

# Real-Time Dilated CNN Denoising for SOEC Stack Operational Monitoring

Elvan Sahin<sup>a</sup>, Nicholas Kane<sup>a</sup>, Jeremy Hartvigsen<sup>a</sup>, Temitayo Olowu<sup>a</sup>, and Micah Casteel<sup>a</sup>

<sup>a</sup>Idaho National Laboratory, Idaho Falls, Idaho, USA, [elvan.sahin@inl.gov](mailto:elvan.sahin@inl.gov), [nicholas.kane@inl.gov](mailto:nicholas.kane@inl.gov), [jeremy.hartvigsen@inl.gov](mailto:jeremy.hartvigsen@inl.gov), [temitayo.olowu@inl.gov](mailto:temitayo.olowu@inl.gov), [micah.casteel@inl.gov](mailto:micah.casteel@inl.gov)

---

**Abstract:** Long-duration solid oxide electrolysis cell stack tests generate multivariate voltage, temperature, flow, and control signals that are essential for degradation assessment and operational monitoring. These measurements contain correlated sensor noise, balance-of-plant disturbances, and slow drift, making it difficult to distinguish meaningful stack-health trends from measurement variability. This paper presents a real-time signal-conditioning framework for SOEC operational data. The framework first characterizes noise across 63 sensor channels using signal-to-noise ratio, autocorrelation time, power spectral density slope, and cross-channel coherence. It then benchmarks simple moving average, Savitzky–Golay, and Butterworth filters under a fixed signal-fidelity constraint. Finally, it applies a compact dilated convolutional neural network with sequence-to-sequence output, per-window normalization, and overlap averaging. The model is trained on two operational files and evaluated on a fully held-out third file from the same operating regime. Across four target channels, the CNN reduces residual noise by 18–40% relative to the best classical filter while preserving high correlation with the pseudo-ground-truth trend. Inference benchmarks show that the model can process data faster than acquisition rates, supporting deployment in online monitoring pipelines. The results demonstrate that structured noise in SOEC stack testing can be learned from operational data, while also motivating future validation across additional operating regimes and degradation scenarios.

---

**Keywords:** Hydrogen Stack Monitoring; Real-Time Signal Denoising; Dilated Convolutional Neural Network; Solid Oxide Electrolysis Cell.

## 1. INTRODUCTION

Solid oxide electrolysis cells (SOECs) are a leading candidate for high-efficiency hydrogen production from renewable electricity, but their commercial deployment depends on demonstrating predictable performance over thousands of hours of operation [1]. Long-duration testing therefore plays a central role in technology qualification: stack voltage, temperature, and gas-flow channels are sampled at high rates over hundreds of hours to detect slow degradation, identify incipient faults, and validate balance-of-plant control. The diagnostic value of such campaigns rests directly on the quality of the recorded signals.

In practice, raw measurements from a long-running SOEC test bench are corrupted by several superimposed disturbances: thermal–electrical fluctuations in the furnace and power-supply electronics, balance-of-plant transients (compressors, heaters, mass-flow controllers), and instrumentation drift [2]. The resulting noise is rarely white. Voltage channels typically exhibit temporally correlated, low-frequency-dominated structure, while thermocouple noise has the  $1/f$  character of thermal-fluctuation processes. The amplitude of this noise is non-trivial relative to the quantities of interest: SOEC stacks at on-load degrade at rates on the order of millivolts per thousand hours [3], whereas raw voltage standard deviations on a single 24-hour window can be one to two orders of magnitude larger. Without effective conditioning, useful trends are masked, and operational decisions risk being driven by noise.

Classical smoothing techniques — simple moving averages (SMA), Savitzky–Golay polynomial filters [4], and Butterworth low-pass filters — are the de facto baseline for signal conditioning in electrochemical testing. They are transparent, computationally trivial, and easy to deploy. However, all three are linear, time-invariant, and parameter-fixed; they cannot exploit the multivariate structure of the measurement chain, and their noise-reduction capability is fundamentally bounded by the spectral overlap between the

noise and the underlying physical dynamics. Aggressive parameterizations remove more noise at the cost of biasing the very degradation rates one is trying to estimate.

Data-driven sequence models offer a route past this trade-off. Dilated causal convolutions, popularized by WaveNet [5] and the temporal convolutional network (TCN) [6], achieve large effective receptive fields with modest parameter counts and admit fully parallel inference, in contrast to recurrent alternatives. They have been used successfully for audio denoising and biomedical signal restoration [7, 8], but applications in industrial energy-system monitoring remain limited, particularly with the rigorous train/test discipline required to demonstrate genuine generalization rather than file-specific memorization.

