Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 53-2 (2020) 10417–10422

LSTM water prediction for feedforward control of moulding sand compressibility ⋆

Alexander Rose * **Alexander Seel** ** **Bennet Luck** ** **Martin Grotjahn** *

* *IKME – Advanced Control, Hochschule Hannover, 30173 Hannover, Germany (e-mail: {alexander.rose, martin.grotjahn}@hs-hannover.de)*
** *Automation Solutions Energy Sector, IAV GmbH, Germany (e-mail: {alexander.seel, bennet.luck}@iav.de)*

Abstract: This paper presents a databased approach for improving the precision of the moulding sand compressibility in the moulding sand mixer of a foundry. In this approach, the deviation between the measured and the target compressibility is reduced by controlling the water addition. The complex dynamic behaviour of the process variables and their influence on the water addition is modelled with a long short-term memory (LSTM) network. Another LSTM network as control path simulates the impact of the water addition on the compressibility. Simulation and experimental results with the applied model for water prediction in a feedforward control yield relevant improvements of the moulding sand compressibility.

Keywords: neural control, neural network model, industrial production system, prediction methods, intelligent control, batch control, target control, feedforward control

1. INTRODUCTION

Process stability is an important criterion to reduce production interruptions at production plants. In green sand casting, the moulding sand compressibility is an essential influencing factor for the quality of casting moulds and cast products [Tilch et al. (2015)]. Influencing factors on the sand compressibility are for example the addition of water and bentonite [1]. These additives are diminished from the moulding sand during the casting process by thermal wear or other losses [VDG (1989b)]. In practice, the compressibility is typically controlled in the moulding sand mixer [Michenfelder et al. (2014), Seeber and Kohler (2012)]. For this purpose different possibilities for control algorithms exist, for instance fuzzy logic [Gemming (2003)].

Exemplified by the foundry Heinrich Meier Eisengießerei GmbH & Co. KG in Rahden, the compressibility is discontinuously controlled by manipulating the water addition. The water control algorithms react on proportional differences between measured and target compressibility. Batch-related, abrupt changes in the compressibility cannot be fully compensated in the following batch. This can lead to typical casting defects such as demolished moulds and cod breakage [Tilch et al. (2015)]. In case of significant defects of this type, a mould cannot be used which may result in production interruptions. The replacement of the defective moulds and thus the higher demand of moulding sand leads to higher energy costs, higher capacity utilization and higher emissions in sand preparation, for example the emission of carbon monoxide [BDG (2019)]. On an international comparison with respect to high energy costs and capacity utilization in Germany [Trinowski (2016)], the reduction of moulding sand offers optimization potential.

Previous work on databased modelling in the sand preparation of the foundry Heinrich Meier Eisengießerei GmbH & Co. KG focused on sand temperature and sand flow. The temporal changes of sand temperature during the sand preparation in the sand cooler could be detected by neural networks [Sommer (2017)]. Furthermore, in this context the prerequisites for a databased approach could be derived [Pätzold (2017)]. An approach for modelling the time characteristic of the moulding sand in the sand mixer could be shown by neural networks [Rose et al. (2020)]. One mature challenge in modelling complex processes are internal dynamics with large time constants, since they require models with storage capabilities. Therefore, this paper presents a databased control approach for the water addition in the moulding sand mixer using a long short-term memory (LSTM) network as prediction model. This prediction model illustrates the dependency of the compressibility and the further process variables on the water addition. The advantage of this method is reducing the deviation from the target compressibility, this is initially simulated with a second model for the control path and then tested as control approach at the production plant.

2. LSTM FOR WATER PREDICTION

Recurrent neural networks (RNN) use feedback loops to store information. Long short-term memory (LSTM) cells are RNNs extended by a long-term memory cell [Hochreiter and Schmidhuber (1997), Manaswi (2018)]. These cell enables LSTM to reproduce correlations in signals with long shifting time constants. LSTM cells combine

⋆ The European Regional Development Fund co-financed this research and publication.

[1] Bentonite belongs to the clays and is responsible for the binding behaviour of the moulding sand [VDG (1989a)].

10.1016/j.ifacol.2020.12.2782

correlation of multiple signals and can be arranged as networks with topologies of e.g. $M \times N$ layers / cells per layer, $M, N \in \mathbb{N}$. A common activation function in the underlying neuronal networks is the sigmoid function

$$\mathrm{sig}(x) := \frac{1}{1+\mathrm{e}^{-x}} = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{x}{2}\right). \quad (1)$$

