MADEV/ADEV/HDEV Report Explanation

This report analyzes a long voltage log from a Keithley 2001 multimeter.
The goal is to estimate the best voltage value, quantify noise, check drift, detect environmental cycles, evaluate quantization, and decide how trustworthy the averaged result is.

Program Start

`Loading C:\Python\7V_long_full.txt`

The script is opening the input file.

`Loaded 279264 samples (0 lines skipped)`

The parser found 279,264 valid voltage readings.

`0 lines skipped` means every line in the file matched the expected format.

`tau0 = 0.506221 s`

This is the estimated sampling interval. It means readings were taken about every 0.506 seconds.

`median timestamp interval`

The script used the median spacing between timestamps, not the mean.
This is more robust if there are a few bad or delayed timestamps.

`timing jitter ~0.06%`

This measures how much the sample spacing varies. 0.06% is very small, so the data can be treated as nearly evenly spaced.

Variance Computation

`Computing MADEV`

The script calculates Modified Allan Deviation. MADEV is useful for identifying noise type and separating phase/frequency noise behavior.

`N=279264`

Number of usable samples.

`max 60 points`

The script does not calculate every possible averaging time.
It selects up to 60 logarithmically spaced tau values to keep the plot readable and computation efficient.

`55 tau values [0.5062s ... 3.882e+04s]`

The MADEV curve has 55 averaging times.

The smallest tau is one sample interval.
The largest tau is about 38,820 seconds, or about 10.8 hours.

`Computing ADEV`

The script calculates regular Allan Deviation. ADEV is the traditional stability metric, but it is less able than MADEV to distinguish some noise types.

`Computing HDEV`

The script calculates Hadamard Deviation. HDEV is useful because it is less sensitive to linear drift.

`Computing statistics`

The script now calculates ordinary statistics, robust statistics, drift, bootstrap confidence intervals, quantization, and environmental-cycle diagnostics.

Main Report Header

`MADEV/ADEV/HDEV Statistical Report -- 7V_long_full.txt`

This identifies the source file and the type of analysis.

`N=279,264`

Total number of cleaned samples used in the report.

`tau0=0.5062 s`

Nominal sample interval.

`Duration=1.414e+05 s`

Total measurement duration in seconds.

`2356.13 min / 39.2688 h`

Same duration converted to minutes and hours. For newer versions, this may also be printed as days / hours / minutes.

Sanitization

This section explains what the script cleaned before doing statistics.

`Spikes flagged : 1083`

The script detected 1,083 suspicious isolated spikes. These are points that differ strongly from their local neighborhood.

`Spikes replaced : 1083`

All flagged spikes were replaced. Usually replacement is done using a local median so the time series remains continuous.

`Non-finite removed : 0`

No `NaN`, `inf`, or invalid numeric values were removed.

`Duplicate times rm : 0`

No duplicate timestamp rows were removed.

`Timing jitter (MAD) : 0.06% of tau0`

The median absolute deviation of sample spacing is only 0.06% of the nominal interval. This is excellent timing regularity.

`Note : replaced 1083 isolated spikes using local median`

This describes the cleaning method.

`window=9`

The local median window used 9 samples.

`nsigma=6.0`

A point had to deviate strongly from the local median before it was considered a spike.

`diff_nsigma=8.0`

The script also checked sudden jumps between neighboring samples.

Most Precise Value

This section gives the best estimate of the measured voltage.

`Best estimate : 7.05652096 V`

This is the grand mean of the cleaned voltage samples. It is the main reported voltage value.

`Uncertainty u(V) : 34.414 nV (k=1, MADEV-derived)`

This is a one-sigma uncertainty estimate derived from MADEV. `k=1` means one standard uncertainty, not an expanded 95% uncertainty.

> **Important:** This is a stability-derived uncertainty, not necessarily the total calibration uncertainty of the measurement system.

`Relative precision : 0.004877 ppm`

This expresses the uncertainty relative to the 7 V signal. 0.004877 ppm is about 4.9 parts per billion.

`95% CI (t-dist)`

A conventional 95% confidence interval based on Student's t distribution. This assumes the ordinary standard error model. For strongly correlated data, the corrected SEM or bootstrap CI is usually more realistic.

`99% CI (t-dist)`

Same idea as the 95% confidence interval, but wider and more conservative.

Basic Statistics

This section describes the cleaned voltage distribution.

`Mean`

Average of all cleaned voltage readings.

`Median`

Middle value of the cleaned voltage readings. If mean and median are close, the distribution is roughly symmetric.

`Std dev sigma`

Standard deviation of individual cleaned samples. This describes the spread of single readings, not the uncertainty of the final mean.

