



Multiple-input neural networks for time series forecasting incorporating historical and prospective context

João Palet¹ · Vasco Manquinho¹ · Rui Henriques¹

Received: 11 February 2022 / Accepted: 3 October 2023 / Published online: 27 October 2023
© The Author(s) 2023

Abstract

Individual and societal systems are open systems continuously affected by their situational context. In recent years, context sources have been increasingly considered in different domains to aid short and long-term forecasts of systems' behavior. Nevertheless, available research generally disregards the role of prospective context, such as calendrical planning or weather forecasts. This work proposes a multiple-input neural architecture consisting of a sequential composition of long short-term memory units or temporal convolutional networks able to incorporate both historical and prospective sources of situational context to aid time series forecasting tasks. Considering urban case studies, we further assess the impact that different sources of external context have on medical emergency and mobility forecasts. Results show that the incorporation of external context variables, including calendrical and weather variables, can significantly reduce forecasting errors against state-of-the-art forecasters. In particular, the incorporation of prospective context, generally neglected in related work, mitigates error increases along the forecasting horizon.

Keywords Time series forecasting · Situational context · Multiple-input neural networks

Responsible editor: Albrecht Zimmermann.

✉ Rui Henriques
rmch@tecnico.ulisboa.pt

João Palet
joao.palet@tecnico.ulisboa.pt

Vasco Manquinho
vasco.manquinho@tecnico.ulisboa.pt

¹ INESC-ID and IST, Universidade de Lisboa, R. Alves Redol 9, 1000-029 Lisboa, Portugal

1 Introduction

Geophysical, societal, ecological, organizational, agro-food and biological systems are situated systems, whose behavior affects and is affected by the surrounding environment. Considering urban systems, mobility dynamics are inherently dependent on calendar status and affected by large-scale events, roadblocks, and meteorological conditions, impacting demand and transport mode preferences (Cerqueira et al. 2021). Medical emergency needs are another paradigmatic case with well-established links to the calendrical context, weather factors, and public events (Channouf et al. 2007; McCarthy et al. 2008; Kam et al. 2010; Wong and Lai 2012).

As a consequence, research on context-aware predictive models for time series forecasting has received significant attention in recent years (Bi et al. 2022; Jozi et al. 2022; Guiguet et al. 2018; Schürholz et al. 2020). In the neural processing field, relationships between target and context variables have been explicitly modeled via graph neural networks (Xu et al. 2021; Fang et al. 2021), as well as multivariate gating units in recurrent neural networks (Guiguet et al. 2018; Zhu et al. 2021; Ruan et al. 2020) and multi-source embeddings (Kamarthi et al. 2021; Huang et al. 2019). Despite their relevance, most solely focus on the role of historical and static context, neglecting the important role of prospective situational context.

Sources of prospective context, such as planned events, calendrical information or weather forecasts (Sardinha et al. 2021), can arguably be positioned as pivotal in predictive tasks given their potential effect on systems' behavior along the horizon of prediction. Nevertheless, significant challenges should be noticed. First, available sources of prospective context are often structurally heterogeneous (Tiam-Lee et al. 2022; Cerqueira et al. 2021), e.g., calendrical data recorded as georeferenced events while weather forecasts as multivariate time series. Second, prospective sources of context may not be accessible from structured sources, and the geographical and temporal footprint of prospective events may not be fully known a priori (Leite et al. 2020). Third, some sources of prospective context are incomplete, as well as susceptible to significant levels of noise. Considering weather forecasts, long-term forecasts may be absent, while short-term forecasts are subjected to arbitrary uncertainty levels (Cerqueira et al. 2021). Fourth, associations between prospective context and forecasts are hypothesized to be more relevant than those associations between prospective context against historical data, raising additional challenges to the development of effective neural processing principles and architectural choices.

This work proposes effective neural processing principles to incorporate both historical and prospective sources situational context in time series forecasting models. To this end, two major contributions are drawn. First, a multiple-input neural architecture consisting of a sequential composition of long short-term memory (LSTM) networks is proposed to incorporate heterogeneous sources of context. While historical context is inputted as auxiliary covariates to identify and correct externalities in the forecasting task, prospective context is incorporated at a later layering stage to act as denoiser of the forecasted series. Second, masking principles are drawn for the normative processing and integration of available context.

The inherent simplicity of the aforementioned principles is purposeful as our primary aim is to show that minimalist and elegant principles - multi-input sequential layering and masking - are sufficient to yield statistically significant improvements in forecasting tasks augmented with prospective context. As the proposed multi-input framework is parameterizable, LSTMs can be easily replaced by stacked recurrent units (Li et al. 2021), multivariate graph convolutional networks (GCNs) (Rico et al. 2021) or deep temporal convolutional networks (TCNs) (Chen et al. 2020). Although statistically significant differences ($\alpha=1E-3$) were not consistently observed with (regularized) deep layering, domain-specific adaptations are supported under the proposed architectural principles.

The proposed principles are integrated within a methodology that is experimentally validated using the demand for Lisbon's emergency medical services (EMS) as the primary case study, and Lisbon's bike sharing system (BSS) demand as a complementary application. Results show that the incorporation of external context variables, including calendrical and weather variables, can significantly reduce predictive errors, providing compelling empirical evidence in favor of the proposed forecasters against state-of-the-art alternatives. In particular, the incorporation of prospective context is the primary driver of efficacy gains, and essential to mitigate error increase along the horizon of prediction.

The manuscript is structured as follows. Section 2 introduces essentials of (multivariate) time series forecasting, while Sect. 3 surveys related contributions on context-aware predictive modeling. Section 4 introduces the proposed multi-input neural processing principles. Section 5 experimentally assesses the proposed methodology in urban data domains, discussing the gathered results. Finally, implications and major concluding remarks are drawn in Sect. 6.

2 Background

Problem formulation. The behavior of systems can be subject to a form of sensorization. A time series is a sequence of observations, usually measured at equally spaced points in time, $\mathbf{y}_{1..T} = (y_1, \dots, y_T)$, where each observation, y_t , recorded at a given time step t , is either univariate, $y_t \in \mathcal{Y}$, or multivariate, $\mathbf{y}_t \in \mathcal{Y}_1 \times \dots \times \mathcal{Y}_n$, depending on the number of the monitored behaviors. Given a (multivariate) time series $\mathbf{y}_{1..T}$ and a target variable, \mathcal{Y}_k , *time series forecasting* aims at estimating the measurements of the target variable along the next h steps, $(y_{T+1}, \dots, y_{T+h})$, where $y_{T+i} \in \mathcal{Y}_k$ and h is the horizon of prediction. Multivariate forecasts, \mathbf{y}_{T+i} , can be complementarily pursued in the presence of multiple target variables.

Situational context can be further monitored by measuring auxiliary behavior of a given system or the properties of its environment. Such endogenous and exogenous features can be captured using dependent variables, well-established in the previous multivariate time series formulation, $\mathbf{y}_t \in \mathcal{Y}_1 \times \dots \times \mathcal{Y}_n$. In addition to historical context, monitored along $1..T$ steps, prospective situational context, such as planned or forecasted events falling in the horizon of prediction, may be available. In this context, given a context-enriched multivariate time series, $\mathbf{y}_{1..T}$, and available prospective time series, $\mathbf{z}_{T+1..T+h}$, the forecasting task can be augmented to estimate the

targets of a variable of interest along a given horizon, $(y_{T+1}, \dots, y_{T+h})$ where $y_{T+t} \in \mathcal{Y}_k$, from both $\mathbf{y}_{1..T}$ and $\mathbf{z}_{T+1..T+h}$.

Context data consolidation. Context data can be acquired from unstructured, semi-structured or structured sources. Social media, public administration repositories, weather portals, online calendars, cultural agendas, theatre sites, and online news can be periodically explored with the aim of retrieving historical or prospective context. Despite the presence of principles to this end (Wibisono et al. 2012; Tempelmeier et al. 2019; Tang et al. 2019), context acquisition from unstructured sources is generally subjected to uncertainties related with data quality and availability. As a result, municipalities and other entities have established efforts towards a more normative gathering and provision of (semi-)structured repositories with situational context (Lemonde et al. 2021). In the Lisbon city, the reservation of public spaces, including stadiums, auditoriums, large halls, arenas, amphitheatres, amongst others, is thoroughly updated at the Lisboa Aberta portal, and can be periodically inquired for context-aware urban data analytics (Leite et al. 2020).

Historical context may not always be effectively used as a proxy to infer prospective context data (Kuijpers et al. 2022), hence the relevance of sourcing available prospective context data. For instance, available weather forecasts by meteorology institutes may not be readily predicted from historical data acquired at meteorology stations as professional forecasts often rely on remote sensing inputs, atmospheric models, and background knowledge that may not be readily available. Complementarily, calendrical information is generally available in advance and not always predictable from historical calendars (e.g., moveable holidays). Similarly, prospective public events (e.g., large-scale concerts, symposiums, summits, sport matches) depend on extensive externalities and thus are hardly predictable from past event data.