In an LSTM network the state of the cells at the discrete times $t \in \mathbb{N}$ are denoted as $\mathbf{s}_t \in \mathbb{R}^N$. With the input sequence $\mathbf{u}_t \in \mathbb{R}^\ell$, the output sequence $\mathbf{h}_t \in \mathbb{R}^N$ and the biases $\mathbf{b_f}, \mathbf{b_g}, \mathbf{b_s}, \mathbf{b_o} \in \mathbb{R}^N$, the information from the last cell state $\mathbf{s}_{t-1}$ will be forgotten by

$$\mathbf{f}_t = \mathbf{sig}(\mathbf{H_f h}_{t-1} + \mathbf{U_f u}_t + \mathbf{b_f}),\ \mathbf{f}_t \in \mathbb{R}^N. \quad (2)$$

$\mathbf{U_f}, \mathbf{U_g}, \mathbf{U_s}, \mathbf{U_o} \in \mathbb{R}^{N\times\ell}$ and $\mathbf{H_f}, \mathbf{H_g}, \mathbf{H_s}, \mathbf{H_o} \in \mathbb{R}^{N\times N}$ denote the weighting factors for the input and output sequences respectively. The common mapping of scalar functions to vector representations is used and the elementwise *Hadamard* multiplication is denoted as $\odot$. Information from the last output sequence and the input sequence will be ignored by

$$\mathbf{g}_t = \mathbf{sig}(\mathbf{H_g h}_{t-1} + \mathbf{U_g u}_t + \mathbf{b_g}),\ \mathbf{g}_t \in \mathbb{R}^N. \quad (3)$$

For the discrete time t the cell state is

$$\mathbf{s}_t = \mathbf{f}_t \odot \mathbf{s}_{t-1} + \mathbf{g}_t \odot \mathbf{tanh}(\mathbf{H_s h}_{t-1} + \mathbf{U_s u}_t + \mathbf{b_s}) \quad (4)$$

and the output sequence reads

$$\mathbf{h}_t = \mathbf{sig}(\mathbf{H_o h}_{t-1} + \mathbf{U_o u}_t + \mathbf{b_o}) \odot \mathbf{tanh}(\mathbf{s}_t). \quad (5)$$

The LSTM network output is condensed to a single time series using an extra neuron. The mean squared error is chosen to be the loss function for evaluation. A popular decision criterion for terminating the learning process is the minimum of the loss function which can be found by *early stopping* [Gal and Ghahramani (2016)]. For regularization a *dropout layer* is implemented in the LSTM network to avoid overfitting [Zaremba et al. (2014)].

2.1 Data preprocessing

Production interruptions, sensor errors and data collection failures are eliminated by preprocessing. Measurement data outside the plausible range, such as negative measurement values for masses, are removed from the data set. As the result of the mixing process, the data collection step size varies within 205 ± 110 s. The transformation of the data from time domain to batch domain accumulates the data to batches. The measured water addition $\mathbf{W}$ and further process variables are stored when the sand mixer is emptied. High-frequency components [2] are filtered using the convolution $(x * k)_n = \sum_m k_{n-m} x_m$ of T data points $\mathbf{x} = (x_1, x_2, \ldots, x_T)$ from a data set with the half normalized gaussian kernel

$$k_n \propto \mathrm{e}^{-n^2/2\sigma^2},\ n \geq 0 \,\text{und}\, k_n \equiv 0,\ n < 0 \quad (6)$$

and its width σ. Setting $k_{n<0} \equiv 0$ accounts for causality. The following examinations are based on a data set of approximately 100 days. The data set is separated by days and divided into 80 % training, 10 % validation and 10 % test data. An influence of randomized data sets from different days could not be clearly identified. The training and validation data are used for learning and evaluating the LSTM network. The validation and test data are used for prediction. The data are linearly scaled to a continuous interval of $[0, 1]$ for reducing the learning time. Table 1 shows the unscaled process variables with mean and standard deviation after preprocessing.

[2] High-frequency components are specific for the measuring procedure of compressibility. Pre-compressed sand samples result low measured compressibilities. Air between sand clots yields high measured compressibilities.

Table 1. List of investigated process variables

	process variable	unit	mean	standard deviation
$\mathbf{S}_M$	return sand mass	kg	5899	28.3
$\mathbf{B}_M$	bentonite mass	kg	16.0	2.05
$\mathbf{C}_M$	coal dust mass	kg	14.1	1.63
$\mathbf{S}_{Mo}$	return sand moisture	%	3.17	0.0720
$\mathbf{S}_T$	return sand temperature	$°C$	38.0	3.91
$\mathbf{C}$	sand compressibility	%	39.9	2.34
$\mathbf{A}_H$	air humidity	%	28.7	6.26
$\mathbf{A}_T$	air temperature	$°C$	31.5	5.40
$\mathbf{W}$	water addition	$10^{-3}\ m^3$	54.6	12.0

