Deep Echo State Networks in Industrial Applications

Stefano Dettori[1](✉), Ismael Matino[1], Valentina Colla[1], and Ramon Speets[2]

[1] Scuola Superiore Sant'Anna - TeCIP Institute – ICT-COISP, Pisa, Italy
{s.dettori,i.matino,v.colla}@santannapisa.it
[2] Tata Steel, IJmuiden, The Netherlands
ramon.speets@tatasteeleurope.com

Abstract. This paper analyzes the impact of reservoir computing, and, in particular, of Deep Echo State Networks, to the modeling of highly non-linear dynamical systems that can be commonly found in the industry. Several applications are presented focusing on forecasting models related to energy content of steelwork byproduct gasses. Deep Echo State Network models are trained, validated and tested by exploiting datasets coming from a real industrial context, with good results in terms of accuracy of the predictions.

Keywords: Reservoir computing · Deep Echo State Networks · Industrial application

1 Introduction

These last few years have shown a dizzying growth trend in the application of Artificial Intelligence (AI) techniques, not only in the academic world, in which AI is considered a standard technique, with many fields that have reached a satisfactory level of development, but also outside the academia and, in particular, in industrial context. The new consciousness of some industrial sectors, and in particular the steel industry, towards the concepts of circular economy and sustainable development requires to tackle challenging technological and scientific problems, such as the optimization of the use and reuse of energy sources. At the same time, the need arises to improve the product quality in order to face the increasing competition in the markets of goods and services. Nowadays stakeholders are converging towards the idea that AI is a clear turning point for optimally addressing current and future challenges [1] and have begun, in the last decade, a race to cancel the knowledge gaps that exist in the industrial world between the engineering experience, which is currently based on standard modeling, optimization and control techniques and the so-called *data-driven* tools and techniques. Among these latter ones, surely AI plays a fundamental role in facing the challenge of digitalization [2] and nowadays the level of understanding and, above all, of acceptance of AI as an effective and reliable technique [3] is surely increased.

Published by Springer Nature Switzerland AG 2020
I. Maglogiannis et al. (Eds.): AIAI 2020, IFIP AICT 584, pp. 53–63, 2020.
https://doi.org/10.1007/978-3-030-49186-4_5

In this context, reservoir computing is inserted, a tool capable of making the philosophy and technique underlying the AI, aimed at the study and modeling of timeseries and dynamic processes, more understandable and affordable in terms of computation skills, efficiency and quality of obtained results. Reservoir computing has been introduced by the work of Maass et al., in which has been described a particular Recurrent Neural Network (RNN) architecture, called *Liquid-state machine* [4].

The literature in the field of reservoir computing applied to industrial processes is quite extensive, and in particular among the various techniques emerges that based on the Echo State Network, thanks to characteristics that make this approach more attractive and effective. Interesting examples of their application are shown in the work of Wang [5] in which ESN and sparse adaboost method are exploited to forecast the electricity consumption in industrial areas; Matino presented a work related to the prediction of blast furnace gas through ESN techniques [6] and Dettori applied several AI methodologies aiming at modelling energy transformation equipment [7]. Colla et al. presented a work related to the use of outlier detection and advanced variable selection to reservoir application in industry [8]. Another interesting work related to the application of ESN to Model predictive control methodologies is presented by Pan [9].

This work presents some models developed within the European project entitled “Optimization of the management of the process gas network within the integrated steelworks” (GASNET), which aims at supporting the optimal use and distribution of valuable energy resources and byproduct gasses while minimizing the environmental impact. This problem is of utmost importance for integrated steelworks, i.e. the industrial plants which produce steel from virgin raw material, as considerable savings in CO_2 emissions as well as in natural gas consumptions can be achieved by means of an optimal distribution of the off-gases, such as discussed in [10–12]. In particular, in the present work, the models are developed through a recent reservoir computing methodology called Deep Echo State Network (DESN). The novelty of the work consists in the application of this novel Neural Network (NN) architecture, which allows modelling complex nonlinear dynamics that can be typically found in industrial processes. The effectiveness of the proposed methodology has ben compared respect to other state of art neural network architectures.