This paper makes three contributions:

1. A reusable noise-characterization protocol for SOEC test data, combining per-channel signal-to-noise ratio (SNR), autocorrelation time, power spectral density (PSD) slope, and cross-channel coherence to inform model design choices.
2. A constrained classical-filter benchmark that reports noise reduction under an explicit signal-fidelity constraint, making the comparison reproducible and fair.
3. A compact dilated CNN with sequence-to-sequence output, per-window normalization, and 16-fold overlap averaging, evaluated under a fully held-out file split.

The framework is demonstrated on operational data from a fuel cell electrolysis (FCE) stack tested at Idaho National Laboratory (INL), with 2,798 hours of on-load operation.

## 2. EXPERIMENTAL DATA AND OPERATING REGIMES

Operational data were recorded from FCE stack experiments at INL. The data acquisition system logged 63 sensor channels at a uniform 3-second sampling interval, covering cell-group and stack voltages, power-supply current and voltage, furnace and gas-stream thermocouples, gas mass-flow rates, and control setpoints.

### 2.1. Operating-Regime Classification

Files were classified by temperature, current and voltage profile to isolate on-load operations suitable for denoising analysis. Of the 21 files, 13 contained full on-load operation, 4 were partially on-load, and the remainder were dominated by start-ups, shutdowns, or ramps and were excluded.

Three on-load regimes emerged. Group A operated at 48.6 A and  $\approx 709$  °C across six files spanning 1,263 hours (June–August) and exhibited a measurable voltage drift consistent with early-life degradation. Group B operated at 48.6 A and  $\approx 716$  °C across three files totaling 836 hours (September), with a current standard deviation effectively at zero and the cleanest signal quality of the campaign. Group C operated at 56.7 A and  $\approx 722$  °C across two files totaling 601 hours (November).

Group B was selected as the development and evaluation dataset because (i) its tight current control eliminates current-driven contributions to voltage variance, isolating the noise of interest; (ii) its three constituent files (SEP 6, SEP 16, SEP 29) cover comparable durations and thus permit a leave-one-file-out evaluation; and (iii) its higher temperature minimizes residual drift effects, allowing the pseudo-ground-truth construction.

## 3. NOISE CHARACTERIZATION

All 63 channels in the Group B dataset were first screened and categorized based on signal variability and data quality. Eleven channels corresponding to constant set-points (e.g., PS1 current limit, H<sub>2</sub> and air flow set-points, furnace temperature set-points) exhibited zero variance and were excluded, as they contain no measurable noise. Three near-constant channels (e.g., PS1 current) were retained for contextual reference but not included in detailed noise analysis. Five channels were identified as invalid due to data anomalies

— specifically, three duplicated furnace internal readings and two cell-rate channels containing non-physical infinity values — and were also excluded. The remaining 40 channels were considered suitable for analysis.

To provide an initial assessment of variability, the usable channels were grouped by coefficient of variation (CV) into three categories: high-noise (CV > 10%, 9 channels), moderate-noise (CV between 1% and 7%, 7 channels), and low-noise (CV < 1%, 24 channels). These thresholds were selected to separate distinct variability regimes observed in the dataset.

### 3.1. Per-Signal Noise Metrics

Four noise metrics were computed for each usable channel. The SNR was estimated by separating the signal into a slow trend (60-second median filter) and a noise residual; reported in dB as:

$$SNR_{dB} = 10 \log_{10} \left( \frac{\sigma_{signal}}{\sigma_{noise}} \right) \quad (1)$$

The autocorrelation function of the residual yielded the noise correlation time, defined here as the lag at which autocorrelation falls below  $1/e \approx 0.368$ . Welch’s method [9] was used to estimate the residual PSD, and the log–log slope of the noise floor was fitted to classify the noise as white (slope  $\approx 0$ ), pink/1/f (slope  $\approx -1$ ), or red/brown (slope  $\approx -2$ ) [10]. Slopes between -0.8 and -1.5 were classified as red/brown-like, and slopes between -0.2 and -0.7 as pink-like, reflecting continuous variation in real sensor noise. Finally, the ratio of maximum to minimum 10-minute rolling standard deviation quantified noise stationarity.

Two dominant noise regimes emerged in Table 1. The electrochemical channels — Stack Voltage 1, Cell Group 1-01 through 1-05, and PS1 Voltage — exhibited red/brown noise with PSD slopes near -1.3, indicating temporally correlated, drift-like fluctuations rather than thermal-Johnson noise. The thermal channels — furnace and stack thermocouples and gas-stream temperatures — exhibited pink (1/f) noise with slopes near -0.4, consistent with thermal-fluctuation processes in heated systems [11]. Three tightly regulated set-points (Steam Temperature, Air Flow, H<sub>2</sub> Flow) had negative SNR, confirming they are not denoising targets. The noise correlation time was uniformly short (3–6 s, 1–2 lags), establishing a hard lower bound on the temporal resolution any useful denoiser must preserve.