Planned events in some context sources can be automatically annotated in accordance with their typology and duration (Cerqueira et al. 2021). The spatial extent, as well as the historical and prospective duration of some of these events (e.g., infrastructural interventions in the city) can be maintained in some of these repositories (Lemonde et al. 2021). In contrast, for public events without such information, rules can be dynamically inferred with expectations on the average event duration in accordance with its typology (Cerqueira et al. 2021). Illustrating, a concert can approximately impact urban mobility 60 min before its start and up to 40 min after its end. Context-specific deviations can be historically assessed in the presence of comparable events against expectations to determine the spatiotemporal footprint of an event in accordance with the principles placed by Cerqueira et al. (2021).

Context data, whether represented by events annotated with spatiotemporal footprints (e.g., gatherings) or by periodically collected/forecasted records (e.g., weather), can then be inputted to a learning system as-is or, in alternative, mapped onto a fused data structure more conducive to the subsequent learning needs. Principles for mapping georeferenced event sets as multivariate time series data structures have been explored in former works (Tiam-Lee et al. 2022; Neves et al. 2020).

Essentials on forecasting. Classical statistical methods, such as regression, estimate a given target, y_{T+t} , from available pairs $(\mathbf{x}^{(t)}, y_t)$ where $t \leq T$ and $\mathbf{x}^{(t)}$ are features drawn from available data at time step t , i.e., $\mathbf{y}_{1..t}$, $\mathbf{z}_{t+1..T}$ and $\mathbf{z}_{T+1..T+h}$. Although

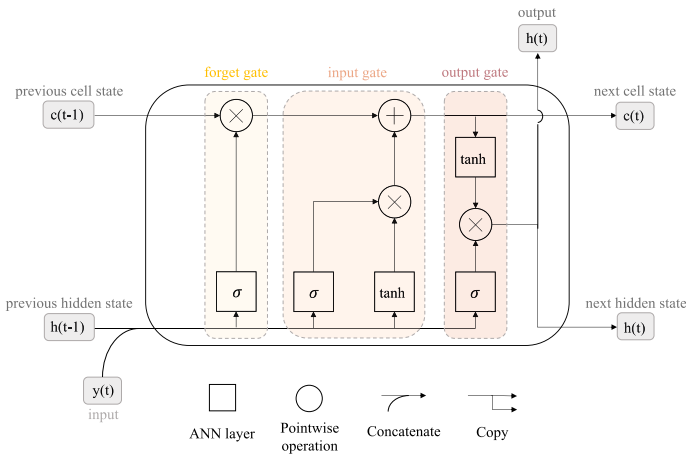


Fig. 1 Composition of an LSTM memory cell

regression methods are a natural context-aware candidate as $\mathbf{x}^{(t)}$ can capture available context at multiple periods, they generally neglect the rich temporal nature of the forecasting problem. Classic time series models are generally parametric descriptors of time series behavior. Series can be decomposed into major components, including trend and seasonality, then projected along the horizon of prediction to produce forecasts (Box et al. 2015). Complementarily, triple exponential smoothing, auto-regression, moving-average and differencing operations can be pursued to model and forecast non-stationary series (Holt 2004; Chatfield 2000). Although classic time series models have been extended to incorporate auxiliary variables, only historical covariates are accounted to assist forecasting (Pfeffermann and Allon 1989; Szeto et al. 2009).

To deal with the complex non-linearities underlying the behavior of real-world systems, artificial neural networks (ANNs) are currently the paradigmatic option to forecast the behavior of such systems. Recurrent neural networks (RNNs) (Rumelhart et al. 1985) are a class of ANNs with feedback (loop) connections, where an output from the previous step is fed as input to the current step. Long short-term memory (LSTM) networks (Hochreiter and Schmidhuber 1997) are specialized RNNs, developed to capture long-term serial dependencies (Fig. 1), thus natural candidates for time series forecasting. LSTMs, as well as gated recurrent units, are inherently prepared to process multivariate time series, being able to elegantly incorporate historical context variables to potentially aid forecasts.

3 Related work

Context-aware forecasting. In the neural processing field, contributions for context-aware forecasting have been mainly propelled by the modeling of historical relationships between the target variable and auxiliary context variables. In this

context, temporal convolution networks (TCN) (Ruan et al. 2020) and graph neural networks (Rico et al. 2021) are the paradigmatic options to capture spatial and temporal dependencies between the target series and the surrounding context. In social and recommendation domains, auxiliary context variables can encode correlated information from related users, objects, and evaluations. In urban domains, auxiliary context variables often correspond to measurements at different points in the city, and their spatial relationships are captured within a graph structure. Considering urban mobility as a guiding case, auxiliary context variables may correspond to traffic measured at different locations. STJLA (Fang et al. 2021) is a context-aware neural architecture that applies linear attention to the spatiotemporal joint graph to capture correlations between all nodes. A graph convolutional network component based on spatial adjacency and functional similarity with context variables has been recently proposed (Xu et al. 2021) to incorporate transportation supply, demographic profile, and historical weather data in scooter-sharing demand prediction. While graph convolutions capture forms of node dependence from spatial and functional similarity, the extracted associations at different time points are then inputted to gated recurrent units to capture the temporal dependencies that are at the basis of the targeted predictive task.

Leveraging on the inherent ability of recurrent neural networks to handle multivariate time series (i.e., cross-series training), Zhu et al. (2021) tackled product demand forecasting in pharmaceutical domains by seeing the demand of related products as auxiliary context variables and further incorporating available domain knowledge.

Motivated by tensor factorization for context-aware recommender systems, Bi et al. (2022) propose a latent factor approach to sales forecasting that leverages on a single tensor factorization model across multiple products and stores. The results gathered from this recent work evidence the presence of synergistic principles between the targeted context-aware learning tasks and multi-source, multi-view, and multi-task learning (Zhang and Yang 2018; Ruder 2017). Probabilistic multi-view neural processing has been considered to learn intermediate representations from multiple data sources in forecasting tasks (Kamarthi et al. 2021). Classic series segmentation and decomposition principles have been also proved useful to assist context-aware forecasts (Ruan et al. 2020).

Context-aware forecasting has been applied across multiple domains. In the energy domain, Jozi et al. (2022) extended consumption forecasting models with context variables such as energy generation, temperature, and occupancy from building sensors to aid energy management in buildings. Cuncu et al. (2022) further exploited the role of inhabitant activities and the use of household appliances to assist this forecasting task. In computer vision, the forecast of future activities from video data has been augmented using activity and scene context (Chakraborty and Roy-Chowdhury 2014). In procurement and logistics, qualitative expert opinion has been integrated onto product forecasting tasks (Arvan et al. 2019). In trajectory prediction, road users and environments have been used to anticipate obstacles in complex driving scenarios (Schäfer et al. 2022), and human trajectories forecasted in crowded spaces considering the dynamics of other moving agents in the scene and static elements that might be perceived as points of attraction (Bartoli et al. 2018).

Domain knowledge has been integrated into deep neural networks to aid traffic prediction in 5 G networks (Garrido et al. 2021). Context-aware embedding modules were explored in recurrent neural networks to include discrete exogenous features when forecasting smart card validations in public transport (Guiguet et al. 2018), as well as in hierarchical networks for forecasting traffic accidents (Huang et al. 2019). Air pollution forecasts have been augmented with context information from both surrounding pollution sources (e.g., bushfire incidents, traffic volumes) and user's health profile (Schürholz et al. 2020).

Despite the inherent relevance of the aforementioned works to context-aware forecasting, they do not explore the role that sources of prospective context may yield in the predictive tasks. Furthermore, the available contributions are not easily extensible towards this end since prospective context variables cannot be straightforwardly modeled as serial covariates or adjacent nodes in graph structures given their disjoint occurrence from historical variables.

Multi-input network layering. The available measurements of a given system are in multiple input neural networks partitioned in accordance with their inherent properties, where each partition is inputted into one or more components of the network for dedicated processing and later merged for joint processing.

Different architectural principles can be found in this, with parallel, sequential and hybrid processing being common options. Naglah et al. (2021) propose a parallel multi-input convolutional neural network (CNN) to perform fusion of two magnetic resonance imaging modalities (diffusion weighted image and apparent diffusion coefficient map) as to enable independent convolution processes for each modality, which can increase the likelihood of detecting deep texture patterns. Oktay et al. (2016) further explored parallel multi-input CNNs by exploring dedicated processing paths for different viewing planes of three-dimensional cardiac imaging for morphology analysis. Related contributions can be found, including the exploration of implicit spectral-spatial information in hyperspectral images for feature extraction ends (Zhong et al. 2022), multi-input analysis of different medical exams for COVID-19 diagnosis (Zhang et al. 2021), or the exploitation of high degrees of correlation and complementary information among neighboring tomography images for denoising ends (Abbasi et al. 2019). Based on the U-Net model, Shi et al. (2021) proposes a multi-input fusion network based on the extraction and fusion of imaging features at different input resolution scales.

In sequential multi-input networks, some of the available measurements are integrated at later stages in the neural processing pipeline. Sánchez-Cauce et al. (2021) propose a multi-input network for cancer diagnosis where former layers are used to process imaging data for the extraction of features while later layers further received available complementary clinical and demographic data. Similar principles are explored by Apostolopoulos et al. (2021) for cardiovascular disease diagnosis using myocardial perfusion imaging and clinical data.

Hybrid architectural variants are also available. Wang et al. (2021) considered parallel processing of time-domain signals, frequency-domain signals, and time-frequency graph inputs for fault diagnosis. These isolated processing paths are connected using fully connected layers to process the previously processed features with additionally inputted bearing and damage features.