The paper is organized as follows: Sect. 2 describes the DESN architecture, Sect. 3 presents processes, models and datasets used, Sect. 4 describes methods and results, while Sect. 5 provides some concluding remarks.

2 Deep Echo-State Network Architecture

ESN is an efficient tool and a universal uniform approximator [13], well known for its intrinsic capability of reconstructing complex dynamical input/output relationships. The concept behind the ESN approach is to generate within a reservoir a rich set of dynamics starting from the exciting input. The frequency information content of the input is somehow distorted and enriched through the non-linear reservoir filter and then used to compute a regression on a target. In the last decade, the research in the field of reservoir computing has deepened the study of the characteristics of ESNs up to a further evolutionary step in its architecture, which borrows the concepts introduced by

Deep Learning and amplifies its effectiveness through the use of algorithms that do not use heavy backpropagation routines. The resulted DESN approach introduced by Gallicchio [14, 15] consists of the exploitation of N reservoirs r connected in series in a deep learning fashion and a readout that collects all the reservoirs dynamics to compute the output of the network, as shown in Fig. 1.

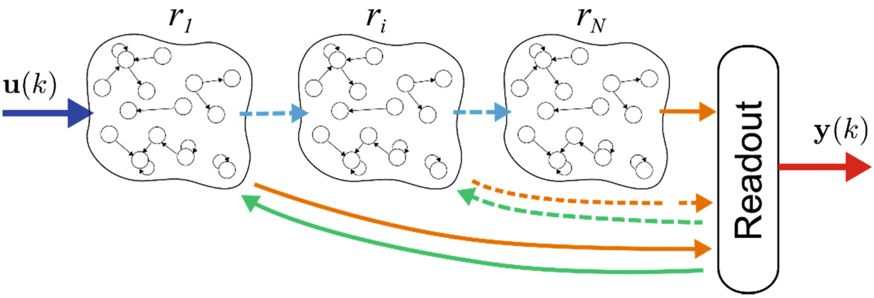

Fig. 1. Architecture of a DESN.

At time k, the state of neurons in each reservoir layers $\boldsymbol{x}_i(k)$ is updated due to the action of its exciting input. The first reservoir layer receives in input the vector $\boldsymbol{u}(k)$, the following layers receive in input the updated state of the previous reservoir layer $\boldsymbol{x}_{i-1}(k)$, whose dynamic is updated according to the following equations:

$$\boldsymbol{x}_1(k) = f\left(c_{\text{in}} \boldsymbol{W}_{\text{in}_1} \boldsymbol{u}(k) + \boldsymbol{W}_{r_1} \boldsymbol{x}_1(k-1) + \boldsymbol{v}_1(k)\right) \tag{1}$$

$$\boldsymbol{x}_i(k) = f\left(c_{\text{is}} \boldsymbol{W}_{in_i} \boldsymbol{x}_{i-1}(k) + \boldsymbol{W}_{r_i} \boldsymbol{x}_i(k-1) + \boldsymbol{v}_i(k)\right) \tag{2}$$

where i is the i-th reservoir layer, f is typically a tanh function, c_{in} and c_{is} are respectively the input scaling and inter-scaling factors, $\boldsymbol{W}_{\text{in}_1}$ and $\boldsymbol{W}_{in_i}$ are, respectively, the input matrix of the first and i-th reservoir layer with dimensions $n_1 \times n_{\text{in}}$ and $n_i \times n_{\text{i}-1}$, n_i is the number of neurons of the i-th reservoir, $\boldsymbol{W}_{r_i}$ is the reservoir matrix of i-th reservoir, $\boldsymbol{y}$ is the output of the DESN and $\boldsymbol{v}_i$ is a small amplitude white noise. The output of the readout is updated as:

$$\boldsymbol{y}(k) = f_{\text{out}}(\boldsymbol{W}_{\text{out}} \boldsymbol{x}(k)) \tag{3}$$

where $\boldsymbol{x}(k)$ is the complete vector of states, f_{out} is the readout neuron function, typically the identical function for time series regression tasks, $\boldsymbol{W}_{\text{out}}$ a $n_{\text{out}} \times n_T$ matrix and n_T is the total number of reservoir neurons of the DESN.

