Science Advances
AAAS

Supplementary Materials for

Machine learning for industrial processes: Forecasting amine emissions from a carbon capture plant

Kevin Maik Jablonka *et al.*

Corresponding author: Berend Smit, berend.smit@epfl.ch; Susana Garcia, susana.garcia@hw.ac.uk

Sci. Adv. **9**, eadc9576 (2023)
DOI: 10.1126/sciadv.adc9576

This PDF file includes:

Supplementary Note 1 Intermittency scenarios

Due to electricity grid load following requirements and the increasing penetration of intermittent renewables, fossil fuel-based electricity generation can play an important role in load balancing the grids. In such a flexible operation it is important that the capture unit can follow these dynamics. Most of the studies so far focused on technical and economic challenges associated with the flexible operation of the capture plant (10, 35, 36). However, a key aspect when assessing the implications of CCS power plant flexibility is how the CO_2 capture plant needs to be operated in order to comply with future environmental legislation with respect to amine emissions.

In our study, we focused on intermittency scenarios that could potentially have a strong effect on amine emissions. As a baseline, we assume that the capture plant operates with the power plant at full load. Therefore, each of the scenarios is looking at the different consequences of reducing the load of the power plant on a daily basis, and hence expecting a lower flow rate of flue gas entering the absorber column.

For the emissions not only the reduced load of the flue gas is relevant, but at the same time, there can be other effects. For example, a possible consequence of a part-load operation can be a reduction in the steam flow entering the reboiler unit or other changes. By combining the possible consequences, our stress test consists of the following eight scenarios:

1. *Water wash temperature increase*; one of the possible side effects of fluctuations in the flue gas flow is that it can result in changes in the temperature in the absorber, which impacts directly the temperature of the water wash. Also, a sudden change in CO_2 concentration caused by unstable operation of the power plant can have a direct effect on the water wash temperature.

2. *Water wash flow rate decrease*; increasing the water wash flow rate is most frequently used to control amine emissions (49). It is therefore important to see if this control mechanism functions sub-optimally.

3. *Flue gas temperature increase*; a variation of the flue gas temperature might be a result of unstable plant operation.

4. *Lean solvent flow rate decrease*; one of the mechanisms to control variations in the capture rate is to change the solvent flow rate (50).

5. *Lean solvent and flue gas flow rate decrease*; a possible part-load operation of the power plant due to less electricity production (51).

6. *Lean solvent temperature increase*; a side effect of the variation of the steam supply to the reboiler can impact the solvent temperature.

7. *Increase of both the lean solvent and water wash temperatures*; one of the possible side effects of both variations in the steam and solvent flow rate and in the flue gas flow.

8. *Capture rate decrease*; is a side effect of a reduced steam availability might happen when the power plant operates at part load and is expected when a higher amount of electricity supply to the grid is required, which can cause a drop in the capture rate.

Each of these scenarios are translated into a change of different process variables that can cause the desired effect. For example, the water wash temperature is changed by changing the flow rate of the cooling water in the heat exchanger of the water wash section. The details are given in Supplementary Table 1.

In this testing phase, no regime of self-accelerating degradation occurred, even if phases with increased reboiler temperature and increased oxygen concentration in the flue gas give reason to expect an increased degradation rate of CESAR1. Note that due to the design of the experimental campaign we cannot directly capture changes that occur over a timescale longer than 24 h.

Tab. S1. Parametric tests with the CESAR1 solvent at PCC pilot plant at Niederaussem. Experiments were performed as step changes and selected to cover relevant regions of interest in the space of operating conditions. The experiments and a preliminary analysis have already been described in Charalambous et al. (12)

	region of interest				
exp.	WW	FG	LS	CRate	description
1	■				Water wash temperature (TI-19) increase from ≈ 45 °C to 55 °C [1]
2	■				Water wash flow (FI-19) decrease from 6000 kg h^{-1} to 5000 kg h^{-1} [2]
3		■			Flue gas temperature (TI-3) increase from 45 °C to 55 °C [3]
4			■		Lean solvent flow (FI-11) decrease from ≈ 2400 kg h^{-1} to 2000 kg h^{-1} [4]
5		■			Lean solvent flow (FI-11) decrease from ≈ 2400 kg h^{-1} to 2000 kg h^{-1} and flue gas flow (FI-2) decrease from 1500 kg h^{-1} to 1300 kg h^{-1}
6		■			Lean solvent temperature (TI-12/TI-13) increase from 43 °C to 52 °C [5]
7	■		■		Lean solvent temperature (TI-12/TI-13) increase from 43 °C to 53 °C and water wash temperature (TI-19) increase to 42 °C
8				■	Capture rate decrease from 90 % to 80 % [7]

[1] This is achieved by changing the temperature of the cooling medium (i.e., water) in the heat exchanger (HEX) located in the water wash section.

[2] This is achieved by changing the speed of the pump located in the water wash section.

[3] This is achieved by changing the temperature of the cooling medium in the HEX located in the direct contact cooler (DCC) unit.

[4] This is achieved by modifying the amount of the solvent in the pilot plant.

[5] This is achieved by controlling the temperature of the cooling medium in the HEX located after the pump which sends the lean solvent from the stripper to the absorber.

[6] In this experiment the water wash baseline temperature was 33 °C.

[7] This is controlled by the steam flow in the reboiler and the reboiler level.

Abbreviation: WW, water wash; FG, flue gas; LS, lean solvent; CRate, CO_2 capture rate.

The actual experimental campaign included two additional experiments (see Supplementary Note 3) which we did not include in our analysis due to the instability of the plant after a shutdown of the power plant.

Supplementary Note 2 Experimental methods

The plant is equipped with online gas-analysis systems to continuously monitor the composition of the inlet and outlet gas streams. The online gas-analysis systems are used to quantify the composition of: (i) flue gas at the absorber inlet (i.e., CO_2, CO, NO: BA5000 Bühler infrared spectroscope; O_2: BA3500, Bühler, paramagnetic detection, SO_2: MCS 100E, Sick/Maihak, photometric detection limit <4 mg m^{-3}), (ii) CO_2-lean flue gas downstream of the water wash outlet, and (iii) CO_2 product stream.

The GasMET analysers used for monitoring solvent emissions exiting the water wash section have been calibrated for standard inorganic components (i.e., NH_3, SO_2, NOx, CO, CO_2) and for AMP and Pz. The detection limit for amines is approx. 1 mg m^{-3} (STP). Both GasMET analysers were zeroed with nitrogen at the beginning of the experimental campaign and re-zeroed once a week during the measuring period.

All liquid samples were analysed by ATR-FTIR spectroscopy for which two calibration sets have been used. One calibration was performed for CO_2 loadings and amine concentrations within the range expected for lean and rich samples. The other calibration was performed for very low amine concentrations and was used to monitor the CO_2 loading and amine concentrations in the water wash. The detection limit for AMP and Pz are around 0.3 wt%.

Supplementary Note 3 Experimental Data

Supplementary Table 2 lists all dynamic tests and the steady-state performance of the pilot plant in terms of the CO_2 capture rate, CO_2 loadings, amine and CO_2 emissions, and reboiler duty. Supplementary Figure 1 illustrates the water wash performance comparing the water wash temperature and the emissions (i.e. CO_2, AMP, Pz, H_2O, NH_3) at the exit of the water wash section. The grey areas in the figure represent the time when no dynamic tests were performed (days 4–5 and 11–16).

Tab. S2. Summary of pilot plant emissions and pilot plant performance. The parametric tests are indicated with the symbol 'X'.

Run		Step-Change Parameter	L/G ratio	Loading / wt%			CO_2 Capture Rate / %	Emissions / mg/m^3 ǂ			Reboiler duty MJ/kg CO_2
				Lean	Rich	WW		AMP	Pz	CO_2ǂ	
1		Baseline	1.56	0.09	0.29	0.53	86.3	64.2	10.2	1.16	2.8
2	X	Water wash (WW) temp.	1.56	0.07	0.30	0.42	86.5	152.1	19.7	1.18	4.2
		Average over 24h	1.56	0.08	0.30	0.45	87.0	102.7	24.9	1.22	3.2
3		Baseline	1.53	0.07	0.32	0.57	86.6	60.2	24.9	1.41	3.1
4	X	WW flow rate	1.53	0.07	0.31	0.48	86.3	67.3	14.3	1.40	3.4
		Average over 24h	1.53	0.07	0.31	0.51	85.7	64.5	21.1	1.46	2.9
5		Baseline	1.53	0.09	0.31	0.58	84.1	63.2	18.8	1.43	3.0
6	X	Flue gas (FG) temperature	1.53	0.09	0.34	0.52	83.8	65.3	12.9	1.49	3.1
		Average over 24h	1.53	0.09	0.30	0.56	85.4	72.2	23.0	1.35	2.7
7		Baseline	1.53	0.07	0.31	0.58	84.4	77.1	59.4	1.63	2.3
8	X	Lean solvent (LS) flow rate	1.30	0.05	0.26	0.67	77.3	48.5	20.6	2.78	2.8
		Average over 24h	1.48	0.06	0.31	0.62	82.8	63.0	28.0	1.96	3.0
9		Baseline	1.53	0.07	0.28	0.54	84.6	34.8	20.6	1.90	3.2
10	X	LS and FG flow rates	1.53	0.07	0.27	0.57	83.5	50.6	18.0	1.07	3.9
		Average over 24h	1.53	0.07	0.29	0.55	83.3	60.1	24.5	1.78	3.0
11		Baseline	1.53	0.07	0.28	0.50	85.0	56.8	16.0	1.11	4.2
12	X	LS temperature	1.53	0.07	0.27	0.41	87.7	97.4	8.0	1.07	2.6
		Average over 24h	1.53	0.07	0.30	0.45	86.3	70.3	17.8	1.08	2.8
13		Baseline	1.53	0.08	0.27	0.53	81.3	62.6	23.0	1.47	2.6
14	X	LS and WW temperature	1.52	0.07	0.28	0.39	84.2	126.6	3.5	1.25	2.7
		Average over 24h	1.53	0.07	0.29	0.44	85.5	82.0	15.8	1.15	3.0
15		Baseline	1.56	0.08	0.30	0.52	82.6	75.8	28.3	1.43	2.4
16	X	Capture rate	1.36	0.07	0.30	0.64	75.5	51.1	23.2	2.25	2.4
		Average over 24h	1.51	0.07	0.28	0.60	83.2	60.2	20.8	1.46	3.3

† At standard temperature and pressure (STP).

ǂ CO_2 is given in vol%.

Abbreviation: WW, water wash; FG, flue gas; LS, lean solvent.