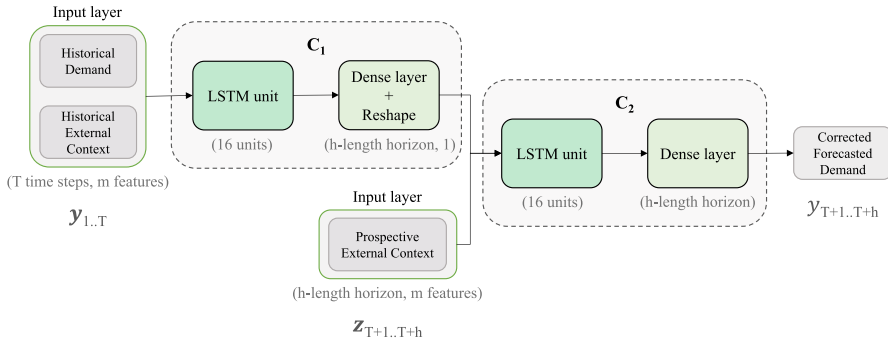


Fig. 2 RNN architecture able incorporate historical and prospective external context variables through multiple inputs

In the context of forecasting tasks, multiple-input networks have been used to combine temporal and static data, as well as to mitigate the issues related to generalization and meteorological effects (Madhiarasan et al. 2021). Xiong et al. (2021) considered ground motion sequences and building features as heterogeneous inputs to a multiple-input convolutional neural network for seismic damage assessment. Despite the relevance of available work, the absence of multi-input neural processing principles to standardly learn from heterogeneous sources of context data is notorious.

4 Multiple-input context-aware neural networks

To aid forecasting tasks in the presence of historical and prospective sources of context, this section proposes a simple yet effective multiple-input neural network architecture (Sect. 4.1), and further establishes masking principles for the effective incorporation of available context (Sects. 4.2 and 4.3).

4.1 Context-aware neural networks

The proposed architecture, schematized in Fig. 2, is a sequential composition of two components, C_1 and C_2 , each composed of a default LSTM cell with 16 units followed by a dense layer. Considering available historical data, the C_1 component takes the context-enriched multivariate series as input, $y_{1..T}$, and returns a forecasted series as the output.

Considering available prospective context, the C_2 component takes as input the forecasted series from C_1 , $\hat{y}_{T+1..T+h|C_1}$, and prospective sources of context along the horizon of prediction, $z_{T+1..T+h}$, returning an adjusted forecasted series, $\hat{y}_{T+1..T+h|C_2}$.

Components C_1 and C_2 can be parameterized with alternative network layering able to process multivariate time series, including stacked recurrent units, (deep) temporal convolutional networks, and graph convolutional networks.

Considering the given architecture, three major training possibilities are selected for this study: *separate*, *alternating* and *joint* training. In the joint training setting, a single loss function is considered at the end of the network. In separate and alternating training settings, a loss function is applied at both the end of component C_1 (history-aware forecasting) and the end of component C_2 (prospective context regularization). In the separate training setting, the parameters of component C_1 are first optimized, followed by the optimization of the parameters of component C_2 , corresponding to the fully independent adjustment of the forecasted series using prospective context. In the alternating training setting, the optimization alternates between components for every iteration of the learning process under a fixed batch size.

Mean absolute error (MAE), mean squared error (MSE) and cosine are tested as viable loss functions for each setting. Variants of the proposed architecture, including the replacement of LSTM units by Gated Recurrent Units (GRU) are also considered. Adam optimizer with early stopping is selected to learn the target networks. Remaining relevant parameters, including the selected activation functions and applied forms of regularization, are subjected to hyperparameterization (Sect. 5).

4.2 Incorporating historical context

Taking both the target and auxiliary historical data as input, C_1 is able to capture significant cross-variable dependencies, as well as their relationship with future targets, via multivariate memory cells and subsequent dense layering. An arbitrarily-high number of historical context variables can be integrated to guide the learning task. To this end, masking principles are necessary to compose the input multivariate time series, $\mathbf{y}_{1..T}$. For the purpose of illustrating the principles introduced along this section, consider the hourly forecast of the number of medical emergencies along a given region, $\mathbf{y}_{example}=(\dots, 11, 9, 8, \dots)$.

In addition to the target variable, calendrical, event and weather variables are paradigmatic context sources in urban domains that can be integrated to guide the prediction. Calendrical variables inform about meaningful information related to the calendar, including:

- *time within the day* to help capture daily seasonal patterns. Taking the illustrative series of hourly medical emergencies, we can enrich it by adding hour information, e.g., $\mathbf{y}_{example}=(\dots, (11, 10pm), (9, 11pm), (8, 12am), \dots)$;
- *weekday* information, in which the day when events occurred is incorporated in the series as, for instance, a nominal encoding of Monday to Sunday. Complementarily, *weekend* information may also be included for a coarser differentiation between weekdays and weekends. By doing so, we can extend the previous series, $\mathbf{y}_{example}$, to further incorporate weekday information, e.g., $\mathbf{y}_{example}=(\dots, (11, 10pm, weekday), (9, 11pm, weekday), (8, 12am, weekend), \dots)$;
- other types of calendrical information such as *holidays*, *festivities*, and *academic calendars* can be further included.

Similarly to the introduced calendar masks, event masks mark periods where events of interest occur, e.g., large-scale gatherings in urban domains such as *festivals*, *concerts*, and *sport events*. These sources of context are generally circumscribed to a specific geographical area. The corresponding series generally establish a measure of the impact of the event in space and time. In urban domains, the real-time acquisition of events from public repositories, as well as the dynamic inference of their spatiotemporal footprint, has been previously studied (Cerqueira et al. 2021).

Finally, weather conditions exert influence on diverse systems, including the illustrative urban ones. Rain, fog and snow increase the likelihood of traffic accidents (Yannis and Karlaftis 2010; Andreescu and Frost 1998), while extreme weather can negatively affect health, correlated with the demand for medical emergency services (Kjellstrom et al. 2010; Wong and Lai 2012). Meteorological variables collected from weather stations can thus be further incorporated to guide the learning process. These can include *temperature*, *relative humidity*, *wind intensity*, among others. Analogously, series can be augmented with this information, e.g., *precipitation* levels registered in the period of each observation, $\mathbf{y}_{\text{example}}=(\dots, (11, 1.6 \text{ mm}), (9, 0.8 \text{ mm}), (8, 0.7 \text{ mm}), \dots)$.

4.3 Incorporating prospective context

The inclusion of C_2 , which takes the forecasted series from C_1 and the prospective external context along the forecasting horizon as input, allows the model to guide and adjust the forecasts along the forecasting horizon in the presence of prospective context information, producing the final forecasted series. In this context, C_2 can be thought of as a context-aware denoiser or time-dependent regularizer of the forecasts. Hence, besides being able to learn relations between historical context and future behaviour of the target variable, the model can also learn relations between the target variable and prospective context variables under the same future periods. Although empirical analysis shows optimal performance of C_2 under an LSTM unit, a gated recurrent unit (GRU) provides a competitive and less demanding needs of this stage. On the opposite pole, stacking of recurrent units and convolutional layering can be further considered for complex data domains, where prospective context is characterized by a high multivariate order.

Considering the illustrative case of medical emergency forecasting, severe weather conditions, specific calendrical festivities (e.g., Christmas period), seasonality factors (e.g., day time), and planned events (e.g., large-scale gatherings), are known to be strongly correlated with the demand observed for multiple types of medical emergencies (Silva et al. 2021). Prospective calendrical variables can be obtained through calendar information similarly as in the historical setting; prospective weather variables can be obtained through public databases and web APIs; and planned events can be derived from web content and (semi-)structured public repositories, such as cultural agendas and usage planning of public spaces. Principles for the autonomous retrieval and preprocessing of these sources of external context have been discussed in previous works (Cerqueira et al. 2021).

Once collected, masking principles introduced in Sect. 4.2 can similarly be applied for prospective context, to produce a novel input for C_2 component, so that the preliminary forecasts, $\hat{y}_{T+1..T+h|C_1}$, are further subjected to context-dependent corrections to improve predictive accuracy.

As some context sources are subject to varying rates of historical and prospective missingness, the availability and completeness of each source should be assessed. For instance, when considering planned public events, some event categories are extensively complete in advance (e.g. festivals, sport events) and can be used in accordance with the proposed masking principles, while event categories with higher missing predisposition should be further inquired. Incomplete context data can be divided according to whether missingness is predominant in the future or in both historical and future time periods. In the earlier case, historical context data can standardly be used. For time periods with high missing predisposition, the default masking principles can be followed under the premise that the partial set of gathered events can assist the forecaster. Nevertheless, as periods without recorded events can be mistakenly interpreted by periods without occurring events, sources with high missing rates can be excluded or, in alternative, dedicated masking symbols included to signal the presence of periods with incomplete information.

5 Results and discussion

Using case studies in the urban domain, this section experimentally assesses the proposed contribution, answering three major research questions:

- Q1 To which extent does the incorporation of prospective context aid forecasts?
- Q2 What are the benefits arising from each accessible source of context in the targeted urban domain?
- Q3 How do context-aware multiple-input networks compare in terms of predictive accuracy against state-of-the-art alternatives?