Some of the aspects that make the approach particularly effective compared to those based on other recurrent networks are the architectural characteristics of the DESN and the training procedures that focus only in the weights of the readout.

More in detail, as said before, the only scope of the reservoir is to generate a rich set of dynamics, while the readout has to combine them in effective way, in order to minimize the regression error through non-iterative algorithms. The training phase is carried out in two sequential routines: The reservoir and input initializations and the readout training.

In the first phase, input and reservoir matrixes are initialized with the only constraint of obtaining a reservoir characterized by stable dynamics. In other words, the state of the reservoir has to be contractive and each neuron has to gradually forget their previous activation. These properties are summarized in the so-called Echo State Property (ESP), widely studied and deepen in several works of Yildiz and Jaeger [16] and then extended also to the case of DESN in the work of Gallicchio [15]. These works define necessary, sufficient conditions and empirical guidelines for the design of a contractive reservoir.

More in detail, each reservoir matrix $\boldsymbol{W}_{r_i}$ is initialized as a sparse randomized $\hat{\boldsymbol{W}}_{r_i}$, which entries are defined in the range $[-1, 1]$. The sparsity, defined as the percentual of non-zero element over the total, is the percentual number of connections within the reservoir. This parameter is typically set below 5%.

The first step to design a single contractive reservoir is to scale its matrix $\hat{\boldsymbol{W}}_{r_i}$ in order to obtain the desired spectral radius $\tilde{\rho}_i$:

$$\boldsymbol{W}_{r_i} = \tilde{\rho}_i \frac{\hat{\boldsymbol{W}}_{r_i}}{\rho\left(\hat{\boldsymbol{W}}_{r_i}\right)} \tag{4}$$

In the case of DESN, the necessary condition to guarantee the ESP is that the greater of the spectral radius of all the reservoirs is less than one. This condition is a guideline to design a contractive global reservoir, but empirically is also sufficient. The spectral radius of each layer is an important hyperparameter, whose choice is subject to optimization. While the reservoir system is initialized, also each reservoir input matrix $\boldsymbol{W}_{in_i}$ is initialized randomly with weights in the range $[-1, 1]$. In the Eq. (1) and (2) these matrixes are actually scaled by further coefficients c_{in} and c_{is}, namely input scaling and inter-scaling factors, that are important hyperparameters and subject to optimization. Once all these matrixes are initialized and left untrained, it is possible to train the readout by exploiting linear regression routines. A straightforward solution is the Tikhonov regularization least square algorithm, calculated on the training dataset:

$$\boldsymbol{W}_{\text{out}} = \bar{\boldsymbol{Y}}\boldsymbol{X}^T\left(\boldsymbol{X}\boldsymbol{X}^T + \lambda\boldsymbol{I}\right)^{-1} \tag{5}$$

where $\bar{\boldsymbol{Y}}$ and $\boldsymbol{X}$ are respectively the sequences of targets and states of all the reservoirs, the latter calculated by using Eqs. (1) and (2). λ is the regularization coefficient that allows to solve also ill-posed inversions and, in general, could be object of optimization during the hyperparameter selection phase.

3 Models and Datasets

Between the several processes involved during the development of GASNET project, in this paper we present two case studies related to the prediction of energetic content and chemical characteristics of Blast Furnace (BF) off-gas, a particular byproduct gas (BFG) generated during the production of pig iron that can be reused as a valuable energy carrier to produce electrical energy or process steam. In particular, for the aims of GASNET project, it is useful to predict the future behavior and characteristics of the off-gas by using a restrict number of process measurements and future knowledge of the process scheduling for 2 h ahead. Together with the prediction of the BFG production, this paper presents an application in which is consumed, the hot blast stoves, also called Cowpers.

In particular, in this process the preheating of the air takes place, subsequently blown inside the BF for the production of pig iron, through the combustion of BFG and other byproduct gasses typically available in some integrated steelworks, such as Coke Oven Gas (COG).