**Table 1. Noise characterization of representative measurement channels (Group B)**

Channel	SNR (dB)	Corr. Time (s)	PSD Slope	Noise Type
Stack Voltage 1	12.4	6	-1.32	Red/brown-like
Cell Group 1-01	21.2	6	-1.17	Red/brown-like
Cell Group 1-02	7.8	6	-1.32	Red/brown-like
PS1 Voltage	12.3	6	-1.13	Red/brown-like
FCE Stack TC#1	31.0	3	-0.42	Pink-like
Fuel In	15.0	6	-0.50	Pink-like
Air In	31.4	3	-0.42	Pink-like
Furnace Inlet	20.4	9	-1.35	Red/brown-like

### 3.2. Cross-Channel Structure

Correlation analysis of the noise residuals showed a clear pattern among the voltage channels. All seven voltage signals are highly correlated, with Pearson correlation coefficients above 0.93. This suggests that they are influenced by a common source of noise, most likely the shared power supply and current path used by the stack.

Among these signals, Cell Groups 1-01 through 1-05 are almost identical to Stack Voltage 1 in terms of noise behavior ( $r = 0.968\text{--}0.998$ ), indicating strong redundancy. The thermal channels show a different pattern. Most of them are weakly correlated with each other, with over 80% of channel pairs having

correlation coefficients below 0.1. Using a correlation threshold of  $r > 0.7$ , about 20 small clusters can be identified, indicating that thermal noise is largely independent across channels. The few moderate correlations that do appear — for example, between Inlet Stack Delta and Fuel In ( $r = 0.797$ ), and between Furnace Inlet and UPSS Outlet ( $r = 0.670$ ) — can be explained by the physical proximity of sensors along the gas flow path.

These observations directly informed the model’s input feature selection. Because the seven voltage channels are statistically redundant, only Stack Voltage 1 was retained as a voltage representative; the eight-channel input space was instead populated with electrochemical and thermal context channels that span both noise regimes and provide complementary information for denoising.

## 4. METHODOLOGY

### 4.1. Pseudo-Ground-Truth Construction

Because an independent noise-free reference is unavailable for operational stack data, we construct a pseudo-ground-truth trend for each target channel using a 101-sample, 303-second centered median filter. The median filter was selected because it is robust to isolated spikes and preserves step-like changes better than a centered mean. [12]. This pseudo-ground truth should be interpreted as an operational definition of the slow trend rather than as a direct measurement of the true physical state. The 303-second window is substantially longer than the observed residual autocorrelation times and therefore suppresses high-frequency measurement variability, but it may also attenuate genuine dynamics occurring on shorter timescales. To reduce the risk of favoring overly aggressive smoothing, all filters are evaluated under an explicit trend-fidelity constraint.

### 4.2. Classical Filter Benchmarks

Three classical filters were evaluated as baselines: simple moving average (SMA), Savitzky–Golay polynomial filter of order 3 [4], and 4th-order Butterworth low-pass filter. Each was swept over a parameter grid: SMA and Savitzky–Golay over odd window lengths from 5 to 61 samples (15 to 183 s); Butterworth over cutoff periods from 15 to 180 s. The performance metric was the noise reduction ratio,

$$NRR = \frac{\sigma(y_{denoised} - y_{ground\ truth})}{\sigma(y_{raw} - y_{ground\ truth})} \quad (2)$$

subject to Pearson  $r(\text{denoised}, \text{ground truth}) \geq 0.95$  to prevent excessive distortion. Lower NRR is better;  $NRR = 1$  corresponds to no noise reduction. Across all four target signals, SMA with a 61-sample (183-s) window consistently gave the best (lowest) NRR under the fidelity constraint, outperforming Savitzky–Golay and Butterworth in Table 2. This is consistent with the red/brown noise regime identified in Section 3: when noise power is concentrated at low frequencies that overlap the signal’s slow dynamics, simple wide-window averaging is more effective than polynomial fitting or sharp-cutoff frequency-domain filtering. The NRR–distortion Pareto frontiers of the three filters were essentially superimposed, indicating that the bottleneck is not filter family but the spectral overlap itself.