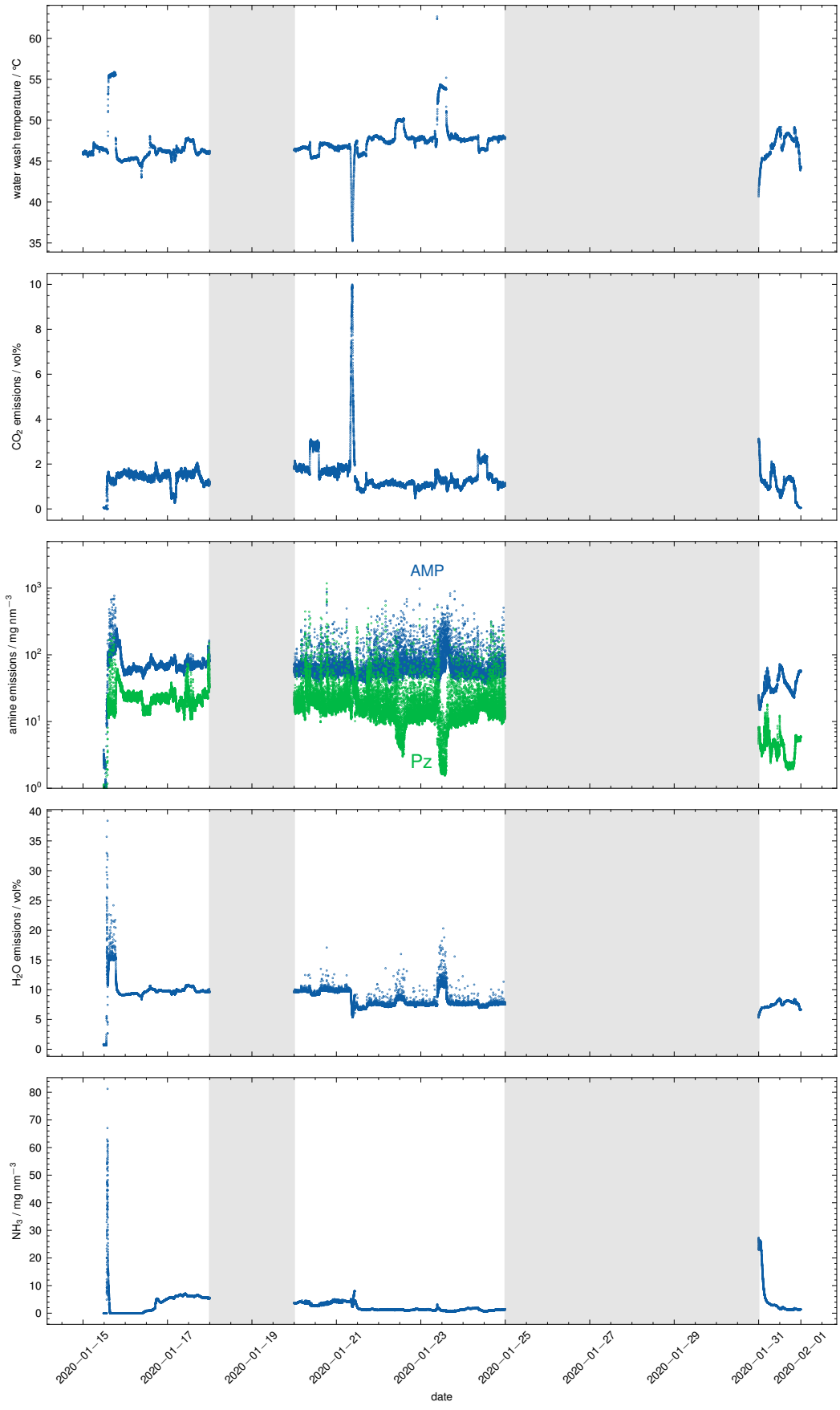


Fig. S1. Water wash temperature and CO_2, AMP, Pz, H_2O vapour, and NH_3 emissions profiles as recorded using the GasMET FTIR at the water wash outlet during the CESAR1 dynamic campaign (i.e., 10 dynamic experiments over 10 operational days). The grey area denotes the period when no dynamic experiments were performed (i.e., days 4–5 and days 11–16).

3.1 Pilot plant parameters

The typical operational parameters and the boundary conditions of the capture plant are provided in Supplementary Table 3. The average value of these parameters are also provided before the start of the dynamic campaign (on day 275 with operation with the CESAR1 solvent).

Tab. S3. Typical operational parameters and boundary conditions used during the testing campaign. Including average values before the start of the campaign.

Parameter	Unit	Value	Value †	Description
Flue gas temperature at DCC inlet	°C	64	63	—
Flue gas temperature at absorber inlet	°C	40–45	44	—
Flue gas flow rate	$m^3\,h^{-1}$ ‡	1150	1150	—
CO_2 content of the flue gas	vol%, dry	12.5	12.5	Measured at absorber inlet
O_2 content of the flue gas	vol%, dry	5.0	5.4	Measured after desulphurization
SO_2 content of the flue gas	$mg\,m^{-3}$ ‡	<1.0	<1.0	Measured at absorber inlet
Dust	$mg\,m^{-3}$ ‡	<2.0	<2.0	—
NOx content of flue gas	$mg\,m^{-3}$ ‡	100–160	100–160	Measured at absorber inlet
NO_2 content of flue gas	$mg\,m^{-3}$ ‡	6–8	6–8	Measured at absorber inlet
Solvent flow rate	$kg\,h^{-1}$	2600	2400	—
Water circulation in the DCC	$kg\,h^{-1}$	8000	9758	—
pH value of water in the DCC	—	7–7.2	7–7.2	—
CO_2-lean flue gas temperature	°C	40-45	46	Measured at water wash outlet
Solvent regeneration temperature	°C	120	120	—
Desorber pressure	bar(a)	1.75	1.75	—
CO_2 capture rate	—	90	92	—
Specific energy demand	GJ/tCO_2	3.0	3.3	For solvent regeneration

† Average values measured before the start of the dynamic campaign (on day 275).

‡ Dry value at standard temperature and pressure (STP).

3.2 Abbreviations

- **PI-2** Pressure of the flue gas stream at the exit of the direct contact cooling (DCC) unit.
- **TI-2** Temperature of the flue gas stream at the exit of the DCC unit.
- **FI-2** Flow rate of the flue gas stream entering the absorber column.
- **PI-3** Pressure of the flue gas stream upstream of the absorber column.
- **TI-3** Temperature of the flue gas stream upstream of the absorber column.
- **CO2-3** CO_2 concentration of the flue gas upstream of the absorber column.
- **O2-3** O_2 concentration of the flue gas upstream the absorber column.

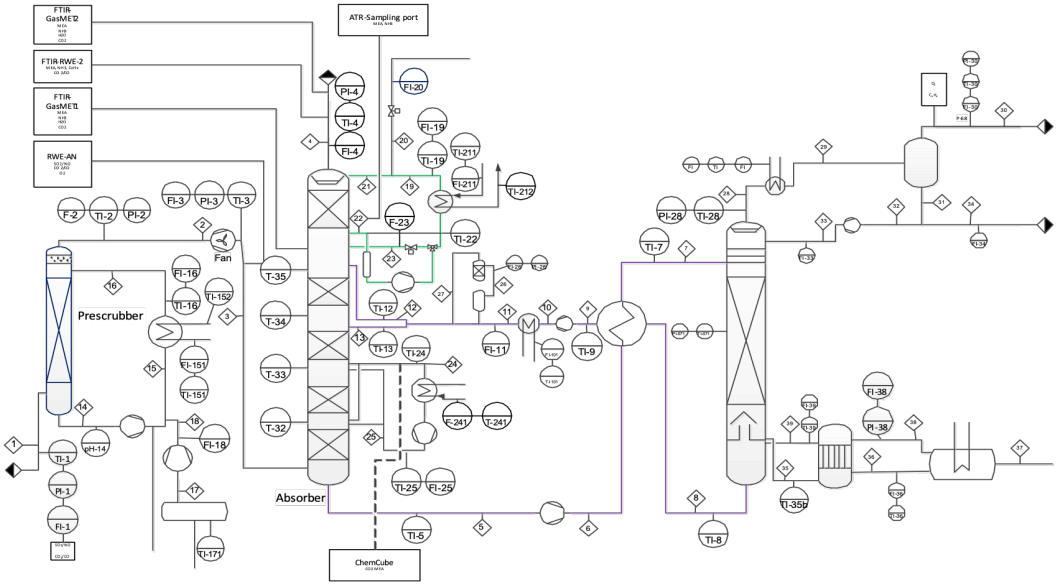


Fig. S2. Detailed PID diagram. Abbreviations are explained in section 3.2.

- **TI-32** Temperature at the top of the 1st bed (counting from the bottom of the absorber).
- **TI-33** Temperature at the top of the 2nd bed (counting from the bottom of the absorber).
- **TI-34** Temperature at the top of the 3rd bed (counting from the bottom of the absorber).
- **TI-35** Temperature at the top of the 4th bed (counting from the bottom of the absorber).
- **PI-4** Pressure of the treated gas exiting the water wash section.
- **TI-4** Temperature of the treated gas exiting the water wash section.
- **FI-4** Flow rate of the treated gas exiting the water wash section.
- **CO2-4** CO_2 concentration of the flue gas at the water wash exit – output (CO_2 emissions).
- **NH3-4** NH_3 concentration of the flue gas at the water wash exit – output (NH_3 emissions).
- **FI-11** Flow rate of the lean solvent entering the absorber (at the top of the column).
- **TI-12** Temperature of the lean solvent entering the absorber (at the top of the column).

- **TI-13** Temperature of the lean solvent entering the absorber.
- **FI-20** Flow rate of fresh water added to the water wash section.
- **FI-211** Flow rate of water in the heat exchanger (HEX) located at the water wash section.
- **TI-211** Temperature of water in the heat exchanger located at the water wash section.
- **TI-212** Temperature of water out of the heat exchanger located at the water wash section.
- **TI-8** Temperature of the lean solvent leaving the desorber sump (upstream of the HEX).
- **TI-9** Temperature of the lean solvent downstream of the HEX.
- **TI-5** Temperature of the rich solvent exiting the absorber.
- **TI-7** Temperature of the rich solvent downstream of the HEX
- **TI-28** Temperature of the CO_2 concentrated stream exiting the top of the stripper.
- **PI-28** Pressure of the CO_2 concentrated stream exiting the top of the stripper.
- **PI-30** Pressure of the CO_2 concentrated stream exiting the condenser (after the stripper).
- **TI-30** Temperature of the CO_2 concentrated stream exiting the condenser.
- **FI-30** Flow rate of the CO_2 concentrated stream exiting the condenser.
- **FI-38** Flow rate of the steam entering the reboiler.
- **PI-38** Pressure of the steam entering the reboiler.
- **FI-36** Flow rate of the stream exiting the reboiler.
- **TI-36** Temperature of the steam exiting the reboiler.
- **Reb. Duty** Reboiler energy use.
- **FI-19** Flow rate of the water wash section (water wash circulation rate).
- **TI-19** Temperature of the water wash.
- **PI-1** Pressure of the flue gas upstream to the DCC unit.

- **TI-1** Temperature of the flue gas upstream to the DCC unit.
- **TI-35b** Temperature of the rich solvent into the reboiler.
- **FI-35** Flow rate of the solvent into the reboiler.
- **TI-39** Temperature of the solvent out of the reboiler.
- **FI-23** Flow rate of excess liquid leaving the water wash (WW) section into the absorber.
- **TI-22** Temperature of the water in the WW section upstream of the WW pump.
- **Level Des.** Level of the liquid into the desorber.
- **Level Reb.** Level of the liquid into the reboiler.
- **TI-24** Temperature of the liquid upstream to the HEX located in the intercooling section.
- **TI-25** Temperature of the liquid downstream to the HEX of the intercooling section.
- **FI-25** Flow rate of the liquid in the intercooling section.
- **FI-16** Flow rate of the liquid downstream of the HEX located in the DCC unit.
- **TI-16** Temperature of the liquid downstream of the HEX located in the DCC unit.
- **FI-151** Flow rate of the water in the HEX of the DCC unit.
- **TI-151** Temperature of the water in the HEX of the DCC unit.
- **TI-152** Temperature of the water out of the HEX of the DCC unit.
- **FI-241** Flow rate of the water in the HEX of the intercooling section.
- **TI-241** Temperature of the water in the HEX of the intercooling section.
- **TI-242** Temperature of the water out of the HEX of the intercooling section.

Supplementary Note 4 Exploratory data analysis

Using the Augmented-Dickey Fuller test (as implemented in the statsmodels library (43)) we tested for stationarity and found that the raw data is non-stationary due to deterministic and non-deterministic trend components (see also autocorrelation functions in Supplementary Fig. 3). For the machine learning analysis, we removed the deterministic trend component using linear regression through the endpoints. We then tested for Granger causalities (Supplementary Figure 4) using the statsmodels library.