The targets of the first two models are the BFG volume flow and its Net Calorific Value (NCV). For each target will be developed a specific model capable of predicting the entire sequence of 2 h ahead dynamic behavior in a one-shot multistep manner.

The targets of the third and fourth model are respectively the consumption of BFG and COG burned in the cowpers. Also in this case a specific model capable of predicting the entire sequence of 2 h ahead dynamic behavior in a one-shot multistep manner.

The models will be trained exploiting real data of industrial partners, through datasets related to a period of 30 days with sampling time of 1 min, sufficient to describe the main dynamics of the process.

Data pre-processing has represented a fundamental step for models design. Unreliable data have been identified through suitable outliers detection techniques [17, 18]. Moreover, the inputs of each model have been selected by exploiting a variable selection approach based on Genetic Algorithms [19, 20].

All the models have in input the scheduling of the respective process for 2 h ahead, a Boolean information that describes if the respective process is active or not. The model that predicts BFG Flow has in input also the current measurements of the O_2 content in the cold wind, the cold wind volume flow, the pressure of the hot wind and the BFG volume flow. The model that predicts the BFG NCV has in input also the same measurements and in addiction the current CO and H_2 contents in the BFG. The third model, which predicts the BFG consumption in the cowpers, takes in input its abovementioned process scheduling and the current measurements of the cold wind flow and the BFG and COG consumed in the cowpers. The fourth model, which predicts the COG consumption in the cowpers takes the same inputs of the previous model and in addiction the future 2 h predictions of BFG NCV. The input/output architecture of each model is depicted in Fig. 2, while the inputs and target dataset descriptions and Units of Measurement (UoM) are shown in the Table 1.

4 Methods and Numerical Results

In order to evaluate the effectiveness of the DESN architectures for predicting the future behavior of the considered processes, a comparison is proposed between the results achievable through DESN and another rather efficient architecture for modeling time-series characterized by non-linear dynamics, i.e. the Long Short-Term Memory (LSTM) [21]. In this work, the LSTM architecture is configured a series of input layer, a LSTM layer, L_{LSTM} fully connected layers, and a linear readout.

Each model has been developed through a systematic procedure. In first place, it is necessary to define an optimal architecture for each model, by choosing a good set of

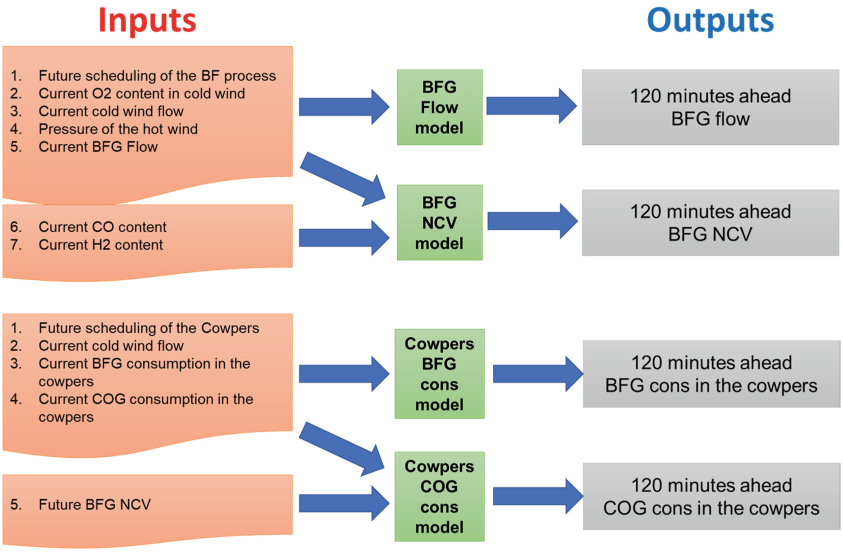
Inputs
Outputs
1. Future scheduling of the BF process
2. Current O2 content in cold wind
3. Current cold wind flow
4. Pressure of the hot wind
5. Current BFG Flow
6. Current CO content
7. Current H2 content
BFG Flow model
BFG NCV model
120 minutes ahead BFG flow
120 minutes ahead BFG NCV
1. Future scheduling of the Cowpers
2. Current cold wind flow
3. Current BFG consumption in the cowpers
4. Current COG consumption in the cowpers
5. Future BFG NCV
Cowpers BFG cons model
Cowpers COG cons model
120 minutes ahead BFG cons in the cowpers
120 minutes ahead COG cons in the cowpers