**Table 2. Best classical-filter NRR per channel under  $r \geq 0.95$**

Channel	Raw $\sigma$	SMA NRR	SavGol NRR	Butter NRR	Best
Stack Voltage 1	0.1069	0.1554	0.2374	0.2234	SMA
FCE Stack TC#1	0.0220	0.1921	0.3168	0.3183	SMA
Fuel In	0.0290	0.1592	0.2498	0.2447	SMA
Furnace Inlet	0.0911	0.2643	0.4690	0.4660	SMA

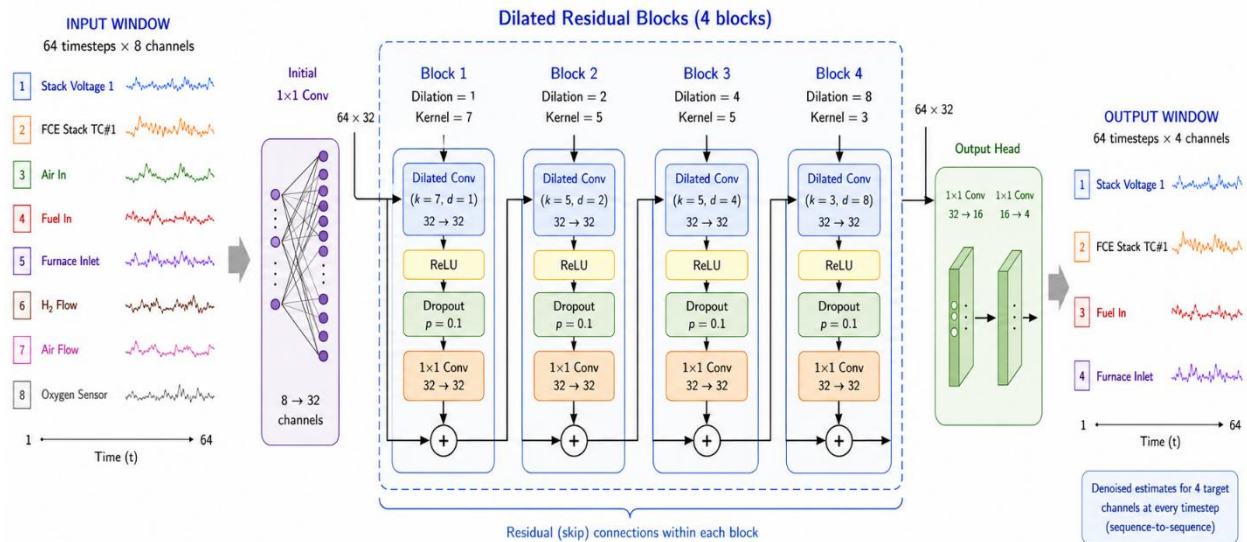
### 4.3. Dilated CNN Denoiser

#### 4.3.1. Architecture

The proposed denoiser is a fully convolutional network with dilated causal convolutions and residual skip connections, inspired by WaveNet [5] and the temporal convolutional network [6]. Each input window contains 64 timesteps across 8 input channels (Stack Voltage 1, FCE Stack TC#1, Air In, Fuel In, Furnace Inlet, H<sub>2</sub> Flow, Air Flow, and Oxygen Sensor). The network outputs denoised estimates for 4 target channels (Stack Voltage 1, FCE Stack TC#1, Fuel In, Furnace Inlet) at every timestep in the window — a sequence-to-sequence design.

An initial  $1 \times 1$  convolution projects the 8-channel input to a 32-dimensional feature space. Four residual blocks follow, with dilation rates  $\{1, 2, 4, 8\}$  and kernel sizes  $\{7, 5, 5, 3\}$ . Each block applies a dilated convolution with ReLU activation, dropout ( $p = 0.1$ ), and a  $1 \times 1$  pointwise convolution, then adds the input via skip connection. The dilation stack yields an effective receptive field of approximately 48 timesteps ( $\approx 144$  s), comfortably larger than the longest observed noise correlation time. The output head consists of a  $1 \times 1$  convolution to 16 features followed by a final  $1 \times 1$  convolution to the 4 output channels. The model has 25,716 trainable parameters in total. The full forward pass is summarized in Figure 1.

Figure 1. Architecture of the dilated CNN denoiser



To make the model invariant to absolute signal level — which differs systematically between files due to slow degradation and operating-condition shifts — each input window is independently zero-centered by subtracting the per-channel mean computed over its 64 timesteps. The window means are stored at inference and added back to recover the original scale. This step proved essential for cross-file generalization: an earlier LSTM-based variant without per-window normalization learned file-specific mean voltage levels and produced a constant offset on the held-out file.