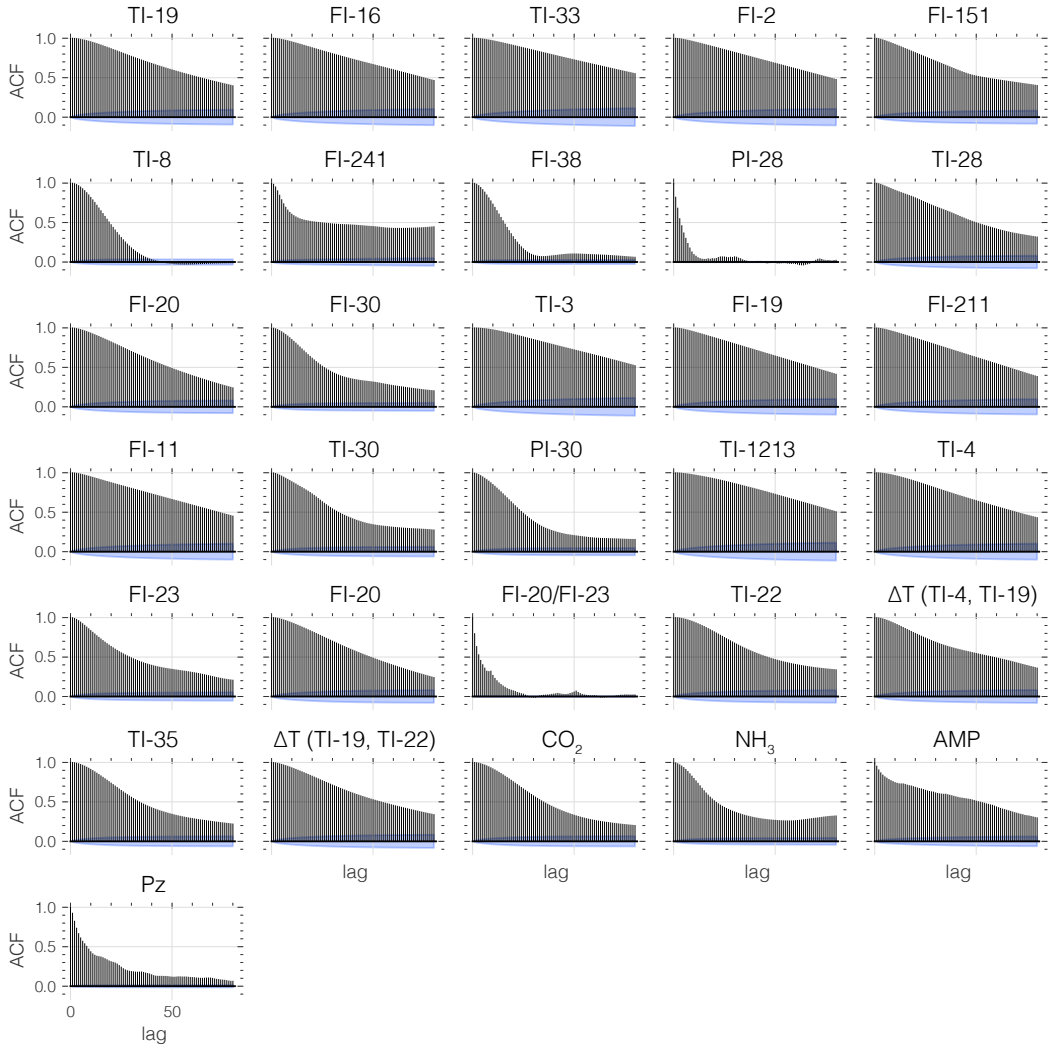


Fig. S3. Autocorrelation functions (ACF). Calculated up to a maximum lag of 80 (i.e., 160 min). Blue region shows the .95 confidence interval.

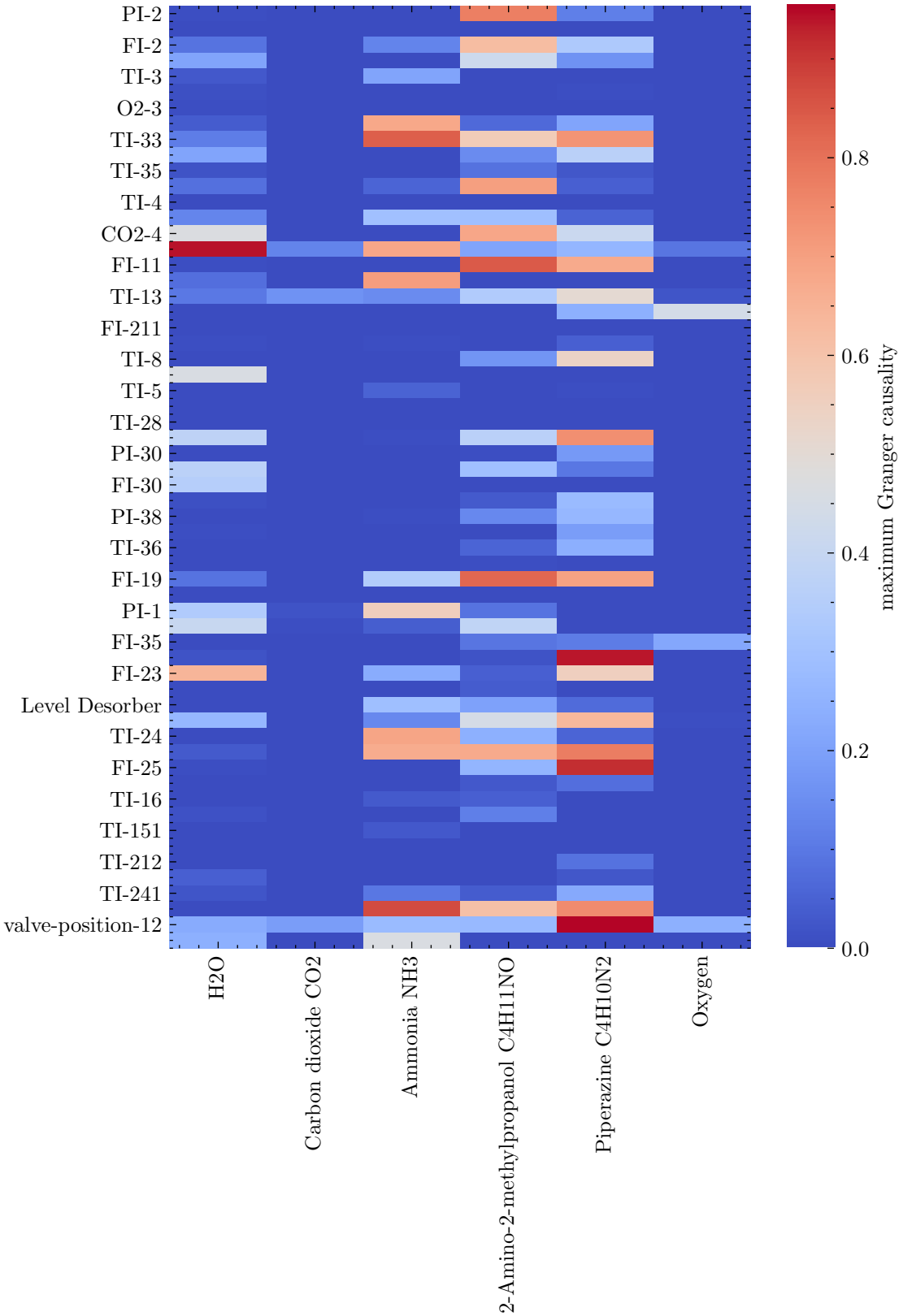
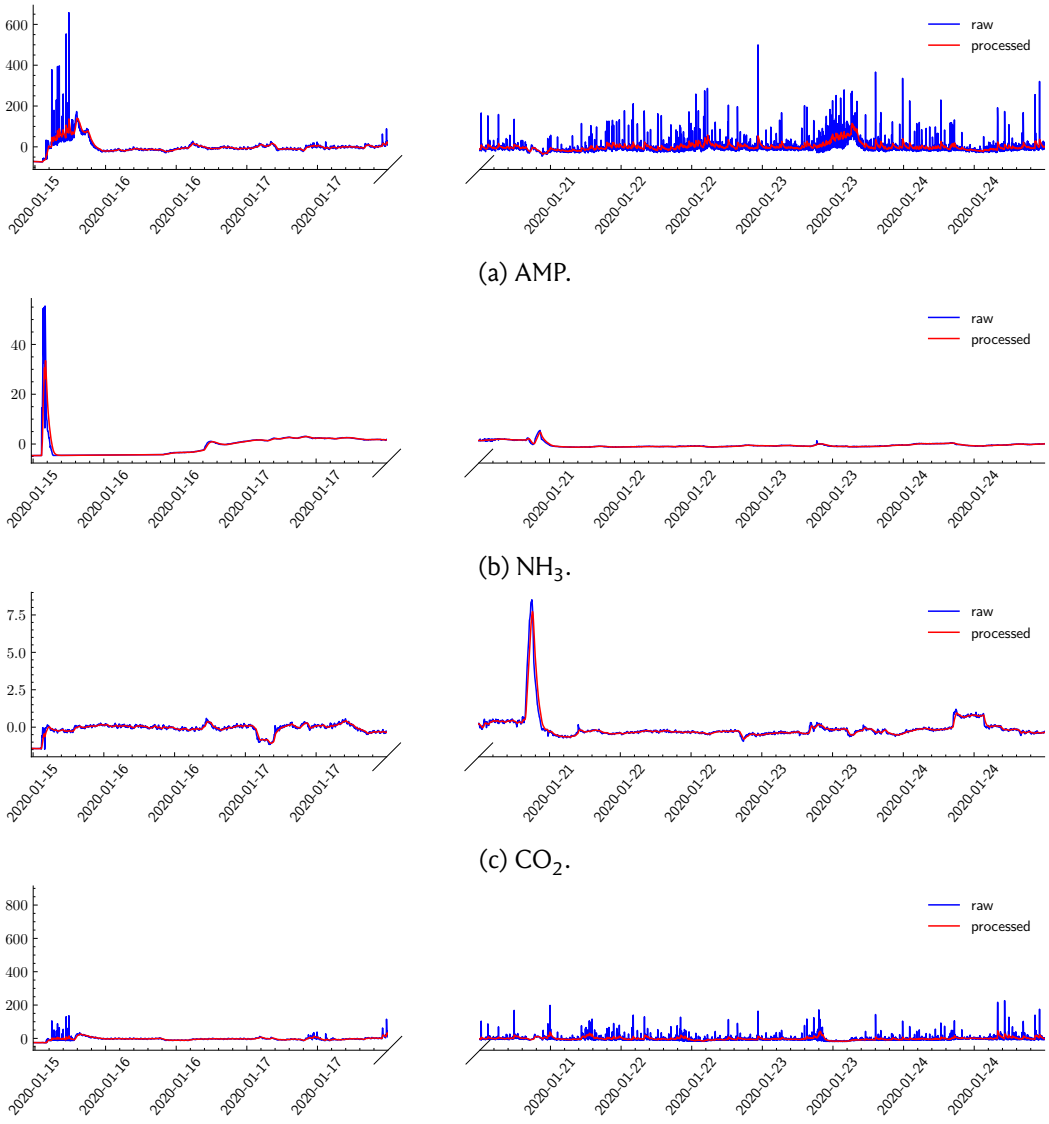


Fig. S4. Granger causalities. Strongest Granger causalities between features and labels in the dataset, computed up to a lag of 20.

Supplementary Note 5 Data pre-processing

To make the data amenable to machine learning, we performed a range of pre-processing steps. Their impact on the time series is illustrated in Figure 5. One can observe that we preserve the shape of the time series but reduce the intensity of the spikes (but we still preserve their presence and location).



(a) AMP.

(b) NH_3.

(c) CO_2.

(d) Pz.

Fig. S5. Impact of data pre-processing. The subfigures show impact on different objectives.

Supplementary Note 6 Feature selection

Feature selection was guided by domain knowledge (i.e., excluding parameters that are expected to have no causal relation with the emissions) and aided by Granger causality heatmaps shown in Figure 4.