Fig. 2. Input/Output architecture of the models

Table 1. Input/output descriptions of each model

Variables	Description	UoM
Scheduling of BFG process	Boolean variable aimed at describing the status of the BF process. On (1), Off (0)	–
Scheduling of cowpers	Boolean variable aimed at describing the status of the cowpers. On (1), Off (0)	–
O_2 content in cold wind	Volume percentage of oxygen in the cold wind in input to the BF after heating	%
Cold wind flow	Volume flowrate of cold wind in input to the BF after heating	m^3/h
Hot wind pressure	Pressure of hot wind in input to the BF	bar
BFG production	Volume of produced BF gas	m^3/h
CO content	Volume percentage of carbon monoxide in BF gas	%
H_2 content	Volume percentage of hydrogen in BF gas	%
BFG consumption in the cowpers	Consumption of BF gas in cowpers	m^3/h
COG consumption in the cowpers	Consumption of COG gas in cowpers	m^3/h

hyperparameters. For the DESN the considered hyperparameters are the of number of layers and neurons of each reservoir, the spectral radius, the input scaling factor and the inter-layer scaling factor. For the LSTM the considered hyperparameters are the number of fully connected layers in series L_{LSTM}, and the number of neurons of each fully connected layer, with the i-th layer characterized by N_{LSTM_i} neurons.

For the sake of simplicity, each layer of the DESN has the same number of neurons N_{DESN} and the same spectral radius ρ_L. Moreover, each layer after the first has the same inter-layer scaling factor. The selection of the hyperparameters is an open topic for scientific research, and several works in literature give guidelines or recommend some particular algorithms for their choice, such as [22]. In our work their optimization is carried out through a *random search technique* [23] during which 1000 training trials have been carried out using a uniform distribution in the intervals: number of reservoir layers $L_{DESN} = [2\ 12]$; Total Number of Reservoir Neurons = [200 2000], $\rho_L = [0.1\ 1]$; Input scaling $c_{\text{in}} = [0.01\ 10]$; Inter Scaling $c_{\text{is}} = [0.1\ 1]$.

In the case of LSTM, the random search is performed over 1000 training trials with a uniform distribution of the hyperparameters in the intervals: $L_{LSTM} = [1\ 10]$, $N_{LSTM_i} = [30\ 300]$. The LSTMs have been trained through the ADAM training method [24]. The random search algorithm does not allow finding the best set of hyperparameters, but in general a good approximation of the optimal solution.

The optimal set of DESN and LSTM hyperparameters minimizes the mean value of the Normalized Root Mean Square Error (NRMSE) of all the outputs of the model ***mNRMSE***, evaluated on the validation set. The comparison between the two different architectures is assessed by evaluating the *mNRMSE,* which is a particularly robust index with respect to the Mean Absolute Percent Error (MAPE) or other common metrics, due to the formulation that takes into account the overall range of the targets. Furthermore, the MAPE is not an adequate measure when intermittent target values (too many values equal or near to 0) are treated.

$$\boldsymbol{mNRMSE} = 100\frac{1}{n_y}\sum_{j=1}^{n_y}\left(\frac{\sqrt{\frac{1}{N_s}\sum_{k=1}^{N_s}\left(\bar{y}_j(k)-y_j(k)\right)^2}}{\max(y_j)-\min(y_j)}\right) \tag{6}$$

The dataset is composed of about 45000 samples, divided in two parts: the first 50% is used for the choice of hyperparameters and subsequently for training the optimal networks, the remaining 50% is used for the test of the trained models. During the phase of the hyperparameters selection, the first fraction of the overall dataset is in turn divided into 60% training and 40% validation.