2.2 Schematic control system

The existing control system for the sand preparation with target compressibility $\mathbf{C}_{\text{target}}$ contains the controller and the pre-dosing, see Fig. 1 (solid lines). The further process variables are summarized here as $\mathbf{D}$. The water addition $\mathbf{W}$ is controlled by the $d \in \mathbb{N}_0$ doses every batch which are classified in the two steps pre-dosing $\mathbf{W}_{\text{pre}}$ with $d = 0$ and post-dosing $\mathbf{W}_{\text{post}}$ with $d > 0$. The counter for d is reset after every batch. For pre-dosing, the value of the last water addition is taken into account by an offset. The size of the offset depends on the difference between the target and measured compressibility $\mathbf{C}$ of the previous batch. Additionally required water can be added by post-dosing after measuring the compressibility. Especially the compressibility measurement is influenced by measure noise. However, to ensure the target mixing time in the sand mixer, only one additional dose per batch can actually be added. In contrast to the existing control system, an LSTM network predicts the water addition with regard to the further process variables $\mathbf{D}$ and a longer time horizon, see Fig. 1 (dashed lines). The LSTM prediction model for water addition is denoted as *Model 1*. In the simulation, the control path is described by *Model 2*.

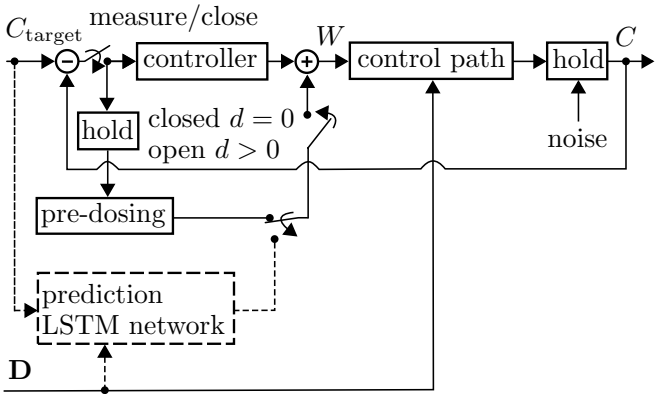

Fig. 1. Schematic control system, additional path for water prediction using LSTM shown in dashed lines

3. MODELLING PROCEDURE

Software basis for implementation of the LSTM networks is *Tensorflow* [Abadi et al. (2015)]. The following procedure is chosen to illustrate the improvement by the one-step-ahead predicted water addition on the compressibility. The steps are:

(1) Sensitivity analyses to figure out the dependency of the water addition $\mathbf{W}$ with respect to the process variables and the network topology, respectively.
(2) Prediction of the water addition $\mathbf{W}_{\mathrm{pred}}$ based on the sensitive process variables as input (*Model 1*).
(3) Simulation of the compressibility $\mathbf{C}_{\circ,\mathrm{sim}}$ with *Model 2* as the control path using the predicted water addition $\mathbf{W}_{\mathrm{pred}}$, the measured water additions $\mathbf{W}_{\mathrm{pre}}$ and $\mathbf{W}_{\mathrm{post}}$ with $\circ = \{\mathrm{pre}, \mathrm{post}, \mathrm{pred}\}$ as inputs.
(4) Test of the prediction model for water addition $\mathbf{W}_{\mathrm{pred}}$ (*Model 1*) at the production plant and benchmarking against the actual water addition for pre-dosing $\mathbf{W}_{\mathrm{pre}}$ by comparing the compressibilities.

The prediction quality results from the comparison of the predicted and measured values and is described by a conformity index. The quantification of this conformity index can be calculated with the Pearson correlation

$$p(\mathbf{a},\mathbf{b}) = \frac{\sum_i (a_i - \overline{a})(b_i - \overline{b})}{\sqrt{\sum_i (a_i - \overline{a})^2 \sum_j (b_j - \overline{b})^2}} \tag{7}$$

for arbitrary $\mathbf{a}, \mathbf{b} \neq \mathbf{0}$. It is rated by the mean value accuracy

$$r(\overline{a},\overline{b}) = 1 - 2\left|\frac{\overline{a} - \overline{b}}{\overline{a} + \overline{b}}\right|, \tag{8}$$

with the progression $\overline{m} := \frac{1}{T}\sum_{i=1}^{T} m_i, \mathbf{m} = \{\mathbf{a}, \mathbf{b}\}$ and ranges $-1 \leq p, r \leq 1$.

3.1 Correlation analysis

For comparison only the preliminary correlation analysis of the signals is based on a covariance matrix calculated with the Pearson correlation (7) on a common time base, see Fig. 2. These results indicate a correlation between water addition $\mathbf{W}$ with return sand moisture $\mathbf{S}_{\mathrm{Mo}}$ (-0.64), return sand temperature $\mathbf{S}_{\mathrm{T}}$ (-0.40) and air temperature $\mathbf{A}_{\mathrm{T}}$ (-0.39). The correlation between the compressibility $\mathbf{C}$ and the water addition (-0.22) is quite small. For the following approaches, filtering the data set with the half normalized gaussian kernel (6) of width $\sigma = 3$ reduces the effect of high-frequency noise.

