The National Institutes of Health | Division of Technical Resources | Office of Research Facilities

Design Requirements Manual **News to Use**

April 2026
Vol. 03, No. 4

The formulae $\frac{\partial \rho U_i}{\partial t}+\frac{\partial}{\partial x_j}(\rho U U_j)=-\frac{\partial P}{\partial x_i}+\frac{\partial}{\partial x_j}\left(\mu\frac{\partial U_i}{\partial x_j}\right)+g_i(\rho-\rho_0)$ for building $\frac{\partial}{\partial x_j}(\rho\overline{U_i}\overline{U_j})=-\frac{\partial P}{\partial x_i}+\frac{\partial}{\partial x_j}\left(\mu\frac{\partial \overline{U_i}}{\partial x_j}-\rho\overline{u_i'u_j'}\right)+g_i(\rho-\rho_0)$ state of the art $\frac{\partial}{\partial x_i}(\rho\overline{U_i}\overline{H})=\frac{\partial}{\partial x_i}\left(\lambda\frac{\partial \overline{T}}{\partial x_i}-\rho\overline{u_i'h'}\right)$ biomedical research facilities.

Improving Operational Data Quality: Wavelet-GARCH Signal Processing Pipeline

Raw sensor data usually contains data-quality issues such as noise and anomalies that make it ill-suited for direct use in business intelligence and machine learning applications. To address this, the medallion architecture is used to organize data into a series of data layers – Bronze, Silver, and Gold – that represent progressively higher levels of data quality within a data lakehouse. The Bronze layer contains raw ingested data, while the Gold layer consists of dashboard and front-end ready datasets. Between these two lies the Silver layer, where data has been cleaned, validated, and transformed through advanced analytical processes to produce high-quality datasets suitable for analysis. This article explores the Wavelet-GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) pipeline, a Silver layer sensor-cleaning methodology that is more effective than traditional Fourier-based methodologies.

Discrete Wavelet Transform (DWT) Denoising

Noise is unwanted signal that interferes with the measurement of sensor signals. Sensor noise z_i is mostly composed of high-frequency signals that are layered on top of the true sensor value $f(t)$ over time:[1]

$$z(t) = s(t) + n(t) \qquad \text{where: } n(t) \sim \mathcal{N}(0, \sigma^2)$$

Although it is tempting to think that a Fourier transform low-pass filter will remove high-frequency signals from the data and effectively eliminate the noise, the true sensor value contains high-frequency signals for describing sharp structures (e.g., edges or spikes), which would be lost if we universally removed high-frequency components from the raw data the way a Fourier-based low-pass filter would. Thus, an effective denoising strategy requires localized and adaptive methods, like the discrete wavelet transform (DWT). DWT decomposes a signal into separate wave-like pieces that highlight short-term details and long-term trends; it uses wavelets $\psi(t)$, which have their energy concentrated in both time and scale, enabling the analysis of transient, nonstationary, and time-varying signal phenomena. There are several types of mother wavelets ψ to choose from based on their properties and applications (e.g., Haar wavelet, Daubechies). The mother wavelet is scaled and translated (controlling its width and position, respectively) to generate a family of wavelets $\psi_{j,k}(t)$, which are used to project the signal and obtain coefficients at different scales and locations.

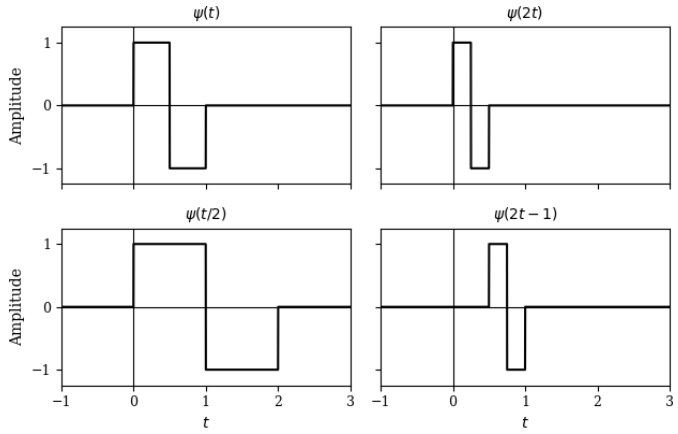


FIGURE 1. EXAMPLE OF A SCALED AND TRANSLATED HAAR MOTHER WAVELET[2]