Successive input windows are extracted with a stride of 4 samples, so each timestep is covered by 16 overlapping windows that view it from different positions (early, middle, or late within their 64-step receptive field). The final denoised estimate is the simple arithmetic mean of these 16 predictions. This overlap averaging is fundamentally different from a uniform-weight moving average: each of the 16 contributing predictions is itself a non-linear, learned function of a different temporal context, and the average therefore combines diverse non-linear smoothing rather than rectangular-window arithmetic. The

16-fold overlap also acts as a Monte-Carlo-style variance reducer on the model’s point predictions, yielding visibly smoother trajectories than the network’s single-window output.

### 4.3.2. Training Configuration

The model was trained to minimize the Huber loss [13] with  $\delta = 0.5$  between predicted and pseudo-ground-truth sequences. Huber was preferred over MSE because its quadratic-to-linear transition penalizes typical residuals strongly while remaining robust to occasional spikes — a desirable property when the target contains the residual noise of the median pre-filter at window boundaries. Optimization used Adam [14] with an initial learning rate of  $10^{-3}$ , a batch size of 512, ReduceLRonPlateau (factor 0.5, patience 5), and a 50-epoch budget. Training data came from two Group B files (SEP 6 and SEP 29), with a time-block 15% validation split inside each file. The third file (SEP 16) was held out entirely and used only for the final test.

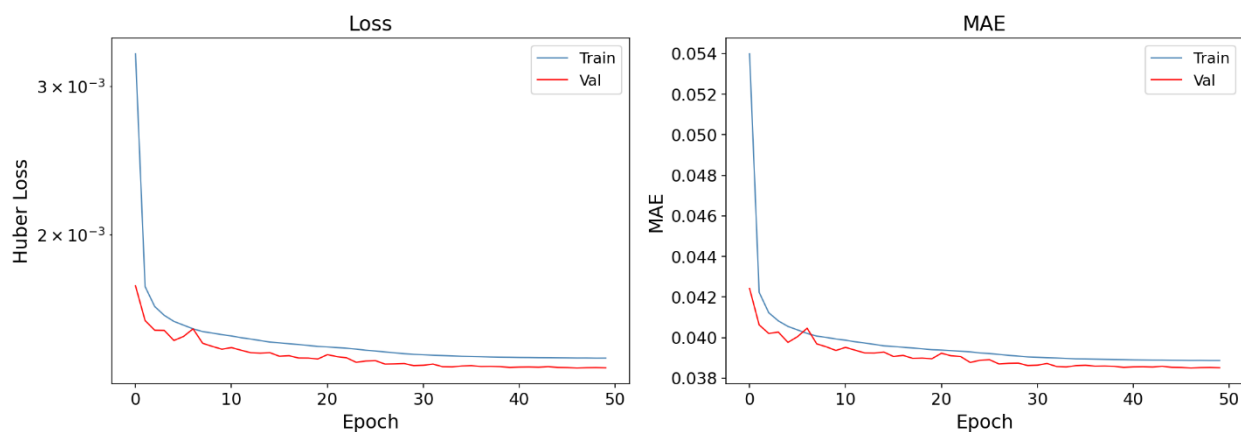
Two design decisions support honest evaluation under this protocol. First, the validation split is taken as a contiguous block at the end of each training file rather than randomly across windows; random splits would leak information across the 64-step receptive field and inflate validation performance. Second, per-window normalization is applied identically to training, validation, and test windows; no test-set statistics are used at any stage.

## 5. RESULTS AND DISCUSSION

### 5.1. Training Performance

Training loss and MAE curves are shown in Figure 2. The model trained for the full 50-epoch budget; ReduceLRonPlateau triggered six times, taking the learning rate from  $10^{-3}$  down to  $2 \times 10^{-6}$  by epoch 49. Best validation performance ( $\text{val\_mae} = 0.0385$ ,  $\text{val\_loss} = 1.4 \times 10^{-3}$ ) was reached at epoch 47, and the EarlyStopping callback restored those weights at termination. The validation curves track below the training curves throughout — the expected pattern when dropout is active during training but disabled during validation.

**Figure 2. Training and validation curves of the dilated CNN denoiser on Group B**



### 5.2. Held-Out Test Performance

The trained model was applied to SEP 16, a Group B file that was withheld from both training and validation, ensuring that all performance metrics reflect generalization rather than memorization of training data patterns.

Table 3 reports the NRR, the percentage improvement over the best classical filter (SMA), and the Pearson correlation of the CNN output with the pseudo-ground-truth for all four output channels. The CNN reduces

NRR by 18.1–39.8% relative to the SMA baseline across all channels, while maintaining correlation with the slow degradation trend.