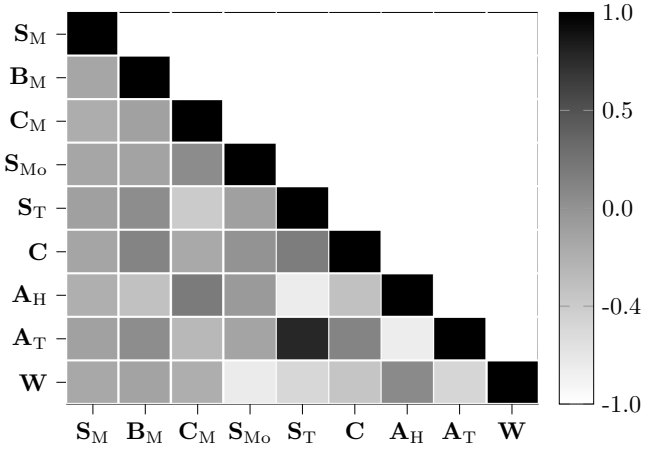

Fig. 2. Heatmap of the Pearson correlation coefficients of the process variables

The Pearson correlation fails with time-shifted signal correlations. Thus, the selection of sensitive process variables is additionally performed by training the LSTM network of *Model 1* by including further process variables, see table 2. The main reason is to achieve the target compressibility. Hence, the compressibility is always included in the variations of *Model 1*. The reference value is the conformity index with 0.27 ± 0.20 when only the compressibility is learned. A comparatively simple network topology with 21 cells in one layer proves to be sufficient and is used to investigate the sensitivity of the process variables to the water addition aiming to short learning times. This topology results from the following hyperparameter optimization (see section 3.2). The conformity index is determined for each individual day and therefore scatters. By comparing two process variable combinations, a significant effect on the prediction quality is assumed when an improvement of at least 0.1 in the mean conformity index is achieved. This value corresponds to the standard deviation of the conformity index of the best results. Similar to the correlation analysis the return sand moisture is clearly the most sensitive process variable to *Model 1*, see table 2 ($id = 4$). Furthermore, the return sand temperature ($id = 5$), the humidity ($id = 6$) and the air temperature ($id = 7$) seem to be also sensitive process variables. Further combinations of the return sand moisture with return sand temperature, humidity and the air temperature shows nearly constant prediction results ($id = 24, 25, 26, 62$). The best prediction result is $id = 26$. The mean value accuracy of this combination is with $r = 0.98 \pm 0.02$ high. If all process variables are learned for comparison ($id = n$), the prediction results are similar. By comparing the LSTM network with the covariance matrix the time-shifted behaviour of the process variables is included and the return sand moisture $\mathbf{S}_{\mathrm{Mo}}$ is the most sensitive signal.

Table 2. Conformity index p for variation id of the process variables with fixed network topology of 21 cells in one layer for *Model 1*

id	p	$\mathbf{S}_{\mathrm{M}}$	$\mathbf{B}_{\mathrm{M}}$	$\mathbf{C}_{\mathrm{M}}$	$\mathbf{S}_{\mathrm{Mo}}$	$\mathbf{S}_{\mathrm{T}}$	$\mathbf{A}_{\mathrm{H}}$	$\mathbf{A}_{\mathrm{T}}$
1	0.22 ± 0.08	+	−	−	−	−	−	−
⋮	⋮							
4	0.71 ± 0.17	−	−	−	+	−	−	−
5	0.38 ± 0.24	−	−	−	−	+	−	−
6	0.35 ± 0.25	−	−	−	−	−	+	−
7	0.43 ± 0.21	−	−	−	−	−	−	+
⋮	⋮							
24	0.84 ± 0.10	−	−	−	+	+	−	−
25	0.83 ± 0.10	−	−	−	+	−	+	−
26	0.86 ± 0.08	−	−	−	+	−	−	+
⋮	⋮							
62	0.85 ± 0.09	−	−	−	+	+	−	+
⋮	⋮							
n	0.85 ± 0.08	+	+	+	+	+	+	+