DWT is performed in three steps:

Step 1: Forward DWT – The first step applies an interval-adapted pyramidal filtering algorithm consisting of successive low-pass and high-pass filtering followed by downsampling. This process decomposes the raw sensor signal into *approximation (*or *scaling) coefficients* (low-frequency components) and *detail (wavelet) coefficients* (high-frequency components) across multiple scales (levels), which can be interpreted as inner products (a measure of how strongly the signal matches scaled and shifted wavelets).[3]

At each decomposition level j, the DWT is implemented using a pair of filters consisting of a low-pass filter $h[n]$ and a high-pass filter $g[n]$:

[1] Donoho
[2] Najmi pg. 112
[3] Najmi pg. 222

'Design Requirements Manual (DRM) News to Use' is a monthly ORF publication featuring salient technical information that should be applied to the design of NIH biomedical research laboratories and animal facilities. NIH Project Officers, A/E's and other consultants to the NIH, who develop intramural, extramural and American Recovery and Reinvestment Act (ARRA) projects will benefit from 'News to Use'. **Please address questions or comments to:** megan.teague@nih.gov

The National Institutes of Health | Division of Technical Resources | Office of Research Facilities

Design Requirements Manual **News to Use**

April 2026
Vol. 03, No. 4

The formulae $\frac{\partial \rho U_i}{\partial t}+\frac{\partial}{\partial x_j}(\rho U U_j)=-\frac{\partial P}{\partial x_i}+\frac{\partial}{\partial x_j}\left(\mu\frac{\partial U_i}{\partial x_j}\right)+g_i(\rho-\rho_0)$ for building $\frac{\partial}{\partial x_j}(\rho\overline{U_i}\overline{U_j})=-\frac{\partial P}{\partial x_i}+\frac{\partial}{\partial x_j}\left(\mu\frac{\partial \overline{U_i}}{\partial x_j}-\rho\overline{u_i u_j}\right)+g_i(\rho-\rho_0)$ state of the art $\frac{\partial}{\partial x_i}(\rho\overline{U_i}\overline{H})=\frac{\partial}{\partial x_i}\left(\lambda\frac{\partial \overline{T}}{\partial x_i}-\rho\overline{u_i h}\right)$ biomedical research facilities.

$$cA_j[n] = c_j[n] = \sum_k c_{j-1}[k]h[2n-k]$$

$$cD_j[n] = d_j[n] = \sum_k c_{j-1}[k]g[2n-k]$$

Where:

$$c_0[n] = x[n] \quad the\ sampled\ version\ of\ the\ raw\ signal\ z(t)$$

The approximation coefficients c_j are further decomposed at the next level, forming a multi-resolution pyramid structure:

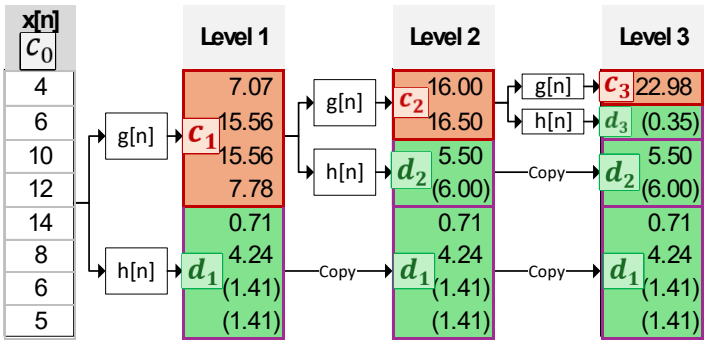


FIGURE 2. THREE-LEVEL FORWARD DWT DECOMPOSITION USING THE HAAR WAVELET

Each level captures progressively lower-frequency components of the signal (in red), while the detail coefficients capture localized high-frequency features (in green).