The proposed context-aware forecasters are implemented in Python and made available via GitHub at:

<https://github.com/joopalet/multiple-input-context-aware-forecaster>.

5.1 Validation methodology

For a robust evaluation of the forecasters, time-aware cross-validation is performed with the necessary care intrinsic to time series data partitioning (Fig. 3). Each dataset produced per iteration is further divided into train, validation and test sets. The training and validation sets comprise 80% of each dataset, from which 80% is training data and 20% is validation data. The remaining 20% of the datasets was used for

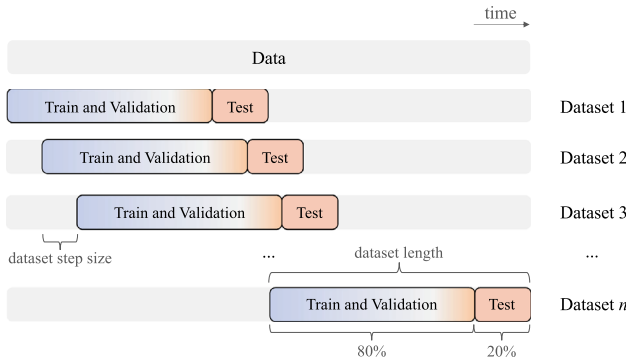


Fig. 3 Creation of datasets for cross-validation of the time-series data

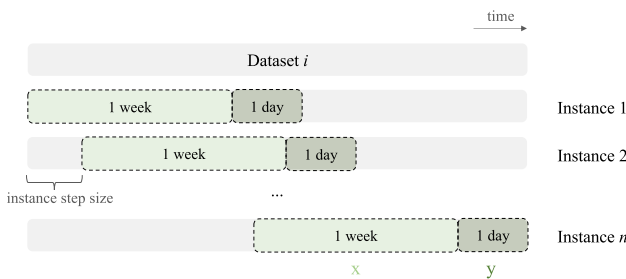


Fig. 4 Input–output pair creation for the learning of the forecasting models

Table 1 Context data sources for the target urban case studies

Source	Masks
Calendrical	Hour; weekday; weekend-week; holiday
Weather	Temperature; precipitation; humidity

testing. The partitions are preserved for the training, validation and testing of each forecaster.

For the learning of the proposed models, each partition is further segmented into a set of data instances, each in the form of an input–output pair. For the given urban scenarios, the input series is a full week of data (168 hour periods), and the output is the subsequent series with the length of the forecasting horizon, i.e., 24 h. Figure 4 illustrates the creation of these instances, ensuring that no testing instances precede training instances.

Table 1 lists the sources of context considered along the target case studies.

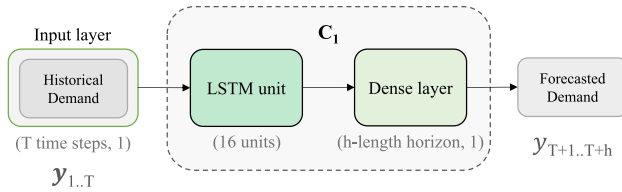


Fig. 5 Univariate RNN architecture with parameterizable C_1 component

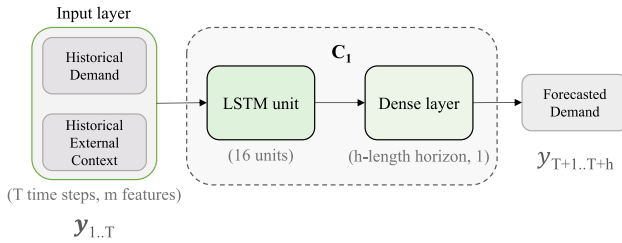


Fig. 6 Multivariate RNN architecture, able to incorporate historical context variables, with parameterizable C_1 component

5.2 Baseline models

The performance of the proposed multi-input neural network is assessed against an equivalent model that only incorporates historical context and a univariate model, both depicted in Figs. 5 and 6, respectively. The proposed models are further compared against Holt Winters’ exponential smoothing and a flexibly optimized feed forward neural network (FFNN). In addition, spatiotemporal graph convolutional networks¹ are selected as reference state-of-the-art forecasting baselines (Yu et al. 2017) given their inherent ability to account for complex interactions between the target and context variables expressed within a graph structure via graph convolutions, as well as to capture temporal dependencies through the incorporation of LSTM layers to perform forecasting on the graph. In the context of our work, the graph captures the pairwise cross-correlation between input variable series and is then postprocessed to either consider all associations (dense graph) or discard uncorrelated variables (sparse graph) when performing graph convolutions. A minimum graph density of 1/3 is fixed to avoid overly sparse graph representations so that dependencies with the available context data sources can be more comprehensively explored. Deep temporal convolutional networks (deep TCNs)² (Chen et al. 2020; Ruan et al. 2020) are further assessed in the presence and absence of historical and prospective sources of context. As TCN networks are not inherently prepared to handle prospective context, two settings are considered to this

¹ https://github.com/keras-team/keras-io/blob/master/examples/timeseries/timeseries_traffic_forecasting.py

² <https://github.com/flaviagiammarino/deep-tn-tensorflow>

Table 2 Baseline forecasting models

Model	Description
LSTM (univariate)	Architecture: LSTM cell with 16 units and output layer with 24 neurons Activation functions: ReLu Optimizer: Adam
LSTM (historical)	Architecture: LSTM cell with 16 units and output layer with 24 neurons Activation functions: ReLu Optimizer: Adam
GCN	Architecture: graph convolution layer followed by a 64 unit LSTM layer (convolution of node tensors is the multivariate input passed to the recurrent layer along time) Graph: adjacency matrix from pairwise input variable cross-correlation in the absence and presence of graph density filter (dense and sparser versions) Optimizer: Adam
Deep TCN	Architecture: deep temporal convolutional network graph: 32 channels with 2-length kernels in the encoder layers, dilation rates ranging from 1 to 8, 64 units per dense layer in the decoder module, non-parametric loss function Optimizer: Adam
FFNN	Architecture: one hidden layer with 16 neurons, output layer with 24 neurons Activation functions: ReLu Optimizer: Adam
HW	Optimizer: limited-memory BFGS

end: i) the proposed multi-input architecture with its components parameterized with the reference TCNs (Chen et al. 2020), and ii) single deep TCN with prospective context data given as complementary input series, convex with historical series whenever applicable.

The assessed neural network models were parameterized with the mean squared error (MSE) loss and Adam optimizer (Kingma and Ba 2014). Sensitivity analysis was performed through manual search with different initial learning rates ($1e-3$ to $1e-6$), values for the batch size (1, 2, 4, 8, and 16), and types of regularization ($l1$ and $l2$). All models were trained for a total of 400 epochs, using early stop criterion based on training versus validation loss. Table 2 describes all baseline models in greater detail.

When considering the proposed learning settings for the target sequential multi-input network – joint, separate and alternating training –, we observed that the joint training setting (single loss function) consistently yields the best results, an observation that is hypothesized to be driven by the inherent simplicity of the proposed architecture, together with the fact that the joint training offers greater sensitivity to the dependencies between the two components. Hence, unless stated otherwise, the displayed results for the target multi-input sequential network are in reference to the joint learning setting.

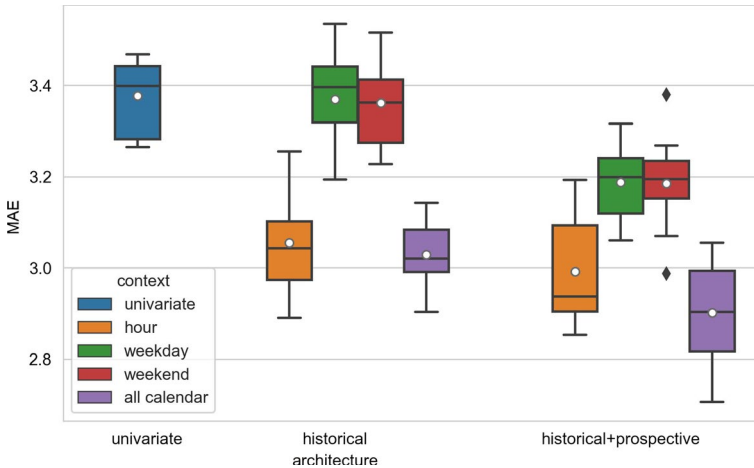


Fig. 7 Impact of incorporating calendrical context in the EMS demand forecasting models

5.3 Case study 1: medical emergencies in Lisbon

The analysis of medical emergencies in Lisbon is introduced as the primary case study to validate the contributions. The pre-hospital medical emergencies in the Lisbon city were provided by Instituto Nacional de Emergência Médica (INEM) – the medical emergency service (EMS) provider in mainland Portugal. The data consists of whole registered emergency cases in Lisbon from 2016 to 2017, comprising a total of 180,234 medical emergencies. Following the principles illustrated in Fig. 3, nine datasets comprising a full year of data were created, with a step size of six weeks between them.

Calendrical context data were gathered from Lisboa Aberta,³ while Lisbon's weather records acquired from meteoTécnico.⁴ The collected variables include temperature, recorded in degrees Celsius (*circ*C), precipitation, recorded in mm, and relative humidity, i.e., the concentration of water vapour present in the air (%). In addition to weather variables, we further compared the performance of models incorporating information regarding the hour of the day (*hour*), day of the week (*weekday*), weekend status (*weekend*), as well as a model that incorporates all three of these masks (*all calendar*).

