

When Does a Kalman Filter Beat an LSTM?

**Arthur Debaugé¹, Noria Foukia¹
and Clément Denis²**

¹ HESSO - HEPIA - ISC - 4, rue de la Prairie, 1202, Geneva, Switzerland

² HESSO - HEPIA - GC - 4, rue de la Prairie, 1202, Geneva, Switzerland

E-mail: arthur.debaugé@hesge.ch, noria.foukia@hesge.ch, clement.denis@hesge.ch

Summary: We benchmark Cubic Splines, Kalman Filter, LSTM, and Kalman + LSTM on the UCI Household Power Consumption dataset. The Kalman filter achieves $R^2 = 0.957$, $\|E\|_F = 6.95$, outperforming all LSTM variants and a naive persistence baseline. Kalman pre-processing reduces LSTM's Frobenius error by 13.4% at no architectural cost. SHAP confirms the Kalman-filtered signal as the dominant predictor (54.9%), justifying the hybrid architecture. We characterise when this holds and provide a principled comparison framework using the Frobenius norm as optimisation criterion.

Keywords: Energy Forecasting, Kalman Filter, LSTM, Frobenius Norm, SHAP, Microgrid.

1. Introduction

Short-term energy load forecasting underpins P2P trading decisions in smart microgrids. The literature favours LSTM-based architectures [2, 7], yet their advantage over classical estimators is rarely quantified on real household data [6]. Our central finding is counter-intuitive: a well-tuned Kalman filter beats all LSTM variants on the Frobenius norm, because the underlying signal is quasi-Markovian ($\rho_{t,t+1} \approx 0.97$). We characterise when this holds and show that, when LSTM is warranted, Kalman pre-processing is an improvement at no architectural cost.

2. Dataset & Setup (Figure 1)

UCI Household Electric Power Consumption [1] (France, 2007–2010, 1-min $\rightarrow$ 2 h), enriched with ERA5 meteorological covariates. Regression target : *Global active power(t+1)* in kW. Features: 7 electrical + 4 meteorological + 4 cyclic time encodings. Temporal split: 70/15/15 (train/val/test), strictly causal.

3. Methods (Figure 2)

3.1. Persistence Baseline

Naïve baseline: $\hat{y}_{t+1} = y_t$. With $\rho_{t,t+1} = 0.972$ on the test set, this is a strong lower bound against which all methods are compared. Any model failing to outperform persistence provides no practical value.

3.2. Kalman Filter (order 2)

Constant-velocity state $x_k = [\text{pos.}, \text{vel.}]^T$:

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H \cdot \hat{x}_k^-), \quad K_k = \frac{P_k^- \cdot H^T}{H \cdot P_k^- \cdot H^T + R}$$

3.3. LSTM

Single-layer LSTM [2]: T=24 steps, Adam ($3 \cdot 10^{-3}$, batch 64), early stopping (patience 6).

3.4. Kalman + LSTM

Raw dataset replaced by its Kalman-filtered version.

No architectural change, no added parameters.

3.5. Frobenius Optimisation

Grid search minimising $\|E\|_F = \sqrt{\sum_i e_i^2}$ on validation. $H \in \{16, 32, 64\}$, $l_r \in \{1 \cdot 10^{-3}, 3 \cdot 10^{-3}, 7 \cdot 10^{-3}\}$. $\|E\|_F$ penalises large errors more heavily than MAE and measures a distance in the error space, making it a natural criterion for comparing models on a fixed test set.

4. Results & Discussion (Figure 2 & Table 1)

Table 1 and Figure 2 summarise all results.

Kalman beats LSTM. The Kalman predictor achieves $R^2 = 0.957$, $\|E\|_F = 6.95$ because $\rho_{t,t+1} \approx 0.97$: a quasi-Markovian signal has no systematic information beyond what a Bayesian recursive estimator already captures [3,5]. *Rule:* benchmark Kalman before LSTM whenever $\rho_{t,t+1} > 0.9$. LSTM adds value for multi-step horizons or heterogeneous feature sets. **Persistence baseline.** The naive persistence predictor yields $\|E\|_F = 32.35$, $R^2 = 0.062$, confirming that the high autocorrelation alone does not trivially solve the forecasting problem. The Kalman filter outperforms persistence by a factor of ≈ 4.6 in $\|E\|_F$, while the raw LSTM offers only a modest improvement ($\|E\|_F = 28.63$).

Kalman pre-processing improves LSTM. Replacing raw by filtered data reduces $\|E\|_F$: 28.0 $\rightarrow$ 24.2 (-13.4%), R^2 : 0.292 $\rightarrow$ 0.469 ($+60.5\%$), at no architectural cost [6].

SHAP. Filtered data holds 54.9% of feature importance [4]. SHAP top-8 ($\|E\|_F = 24.27$) equals Frobenius-optimised tuning ($\|E\|_F = 24.50$) with 47% fewer features.

Takeaway. Benchmark Kalman before LSTM ($\rho_{t,t+1} > 0.9$); Kalman pre-processing is architecturally free; tune with $\|E\|_F$; audit with SHAP.