Step 2: Thresholding – Since important features produce large coefficients and coefficients related to noise become small, suppressing lower-scale coefficients produces a less noisy signal. This suppression is accomplished by applying either a hard thresholding (where coefficients below a threshold are set to zero):

$$\hat{d}_{j,k} = d_{j,k} \cdot I(|d_{j,k}| > \lambda)$$

or soft thresholding (where coefficients are reduced by a quantity equal to the threshold value):

$$\hat{d}_{j,k} = sgn(d_{j,k}) \cdot max(|d_{j,k}| - \lambda, 0)$$

Step 3: Inverse DWT – The signal is reconstructed from the coefficients using an inversion of the original pyramid filtering:[4]

$$\tilde{x}[n] = \sum_{k=-\infty}^{\infty} c_L[k]\varphi_{L,k}[n] + \sum_{j=1}^{L}\sum_{k=-\infty}^{\infty} \hat{d}_j[k]\psi_{j,k}[n]$$

Where $j = 1, \dots, L$ indexes the decomposition levels and L is the final scale level.

Since the thresholding in Step 2 suppressed the noise coefficients, the resulting reconstructed signal is denoised yet preserves transient events.

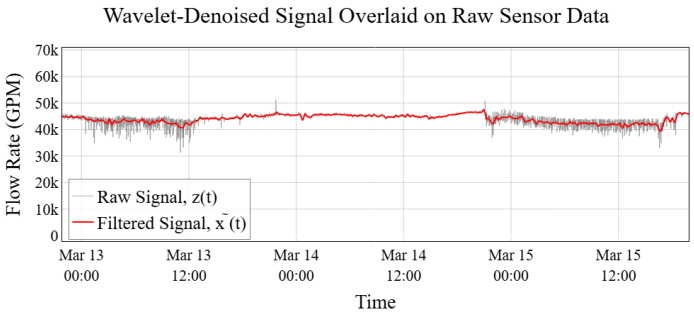


FIGURE 3. WAVELET DENOISED SIGNAL OVERLAID ON RAW SIGNAL

GARCH Model (Volatility Tracking)

After denoising, the signal may still contain anomalous periods that should be flagged as unreliable or non-representative operations conditions (e.g., power spikes or transient disturbances). While the definition of such anomalies may vary by application, they are commonly characterized by continuous periods of clustered volatility. A key property of these signals is heteroskedasticity, where the variance is not constant over time but instead evolves dynamically.

To flag these anomalous periods, the *GARCH* model can be applied to the denoised wavelet signal. The GARCH model captures this clustered volatility through the recursive relationship:

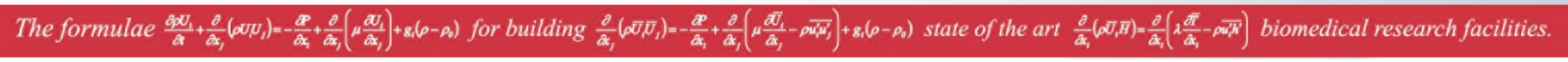
$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

[4] Najmi pg. 140

The National Institutes of Health | Division of Technical Resources | Office of Research Facilities

Design Requirements Manual **News to Use**

April 2026
Vol. 03, No. 4



where ω is the long-run variance constant, α captures the impact of prior shocks, and β captures volatility persistence. The usual parameter restrictions

$$\omega > 0,\ \ \alpha \geq 0,\ \ \beta \geq 0,\ \ \alpha + \beta < 1$$

ensure that the conditional variance remains positive and that the process is covariance stationary (meaning its average behavior stays stable over time). Unlike models that assume constant variance, GARCH explicitly models the conditional variance of the error term (volatility) changes over time and is dependent on past variance and error. This enables it to capture volatility clustering, where periods of high volatility tend to be followed by more high volatility and, conversely, periods of low volatility by more low volatility. Thus, the GARCH value spikes, then slowly decays after a volatility spike. The rate and persistence of this decay depend on user-defined hyperparameters $p\ and\ q$, which control the influence of past variances and past shocks, respectively.