In the final models, we used the following feature set:

- TI-19: Temperature of the liquid in the water wash.
- FI-19: Flow rate of the liquid in the water wash.
- FI-11: Flow rate of the lean solvent entering the absorber.
- TI-3: Temperature of the flue gas entering the absorber.
- TI-12/TI-13: Temperature of the lean solvent entering the absorber
- TI-35: Temperature at the topmost bed of the absorber

For causal impact analysis, we also included (TI-35 - TI-4), i.e., the temperature difference of the temperature at the topmost bed of the absorber and the temperature of the treated gas exiting the water wash section

Supplementary Note 7 Quantile Regression using Gradient Boosted Decision Trees

Gradient-boosted decision trees can be used for time series forecasting by using a concatenated lagged time series as input. One new regressor can then be trained for every point in the forecasting horizon. Here, we use the LightGBM implementation (37).

To produce uncertainty intervals we regress on the quantile loss, $_\alpha(x)$, which for quantile α reads

$$l_\alpha(x) = \begin{cases} -(1-\alpha)x & \text{if} x \leq 0 \\ \alpha x & \text{if} x \geq 0 \end{cases} \tag{1}$$

7.1 Performance

We use multiple metrics to measure the predictive performance of our models given two timeseries y^1 and y^2 of length T.

Mean absolute error (MAE)

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}(|y_t^1 - y_t^2|). \tag{2}$$

Mean absolute percentage error (MAPE)

$$\text{MAPE} = 100 \cdot \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t^1 - y_t^2}{y_t^1}\right|. \tag{3}$$

where y_t^1 is the actual time series (ground truth).

Overall percentage error (OPE)

$$\text{OPE} = 100 \cdot \left|\frac{\sum_{t=1}^{T} y_t^1 - \sum_{t=1}^{T} y_t^2}{\sum_{t=1}^{T} y_t^1}\right|. \tag{4}$$

where y_t^1 is the actual time series (ground truth).

Tab. S4. Performance metrics. Metrics for the median historical forecasts of the gradient-boosted decision tree model for different forecasting horizons.

	AMP			*Pz*		
	2 min	1 h	2 h	2 min	1 h	2 h
MAE / a.u.	0.0089	0.039	0.036	0.0095	0.050	0.040
MAPE / %	2.4	11	9.5	4.3	23	21
OPE / %	0.38	0.34	3.8	2.0	17	10

7.2 Hyperparameter Optimisation

For all models except the ones used for the causalimpact analysis, we searched hyperparameters on the grid presented in Table 5. For the causal impact analysis, we did not optimise every feature lag separately but rather used the same lag for all covariates. We used the Bayesian optimisation implemented in the weights and biases platform (`https://docs.wandb.ai/guides/sweeps`) for all searches. To limit computational cost, we only performed a hyperparameter search for the 0.5 quantile and reused the same hyperparameters for the other quantiles. However, we separately optimised hyperparameters for AMP and Pz. Also, for the first step change we had, due to the short window preceding the step change, limit the search to lags smaller than 40.

Tab. S5. Parameter ranges for hyperparameter search. Hyperparameter ranges considered for the GBDT models.

parameter name	range
`lag`	uniform in (0, 200)
`feature lag`	uniform in (-200, 0)
`n_estimators`	uniform in (50, 1000)
`bagging_freq`	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
`bagging_fraction`	uniform in (0.001, 1.0)
`num_leaves`	integers in (1, 200)
`extra_trees`	(True, False)
`max_depth`	(-1, 10, 20, 40, 80, 160, 320)

Tab. S6. **Hyperparameter settings.** Hyperparameter settings for AMP for different output horizons.

parameter	2 min	60 min	120 min
`bagging_fraction`	0.572	0.989	0.948
`bagging_freq`	1	4	10
`extra_trees`	False	False	False
`lag_1`	-1	-158	-169
`lag_2`	-134	-70	-118
`lag_3`	-55	-196	-38
`lag_4`	-10	-125	-97
`lag_5`	-23	-102	-7
`lag_6`	-176	-113	-141
`lags`	100	70	174
`max_depth`	-1	20	40
`n_estimators`	743	375	559
`num_leaves`	7	177	74

Tab. S7. **Hyperparameter settings.** Hyperparameter settings for Pz for different output horizons.

parameter	2 min	60 min	120 min
`bagging_fraction`	0.331	0.121	0.180
`bagging_freq`	10	8	7
`extra_trees`	False	False	False
`lag_1`	-93	-136	-81
`lag_2`	-182	-144	-170
`lag_3`	-111	-146	-62
`lag_4`	-106	-114	-77
`lag_5`	-56	-55	-154
`lag_6`	-90	-83	-75
`lags`	100	171	195
`max_depth`	40	40	80
`n_estimators`	126	275	193
`num_leaves`	4	90	113

For the causal impact analysis we used a forecasting horizon of twice the duration of the step change.

Tab. S8. **Hyperparameter settings.** Hyperparameter settings for AMP for different step changes (causal impact analysis).

parameter	step 0	step 1	step 2	step 3	step 4	step 5	step 6
bagging_fraction	0.334	0.475	0.742	0.526	0.955	0.947	0.847
bagging_freq	10	8	3	5	10	0	2
extra_trees	True	False	True	False	False	True	False
covariates lag	-51	-172	-30	-71	-184	-23	-126
lags	47	5	145	167	194	173	121
max_depth	80	160	320	160	-1	80	160
n_estimators	123	866	816	348	880	394	511
num_leaves	40	178	177	42	162	167	130

Tab. S9. **Hyperparameter settings.** Hyperparameter settings for Pz for different step changes (causal impact analysis).

parameter	step 0	step 1	step 2	step 3	step 4	step 5	step 6
bagging_fraction	0.817	0.959	0.685	0.241	0.785	0.796	0.900
bagging_freq	9	2	10	3	0	2	10
extra_trees	True	True	True	True	False	True	False
covariates lag	-27	-10	-106	-19	-10	-79	-150
lags	13	185	170	46	77	10	130
max_depth	20	320	20	160	40	40	160
n_estimators	69	221	92	130	979	975	933
num_leaves	48	109	84	67	73	44	9

Supplementary Note 8 Temporal Convolutional Neural Networks

Supplementary Fig. 6 gives an overview of the modeling process using temporal convolutional neural networks.

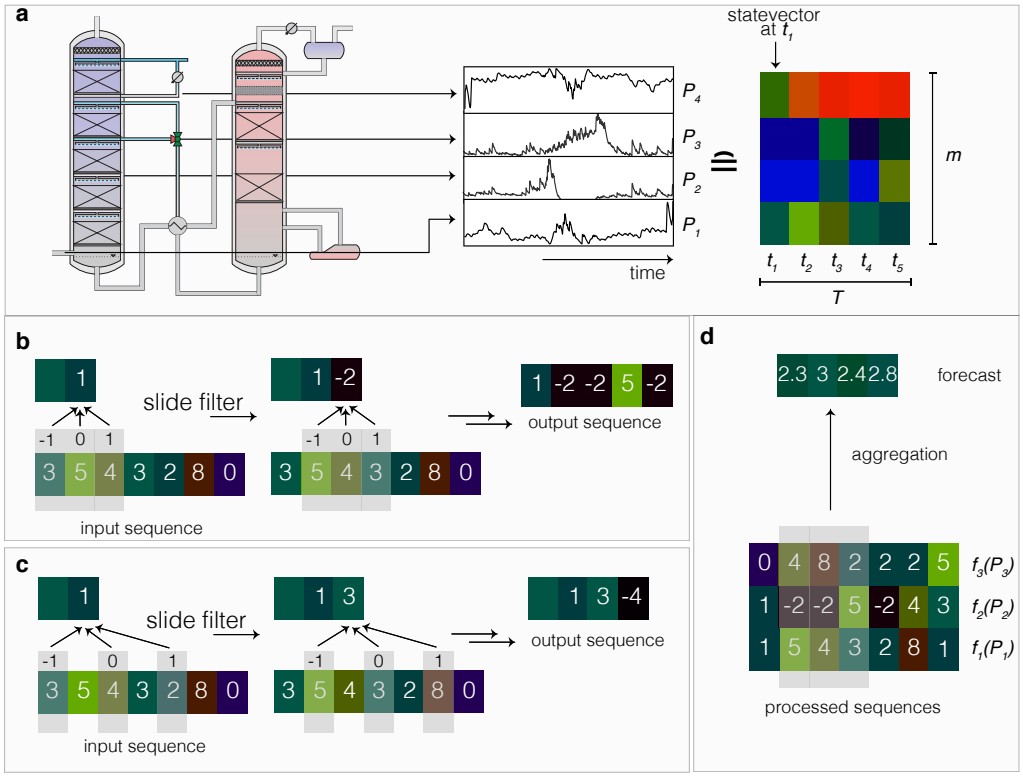


Fig. S6. Schematic illustration of the modeling approach using temporal convolutional neural networks. Mapping of the data of the plant onto an image; the data set can be thought of an "image" with "width" equal to length of the input sequence (T) and "height" equal to the number of parameters, m. We represent with colours the value of parameter P_j at a time t_i. As the predictions should be invariant to the order of the rows, we only apply the pattern learning via convolutional filters (light grey) in the time direction, therefore the image of the plant should be seen as m one-dimensional image. **b**, Convolutional kernels are slid over the m images as part of the pattern recognition algorithm. The weights of the kernels (of the fth filter), $W_f = [w^f_{-1}, w^f_0, w^f_1]$, are initially set according to a conventional initialisation scheme and learned in the training procedure. In the first layer, the kernel operates on directly neighbouring values. In order to allow for the model to learn different representations (patterns) we use multiple learnable filters per layer of the neural network (38, 52), i.e., the layer outputs multiple (e.g., 64) one-dimensional "images". The output of one layer is fed into the next layer as an input. **c**, To allow learning of correlations across large time scales, as they are expected to be relevant in industrial processes, we add "holes" to the kernels (dilated convolutions) that operate on the output of preceding convolutional layers. **d**, The results of all the kernel operations (after applying operations of the forms of **b** and **c** multiple times) are all collected via a "2D" convolution into a predicted emission. From this schematic, it follows that our output sequence cannot be longer than T, the length of the input sequence. To deal with the "edges", we apply (causal) padding with zeros at the front of the input sequence (not shown in the figure).

8.1 Model architecture

Temporal convolutional neural networks contain multiple key design elements (see Fig S. 7):

- *layers of dilated causal convolutions:* convolutional layers found widespread use in computer vision applications and a key factor contributing to the success is the concept of weight sharing. In practice, it has been found that such models are much easier to train than recurrent neural networks. One problem with convolutional layers is that they usually only have a local receptive field. To remedy this problem, dilated convolutions have been developed which differ from "normal" convolutional kernels by having "holes" in the kernel. By increasing the size of the holes one can achieve exponential increases in the receptive field of the model. The term "causal" refers to the fact that one wants to avoid lookahead, i.e., a forecasting model should not depend on future data. For this reason, zero padding is only applied to one side of the time series.

- *residual connections:* also this technique was first developed in the context of computer vision (53), and it was empirically found that the option to skip some layers via an "identity mapping" can stabilise training and boost predictive performance.

- *weight normalisation:* is a reparametrization trick that was found to stabilise and speed up convergence of training.

- *dropout:* is a well known technique for regularising neural networks that works by randomly disabling certain weights.

Note that this is not a model architecture we specifically design for this work. It is already available in the darts library (38).