3.2 Hyperparameter analysis

The influence of the network topology on the prediction quality is investigated by a hyperparameter optimization, see Fig. 3. The optimization strategy in this paper is carried out by a *Tree of Parzen Estimator*. The software for the related implementation is the *hyperopt* package [Bergstra et al. (2013)]. Decision criterion for the best network topology is the validation loss minimum. This value results from the mean squared error of the data predicted by the network in comparison to the validation data. The euclidean norm $\|\mathbf{c}\|_2 := \sqrt{\sum_{i=1} c_i^2}$ is chosen to cumulate the number of cells over all layers for each variation. The number of cells in each layer is selected from a progressive sequence, here for example from the subset $\{2, 3, 5, .., 34\}$ of the *Fibonacci series* to keep the variations low at high cell numbers. The number of layers are varied in the set $\{1, 2, 3, 4\}$. Preliminary studies show an optimal sequence length of three batches in this case. The calculation results, see Fig. 3 (black points), are projected on each coordinate plane, see Fig. 3 (blue, red and green markers). Instead of an expected clear identifiable minimum, the validation losses form a plateau of comparable network topologies with similar values, see Fig. 3 (sectors A and B). The values of the validation loss converge to approximately $2.9 \cdot 10^{-3}$. Therefore, an LSTM of 1×21 is chosen as a proper and simple optimal network topology. The influence from dropout on the validation loss is tested by a random variation of the dropout rate from ten samples in the continuous range $[0, 1]$ for a fixed LSTM network topology. This topology is identified by another hyperparameter optimization for the correlation between compressibility and water addition. In this setting, the optimization algorithm selects a dropout rate in the continuous range $[0.25, 0.89]$. The ratio of the standard deviation to the mean is defined as the coefficient of variation. The coefficient of variation of the validation losses is relatively low at 1.50 %, so the influence of dropout rate seems to be constant in this context.

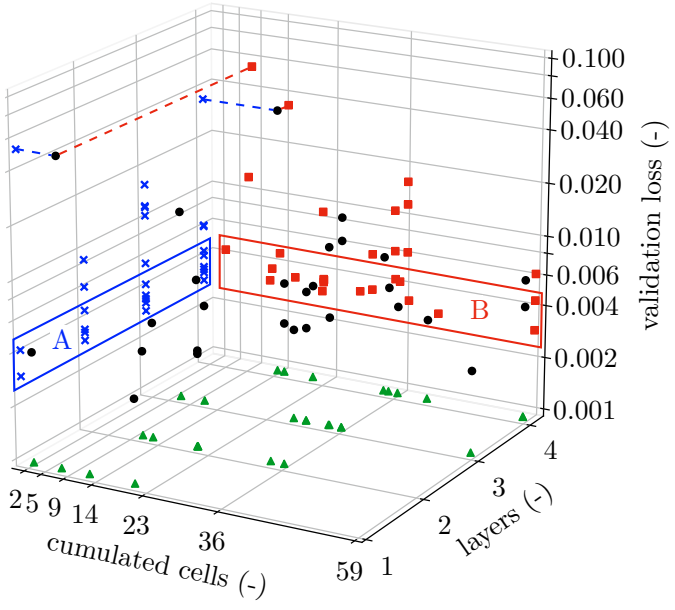

Fig. 3. Hyperparameter optimization results depending on the layers and the cumulated cells

4. PREDICTION RESULTS

The prediction of the water addition is performed with *Model 1* ($id = 26$) as inverse model of the control path. For one sample day with an average prediction result, the data of the process variables are scaled to its maximum, see Fig. 4 (top). The model learns the water addition after post-dosing $\mathbf{W}_{\text{post}}$ to reach the target compressibility. Fig. 4 (bottom) shows the comparison of the measured water addition with the predicted water addition $\mathbf{W}_{\text{pred}}$ including both means. The principal characteristics are well reproduced in particular the dynamic behaviour, see Fig. 4 (bottom, sector A). Incompletely formed peaks typically result in significant losses in the prediction quality, see Fig. 4 (bottom, sector B). A possible cause for incompletely formed peaks can be the influence of an unknown process variable. Over the investigated days, exceptions in the prediction quality are caused by individual daily data that have a clear offset above one percentage point between the means or do not form peaks correctly.

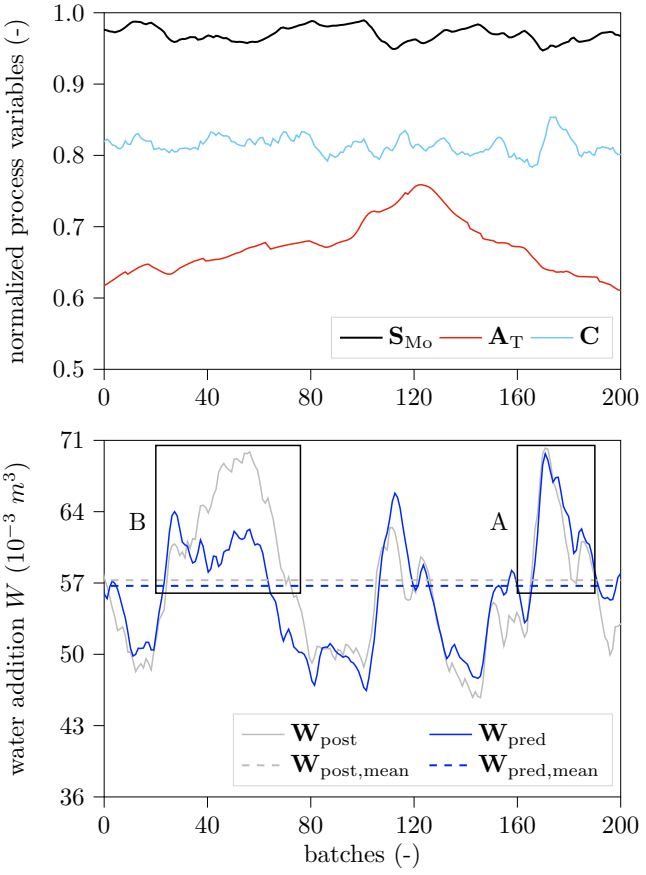

Fig. 4. Predicted water addition using *Model 1* and measured water addition after post-dosing to the time series from 22.11.2018 with $p = 0.841$ and $r = 0.995$