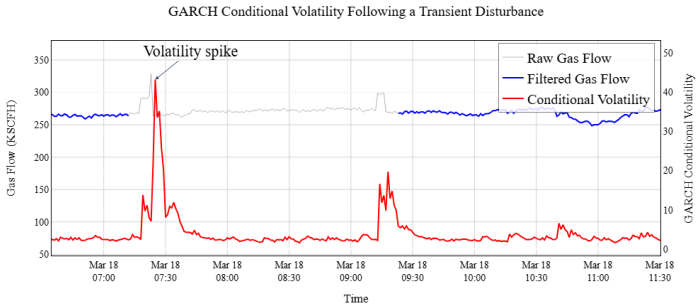


FIGURE 4. GARCH CONDITIONAL VOLATILITY SERIES ILLUSTRATING VOLATILITY SPIKES AND SUBSEQUENT DECAY

An adaptive threshold on the GARCH value is set up to flag and scrub the anomalous periods, before finally publishing the data - free of noise and anomalies - to the Silver table.

Conclusion

The Wavelet-GARCH pipeline provides an effective methodology for transforming raw sensor data into high-quality analytical datasets suitable for the Silver layer of medallion architecture. By first applying DWT, noise is removed from the signal in a localized and adaptive manner while preserving important transient features. The subsequent GARCH model enables the identification of anomalous periods through its ability to capture clustered volatility. Together, these steps produce a clean and reliable signal. Compared to traditional filtering techniques (i.e., Fourier low-pass filters), this pipeline offers greater adaptability to the data source and can be applied to diverse data applications. This approach has been applied to the NIH Bethesda Campus Central Utility Plant (CUP) data streams, where identifying true demand peaks can influence operational strategies and cost optimization on the Cogen gas purchasing model.

References

1. Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, *41*(3), 613–627.
2. Najmi, A.-H. (2012). *Wavelets*. JHU Press.

Further Reading

1. Barclay, V. J., Bonner, R. F., & Hamilton, I. P. (1997). Application of Wavelet Transforms to Experimental Spectra: Smoothing, Denoising, and Data Set Compression. *Analytical Chemistry*, *69*(1), 78–90.
2. Engle, R. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives*, *15*(4), 157–168.

Nomenclature

$z(t)$: $observed/raw\ sensor\ signal$
$s(t)$: $true\ underlying\ sensor\ signal$
$n(t)$: $noise\ component$
$\mathcal{N}(0,\sigma^2)$: $normal\ (Gaussian) distribution\ with\ mean\ 0\ and\ variance\ \sigma^2$
φ: $Scaling\ function$
ψ: $Mother\ wavelet\ function$
$\psi_{j,k}[n]$: $wavelet\ scaled\ by\ j, translation\ by\ k$
$cD_j[n]$: $detail(wavelet)\ coefficient$
$cD_j[n]$: $scaling\ coefficient$
$\hat{d}_{j,k}$: $thresholded\ detail(wavelet)\ coefficient$
$\tilde{x}[n]$: $reconstruced/denoised\ signal$
λ: $wavelet\ coefficient\ threshold$
J: $final\ decomposition\ scale\ level$
σ_t^2: $Conditional\ variance\ in\ GARH\ at\ time\ t$
ω: $GARCH\ baseline\ variance$
α_i: $GARCH\ impact\ of\ past\ shocks$
β_j: $GARCH\ persistence\ of\ volatility$
ϵ_t: $Residuals\ in\ the\ GARCH\ model\ at\ time\ t$
p: $Order\ of\ GARCH\ terms\ in\ the\ model$
q: $Order\ of\ ARCH\ terms\ in\ the\ GARCH\ model$

'Design Requirements Manual (DRM) News to Use' is a monthly ORF publication featuring salient technical information that should be applied to the design of NIH biomedical research laboratories and animal facilities. NIH Project Officers, A/E's and other consultants to the NIH, who develop intramural, extramural and American Recovery and Reinvestment Act (ARRA) projects will benefit from 'News to Use'. **Please address questions or comments to:** megan.teague@nih.gov