**Table 3. Held-out test denoising performance**

Channel	Raw $\sigma$	SMA NRR	CNN NRR	Improvement	CNN Corr.
Stack Voltage 1	0.1069	0.1554	0.1026	+34.0%	0.9979
FCE Stack TC#1	0.0220	0.1926	0.1577	+18.1%	1.0000
Fuel In	0.0290	0.1592	0.1163	+26.9%	0.9998
Furnace Inlet	0.0911	0.2643	0.1590	+39.8%	0.9998

The performance differences across channels follow the noise characteristics identified earlier. Stack Voltage 1 and Furnace Inlet show the largest improvements (34.0% and 39.8%). Both signals contain strongly correlated, structured noise (PSD slopes  $\approx -1.3$ ), which the CNN can learn and remove more effectively than a simple moving average. In contrast, SMA treats this structure as part of the signal and cannot separate it cleanly.

Fuel In shows a moderate improvement (26.9%). Its noise is closer to pink (PSD slope  $-0.50$ ), which still has structure but is less pronounced. As a result, CNN provides benefit, but the gain is smaller.

FCE Stack TC#1 shows the smallest improvement (18.1%). This channel already has low noise and high SNR (31.0 dB), and its noise is closer to white (PSD slope  $-0.42$ , short correlation time). In this case, there is less structure to exploit, so both SMA and CNN perform similarly. CNN still reduces residual noise slightly while preserving the signal.

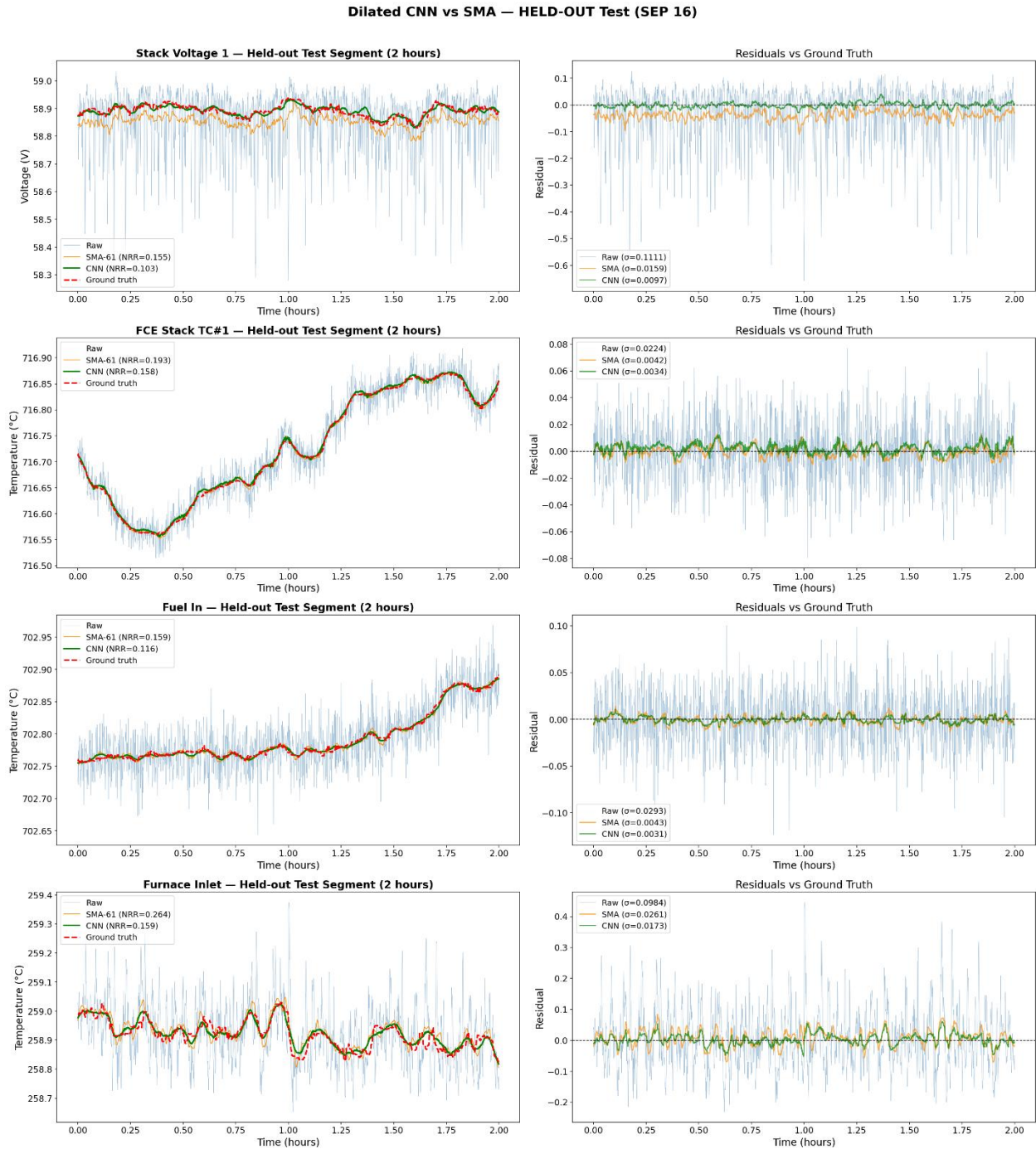
Figure 3 shows the time-domain output for a representative two-hour segment of the held-out SEP 16 experiment. For each channel, the left panel overlays raw data (blue), SMA (orange), CNN (green), and pseudo-ground-truth (red dashed); the right panel plots residuals relative to the ground truth, with residual standard deviations annotated in the legend.