The DESN results related to the trained optimal network are summarized in Table 2, which shows the results on the test dataset and the optimal architecture of each model. The comparison between the DESN- and LSTM-based architectures is reported in Table 3, in terms of ***mNRMSE*** on the training and test dataset.

In the one-shot multistep ahead forecast approach, the k-th output is referred to k sample ahead prediction. For all the models, the error is low in the first 10–20 samples of prediction and, as it can normally be expected, it tends to increase as we want to predict the phenomenon in the distant future. An example of prediction 2 h ahead of the BFG and COG consumption in the cowpers through DESN and LSTM architectures are shown in Fig. 3 and Fig. 4, respectively, which allow a comparison between the measured-target values (in blue) and the LSTM and DESN forecasted values (respectively in orange and yellow). In deeper detail, the figures show an example of one-shoot multistep prediction of the abovementioned process for a specific instant of prediction, during which the

Table 2. Optimal DESN architectures and test results of each model.

Model	L_{DESN}	N_{DESN}	ρ_L	c_{in}	c_{is}	Test mNRMSE
BFG flow	7	90	0.549	0.154	0.952	6.7
BFG NCV	5	139	0.991	0.214	0.031	7.3
Cowpers BFG cons	5	293	0.447	5.10	0.022	6.09
Cowpers COG cons	5	180	0.722	0.073	0.691	9.87

Table 3. Comparison between DESN and LSTM architectures

Model	Architecture	Training mNRMSE	Test mNRMSE
BFG flow	DESN	5.02	6.70
	LSTM	7.95	10.92
BFG NCV	DESN	5.93	7.31
	LSTM	8.36	16.6
Cowpers BFG consumption	DESN	4.60	6.09
	LSTM	4.95	6.38
Cowpers COG consumption	DESN	8.08	9.87
	LSTM	11.6	13.4

target is characterized by a rich dynamic content. For confidentiality constraints, in the figures the absolute ranges of the measured and predicted values are normalized.

The results achieved during the test of the modelling approach are very encouraging. In particular, the models related to the BFG production and its energy contents are characterized of errors around 7% that, considering the heavy nonlinearity of the multi input multi output BF process, for control application are very low. This allows predicting the BFG production with good accuracy, sufficient to optimize its use in an energy control and optimization strategy, and to provide a support to process operators for the following 2 h.

In the case of BFG consumption in the cowpers, the prediction errors are satisfactory. The model for the prediction of COG consumption in the cowpers show a greater error, but also in this case the model can be considered useful in an energy control and optimization strategy.

The comparison between DESN and LSTM shows a clear difference between the results and the quality of the prediction obtainable with the two different architectures. In each proposed modelled process, DESNs outperform LSTMs on both training and test dataset. In particular, with respect to LSTM architecture, DESNs allow obtaining an improvement of the performances of the 4 models, equal to 4.22%, 9.29%, 0.29% and 3.53%, respectively.

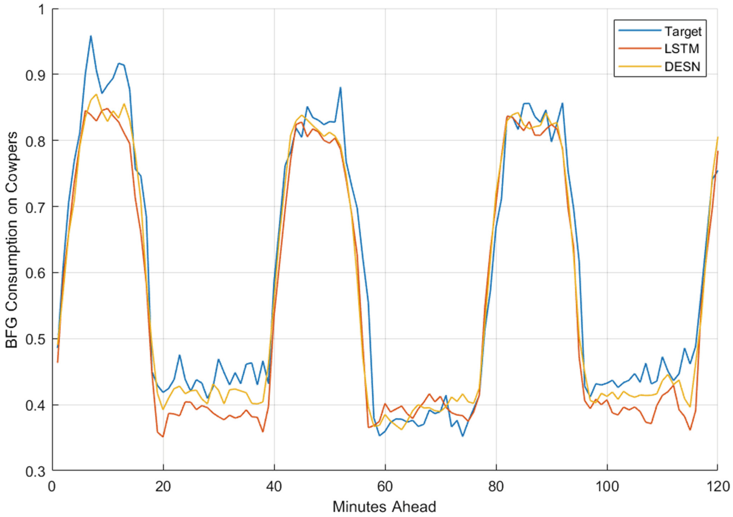

Fig. 3. Prediction example of BFG consumption in the cowpers.