3. Figures and Table

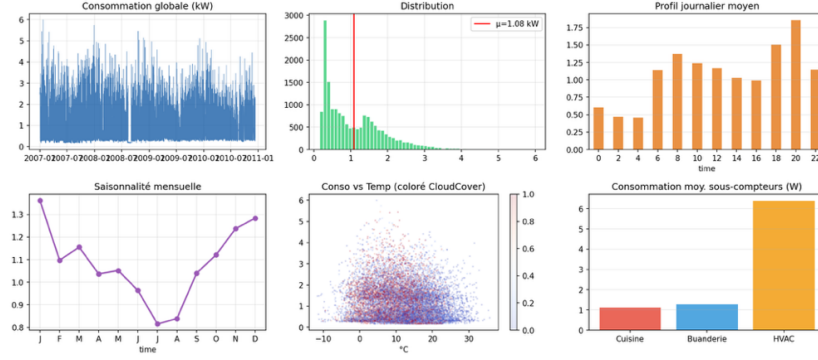


Fig. 1. Exploratory Data Analysis of the electrical consumption dataset. Highlights include the global time series (2007–2011), daily load profiles, monthly seasonality, and the distribution of sub-metering, with HVAC showing the highest energy demand.

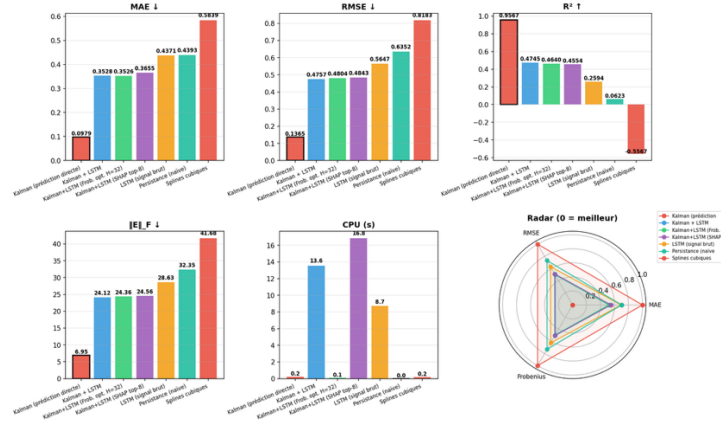


Fig. 2. Comparative benchmark of prediction models across multiple metrics (MAE, RMSE, R^2 , $\|E\|_F$, and CPU time). The Kalman filter (direct prediction) outperforms LSTM and Spline-based approaches in both accuracy ($R^2 = 0.9567$) and computational efficiency.

Method	MAE (kW)	RMSE (kW)	R^2	$\ E\ _F$	CPU (s)
Kalman filter (direct)	0.0979	0.1365	0.9567	6.952	0.2
Kalman + LSTM	0.3497	0.4780	0.4694	24.237	19.6
Kalman+LSTM (SHAP top-8)	0.3574	0.4787	0.4678	24.273	10.1
Kalman+LSTM (Frob. opt.)	0.3591	0.4833	0.4577	24.503	0.1 ^a
LSTM (raw signal)	0.4227	0.5521	0.2921	27.996	10.1
Persistence ($\hat{y}_{t+1}=y_t$) ^b	0.439	0.635	0.062	32.35	<0.1
Cubic Splines	0.5839	0.8183	-0.5567	41.684	0.2

^a Inference latency only (best model already selected by grid search); training cost of the 3×3 grid search is excluded from this value.

^b Added as lower-bound baseline; lag-1 autocorrelation on the test set: $\rho_{t,t+1}=0.972$.

Table 1: Test-set benchmark sorted by $\|E\|_F$ (ascending). **Bold** = best per metric. **Green** = dominant method. **Blue** = persistence baseline (added per reviewer request). $\|E\|_F = \sqrt{\sum_i e_i^2}$. CPU measured on a single core.

References

- [1]. G. Hébrail & A. Bérard, “Individual Household Electric Power Consumption Data Set,” UCI ML Repository, 2012.
- [2]. W. Kong et al., “Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network,” IEEE Trans. Smart Grid, 10(1):841–851, 2019. doi:10.1109/TSG.2017.2753802
- [3]. R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” Trans. ASME, J. Basic Eng., 82(1):35–45, 1960.
- [4]. S. M. Lundberg & S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” NeurIPS, 30:4765–4774, 2017.
- [5]. S. Sharma, A. Majumdar, V. Elvira & E. Chouzenoux, “Blind Kalman Filtering for Short-term Load Forecasting,” IEEE Trans. Power Syst., 35(6):4916–4919, 2020. doi:10.1109/TPWRS.2020.3018623
- [6]. T. Y. Kim & S. B. Cho, “Predicting Residential Energy Consumption using CNN-LSTM Neural Networks,” Energy, 182:72–81, 2019. doi:10.1016/j.energy.2019.06.009
- [7]. Y. Wang, N. Zhang & X. Chen, “A Short-Term Residential Load Forecasting Model Based on LSTM Recurrent Neural Network Considering Weather Features,” Energies, 14(10):2737, 2021. doi:10.3390/en14102737