Three qualitative observations emerge from the time-domain visualization that are not captured by the aggregate NRR metric alone. First, CNN residuals are centered tightly on zero with no apparent low-frequency bias across the full two-hour window. This confirms that the per-window normalization strategy successfully eliminated the systematic offset that defeated the initial LSTM model. The zero-mean residual property is essential for degradation rate estimation, because any persistent bias in the denoised signal would directly corrupt slope estimates computed over multi-hour windows.

Second, CNN demonstrates superior handling of transient noise excursions compared to SMA. On Stack Voltage 1, the raw signal contains intermittent downward spikes of 0.3–0.5 V below the trend, likely caused by momentary disturbances in the power electronics. The SMA attenuates these spikes but spreads their effect over its 183-second window, creating a prolonged dip in the filtered output. The CNN, by contrast, suppresses these spikes more sharply, returning to the trend within seconds rather than minutes. This difference is most visible in the Furnace Inlet channel, where the CNN residual standard deviation (0.0173°C) is 34% lower than SMA (0.0261°C), and spike-like excursions that the SMA only partially attenuates are effectively removed by the CNN.

Third, on FCE Stack TC#1, the two-hour segment reveals a genuine thermal drift of approximately 0.3°C. Both the SMA and CNN track this drift faithfully — their filtered outputs overlap closely with the ground truth throughout the transition. The CNN residual  $\sigma$  of 0.0034°C is modestly lower than SMA's 0.0042°C, consistent with the 18.1% NRR improvement in Table 3. The fact that both methods preserve this slow drift validates the ground truth construction: the 303-second median filter captures real thermal dynamics at the multi-minute timescale while rejecting higher-frequency noise.

**Figure 3. Held-out test segment performance (SEP 16, two hours).**



### 5.3. Real-Time Deployment Characteristics

A critical requirement for operational monitoring is that the denoising model must process incoming data at least as fast as it is acquired. To evaluate real-time viability, we benchmarked inference latency, memory footprint, and streaming processing performance, and compared the CNN against the causal (real-time capable) variant of the SMA baseline.

Single-window inference on a standard GPU required a mean of 113.5 ms (P99 = 160.7 ms), well within the 12,000 ms budget defined by the stride interval (4 samples  $\times$  3 seconds). This provides 96.2% computational headroom, meaning the model could sustain real-time operation even under significant system load. Batch processing further improves efficiency: at batch size 256, throughput reaches 1,633 windows per second with a per-window latency of 0.61 ms, corresponding to a real-time factor of approximately 19,600 $\times$ . Table 4 summarizes the inference performance across batch sizes.

**Table 4. Inference Latency and Throughput at Different Batch Sizes**

Batch Size	Total (ms)	Per Window (ms)	Throughput (win/s)	Real-Time Factor
1	134.8	134.8	7.4	89 $\times$
4	128.5	32.1	31.1	374 $\times$
16	108.7	6.8	147.2	1,766 $\times$
64	111.9	1.7	571.9	6,863 $\times$
256	156.8	0.6	1,633.1	19,598 $\times$
1024	640.2	0.6	1,599.5	19,194 $\times$

To validate real-time viability under realistic conditions, a streaming simulation processed 16.7 hours of held-out test data (20,000 samples) sequentially, issuing one window prediction at each stride interval. The simulation processed all 4,984 windows in 678.5 seconds of wall-clock time, representing a processing ratio of 0.011 $\times$  — meaning the model runs 88 $\times$  faster than real-time even in single-window (non-batched) mode. The overlap count builds from 1 to the maximum of 16 over the initial 252 seconds (84 samples), after which the system operates at full denoising capacity.