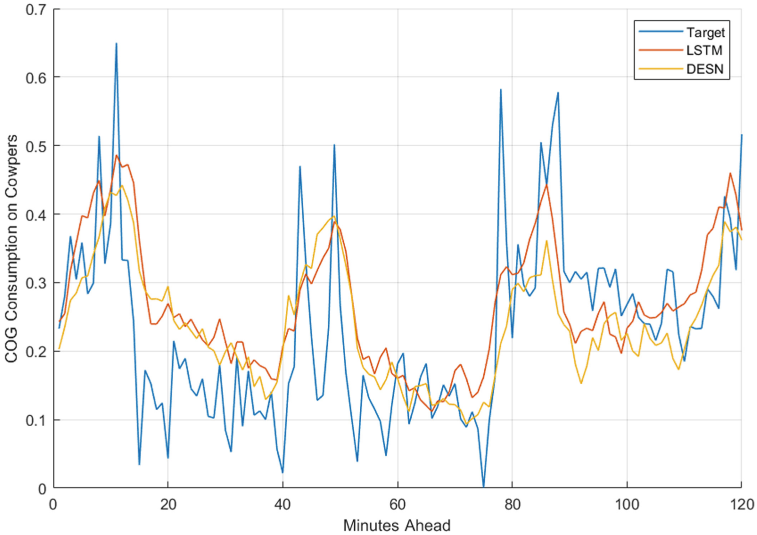

Fig. 4. Prediction example of COG consumption in the cowpers.

5 Conclusions

This paper describes the application of a particular reservoir computing technique called Deep Echo State Network to the modelling of nonlinear dynamics typical of complex industrial processes. The presented case studies are the forecast of energetic content in blast furnace gasses, produced during the production of pig iron in steelworks and one application of its consumption in the process called Hot Blast Stoves (Cowpers). The models have been trained, validated and tested by using real data. In particular, the hyperparameters of the DESN-based models are optimized through a random search approach that aims to minimize the validation error. The proposed DESN-based methodology has been compared with an LSTM-based architecture in order to assess the accuracy with respect to the state of art. The results show a great advantage in using DESNs to model the dynamic behavior of the considered processes, with respect to the LSTM architecture. The achieved results are satisfactory and the trained models are effectively used inside a control strategy for the optimal distribution of byproduct gasses aiming at minimize the environmental impact of steelworks.

Acknowledgments. The work described in the present paper was developed within the project entitled "Optimization of the management of the process gases network within the integrated steelworks - GASNET" (Contract No. RFSR-CT-2015-00029) and received funding from the Research Fund for Coal and Steel of the European Union, which is gratefully acknowledged. The sole responsibility of the issues treated in the present paper lies with the authors; the Union is not responsible for any use that may be made of the information contained therein.