Figure 7 shows the results obtained when incorporating historical context (*historical architecture*) with the ones obtained when considering both historical and prospective context (*historical+prospective architecture*), as well as no context (*univariate architecture*). Compared to the univariate architecture, all models produce significantly better results, with the exception of the models that only incorporate historical weekday and weekend information, producing comparable results. These

³ <https://lisboaaberta.cm-lisboa.pt/index.php/pt/>.

⁴ <https://meteo.tecnico.ulisboa.pt>.

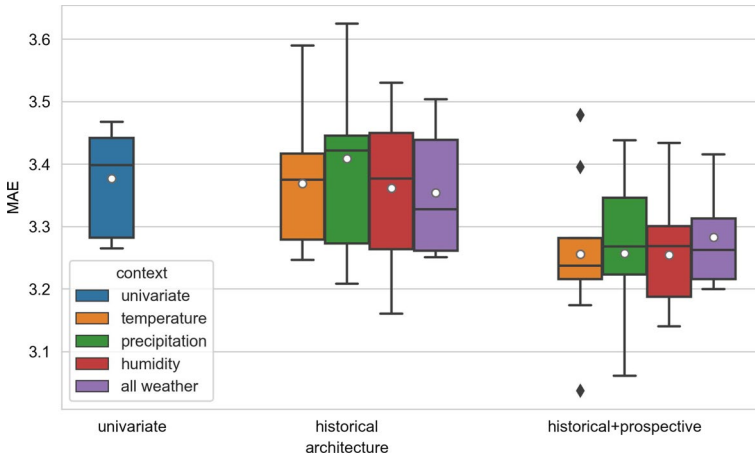


Fig. 8 Impact of incorporating weather context in the EMS demand forecasting models

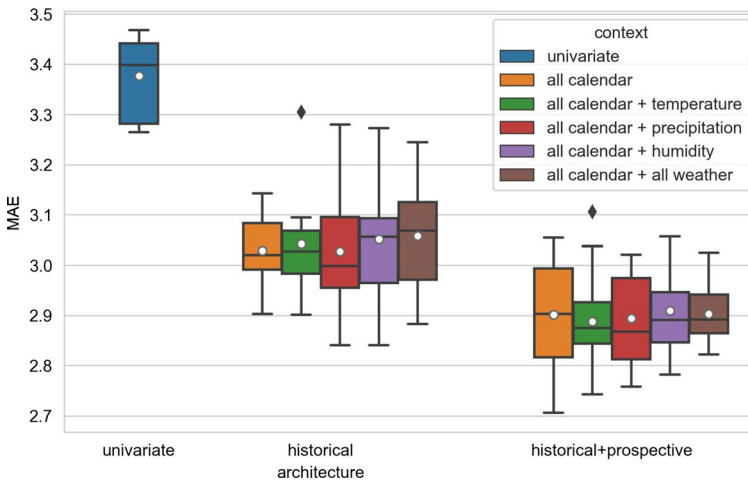


Fig. 9 Impact of incorporating calendrical context together with weather context in the EMS demand forecasting models

results suggest that the hour of the day is the calendrical variable that most affects the volume of EMS demand. Incorporating prospective calendrical context along the forecasting horizon is shown to further improve the performance in comparison to the models that only incorporate context in the historical data. Incorporating all three calendrical variables on both historical and prospective data yield the best results.

Figure 8 presents the results obtained when incorporating weather variables in the models, while Fig. 9 presents the results obtained when incorporating both weather and calendrical variables. Temperature, precipitation, relative humidity (*humidity*),

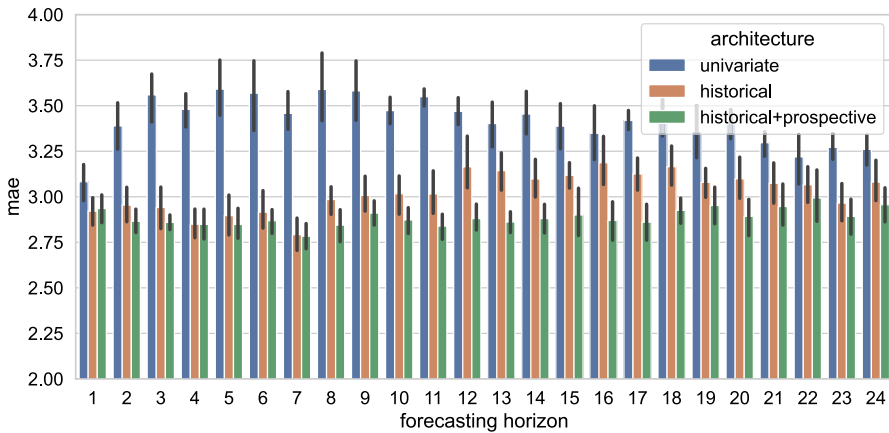


Fig. 10 Comparison of predictive errors (MAE) per time step along the 24-hforecasting horizon when comparing the univariate architecture, the architecture that only incorporates historical context, and the one that also incorporates prospective context

and all previous weather variables (*all weather*) are considered. We observe that, when incorporated alone, historical weather variables do not produce a significant impact on the results. However, improvements are achieved when the models consider prospective weather variables. When added on top of all calendrical variables, temperature and precipitation variables further moderately improve the performance of the models. Once again, results show that models also incorporate prospective context perform considerably better across the reference ones.

Considering that we are working with a forecasting horizon of 24 hours, one of the main reasons to incorporate prospective context is to try to mitigate the increase in uncertainty as we move forward in the predictions along the forecast horizon. Taking the best performing model of each presented architecture, we analyzed how predictive errors evolve with the forecasting horizon. According to the collected results, the best forecasters sensitive to historical context are the ones incorporating hour, weekday, weekend, and precipitation context, while the best forecasters sensitive to both historical and prospective context incorporate hour, weekday, weekend, and temperature context. The assessment of predictor errors along the forecasting horizon is shown in Fig. 10. We can see that, in the first few horizon hours, the errors of the univariate architecture are already much higher, since the model has no sense of the situational context, while both context-aware models are able to keep the errors at a fairly constant rate. As we move forward in the predictions and the current context may start to differ from the historical context, we can observe that starting around the 8th hour period of the forecasting horizon, the errors of the model that only considers historical context start to increase while the model incorporating prospective context are still kept fairly constant.

Table 3 presents the forecasting errors (MAE and RMSE) of the proposed models and all suggested baseline methods for each combination of context variables. Results suggest that incorporating calendrical and weather information, particularly both historical and prospective, improve the performance of the forecasting models.

Table 3 Summary of results obtained (MAE and RMSE), for all EMS demand forecasters implemented, as *mean ± standard deviation* ()

Model	Context	MAE	RMSE	
HW	–	3.18 ± 0.18	4.13 ± 0.28	
FFNN	–	2.91 ± 0.09	3.77 ± 0.18	
LSTM	Univariate	3.38 ± 0.08	4.31 ± 0.21	
Deep TCN	Univariate	3.10 ± 0.16	3.86 ± 0.22	
GCN	Univariate	3.02 ± 0.06	3.81 ± 0.09	
	Calendar (sparse)	3.17 ± 0.11	4.21 ± 0.17	
	Calendar (dense)	3.83 ± 0.14	4.52 ± 0.18	
	Weather (sparse)	3.61 ± 0.17	4.59 ± 0.20	
	Weather (dense)	3.89 ± 0.08	4.91 ± 0.24	
	Calendar and weather (sparse)	3.15 ± 0.11	4.14 ± 0.19	
	Calendar and weather (dense)	3.31 ± 0.13	4.27 ± 0.19	
	LSTM (historical context)	Hour	3.06 ± 0.12	3.92 ± 0.20
		Weekday	3.37 ± 0.11	4.28 ± 0.19
		Weekend	3.36 ± 0.09	4.27 ± 0.18
All calendar		3.03 ± 0.08	3.90 ± 0.15	
Temperature		3.37 ± 0.11	4.30 ± 0.21	
Precipitation		3.41 ± 0.14	4.33 ± 0.23	
Humidity		3.36 ± 0.12	4.28 ± 0.20	
All weather		3.35 ± 0.10	4.28 ± 0.17	
All calendar + temperature		3.04 ± 0.12	3.91 ± 0.19	
All calendar + precipitation		3.03 ± 0.12	3.90 ± 0.22	
LSTM (historical and prospective context)	All calendar + humidity	3.05 ± 0.13	3.93 ± 0.23	
	All calendar + all weather	3.06 ± 0.13	3.94 ± 0.21	
	Hour	2.99 ± 0.12	3.86 ± 0.20	
	Weekday	3.19 ± 0.09	4.09 ± 0.16	
	Weekend	3.18 ± 0.11	4.09 ± 0.18	
	All calendar	2.90 ± 0.11	3.74 ± 0.17	
	Temperature	3.26 ± 0.13	4.17 ± 0.20	
	Precipitation	3.26 ± 0.12	4.16 ± 0.19	
	Humidity	3.25 ± 0.09	3.76 ± 0.16	
	All weather	3.28 ± 0.08	4.21 ± 0.16	
Deep TCN (historical context)	All calendar + temperature	2.89 ± 0.08	3.74 ± 0.15	
	All calendar + precipitation	2.89 ± 0.10	3.74 ± 0.17	
	All calendar + humidity	2.91 ± 0.10	3.76 ± 0.16	
	All calendar + all weather	2.90 ± 0.07	3.76 ± 0.15	
	All calendar	3.01 ± 0.14	3.78 ± 0.22	
	All weather	3.02 ± 0.19	3.82 ± 0.25	
	All calendar + all weather	3.12 ± 0.20	4.09 ± 0.24	
	Deep TCN (historical and prospective context)	All calendar	2.82 ± 0.12	3.67 ± 0.18
		All weather	3.04 ± 0.22	3.82 ± 0.29
		All calendar + all weather	2.82 ± 0.14	3.51 ± 0.22