4.1 Compressibility simulation

The simulation of the following compressibilities $\mathbf{C}_{\circ,\text{sim}}$ is performed with *Model 2* as control path. Reference trajectory is the target compressibility $\mathbf{C}_{\text{target}}$. Inputs for *Model 2* are the respective water additions $\mathbf{W}_{\circ,\text{sim}}$, the return sand moisture $\mathbf{S}_{\text{Mo}}$, the return sand temperature $\mathbf{S}_{\text{T}}$

and the air temperature $\mathbf{A}_{\mathrm{T}}$. The predicted water addition $\mathbf{W}_{\mathrm{pred}}$ is calculated using *Model 1*. For a sample day, the data of the process variables including the water additions are scaled to its maximum, see Fig. 5 (top). Fig. 5 (middle) shows the comparison between the target compressibility and the simulated compressibilities $\mathbf{C}_{\circ,\mathrm{sim}}$. The reference trajectory of the compressibility is relatively smooth, therefore the root mean squared error (rmse) is used to compare the compressibilities. The compressibility simulation with the predicted water addition as input $\mathbf{C}_{\mathrm{pred,sim}}$ is approximately 65 % more accurate than the simulated compressibility from the pre-dosing $\mathbf{C}_{\mathrm{pre,sim}}$ and approximately 63 % better than simulation with post-dosing $\mathbf{C}_{\mathrm{post,sim}}$, see Fig. 5 (middle). If the compressibility is too low, the post-dosing compensates the compressibility deviations partially, see Fig. 5 (middle, sector A). The compressibility with the predicted water addition as input is in general slightly below the target compressibility. Therefore, post-dosing seems to be possible. The compressibility drifts probably caused by the return sand temperature and the air temperature can be well predicted, see Fig. 5 (top and middle, sectors B and B*). The root mean squared error of the deconvoluted predicted compressibility $\mathbf{C}_{\mathrm{pred,sim,deconv}}$ is in absolute values 0.69 higher than the simulated convoluted compressibility $\mathbf{C}_{\mathrm{pred,sim}}$, see Fig. 5 (middle and bottom). On average, the rmse for the deconvoluted predicted compressibility is 1.16 ± 0.15 and for the measured compressibility 1.99 ± 0.18. The relative distance of the root mean squared errors between the deconvoluted compressibility $\mathbf{C}_{\mathrm{pred,sim,deconv}}$ and the measured compressibility after post-dosing $\mathbf{C}_{\mathrm{post}}$ is in the same range than the difference between the simulated convoluted compressibilities $\mathbf{C}_{\mathrm{pred,sim}}$ and $\mathbf{C}_{\mathrm{pre,sim}}$.