References

1. Brynjolfsson, E., Rock, D., Syverson, C.: Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics. National Bureau of Economic Research (2017). http://doi.org/10.3386/w24001
2. Branca, T.A., Fornai, B., Colla, V., Murri, M.M., Streppa, E., Schröder, A.J.: The challenge of digitalization in the steel sector. Metals **10**(2), 288 (2020). https://doi.org/10.3390/met10020288
3. Ransbotham, S., Kiron, D., Gerbert, P., Reeves, M.: Reshaping business with artificial intelligence: closing the gap between ambition and action. MIT Sloan Manag. Rev. **59**(1) (2017)
4. Maass, W., Natschläger, T., Markram, H.: Real-time computing without stable states: a new framework for neural computation based on perturbations. Neural Comput. **14**(11), 2531–2560 (2002)
5. Wang, L., Lv, S.-X., Zeng, Y.-R.: Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in China. Energy **155**, 1013–1031 (2018). https://doi.org/10.1016/j.energy.2018.04.175
6. Matino, I., Dettori, S., Colla, V., Weber, V., Salame, S.: Forecasting blast furnace gas production and demand through echo state neural network-based models: pave the way to off-gas optimized management. Appl. Energy **253**, 113578 (2019). https://doi.org/10.1016/j.apenergy.2019.113578
7. Dettori, S., Matino, I., Colla, V., Weber, V., Salame, S.: Neural network-based modeling methodologies for energy transformation equipment in integrated steelworks processes. Energy Procedia **158**, 4061–4066 (2019). https://doi.org/10.1016/j.egypro.2019.01.831
8. Colla, V., Matino, I., Dettori, S., Cateni, S., Matino, R.: Reservoir computing approaches applied to energy management in industry. In: Macintyre, J., Iliadis, L., Maglogiannis, I., Jayne, C. (eds.) EANN 2019. CCIS, vol. 1000, pp. 66–79. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20257-6_6
9. Pan, Y., Wang, J.: Model predictive control of unknown nonlinear dynamical systems based on recurrent neural networks. IEEE Trans. Industr. Electron. **59**(8), 3089–3101 (2011). https://doi.org/10.1109/tie.2011.2169636

10. Maddaloni, A., Porzio, G.F., Nastasi, G., Colla, V., Branca, T.A.: Multi-objective optimization applied to retrofit analysis: a case study for the iron and steel industry. Appl. Therm. Eng. **91**, 638–646 (2015). https://doi.org/10.1016/j.applthermaleng.2015.08.051
11. Colla, V., et al.: Assessing the efficiency of the off-gas network management in integrated steelworks. Materiaux et Techniques **107**(1), 104 (2019). https://doi.org/10.1051/mattech/2018068
12. Porzio, G.F., et al.: Process integration in energy and carbon intensive industries: an example of exploitation of optimization techniques and decision support. Appl. Therm. Eng. **70**(2), 1148–1155 (2014). https://doi.org/10.1016/j.applthermaleng.2014.05.058
13. Grigoryeva, L., Ortega, J.P.: Echo state network are universal. Neural Netw. **108**, 495–508 (2018). https://doi.org/10.1016/j.neunet.2018.08.025
14. Gallicchio, C., Micheli, A., Pedrelli, L.: Deep reservoir computing: a critical experimental analysis. Neurocomputing **268**, 87–99 (2017). https://doi.org/10.1016/j.neucom.2016.12.089
15. Gallicchio, C., Micheli, A.: Echo state property of deep reservoir computing networks. Cogn. Comput. **9**(3), 337–350 (2017). https://doi.org/10.1007/s12559-017-9461-9
16. Yildiz, I.B., Jaeger, H., Kiebel, S.J.: Re-visiting the echo state property. Neural Netw. **35**, 1–9 (2012). https://doi.org/10.1016/j.neunet.2012.07.005
17. Cateni, S., Colla, V., Nastasi, G.: A multivariate fuzzy system applied for outliers detection. J. Intell Fuzzy Syst. **24**(4), 889–903 (2013). https://doi.org/10.3233/ifs-2012-0607
18. Cateni, S., Colla, V., Vannucci, M.: A fuzzy logic-based method for outliers detection. In: Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA 2007), pp. 561–566 (2007)
19. Cateni, S., Colla, V., Vannucci, M.: General purpose input variables extraction: a genetic algorithm based procedure GIVE a GAP. In: 9th International Conference on Intelligent Systems Design and Applications (ISDA 2009), pp. 1278–1283 (2009). https://doi.org/10.1109/isda.2009.190
20. Cateni, S., Colla, V., Vannucci, M.: A genetic algorithm-based approach for selecting input variables and setting relevant network parameters of a SOM-based classifier. Int. J. Simul. Syst. Sci. Technol. **12**(2), 30–37 (2011). https://doi.org/10.1109/ems.2010.23
21. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
22. Gallicchio, C., Micheli, A., Pedrelli, L.: Design of deep echo state networks. Neural Netw. **108**, 33–47 (2018). https://doi.org/10.1016/j.neunet.2018.08.002
23. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. J. Mach. Learn. Res. **13**, 281–305 (2012)
24. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)