Top models are given in bold

Graph convolutional networks show for the target case study an inherent difficulty to exploit relevant associations with historical context given the moderately lower performance against the context-unaware version. In fact, denser graph representations, where a high number of pairwise associations between input variables are considered to guide the learning, are shown to deteriorate performance on the given emergency domain. As the original architecture was proposed for the spatiotemporal forecasting of traffic data, where graph associations capture traffic-wise correlated roads, its adaptation to domains where the input variables are less correlated is shown to be less stable, potentially requiring further architectural tuning or the replacement of graph convolutions by alternative transformations. Complementarily, feed-forward neural networks show higher error variability along the prediction horizon (within-instance variability) than the simplistic recurrent layering. Although unable to adequately explore the temporal dependencies of the available context data under the proposed masking principles, their exhaustive optimization and limited series length ensured a more competitive performance regarding MAE.

Considering the proposed architecture, replacing the LSTM unit in the C_1 component by a deep temporal convolutional network (TCN) produced soft improvements for the combined incorporation of historical and prospective context (statistically significant for MAE, non-significant for RMSE under $\alpha=1E-3$). Similarly, the stacking of an additional LSTM in C_1 produced statistically significant improvements against the single context-aware LSTM counterpart (p -value $<1E-3$). In this context, the exploration of stacking or alternative layering within both components of the proposed architecture is suggested as future direction.

To assess the statistical significance of the previous results, we used the T-test when error estimate samples pass the Shappiro-Wilko normality test ($\alpha=0.01$), otherwise, the Mann–Whitney U test, the non-parametric peer. Table 4 summarizes the p -values of the statistical tests performed, for each combination of context variables tested. The obtained p -values show that, for 21 out of the 24 pairs tested, the p -values were lower or equal than the threshold (0.05), indicating that the difference between the great majority of the results obtained are statistically significant.

Complementarily, Fig. 11 measures the impact of replacing the LSTM units by Gated Recurrent Units (GRU) in the compared models, showing that the adequacy of GRU-based network design become only competitive in the multi-input stacked model due to the inherent simplicity of this unit.

Finally, to measure the upper limits of introducing prospective weather data, Fig. 12 assesses the differences between the gains from incorporating actual weather observations in the future (no noise) versus the observed gains from incorporating professional weather forecasts (inherently susceptible to noise). Most of the observed differences are small to moderate, showing the relevance of weather forecasts as proxies for future weather, while bounded in their potential predictive value for the targeted forecasting problem.

Table 4 *P*-values between the EMS demand forecasting errors of the models incorporating both historical and prospective context against two baseline methods: the univariate architecture and the model only historical context, for the the same context variables

	Historical + prospective	Baseline model	
		Univariate	Historical
Hour		5.352e – 7	0.274
Weekday		2e – 4	0.001
Weekend		8e – 4	0.002
All calendar		1.879e – 8	0.013
Temperature		0.028	0.060
Precipitation		0.024	0.026
Humidity		0.009	0.050
All weather		0.026	0.105
All calendar + temperature		7.815e – 10	0.004
All calendar + precipitation		3.953e – 9	0.021
All calendar + humidity		7.117e – 9	0.016
All calendar + all weather		3.874e – 10	0.007

P-values lower or equal than the threshold (0.05) are highlighted in bold

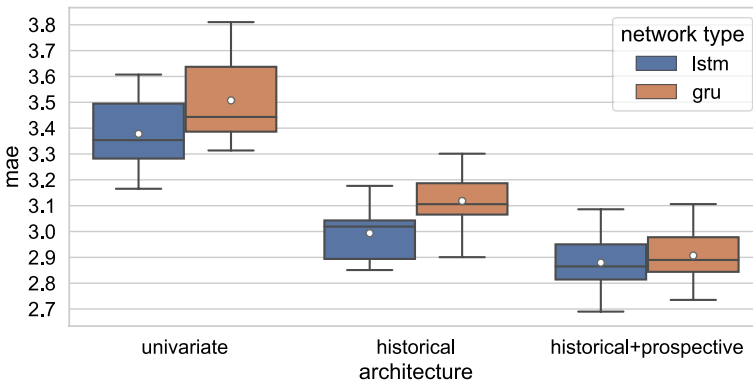


Fig. 11 Comparing the predictive efficacy of GRU versus LSTM units in the introduced context-aware neural architectures

5.4 Case Study 2: lisbon’s bike sharing system

As a complementary case study, the demand for Lisbon’s public bike sharing system (BSS), called GIRA, is introduced. Descriptive statistics of capacity-demand dynamics of GIRA network are publicly available.⁵ Two major data sources are considered. First, a curated sample with all bike trip records from December 2018 until February 2019, along with timestamps for every change in station state and

⁵ <https://lisboaparapessoas.pt/biblioteca/dados/>, <https://www.ciclovias.pt> and <https://dados.cm-lisboa.pt/dataset/giras-docas> (accessed Dec. 1, 2022).

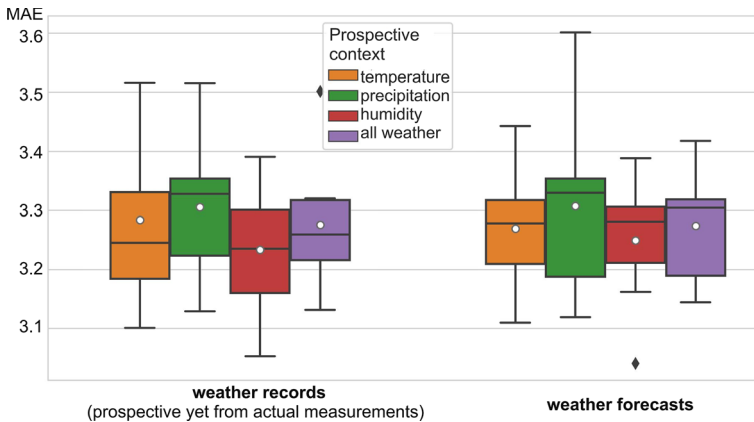


Fig. 12 Potential of prospective weather data for multi-input networks: ground truth prospective weather versus professional weather forecasts

corresponding load value. Second, publicly available station state records⁶ with the load state of each station at specific time points of a day from 2020 to 2022. Accordingly, we target two major forecasting tasks: the prediction of the hourly number of check-in events for the GIRA network along a single day using the first data source, and the hourly station-specific load along a single day using the second data source. Following the principles illustrated in Fig. 3, six (twelve) datasets were derived for the first (second) data source, each comprising six weeks (18 months) of data with a step size of one week (one month), were created.

Considering the forecasting task of hourly check-in events, calendrical information and relative humidity were selected as the sources of context. During the accessible three month period, temperature showed stable daily bounds and oscillation, and precipitation was only observed on four days, hence excluded as understandably insufficient to learn relationships with bike demand. Figure 13 provides a view on the forecasting error (MAE) in the absence and presence of historical and prospective context. In the majority of scenarios, particularly under calendrical variables, we observe that the incorporation of prospective context benefits the results, although not as significantly as in the previous medical emergency domain. Statistical significance improvements against context-unaware forecasters ($\alpha=1E-3$) are only observed when introducing sources of prospective context. The low volume of available data, limiting the generalization ability of the underlying forecasters, is hypothesized as the major cause for the moderate improvements.

Complementarily, Fig. 14 depicts the distribution of the errors along the prediction horizon, highlighting that the gains from incorporating prospective context data are generally statistically significant between 7AM and 8PM. The higher impact of calendrical information on bike demand throughout the daylight period, along with the

⁶ <https://emel.city-platform.com/opendata> (accessed December 1, 2022). <https://dados.cm-lisboa.pt/datas-et/gira-bicicletas-de-lisboa-historico> (accessed Dec. 1, 2022).

