Zhu, Y. Liu, Z. Li, J. Xu, and V. S. Sheng, "Where to go next: A spatio-temporal LSTM model for next POI recommendation," 2018, *arXiv:1806.06671*.
- [34] L. Huang, Y. Ma, S. Wang, and Y. Liu, "An attention-based spatiotemporal LSTM network for next POI recommendation," *IEEE Trans. Services Comput.*, vol. 14, no. 6, pp. 1585–1597, Nov/Dec. 2021.

- [35] D. Kong and F. Wu, "HST-LSTM: A hierarchical spatial-temporal long-short term memory network for location prediction," in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 18, 2018, pp. 2341–2347.
- [36] J. Liu, A. Shahroudy, D. Xu, A. C. Kot, and G. Wang, "Skeleton-based action recognition using spatio-temporal LSTM network with trust gates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 3007–3021, Dec. 2018.
- [37] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," 2015, *arXiv:1503.00075*.
- [38] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [39] R. A. Shayani and M. A. G. de Oliveira, "Photovoltaic generation penetration limits in radial distribution systems," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1625–1631, Aug. 2011.
- [40] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2342–2350.
- [41] Y. Si, N. Korada, R. Ayyanar, and Q. Lei, "A high performance communication architecture for a smart micro-grid testbed using customized edge intelligent devices (EIDs) with spi and modbus TCP/IP communication protocols," *IEEE Open J. Power Electron.*, vol. 2, pp. 2–17, Jan. 2021.
- [42] Y. Weng, M. D. Ilić, Q. Li, and R. Negi, "Distributed algorithms for convexified bad data and topology error detection and identification problems," *Int. J. Electr. Power Energy Syst.*, vol. 83, pp. 241–250, Dec. 2016.
- [43] M. I. Jordan, "Serial order: A parallel distributed processing approach," *Adv. Psychol.*, vol. 121, pp. 471–495, Jan. 1997.
- [44] J. L. Elman, "Finding structure in time," *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, 1990.
- [45] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [46] K. Montano-Martinez *et al.*, "Detailed primary and secondary distribution system model enhancement using AMI data," 2021, *arXiv:2105.14161*.
- [47] S. Thakar, V. Vittal, and R. Ayyanar, "An integrated transmission-distribution co-simulation for a distribution system with high renewable penetration," in *Proc. IEEE Photovoltaic Specialists Conf.*, 2021, pp. 672–679.
- [48] "CYME Power Engineering Software." EATON. [Online]. Available: <https://www.cyme.com/> (Accessed: Aug. 8, 2022).
- [49] "CYME Software Module Reference Guide." EATON. [Online]. Available: <https://www.eaton.com/us/en-us/products/utility-grid-solutions/cyme-power-engineering-solutions.html> (Accessed: Aug. 8, 2022).
- [50] "CYME Integration Capacity Analysis." EATON. [Online]. Available: <https://www.eaton.com/us/en-us/products/utility-grid-solutions/software-modules/integration-capacity-analysis.html> (Accessed: Aug. 8, 2022).
- [51] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: A comprehensive review," *Comput. Social Netw.*, vol. 6, no. 1, pp. 1–23, 2019.



Jiaqi Wu (Student Member, IEEE) received the B.Sc. degree in electrical engineering from Shandong University, Jinan, China, and the M.Sc. degree in electrical engineering from Arizona State University, Tempe, AZ, USA, where he is currently pursuing the Ph.D. degree. His current research interests are in the interdisciplinary area of artificial intelligence applications in power systems, e.g., data-driven machine learning techniques for dynamic hosting capacity analysis. He received University Graduate Fellowship in 2021.



Raja Ayyanar (Fellow, IEEE) received the M.S. degree from the Indian Institute of Science, Bengaluru, India, in 1995, and the Ph.D. degree from the University of Minnesota, Minneapolis, MN, USA, in 2000. He is currently a Professor with Arizona State University, Tempe, AZ, USA.



Jingyi Yuan (Student Member, IEEE) received the B.Sc. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, and the M.Sc. degree in electrical engineering from Arizona State University (ASU). She is currently pursuing the Ph.D. degree in electrical, computer and energy engineering with Arizona State University, Tempe, AZ, USA, where she currently serving as a Research Assistant. Her research interests are in the areas of machine learning and big data applications in power systems, explainable data-driven modeling of cyber-physical systems, and dynamic hosting capacity of photovoltaics. In recognition of her extraordinary academic achievements, she won University Graduate Fellowship and Engineering Graduate Fellowship. She won the Best Paper (Presentation) Award of the 51st IEEE North American Power Symposium in 2019.



Yang Weng (Senior Member, IEEE) received the B.E. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, the first M.Sc. degree in statistics from the University of Illinois at Chicago, Chicago, IL, USA, and the second M.Sc. degree in machine learning of computer science and the M.E. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA. He joined Stanford University, Stanford, CA, USA, as a TomKat Fellow for sustainable energy. He is currently an Assistant Professor of Electrical, Computer, and Energy Engineering with Arizona State University, Tempe, AZ, USA. His research interest is in the interdisciplinary area of power systems, machine learning, and renewable integration. He received the CMU Dean's Graduate Fellowship in 2010, the Best Paper Award at the International Conference on Smart Grid Communication (SGC) in 2012, the First Ranking Paper of SGC in 2013, the Best Papers at the Power and Energy Society General Meeting in 2014, the ABB Fellowship in 2014, the Golden Best Paper Award at the International Conference on Probabilistic Methods Applied to Power Systems in 2016, and the Best Paper Award at IEEE Conference on Energy Internet and Energy System Integration in 2017, the IEEE North American Power Symposium in 2019, and the IEEE Sustainable Power and Energy Conference in 2019.