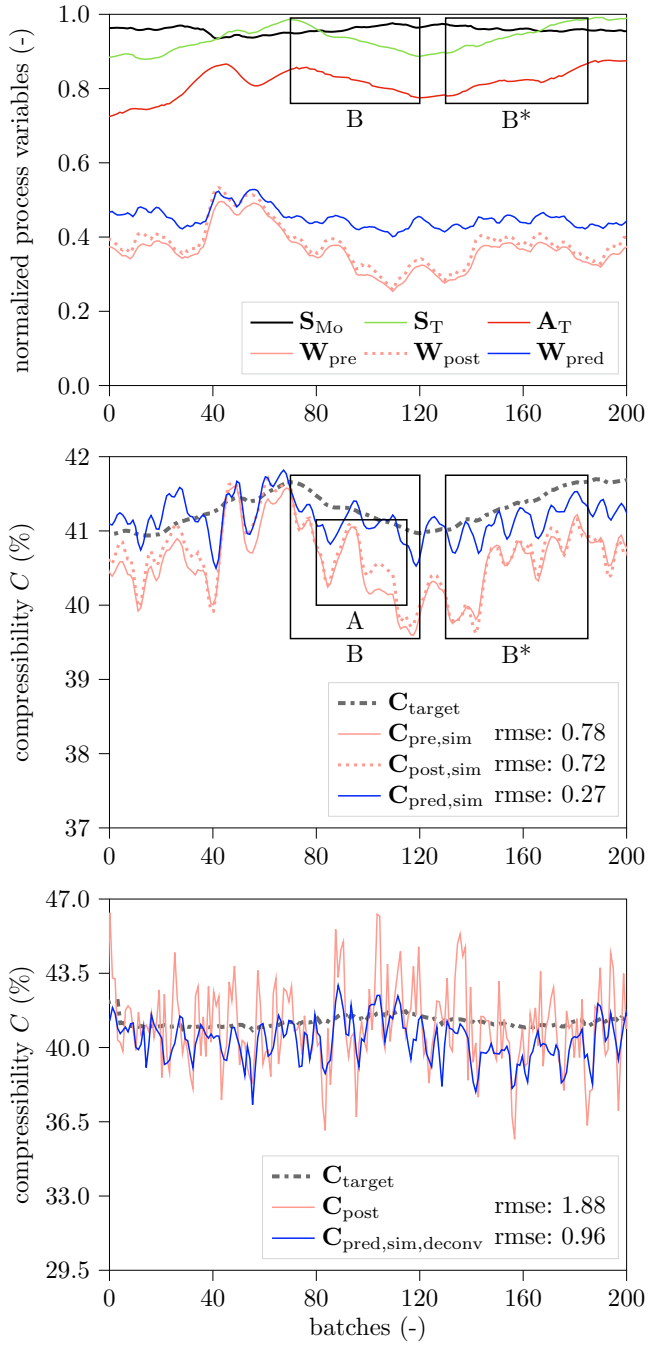

Fig. 5. Measured and simulated compressibilities using *Model 2* with measured and predicted water addition for inputs to the time series from 21.03.2019

4.2 Production plant test

The test of the predicted water addition $\mathbf{W}_{\mathrm{pred}}$ at the production plant runs with a typical casting model for 60 batches (approximately three hours) and is performed

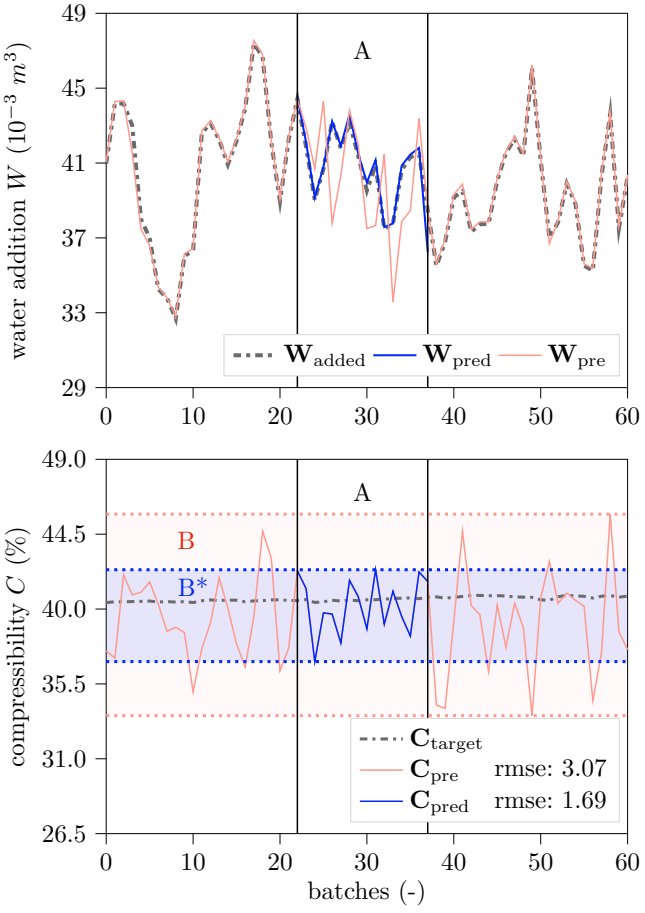

Fig. 6. Test of the water addition for pre-dosing using *Model 1* at the production plant and measured compressibility response to the time series from 16.10.2019

on a day with high fluctuations in the compressibility, see Fig. 6. The calculation of the water addition is based on *Model 1* ($id = 26$) with an input data set consisting of compressibility, return sand moisture and air temperature. The sequence length includes three batches and the prediction is one step ahead. The algorithm for the water prediction is active for 17 batches, see Fig. 6 (sector A). Post-dosings are excluded from the unconvoluted data to evaluate only the pre-dosing effect. The predicted water addition and the water addition for pre-dosing calculated by the existing control $\mathbf{W}_{\text{pre}}$ are shown in comparison, see Fig. 6 (top). The differences between the actual given water addition $\mathbf{W}_{\text{added}}$ and the predicted water addition are specific to this production plant and vary in the typical range $\pm 0.5\,(10^{-3}\,m^3)$. By comparing the total variability of the measured compressibilities, the fluctuation of the compressibility $\mathbf{C}_{\text{pred}}$ is less than the fluctuation of the compressibility $\mathbf{C}_{\text{pre}}$, see Fig. 6 (bottom, sectors B and B*). The root mean squared error in case B* is approximately 45 % lower than in case B. Therefore, the water prediction with *Model 1* is about to reduce the difference between the measured compressibility and the target compressibility at the given production plant.