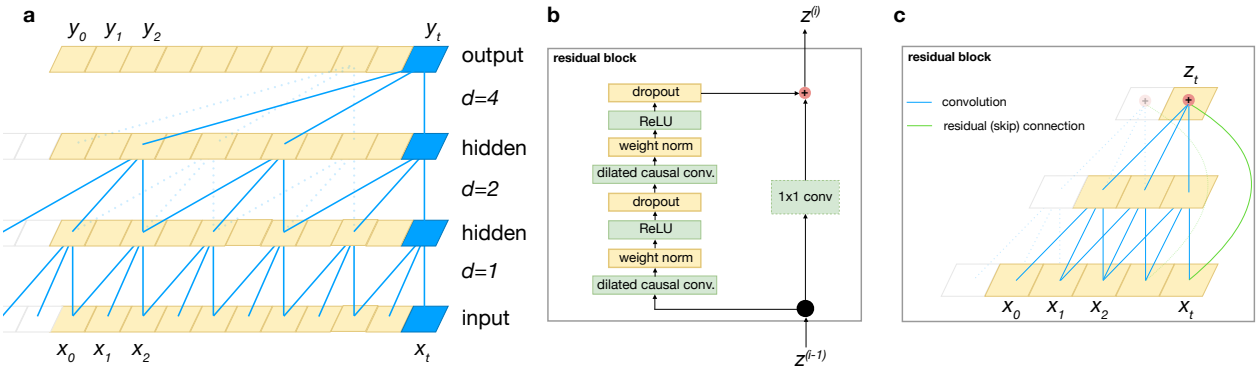


Fig. S7. Building blocks of TCN models. Figure adapted from Bui et al. (1) **a** Dilated convolutions allow exponential increases in temporal resolutions via the use of convolutions with "holes". The parameter d indicates the dilation rate. **b** A residual block contains layers of dilated causal convolutions, dropout, ReLU activation, and weight normalisation. The network also has the possibility to "skip" the dilated convolution. **c** Dilated convolutions with skip connections form a block.

.

Models were trained within a few minutes on an NVIDIA Quadro RTX 6000 graphics card.

8.2 Monte Carlo Dropout

In statistics, one distinguishes between epistemic and aleatoric uncertainty. Aleatoric uncertainty captures noise inherent in the data, whereas epistemic uncertainty is the uncertainty in the model (prior distribution over the weights). With the Monte-Carlo dropout approach, we approximate the latter term by approximating the sampling from the posterior by enabling the dropout layers during inference time. This approach has already been used, for example at Uber (54), to estimate the uncertainty of forecasts. Note that we did not add a term for the aleatoric uncertainty in our uncertainty estimates. Also note that the theoretical justification of this uncertainty estimation procedure is still debated (55, 56).

8.3 Hyperparameter optimisation

For hyperparameter optimisation, we focused on predicting the amine emissions and optimised on a validation set of the AMP emissions. (Implicitly assuming that hyperparameters that perform well for AMP will also perform well for Pz.) For this we performed a time-based split using the first 50 % of the data for training, the subsequent 25 % for validation, and the last 25 % for testing. Note that in our case this is a particularly challenging setting for the model as every day different interventions were performed. We considered the hyperparameter grid in Table

S 10 and optimised for the mean absolute percentage error on historical forecasts on the validation set.

For hyperparameter optimisation we used the Bayesian optimisation approach implemented in weights and biases (https://docs.wandb.ai/guides/sweeps).

Tab. S10. Hyperparameter grid considered in this work as well as final settings. The 60 min, 10 min, and 2 min output sequence length model reached a validation mean absolute percentage error of 10.4 %, 12.05 %, and 12.35 %, respectively.

parameter	range	120 min output	60 min output	10 min output	2 min output
number of convolutional layers	[4, 8, 16]	4	8	4	16
number of filters	[8, 16, 32, 64]	32	64	64	16
weight norm	true / false	false	true	false	false
kernel size	[2, 3, 4, 5]	2	4	3	3
dropout probability	uniform distribution between 0.1 and 0.9	0.5617	0.3668	0.3239	0.1511
batch size	[32, 64, 128]	64	64	128	128
number of epochs	[*100*, **200**, 300, 400]	100	200	100	200
input sequence length	[31, 40, 60, 61, 80, 160] (timestamps)	61	80	31	80
learning rate	uniform sampling in logarithmic space between 10^{-5} and 10^{-1}	0.0192	0.0297	0.0100	0.01197

8.4 Forecasting performance

See Fig. S 7 for historical forecasts.

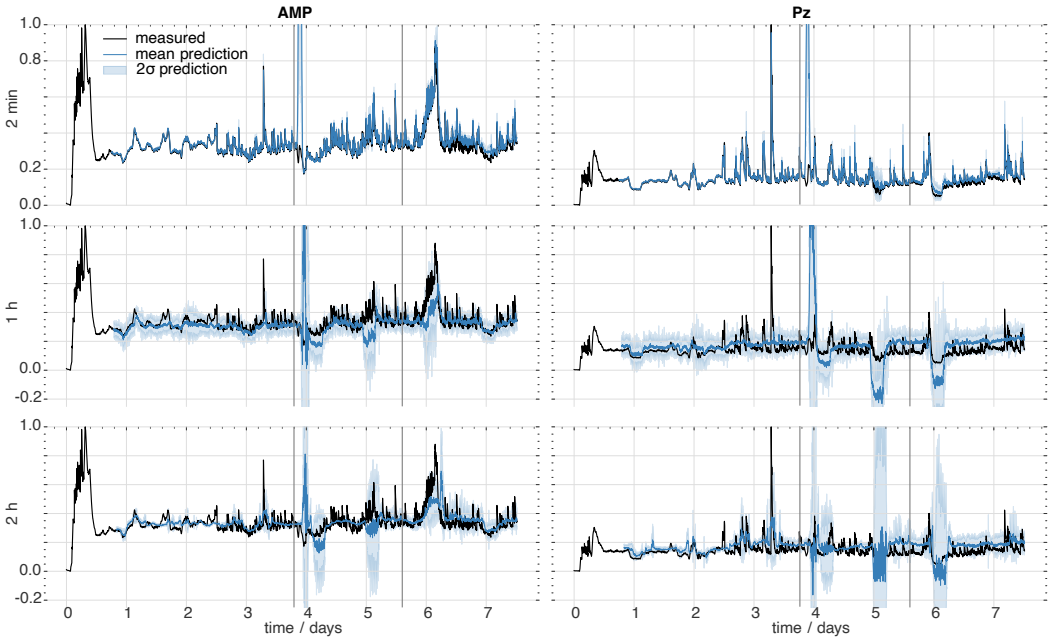


Fig. S8. Historical forecasts for temporal convolutional neural networks for three output sequence length. The 60 min output sequence length model achieved an MAE of 12.05 % on the validation set. The 10 min output sequence length model achieved an MAE of 12.35 % on the validation set. The 2 min output sequence length model achieved an MAE of 12.35 % on the validation set. The vertical lines indicate the validation/test split points. The black curves show the measured emissions, and the blue curves show the forecasts with the solid line indicating the mean and the band indicating the 2σ interval estimated from 50 Monte-Carlo dropout runs.

Supplementary Note 9 VARIMA baseline

We also attempted to use a vector autoregressive moving average with exogenous regressors model (VARIMA) model ($p = q = 5$ with constant and linear trend term[1]) on the same covariates. The predictions are shown in Fig. S 9. We see that also a VARIMA model can learn from the data, however it — as one would expect — struggles to predict the spikes.

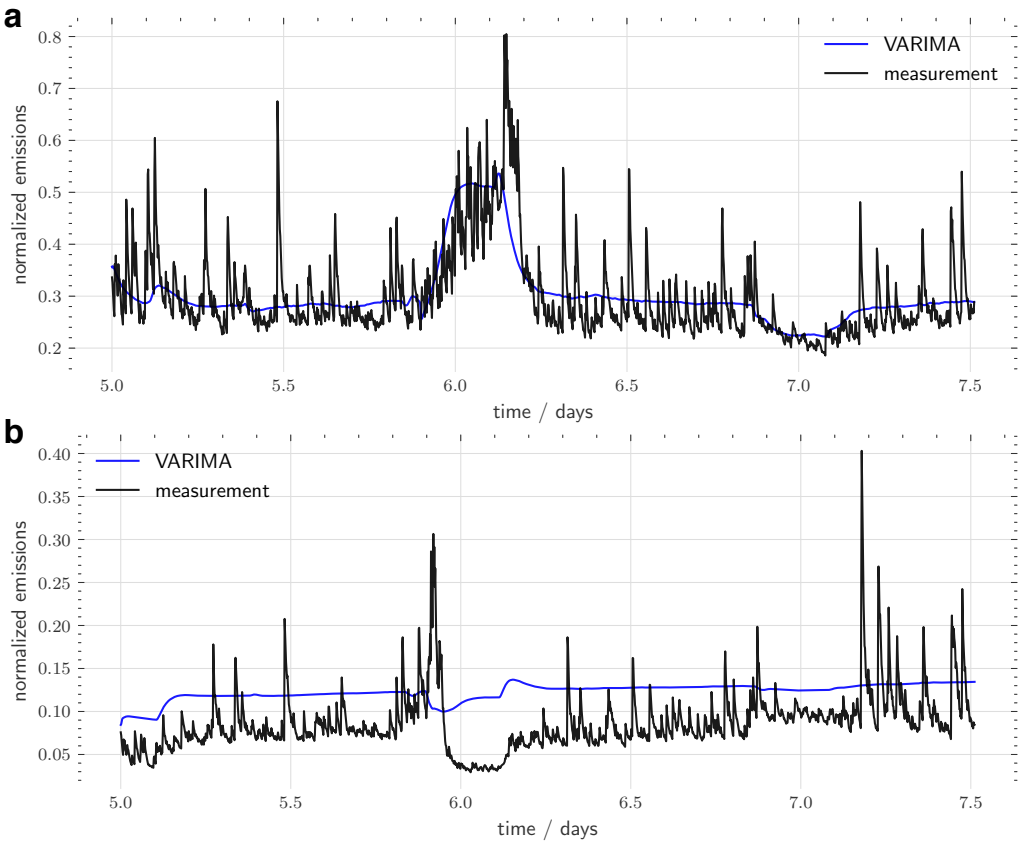


Fig. S9. Forecasts vs. measured AMP (a) and Pz (b) emissions for a VARIMAX model.

Additionally, we investigated the use of the recently proposed Temporal Fusion Transformers (57). However, within our limited tuning (testing hidden layer sizes 8, 16, 32), we did not find them to outperform the temporal convolutional model (while being more expensive in training and inference).

[1]The forecast quality for $p = 20, q = 0$ and $p = 10, q = 0$ are comparable to the ones shown here.

Supplementary Note 10 Causal impact analysis

Tab. S11. **Performance metrics.** Model performance in the pre-intervention periods for AMP.

day	MAPE / %	OPE / %
1	200	nan
2	3.5 1.9	
3	4.4	1.3
4	14	15
5	8.9	3.4
6	9.2	10
7	18	19

Tab. S12. **Performance metrics.** Model performance in the pre-intervention periods for Pz.

day	MAPE / %	OPE / %
1	670	nan
2	9.2	9.4
3	8.6	6.9
4	46	52
5	32	36
6	14	16
7	32	36

10.1 Day 1: Step increase in water wash temperature

For both amines we observe a significant increase in emissions. However, the model is not reliable for this step change as there is only little data preceding the step change.

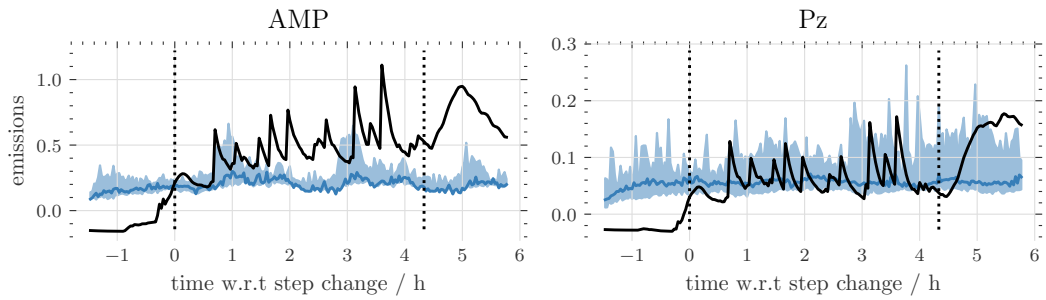


Fig. S10. Causal impact analysis for day 1. Causal impact analysis for the step increase in water wash temperature.