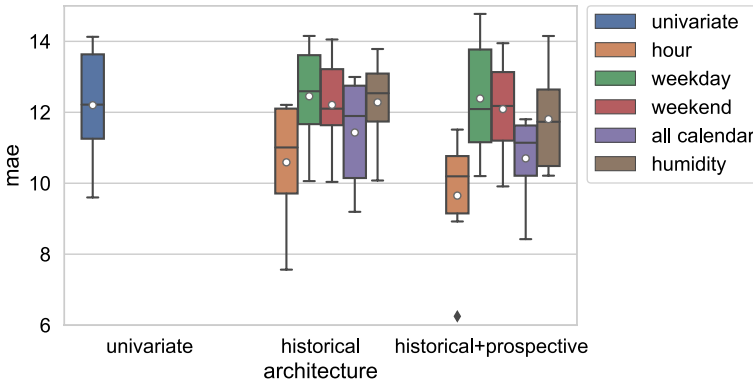


Fig. 13 Impact of incorporating calendrical context together with weather context in the BSS demand forecasting models

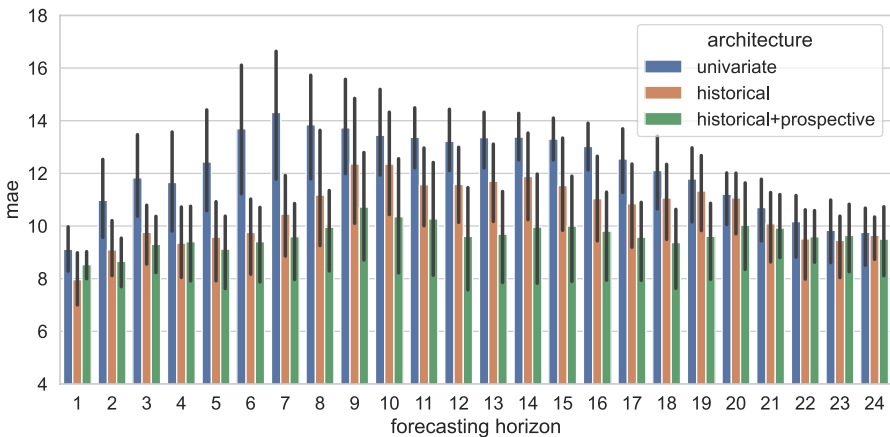
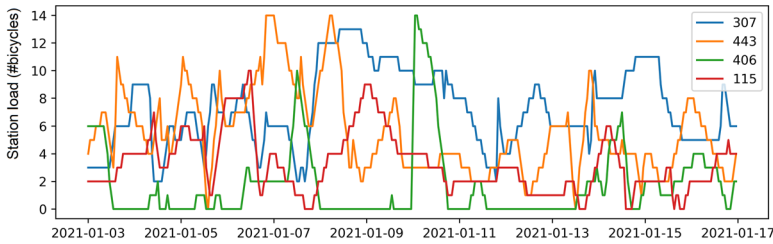


Fig. 14 Comparison of predictive errors (MAE) along the 24-hour forecasting horizon when incorporating historical and prospective context

greater weather influence on transport modal choices due to the adequate offer of public transport alternatives during this period, are arguably linked to this observation.

Moving from check-in forecasts to the prediction of station-specific hourly load states, we observe that the incorporation of prospective context provides statistically significant improvements for over 2/3 of the approximately 100 stations active during the target period (2020-2022). Figure 15a motivates the intrinsic difficulty of this task by disclosing the irregularity of the hourly load state for different stations along two consecutive weeks. For illustrative purposes, 4 stations with distinct profiles are selected: 2 stations at business centers (307 at Marques de Pombal and 406 at Praça Saldanha); 1 station at a central residential area (443 at Av. Roma); and 1 station at the city periphery (115 at Passeio dos Heróis do Mar). Figure 15b assesses the role of incorporating historical and prospective context sources using the proposed



(a) Hourly load state (number of bicycles) of four stations along two early weeks in 2021

<i>statistical significance</i> (at $\alpha=5\%$ for both MAE and RMSE)	$\text{all_context}_{acc} > \text{univariate}_{acc}$	$\text{all_context}_{acc} > \text{hist_context}_{acc}$
307 (Marques de Pombal)	✓	✓
443 (Avenida de Roma)	✓	✗
406 (Praça do Saldanha)	✓	✓
115 (Passeio dos Heróis do Mar)	✓	✓

univariate_{acc} : context-unaware predictive accuracy
 $\text{hist_context}_{acc}$: historical context-aware predictive accuracy
 all_context_{acc} : historical and prospective context-aware predictive accuracy

(b) Statistical improvements of station load forecasts in the absence and presence of context

Fig. 15 Statistical improvements from historical and prospective context incorporation for the load state forecasting of stations with distinct profiles using multi-input networks parameterized with default deep TCNs

multi-input network (parameterized with deep TCNs) for these stations, showing generalized statistically significant improvements.

6 Conclusions

This work proposed neural processing principles to leverage the performance of prediction tasks in the presence of heterogeneous sources of historical and prospective context data. To this end, we introduce a multiple-input recurrent neural architecture that is a serial composition of two LSTM-based components, where the former component is placed to capture historical cross-variable relationships, while the second component uses prospective context for the time-dependent correction of forecasts returned by the former component. Masking principles to derive auxiliary endogenous and exogenous series.

Demand analysis of medical emergencies and public bike sharing in the city of Lisbon was considered as case studies to validate the proposed methodology. Results showed that network models incorporating both historical and prospective context provide significantly more accurate forecasts than the counterparts that do not incorporate context or only consider historical context to guide the forecasts. Statistically significant improvements were further confirmed against state-of-the-art forecasters. The role of calendrical context variables, although often disregarded, are shown to

be critical in the targeted urban domains. Finally, the proposed models are able to maintain the errors low as we move forward in the forecasting horizon.

The proposed neural processing units are simplistic, yet effective, providing a basis to capture relationships with available sources of historical and prospective context data, potentially embed within more complex neural-based predictive models. The introduced principles also form a sound and state-of-the-art reference for the assessment of upcoming contributions in the context-aware forecasting field.

The development of superior context-aware forecasters is highlighted as a future line of work. Similarly, we expect to assess the role of complementary masking principles. Finally, as some sources of context data are inherently sparse and their incorporation increases the multivariate order of the available, we further aim at more closely exploring how much context data are necessary to escape generalization difficulties.

Acknowledgements The authors would like to thank Instituto Nacional de Emergência Médica (INEM) for providing us with the emergency data used in the development of this study. This research was funded by Fundação para a Ciência e Tecnologia (FCT) grant number DSAIPA/AI/0044/2018 (Data2Help), DSAIPA/AI/0033/2019 (LAIfeBlood) and DSAIPA/AI/0111/2018 as well as Instituto de Engenharia de Sistemas e Computadores – Investigação e Desenvolvimento (INESC-ID) grant UIDB/50021/2020.

Funding Open access funding provided by FCTIFCCN (b-on).

Data availability The accompanying software and synthetic data are made available at GitHub for testing purposes. Weather data accessible via Meteo|Técnico (<https://meteo.tecnico.ulisboa.pt/>) and IPMA. Planned events and calendrical information for the Lisbon city available at Lisboa Aberta (<https://lisboaaberta.cm-lisboa.pt/index.php/pt/>). GIRA station load data available at Dados Abertos (<https://dados.cm-lisboa.pt/dataset/gira-bicicletas-de-lisboa-historico>). Access to EMS and GIRA check-in datasets subjected to non-disclosure agreements with INEM (Case Study I) and Lisbon Municipality (Case Study II).

Declarations

Conflict of interest The authors declare that there are no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbasi A, Monadjemi A, Fang L, Rabbani H, Zhang Y (2019) Three-dimensional optical coherence tomography image denoising through multi-input fully-convolutional networks. *Comput Biol Med* 108:1–8
- Andreescu M-P, Frost DB (1998) Weather and traffic accidents in montreal, canada. *Climate Res* 9(3):225–230