CNN requires an initial buildup period of 252 seconds before reaching full 16 $\times$  overlap averaging. During this transient, denoising quality is reduced but still operational. For continuous monitoring systems running over hundreds of hours, this one-time startup latency is negligible. In practice, the system can be initialized during the stack’s thermal ramp-up phase, ensuring full denoising capacity is available when the stack enters the on-load operating regime where degradation monitoring is most critical.

## 6. CONCLUSION

The 18–40% reduction in noise reduction ratio over the best classical filter — achieved on a fully held-out file with no parameter re-tuning — demonstrates that structured noise in SOEC stacks is learnable and that the learned representation generalizes across operating segments. The results suggest that the CNN exploits repeatable correlated residual structure that is not captured by fixed linear smoothers, allowing it to more closely track the pseudo-ground-truth slow trend while reducing residual variability.

The framework is designed for real-time deployment: the 25,716-parameter dilated CNN processes each window in 113 ms (96% computational headroom at the 3-second sampling rate), requires only 0.12 MB of runtime memory, and runs 88 $\times$  faster than real-time even in single-window mode.

Three lessons generalize beyond this dataset. First, per-window normalization is essential when training across files with different operating points; without it, the model risks memorizing file-level means rather than noise structure. Second, sequence-to-sequence output with stride-based overlap averaging is a simple but effective regularizer that exploits temporal redundancy at zero parameter cost. Third, demonstrating data-driven denoising on a held-out file from the same operating regime is a stronger and more honest evaluation than the random-window splits commonly reported in the literature.

Future work will (i) extend the framework to Groups A and C, which carry stronger long-term degradation and current-regime variation; (ii) quantify the impact of CNN denoising on degradation-rate estimator variance and bias under explicit synthetic-trend injection; (iii) explore uncertainty-aware variants (e.g., MC-

dropout) to deliver per-sample confidence intervals; and (iv) integrate the model into the live data acquisition pipeline at INL for online monitoring during the next long-duration test campaign.

### Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Nuclear Energy under DOE Idaho Operations Office Contract DE-AC07-05ID14517. Accordingly, the U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript or allow others to do so, for U.S. Government purposes.

### References

1. Singhal SC, Kendall K (2003) High-temperature Solid Oxide Fuel Cells: Fundamentals, Design and Applications. Elsevier
2. Brisse A, Schefold J, Zahid M (2008) High temperature water electrolysis in solid oxide cells. *Int J Hydrog Energy* 33:5375–5382. <https://doi.org/10.1016/j.ijhydene.2008.07.120>
3. Ebbesen SD, Jensen SH, Hauch A, Mogensen MB (2014) High Temperature Electrolysis in Alkaline Cells, Solid Proton Conducting Cells, and Solid Oxide Cells. *Chem Rev* 114:10697–10734. <https://doi.org/10.1021/cr5000865>
4. Savitzky Abraham, Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* 36:1627–1639. <https://doi.org/10.1021/ac60214a047>
5. Oord A van den, Dieleman S, Zen H, et al (2016) WaveNet: A Generative Model for Raw Audio
6. Bai S, Kolter JZ, Koltun V (2018) An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling
7. Liu L, Ouyang W, Wang X, et al (2019) Deep Learning for Generic Object Detection: A Survey
8. Chen Z, Li C, Sanchez R-V (2015) Gearbox Fault Identification and Classification with Convolutional Neural Networks. *Shock Vib* 2015:390134. <https://doi.org/10.1155/2015/390134>
9. Welch P (1967) The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoustics* 15:70–73. <https://doi.org/10.1109/TAU.1967.1161901>
10. Johnson JB (1925) The Schottky Effect in Low Frequency Circuits. *Phys Rev* 26:71–85. <https://doi.org/10.1103/PhysRev.26.71>
11. Dutta P, Horn PM (1981) Low-frequency fluctuations in solids: 1/f noise. *Rev Mod Phys* 53:497–516. <https://doi.org/10.1103/RevModPhys.53.497>
12. Huang T, Yang G, Tang G (1979) A fast two-dimensional median filtering algorithm. *IEEE Trans Acoust Speech Signal Process* 27:13–18. <https://doi.org/10.1109/TASSP.1979.1163188>
13. Huber PJ (1992) Robust Estimation of a Location Parameter. In: Kotz S, Johnson NL (eds) *Breakthroughs in Statistics*. Springer New York, New York, NY, pp 492–518
14. Kingma DP, Ba J (2017) Adam: A Method for Stochastic Optimization