10.2 Day 2: Step decrease in water wash flow

For both amines, we observe a decrease in emissions.

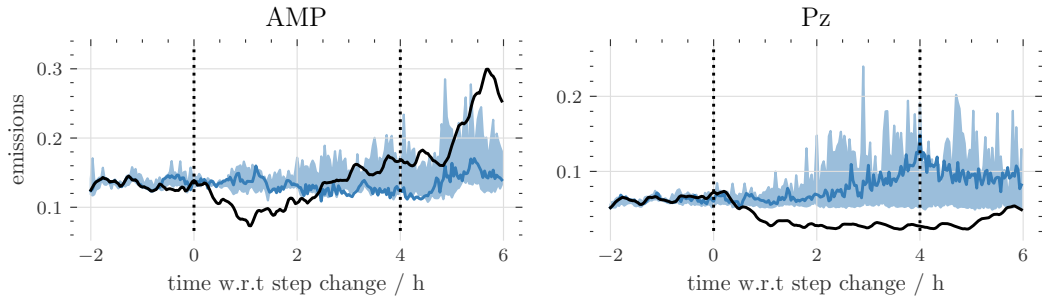


Fig. S11. Causal impact analysis for day 2. Causal impact analysis for the step decrease in water wash flow rate.

10.3 Day 3: Step increase in flue gas temperature

Even though the measured emissions (black) might suggest an increase in emissions, this increase is not statistically significant.

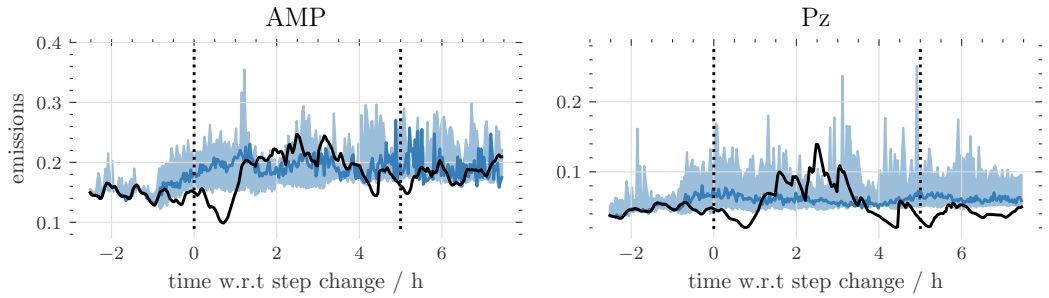


Fig. S12. Causal impact analysis for day 3. Causal impact analysis for the step increase in flue gas temperature.

10.4 Day 4: Step decrease in lean solvent flow

The model show relatively large prediction intervals and we cannot see a significant effect.

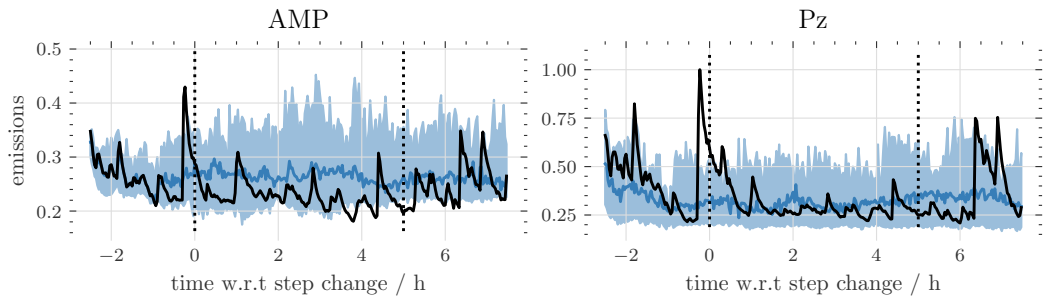


Fig. S13. Causal impact analysis for day 4. Causal impact analysis for the step decrease in lean solvent flow.

10.5 Day 5: Lean solvent flow decrease and flue gas flow decrease

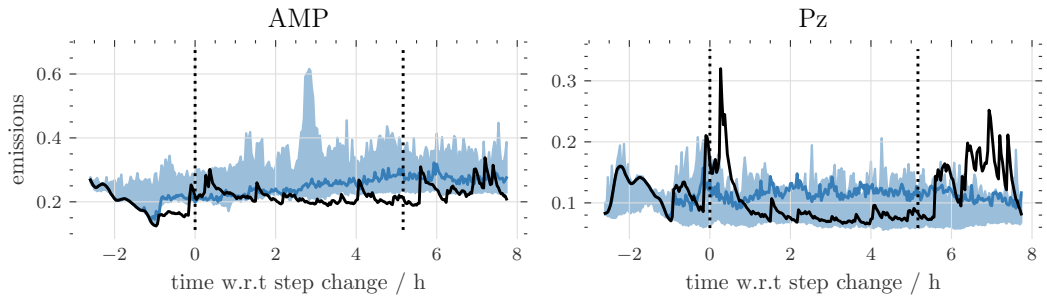


Fig. S14. Causal impact analysis for day 5. Causal impact analysis for the lean solvent flow decrease and flue gas flow decrease.

10.6 Day 6: Step increase in lean solvent temperature

We can observe a slight decrease in Pz emissions compared to the baseline predictions.

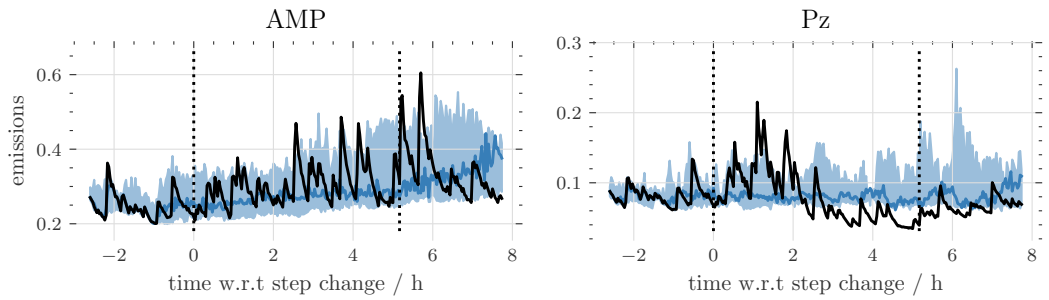
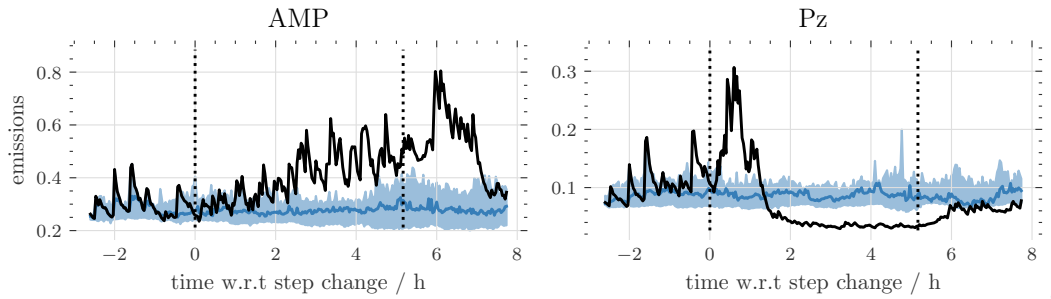


Fig. S15. Causal impact analysis for day 6. Causal impact analysis for the step increase in lean solvent temperature.

10.7 Day 7: Lean solvent and water wash temperature increase

We observe a decrease in emissions for Pz and increase in emissions for AMP compared to the baseline.



conda

activate tfmetal

Fig. S16. Causal impact analysis for day 7. Causal impact analysis for the lean solvent and water wash temperature increase.

Supplementary Note 11 Emission mitigation (scenarios)

11.1 Method

Assumptions Note that the modelling of scenarios is based on several assumptions:

- We assume that the dynamics of the systems are unchanged.
- We assume that other, possibly correlated variables, remain unchanged. (comparably to the assumptions in partial dependency plots) (58). Note that we recompute composite variables such as temperature differences, after performing the perturbations (before running the models).
- We assume that the cumulative change in emissions is a meaningful measure of the impact of the interventions.

Note that the maps shown in the main text are computed based on historical forecasts. That is, we use as inputs for predicting the next timesteps the actual observed emissions together with the changed inputs in process parameters.

We need to impose such harsh assumptions since, due to the time ordering of the inputs, conventional black box explanation methods such as SHAP (59) cannot directly be applied. Note that while this analysis does not provide us with a causal interpretation it still can reveal important patterns in the emission mechanisms.

Algorithm To compute the heatmaps, we create a mesh grid of values, typically of dimension 21×21 ranging from –20 % to 20 % relative change. For every point in the grid, we change the values of the two features on the axis according to the coordinate tuple (p_i, p_j) and run a forecast using our model, resulting in the time series $F(p_i, p_j, t)$. For the heatmap, we then compute

$$E(p_i, p_j) = \sum_t \left(F(p_i, p_j, t) - F(0, 0, t)\right) \tag{5}$$

For visualisation purpose we smooth the matrix $E(p_i, p_j)$ using a Gaussian filter. Note that the Gaussian filter is applied after centering the data according to eq. 5 which might lead the centre of the map not being centred at exactly zero. To remedy this, we recenter the maps after the smoothing step.

11.2 Additional maps

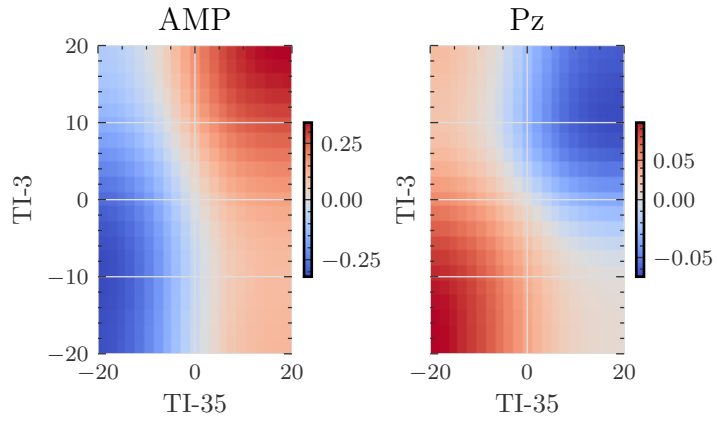


Fig. S17. Scenario in TI-3 and TI-35. Emissions as a function of change (in percent) in TI-3 and TI-35 computed using historical forecasts of one-step-ahead predictions.

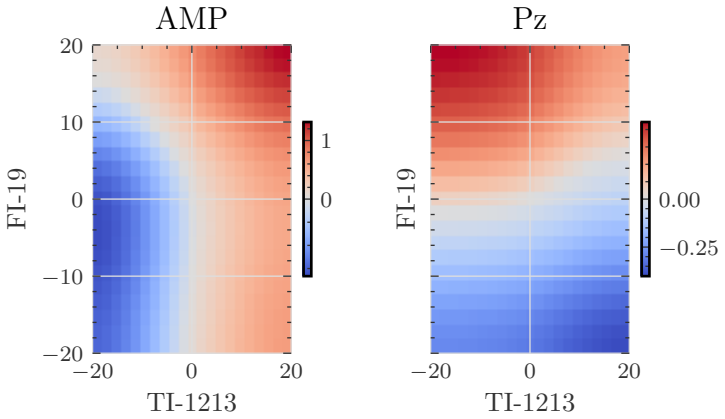


Fig. S18. Scenario in FI-19 and TI-1213. Emissions as a function of change (in percent) in FI-19 and TI-1213 computed using historical forecasts of one-step-ahead predictions.