5. CONCLUSION

This paper illustrates a databased approach to control the moulding sand compressibility using LSTM networks. The compressibility is controlled by the predicted pre-dosing of the water addition. These prediction is possible for the underlying production plant and shows significant sensitivity with respect to the return sand moisture and the return sand temperature or the air temperature. The network topology is essential for the prediction result. However, quite a number of best topologies seem to exist that allow for similar prediction qualities. The utilization of water prediction in a feedforward control scheme leads to a significant improvement of the compressibility. The simulation and the plant operation experiment prove the capability of the presented approach. Thus, the presented method contributes to the qualification of databased control approaches in foundries.

In future studies, the prediction quality will be further increased by considering more process variables. Initial data of the motor currents of the main drive and the agitator in the moulding sand mixer indicate positive effects on the quality of the water prediction. These will be included when sufficient data are available for the learning process. By optimizing the sensitivity of the compressibility to the water addition the accuracy of the water pre-dosing can be further improved. More accurate compressibility measurement devices are provided for this purpose. An endurance test with the predicted water addition at this foundry is planned to evaluate the presented method under a variety of environmental influences.

ACKNOWLEDGEMENTS

We would like to thank the companies Heinrich Meier Eisengießerei GmbH & Co. KG, Rahden and KÜNKEL WAGNER Germany GmbH, Alfeld for their generous support of this research work.

REFERENCES

Abadi, M. et al. (2015). Tensorflow: Large-scale machine learning on heterogeneous systems. `https://www.tensorflow.org/`.

BDG (2019). BDG-Richtlinie R 311 Emission beim Einsatz von Formstoffbindemitteln und Formstoffüberzugsstoffen. *BDG-Umweltausschuss*. `https://www.bdguss.de/fileadmin/content_bdguss/Der_BDG/Richtlinien/R_311.pdf`.

Bergstra, J., Yamins, D., and Cox, D.D. (2013). Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. `http://proceedings.mlr.press/v28/bergstra13.pdf`.

Gal, Y. and Ghahramani, Z. (2016). A Theoretically Grounded Application of Dropout in Recurrent Neural Networks. In D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 29*, 1019–1027. Curran Associates, Inc. `http://papers.nips.cc/paper/6241-a-theoretically-grounded-application-of-dropout-in-recurrent-neural-networks.pdf`.

Gemming, H. (2003). *Prozesssicherheit in der Formstoffaufbereitung mit Hilfe der Fuzzy-Logik*. Dissertation, Technische Universitt Freiberg. `http://webdoc.sub.gwdg.de/ebook/serien/aa/Freiberger_Diss_Online/185.pdf`.

Hochreiter, S. and Schmidhuber, J. (1997). long short-term memory. *Neural Computation*, 1735–1780.

Manaswi, N.K. (2018). *Deep Learning with Applications Using Python*. Apress, Berkeley, CA. doi:10.1007/978-1-4842-3516-4.

Michenfelder, M., Liedtke, A., and Huck, C. (2014). Prozessintegriertes Formstoffmanagement an der Formanlage. *Giesserei*, 54–59.

Pätzold, A. (2017). Extraktion von Mustern und Modellierung zeitlicher Abhängigkeiten für datenbasierte Prozessoptimierung. Abschlussarbeit an der Ostfalia Hochschule für angewandte Wissenschaften.

Rose, A., Seel, A., and Grotjahn, M. (2020). Prädiktion der Verdichtbarkeit von Formsand mittels neuronaler Netze am Beispiel des Formstoffmischers. *3. Formstoff-Forum 2020*, 84–89.

Seeber, R. and Kohler, C. (2012). Method for processing moulding sand. US patent:US8225844B2. `https://patents.google.com/patent/US8225844B2/en`.

Sommer, L. (2017). Modellbildung industrieller Produktionsprozesse mit künstlichen neuronalen Netzen. Masterarbeit an der Universität Bremen.

Tilch, W., Polzin, H., and Franke, M. (2015). *Praxishandbuch bentonitgebundener Formstoff*. Fachverlag Schiele & Schön, 1st edition.

Trinowski, D.M. (2016). Die Gießerei-Industrie in Europa und den USA - ein Vergleich. *Giesserei*, 103(8), 58–63.

VDG (1989a). VDG-Merkblatt R 100 Formstoff-Bindemittel. *Verein deutscher Giessereifachleute*.

VDG (1989b). VDG-Merkblatt R 95 Formstoffkreislauf. *Verein deutscher Giessereifachleute*.

Zaremba, W., Sutskever, I., and Vinyals, O. (2014). Recurrent Neural Network Regularization. `http://arxiv.org/pdf/1409.2329v5`.