- Anirban C, Roy-Chowdhury AK (2014) Context-aware activity forecasting. Asian conference on computer vision. Springer, London, pp 21–36
- Apostolopoulos I, Apostolopoulos D, Spyridonidis T, Papathanasiou ND, Panayiotakis G (2021) Multi-input deep learning approach for cardiovascular disease diagnosis using myocardial perfusion imaging and clinical data. *Phys Med* 84:168–177
- Arvan M, Fahimnia B, Reisi M, Siemsen E (2019) Integrating human judgement into quantitative forecasting methods: A review. *Omega* 86:237–252
- Bartoli F, Lisanti G, Ballan L, Del Bimbo A (2018) Context-aware trajectory prediction. In: 24th International Conference on Pattern Recognition (ICPR), pages 1941–1946. IEEE
- Bi X, Gediminas A, Li W, Qu A (2022) Improving sales forecasting accuracy: a tensor factorization approach with demand awareness. *INFORMS J Comput* 34(3):1644–60
- Box G, Jenkins G, Reinsel G, Ljung G (2015) Time series analysis: forecasting and control. Wiley, New Jersey
- Cerqueira S, Arsenio E, Henriques R (2021) On how to incorporate public sources of situational context in descriptive and predictive models of traffic data. *European Transp Res Rev* 13(1):1–22
- Channouf N, L'Ecuyer P, Ingolfsson A, Avramidis AN (2007) The application of forecasting techniques to modeling emergency medical system calls in calgary, alberta. *Health Care Manag Sci* 10(1):25–45
- Chatfield C (2000) Time-series forecasting. Chapman and Hall/CRC, Boca Raton
- Chen Y, Kang Y, Chen Y, Wang Z (2020) Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing* 399:491–501
- Cuncu E, Manca MM, Pes B, Riboni D (2022) Towards context-aware power forecasting in smart-homes. *Proced Comput Sci* 198:243–248
- Fang Yuchen, Qin Yanjun, Luo Haiyong, Zhao Fang, Wang Chenxing (2021) STJLA: A multi-context aware spatio-temporal joint linear attention network for traffic forecasting. CoRR, [arXiv:2112.02262](https://arxiv.org/abs/2112.02262)
- Garrido Luis A, Mekikis Prodromos-Vasileios, Dalgkitsis Anestis, Verikoukis Christos (2021) Context-aware traffic prediction: Loss function formulation for predicting traffic in 5g networks. In: ICC 2021-IEEE International Conference on Communications, pages 1–6. IEEE
- Guiguet Valentin, Baskiotis Nicolas, Guigue Vincent, Gallinari Patrick (2018) Context-aware forecasting for multivariate stationary time-series
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural computat* 9(8):1735–1780
- Holt CC (2004) Forecasting seasonals and trends by exponentially weighted moving averages. *Int J Forecast* 20(1):5–10
- Huang Chao, Zhang Chuxu, Dai Peng, Bo Liefeng (2019) Deep dynamic fusion network for traffic accident forecasting. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, p 2673–2681
- Jozi A, Pinto T, Vale Z (2022) Contextual learning for energy forecasting in buildings. *Int J Electr Power Energy Syst* 136:107707
- Kam HJ, Sung JO, Park RW (2010) Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Health Inform Res* 16(3):158
- Kamarathi Harshavardhan, Kong Ling kai, Rodríguez Alexander, Zhang Chao, Aditya Prakash B (2021) Camul: Calibrated and accurate multi-view time-series forecasting. CoRR, [arXiv:2109.07438](https://arxiv.org/abs/2109.07438)
- Kingma Diederik P, Ba Jimmy (2014) Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- Kjellstrom T, Butler AJ, Lucas RM, Bonita R (2010) Public health impact of global heating due to climate change: potential effects on chronic non-communicable diseases. *Int J Public Health* 55(2):97–103
- Kuijpers WJP, Antunes DJ, van Mourik S, van Henten EJ, van de Molengraft MJG (2022) Weather forecast error modelling and performance analysis of automatic greenhouse climate control. *Biosyst Eng* 214:207–229
- Leite Inês, Finamore Anna Carolina, Henriques Rui (2020) Context-sensitive modeling of public transport data. Transport Research Arena (TRA2020), Helsinki
- Lemond C, Arsenio E, Henriques R (2021) Integrative analysis of multimodal traffic data: addressing open challenges using big data analytics in the city of lisbon. *European Transp Res Rev* 13(1):1–22
- Li J, Guo F, Sivakumar A, Dong Y, Krishnan R (2021) Transferability improvement in short-term traffic prediction using stacked lstm network. *Transp Res Part C Emerg Technol* 124:102977
- Madhidasan M, Louzazni M, Roy PP (2021) Novel cooperative multi-input multilayer perceptron neural network performance analysis with application of solar irradiance forecasting. *Int J Photoen* 27:1–24

- McCarthy ML, Zeger SL, Ding R, Aronsky D, Hoot NR, Kelen GD (2008) The challenge of predicting demand for emergency department services. *Acad Emergency Med* 15(4):337–346
- Naglah Ahmed, Khalifa Fahmi, Khaled Reem, El-Baz Ayman, et al. (2021) Thyroid cancer computer-aided diagnosis system using mri-based multi-input cnn model. In: 2021 IEEE 18th international symposium on biomedical imaging (ISBI), pages 1691–1694. IEEE
- Neves Francisco, Finamore Anna, Henriques Rui (2020) Efficient discovery of emerging patterns in heterogeneous spatiotemporal data from mobile sensors. In: *MobiQuitous 2020-17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 158–167
- Oktay Ozan, Bai Wenjia, Lee Matthew, Guerrero Ricardo, Kamnitsas Konstantinos, Caballero Jose, de Marvao Antonio, Cook Stuart, O'Regan Declan, Rueckert Daniel (2016) Multi-input cardiac image super-resolution using convolutional neural networks. In: *International conference on medical image computing and computer-assisted intervention*, pages 246–254. Springer
- Pfeffermann D, Allon J (1989) Multivariate exponential smoothing: method and practice. *Int J Forecast* 5(1):83–98
- Rico João, Barateiro José, Oliveira Arlindo (2021) Graph neural networks for traffic forecasting. *CoRR*, [arXiv:2104.13096](https://arxiv.org/abs/2104.13096)
- Ruan T, Dexing W, Chen T, Jin C, Lei X, Zhou S, Jiang Z (2020) Context-aware traffic prediction framework based on series decomposition. *IEEE Access* 8:202848–202857
- Ruder Sebastian (2017) An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*
- Rumelhart David E, Hinton Geoffrey E, Williams Ronald J (1985) Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science
- Sánchez-Cauce R, Pérez-Martín J, Luque M (2021) Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data. *Comput Methods Progr Biomed* 204:106045
- Sardinha Cláudio, Finamore Anna Carolina, Henriques Rui (2021) Context-aware demand prediction in bike sharing systems: incorporating spatial, meteorological and calendrical context. *CoRR*, [arXiv:2105.01125](https://arxiv.org/abs/2105.01125)
- Schäfer Maximilian, Zhao Kun, Bühren Markus, Kummert Anton (2022) Context-aware scene prediction network (casenet)
- Schürholz D, Kubler S, Zaslavsky A (2020) Artificial intelligence-enabled context-aware air quality prediction for smart cities. *J Cleaner Product* 271:121941
- Shi J, Ye Y, Zhu D, Lianta S, Huang Y, Huang J (2021) Automatic segmentation of cardiac magnetic resonance images based on multi-input fusion network. *Comput Methods Progr Biomed* 209:106323
- Silva Leonor, Galhardas Helena, Manquinho Vasco, Henriques Rui (2021) Uniano: robust and efficient anomaly consensus in time series sensitive to cross-correlated anomaly profiles. In: *SIAM International Conference on Data Mining (SDM)*, pages 324–332. SIAM
- Szeto WY, Ghosh B, Basu B, O'Mahony M (2009) Multivariate traffic forecasting technique using cell transmission model and sarima model. *J Transp Eng* 135(9):658–667
- Tang L, Duan Z, Zhao Y (2019) Toward using social media to support ridesharing services: challenges and opportunities. *Transp Plann Technol* 42(4):355–379
- Tempelmeier Nicolas, Rietz Yannick, Lishchuk Iryna, Kruegel Tina, Mumm Olaf, Carlow Vanessa Miriam, Dietze Stefan, Demidova Elena (2019) Data4urbanmobility: Towards holistic data analytics for mobility applications in urban regions. In: *Companion Proceedings of The 2019 World Wide Web Conference*, pages 137–145, New York, NY, US. ACM
- Tiam-Lee Thomas, Henriques Rui, Costa Jose, Manquinho Vasco, Galhardas Helena (2022) Consolidation of massive medical emergency events with heterogeneous situational context data sources. In: *6th IW on Data Analytics solutions for Real-Life Applications*, EDBT/ICDT
- Wang Y, Yang M, Li Y, Zeda X, Wang J, Fang X (2021) A multi-input and multi-task convolutional neural network for fault diagnosis based on bearing vibration signal. *IEEE Sens J* 21(9):10946–10956
- Wibisono Ari, Sina Ibnu, Andri Ihsannuddin M, Hafizh Ahmad, Hardjono Benny, Nurhadiyatna Adi, Jatmiko Wisnu, et al. (2012) Traffic intelligent system architecture based on social media information. In: *2012 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 25–30. IEEE
- Wong HT, Lai PC (2012) Weather inference and daily demand for emergency ambulance services. *Emergency Med J* 29(1):60–64

- Xiong C, Zheng J, Liangjin X, Cen C, Zheng R, Li Y (2021) Multiple-input convolutional neural network model for large-scale seismic damage assessment of reinforced concrete frame buildings. *Appl Sci* 11(17):8258
- Xu Yiming, Paliwal Mudit, Zhao Xilei (2021) Real-time forecasting of dockless scooter-sharing demand: A context-aware spatio-temporal multi-graph convolutional network approach. *CoRR*, [arXiv:2111.01355](https://arxiv.org/abs/2111.01355)
- Yannis George, Karlaftis Matthew G (2010) Weather effects on daily traffic accidents and fatalities: a time series count data approach. In: *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, volume 10, page 14
- Yu Bing, Yin Haoteng, Zhu Zhanxing (2017) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint* [arXiv:1709.04875](https://arxiv.org/abs/1709.04875)
- Zhang Yu, Yang Q (2018) An overview of multi-task learning. *National Sci Rev* 5(1):30–43
- Zhang Y-D, Zhang Z, Zhang X, Wang S-H (2021) Midcan: a multiple input deep convolutional attention network for covid-19 diagnosis based on chest ct and chest x-ray. *Patt Recognit Lett* 150:8–16
- Zhong Huan, Li Li, Ren Jiansi, Wei Wu, Wang Ruoxiang (2022) Hyperspectral image classification via parallel multi-input mechanism-based convolutional neural network. *Multimedia Tools and Applications* 1–26
- Zhu X, Ninh A, Zhao H, Liu Z (2021) Demand forecasting with supply-chain information and machine learning: evidence in the pharmaceutical industry. *Product Operat Manag* 30(9):3231–3252

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.