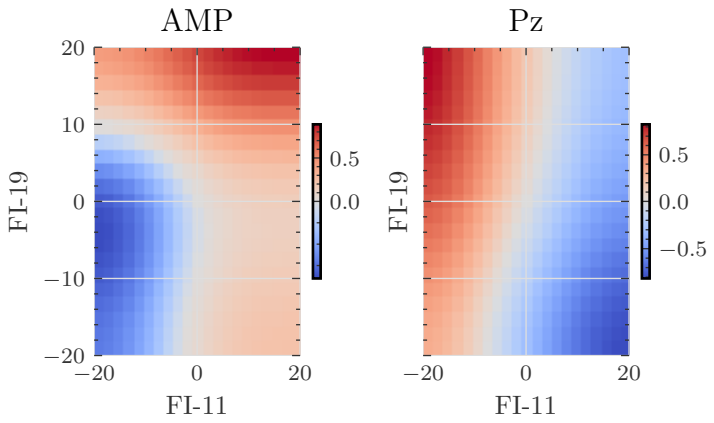


Fig. S19. Scenario in FI-19 and FI-11. Emissions as a function of change (in percent) in FI-19 and FI-11 computed using historical forecasts of one-step-ahead predictions.

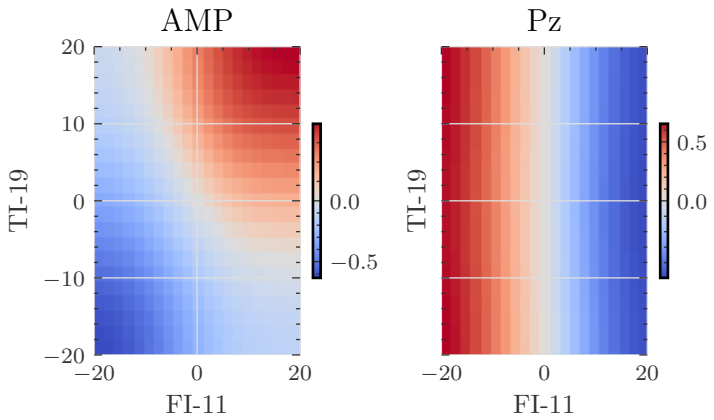


Fig. S20. Scenario in TI-19 and FI-11. Emissions as a function of change (in percent) in TI-19 and FI-11 computed using historical forecasts of one-step-ahead predictions.

11.3 Using forecasts as input for new forecasts

The maps in the main text are computed using historical forecasts for one-step-ahead predictions. This implies that we assume that to assume that the emissions before a prediction are as measured — but with changed values for some covariates.

In this section, we lift this assumption and use the predicted historic emissions to compute forecasts.

From the figures we see that the overarching conclusion that the two amines behave differently for different possible mitigation measures still holds true under this perspective.

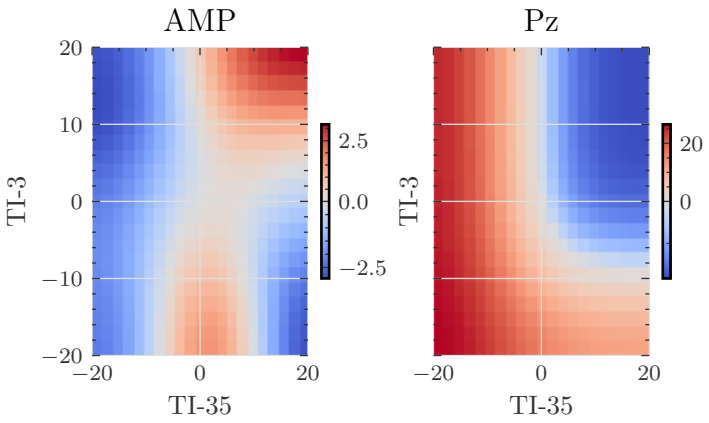


Fig. S21. Scenario in TI-3 and TI-35. Emissions as a function of change (in percent) in TI-3 and TI-35 computed using forecasted emissions as input for subsequent forecasts.

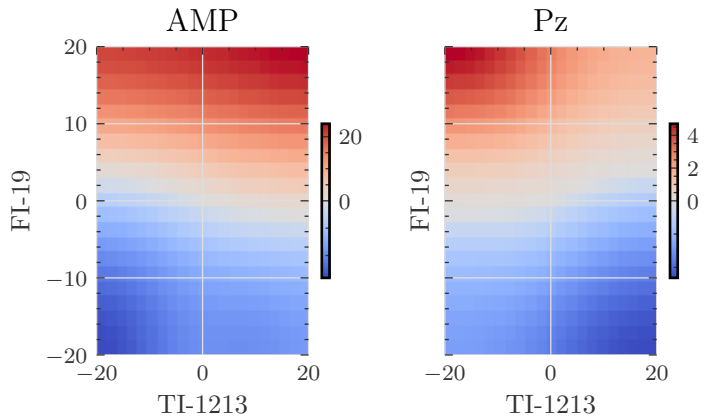


Fig. S22. Scenario in FI-19 and TI-1212. Emissions as a function of change (in percent) in FI-19 and TI-1213 computed using forecasted emissions as input for subsequent forecasts.

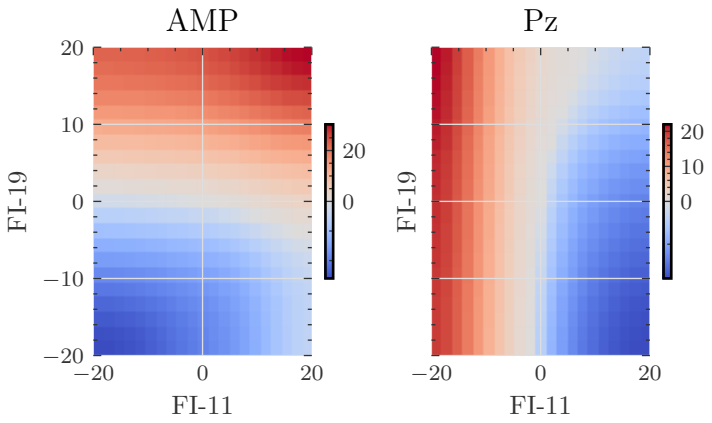


Fig. S23. Scenario in FI-19 and FI-11. Emissions as a function of change (in percent) in FI-19 and FI-11 computed using forecasted emissions as input for subsequent forecasts.

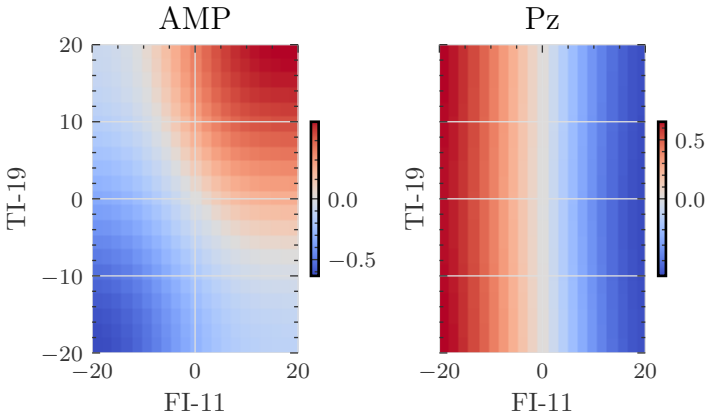


Fig. S24. Scenario in TI-19 and FI-11. Emissions as a function of change (in percent) in TI-19 and FI-11 computed using forecasted emissions as input for subsequent forecasts.

Supplementary Note 12 Caveats of the modelling approach

Clearly, there is no guarantee that our model learned causal relationships (i.e., true cointegration in contrast to spurious correlations) and the process parameter - emissions relationships shown in the heatmaps might be influenced by indirect influences. To counteract this, one could build graphical models (time series chain graphs) (60).

For the interpretation of our results this implies that any strong effect we predict, for example, in our heatmaps might equally well be caused by a variable that is strongly correlated. Note though that we performed careful feature selection guided by discussion with domain experts to mitigate learning of spurious correlations. However, this also implies that the approach presented in this work is no silver bullet that can be applied to any dataset without any data preprocessing.

12.1 Consistency of feature attributions

Our feature attribution techniques does not fulfil all requirements that Lundberg et al. enumerate (59). However, we observed that the trends, in particular the qualitatively different behaviour of AMP and Pz, is consistent between models trained on different feature sets.

Supplementary Note 13 Checklist for reporting and evaluating machine learning models

Checklist according to Artrith et al (61).

13.1 Data sources

1. *Are all data sources listed and publicly available?*
   We provide raw and processed data on Zenodo.

2. *If using an external database, is an access date or version number provided?*
   Not applicable.

3. *Are any potential biases in the source dataset reported and/or mitigated?*
   We discuss the facts that plant was not at steady-state operation and potential degradation of the solvent in the article and Supplementary Material.

13.2 Data cleaning

1. *Are the data cleaning steps clearly and fully described, either in text or as a code pipeline?*
   The data pre-processing is described in the methods section and the impact is described in the Supplementary Material. We provide the code in form of Jupyter notebooks.

2. *Is an evaluation of the amount of removed source data presented?*
   We did not consider data from the two days after the shutdown of the power plant due to instability of the capture plant.

3. *Are instances of combining data from multiple sources clearly identified, and potential issues mitigated?*
   Not applicable.

13.3 Data representations

1. *Are methods for representing data as features or descriptors clearly articulated, ideally with software implementations?*
   Not directly applicable. However, the manual feature selection is documented in the Supplementary Material.

2. *Are comparisons against standard feature sets provided?*
   Not directly applicable. However, we compare a representation learning approach (temporal convolutional neural network) with a gradient boosted decision tree.

13.4 Model choice

1. *Is a software implementation of the model provided such that it can be trained and tested with new data?*
   The code is provided on GitHub. We also provide model checkpoints. Since the code depends on some customizations, we could not provide the models in an interoperable format such as ONNX.

2. *Are baseline comparisons to simple/trivial models (for example, 1-nearest neighbour, random forest, most frequent class) provided?*
   We also show the forecasting performance of VARIMAX models and temporal convolutional neural networks.

3. *Are baseline comparisons to current state-of-the-art provided?*
   Not clear what the current state-of-the-art is.

13.5 Model training and validation

1. *Does the model clearly split data into different sets for training (model selection), validation (hyperparameter optimisation), and testing (final evaluation)?*
   We performed a time-based split into training, validation and testing sets as indicated by vertical lines in our plots.

2. *Is the method of data split (data splitting (for example, random, cluster- or time-based splitting, forward cross-validation) clearly stated? Does it mimic anticipated real-world application?*
   For the timeseries model, the split for the time-based data splitting is indicated in the figure and reflects potential use for forecasting applications.

3. *Does the data splitting procedure avoid data leakage (for example, is the same composition present in the training and test sets)?*
   The time-based split used for the time-series models avoids lookahead.

13.6 Code and reproducibility

1. *Is the code or workflow available in a public repository?*
   We provide the code on GitHub.

2. *Are scripts to reproduce the findings in the paper provided?*
   Notebooks and scripts that can be used to reproduce the findings in the article can be found in the `paper` directory on GitHub.

REFERENCES AND NOTES

1. M. Bui, C. S. Adjiman, A. Bardow, E. J. Anthony, A. Boston, S. Brown, P. S. Fennell, S. Fuss, A. Galindo, L. A. Hackett, J. P. Hallett, H. J. Herzog, G. Jackson, J. Kemper, S. Krevor, G. C. Maitland, M. Matuszewski, I. S. Metcalfe, C. Petit, G. Puxty, J. Reimer, D. M. Reiner, E. S. Rubin, S. A. Scott, N. Shah, B. Smit, J. P. M. Trusler, P. Webley, J. Wilcox, N. M. Dowell, Carbon capture and storage (CCS): The way forward. *Energ. Environ. Sci.* **11**, 1062–1176 (2018).

2. A. J. Reynolds, T. V. Verheyen, S. B. Adeloju, E. Meuleman, P. Feron,Towards commercial scale postcombustion capture of CO_2 with monoethanolamine solvent: Key considerations for solvent management and environmental impacts. *Environ. Sci. Technol.* **46**, 3643–3654 (2012).

3. K. Veltman, B. Singh, E. G. Hertwich, Human and environmental impact assessment of postcombustion CO_2 capture focusing on emissions from amine-based scrubbing solvents to air. *Environ. Sci. Technol.* **44**, 1496–1502 (2010).

4. IEA Greenhouse Gas R&D Programme, Environmental Impacts of Amine Emission During Post Combustion Capture (2010); www.globalccsinstitute.com/archive/hub/publications/106171/environmental-impacts-amine-emissions-post-combustion-capture.pdf.

5. P. Khakharia, J. Mertens, M. Abu-Zahra, T. Vlugt, E. Goetheer, in *Absorption-Based Post-combustion Capture of Carbon Dioxide*, P. H. M. Feron, Ed. (Elsevier, 2016), pp. 465–485.

6. L. Biegler, *Systematic Methods of Chemical Process Design* (Prentice Hall PTR, 1997).

7. IEAGHG, *Valuing Flexibility in CCS Power Plants* (Tech. Rep. 2017-09, 2017); https://ieaghg.org/exco_docs/2017-09.pdf.

8. N. E. Flø, H. M. Kvamsdal, M. Hillestad, Dynamic simulation of post-combustion CO2 capture for flexible operation of the Brindisi pilot plant. *Int. J. Greenh. Gas Control.* **48**, 204–215 (2016).

9. J. Gaspar, J. B. Jorgensen, P. L. Fosbol, Control of a post-combustion CO2 capture plant during process start-up and load variations. *IFAC-PapersOnLine* **48**, 580–585 (2015).

10. H. Chalmers, M. Leach, M. Lucquiaud, J. Gibbins, Valuing flexible operation of power plants with CO2 capture. *Energy Procedia* **1**, 4289–4296 (2009).

11. N. E. Flø, H. M. Kvamsdal, M. Hillestad, T. Mejdell, Dominating dynamics of the post-combustion CO2 absorption process. *Comput. Chem. Eng.* **86**, 171–183 (2016).

12. C. Charalambous, A. Saleh, M. van der Spek, G. Wiechers, P. Moser, A. Huizinga, P. Gravesteijn, J. Ros, J. G. M.-S. Monteiro, E. Goetheer, S. Garcia, Analysis of flexible operation of CO_2 capture plants: Predicting solvent emissions from conventional and advanced amine systems. *SSRN Electron. J.* (2021).

13. A. Kachko, L. V. van der Ham, L. F. G. Geers, A. Huizinga, A. Rieder, M. R. M. Abu-Zahra, T. J. H. Vlugt, E. L. V. Goetheer, Real-time process monitoring of CO_2 capture by aqueous AMP-PZ using chemometrics: Pilot plant demonstration. *Ind. Eng. Chem. Res.* **54**, 5769–5776 (2015)

14. J. Pearl, *Causality* (Cambridge Univ. Press, 2009).

15. P. Moser, S. Schmidt, G. Sieder, H. Garcia, T. Stoffregen, Performance of MEA in a long-term test at the post-combustion capture pilot plant in Niederaussem. *Int. J. Greenh. Gas Control* **5**, 620–627 (2011).

16. P. Moser, S. Schmidt, K. Stahl, Investigation of trace elements in the inlet and outlet streams of a MEA-based post-combustion capture process results from the test programme at the Niederaussem pilot plant. *Energy Procedia* **4**, 473–479 (2011).

17. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS' 17 (Curran Associates Inc., 2017), pp. 3149–3157.

18. J. H. Friedman, Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).

19. K. Das, M. Krzywinski, N. Altman, Quantile regression. *Nat. Methods* **16**, 451–452 (2019).

20. R. Koenker, G. Bassett Jr., *Econometrica* **46**, 33–50 (1978).

21. M. A. Hernán, J. M. Robins, *Causal Inference* (CRC, 2010).

22. K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, S. L. Scott, Inferring causal impact using Bayesian structural time-series models. *Ann. Appl. Stat.* **9**, 247–274 (2015).

23. A. Hartono, H. F. Svendsen, H. K. Knuutila, Impact of absorption kinetics of individual amine components in CESAR1 solvent on aerosol model performance in the absorber, in *Proceedings of the 15th Greenhouse Gas Control Technologies Conference* (SSRN, 2021).

24. J. Mertens, H. Lepaumier, D. Desagher, M.-L. Thielens, Understanding ethanolamine (MEA) and ammonia emissions from amine based post combustion carbon capture: Lessons learned from field tests. *Int. J. Greenh. Gas Control.* **13**, 72–77 (2013).

25. P. Khakharia, L. Brachert, J. Mertens, C. Anderlohr, A. Huizinga, E. S. Fernandez, B. Schallert, K. Schaber, T. J. Vlugt, E. Goetheer, Understanding aerosol based emissions in a post combustion CO_2 capture process: Parameter testing and mechanisms. *Int. J. Greenh. Gas Control.* **34**, 63 (2015).

26. A. F. Ciftja, A. Hartono, E. F. da Silva, H. F. Svendsen, Study on carbamate stability in the Amp/CO_2/H_2O system from ^{13}C-NMR spectroscopy, *Energy Procedia* **4**, 614–620 (2011).

27. K. M. Jablonka, G. M. Jothiappan, S. Wang, B. Smit, B. Yoo, Bias free multiobjective active learning for materials design and discovery. *Nat. Commun.* **12**, 2312 (2021).

28. T. Lookman, P. V. Balachandran, D. Xue, R. Yuan, Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput. Mater.* **5** (2019).

29. K. M. Jablonka, L. Patiny, B. Smit, Making the collective knowledge of chemistry open and machine actionable. *Nat. Chem.* **14**, 365–376 (2022).

30. A. M. Schweidtmann, E. Esche, A. Fischer, M. Kloft, J.-U. Repke, S. Sager, A. Mitsos, Machine learning in chemical engineering: A perspective. *Chem. Ing. Tech.* **93**, 2029–2039 (2021).

31. J. M. Weber, Z. Guo, C. Zhang, A. M. Schweidtmann, A. A. Lapkin, Chemical data intelligence for sustainable chemistry. *Chem. Soc. Rev.* **50**, 12013–12036 (2021).

32. P. Moser, S. Schmidt, K. Stahl, G. Vorberg, G. A. Lozano, T. Stoffregen, F. Rösler, Demonstrating emission reduction – Results from the post-combustion capture pilot plant at Niederaussem. *Energy Procedia* **63**, 902–910 (2014).

33. A. Rieder, S. Dhingra, P. Khakharia, L. Zangrilli, B. Schallert, R. Irons, S. Unterberger, P. Van Os, E. Goetheer, Understanding solvent degradation: A study from three different pilot plants within the OCTAVIUS Project. *Energy Procedia* **114**, 1195–1209 (2017).

34. E. F. da Silva, K. A. Hoff, A. Booth, Emissions from CO_2 capture plants; An overview. *Energy Procedia* **37**, 784–790 (2013).

35. M. Bui, I. Gunawan, V. Verheyen, P. Feron, E. Meuleman, S. Adeloju, Dynamic modelling and optimisation of flexible operation in post-combustion CO_2 capture plants—A review. *Comput. Chem. Eng.* **61**, 245–265 (2014).

36. M. Bui, N. E. Flø, T. de Cazenove, N. M. Dowell, Demonstrating flexible operation of the Technology Centre Mongstad (TCM) CO_2 capture plant. *Int. J. Greenh. Gas Control.* **93**, 102879 (2020).

37. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural. Inf. Process Syst.* **30**, 3146 (2017).

38. J. Herzen, F. Lässig, S. G. Piazzetta, T. Neuer, L. Tafti, G. Raille, T. V. Pottelbergh, M. Pasieka, A. Skrodzki, N. Huguenin, M. Dumonal, J. Kościsz, D. Bader, F. Gusset, M. Benheddi, C. Williamson, M. Kosinski, M. Petrik, G. Grosch, Darts: User-friendly modern machine learning for time series. *J. Mach. Learn. Res.* **23**, 1–6 (2022).

39. G. Van Rossum, F. L. Drake, *Python 3 Reference Manual* (CreateSpace, 2009).

40. W. McKinney, Data structures for statistical computing in python, in *Proceedings of the 9th Python in Science Conference* (SciPy 2010), Austin, Texas, 28 June to 3 July 2010, vol. 445, pp. 51–56.

41. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

42. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro,F. Pedregosa, P. van Mulbregt; SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods* **17**, 261–272 (2020).

43. S. Seabold, J. Perktold, Statsmodels: Econometric and statistical modeling with Python, in *9th Python in Science Conference* (SciPy 2010), Austin, Texas, 28 June to 3 July 2010.

44. J. D. Hunter, Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

45. T. Kluyver, B. Ragan-Kelley, F. Pérez, B. E. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. B. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing; Jupyter Development Team, Jupyter Notebooks-a publishing format for reproducible computational

workflows, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (IOS Press, 2016), vol. 2016.

46. C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with NumPy. *Nature* **585**, 357–362 (2020).

47. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett, Eds. (Curran Associates Inc., 2019), pp. 8024–8035.

48. S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).

49. J. Monteiro, J. Ros, E. Skylogiani, A. Hartono, H. Svendsen, H. Knuutila, P. Moser, G. Wiechers, C. Charalambous, S. Garcia, Accelerating Low carboN Industrial Growth through CCUS Deliverable Nr. D1.1.7 Guidelines for Emissions Control (Tech. Rep., 2021).

50. E. Mechleri, A. Lawal, A. Ramos, J. Davison, N. M. Dowell, Process control strategies for flexible operation of post-combustion CO_2 capture plants. *Int. J. Greenh. Gas Control.* **57**, 14–25 (2017).

51. M. Bui, I. Gunawan, V. Verheyen, P. Feron, E. Meuleman, Flexible operation of CSIRO's post-combustion CO_2 capture pilot plant at the AGL Loy Yang power station. *Int. J. Greenh. Gas Control.* **48**, 188–203 (2016).

52. S. Bai, J. Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271 [cs.LG] (4 March 2018).

53. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. arXiv:1512.03385 [cs.CV] (10 December 2015).

54. L. Zhu, N. Laptev, Deep and confident prediction for time series at Uber. *IEEE Int. Conf. Data Mining Workshops*, 103–110 (2017).

55. J. Hron, A. G. de G. Matthews, Z. Ghahramani, Variational Gaussian Dropout is not Bayesian. arXiv:1711.02989 [stat.ML] (8 November 2017).

56. J. Hron, A. G. de G. Matthews, Z. Ghahramani, Variational Bayesian dropout: Pitfalls and fixes. arXiv 1807.01969 [stat.ML] (5 July 2018).

57. B. Lim, S. O. Arik, N. Loeff, T. Pfister, Temporal fusion transformers for interpretable multi-horizon time series forecasting **37**, 1748–1764 (2021).

58. C. Molnar, *Interpretable Machine Learning* (2019); https://christophm.github.io/interpretable-ml-book/.

59. S. M. Lundberg, S.-I. Lee, in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, Eds. (Curran Associates Inc., 2017), pp. 4765–4774.

60. M. Eichler, Granger causality and path diagrams for multivariate time series. *J. Econom.* **137**, 334–353 (2007).

61. N. Artrith, K. T. Butler, F.-X. Coudert, S. Han, O. Isayev, A. Jain, A. Walsh, Best practices in machine learning for chemistry. *Nat. Chem.* **13**, 505–508 (2021).