`Std error SEM`

Naive standard error of the mean: **SEM = standard deviation / sqrt(N)**

`Strategy : 100k deterministic subsample for JB/AD; full tails`

The dataset is large, so the script does not run every normality test on every sample. Instead:

- Jarque-Bera and Anderson-Darling use a deterministic 100,000-point subsample.
- Tail percentages use the full cleaned dataset.
- QQ correlation uses a smaller capped sample.

This avoids overloading the test while still giving meaningful diagnostics.

`Test sample N : 100,000 (QQ N=20,000)`

100,000 samples were used for formal normality tests. 20,000 samples were used for the QQ-linearity calculation.

`Jarque-Bera p`

Jarque-Bera checks whether skewness and kurtosis match a Gaussian distribution. For very large datasets, even tiny deviations can produce a very small p-value. So a small p-value here does not automatically mean the data is bad.

`large N: sensitive`

This warning means the p-value should not be interpreted too aggressively. With hundreds of thousands of samples, almost no real-world dataset is perfectly Gaussian.

`Anderson-Darling`

Anderson-Darling is another normality test. It is especially sensitive to tail behavior.

`A2=10.98, 5% crit=0.787 (fail)`

The Anderson-Darling statistic is larger than the 5% critical value, so the formal test rejects perfect normality. However, for large N this is expected even for very good practical data.

`QQ correlation`

This measures how straight the quantile-quantile plot would be. Values close to 1 mean the distribution is very close to Gaussian in shape. A value such as 0.999864 is excellent.

`|z| > 3 sigma`

This is the percentage of samples farther than 3 standard deviations from the mean. For a perfect Gaussian distribution, the expected value is about 0.27%. If your result is close to 0.27%, the practical tails are very Gaussian-like.

`|z| > 4 sigma`

This is the percentage of samples farther than 4 standard deviations from the mean. For a perfect Gaussian distribution, the expected value is about 0.0063%. This is a sensitive check for extreme tails.

`Robust |z| > 3`

This is similar to the 3-sigma tail check, but it uses a robust MAD-based sigma estimate instead of ordinary standard deviation. It is less affected by remaining outliers.

`Practical flag : small deviation from Gaussian`

This summarizes the normality diagnostics in practical terms. It means formal tests reject perfect Gaussian behavior, but the distribution is still close enough for most engineering/statistical uses.

Robust / Correlated-Sample Statistics

This section is very important for long time-series data. Many samples are correlated, so the number of statistically independent samples is much smaller than the number of raw readings.

`Trimmed mean (10%)`

This mean is calculated after removing the lowest 10% and highest 10% of samples. It is more resistant to tails and remaining outliers than the ordinary mean. If the trimmed mean and ordinary mean are close, the average is robust.

`Robust sigma (MAD)`

This is a robust estimate of standard deviation based on median absolute deviation. It is less sensitive to outliers than ordinary standard deviation. For Gaussian data:

robust sigma ≈ 1.4826 × MAD

`Winsorized std`

This is a robust standard deviation after replacing extreme values with percentile boundary values. It estimates spread while reducing the influence of tails.

`Effective N`

This is the estimated number of independent samples after accounting for autocorrelation.

Example: Effective N : 528.2 This means the full dataset behaves like about 528 independent readings for estimating the mean, even though it may contain hundreds of thousands of samples.

`autocorr lag used`

This tells how many sample lags were used while estimating autocorrelation.

Example: autocorr lag used: 1000 If the sample interval is about 0.506 seconds, 1000 lags correspond to about 506 seconds, or 8.4 minutes.

`Corr. SEM`

This is the autocorrelation-corrected standard error of the mean. It is calculated approximately as:

Corr. SEM = standard deviation / sqrt(Effective N)

This is usually much more realistic than the naive SEM for long correlated measurement records.

Raw vs Clean Comparison

This section shows how much the sanitization changed the data.

`Raw mean`

Average voltage before spike cleaning.

Average voltage after spike cleaning.

`Mean delta`

Difference between raw mean and cleaned mean. A small mean delta means spike cleaning did not significantly bias the final result.

`Raw std sigma`

Standard deviation before spike cleaning. This is often much larger if there were large spikes.

`Clean/raw std ratio`

Ratio of cleaned standard deviation to raw standard deviation.

Example: Clean/raw std ratio : 0.042x This means the cleaned standard deviation is only 4.2% of the raw standard deviation. That indicates the raw data had large spikes that strongly inflated the ordinary standard deviation.

MADEV Noise Floor

This section summarizes Modified Allan Deviation. MADEV is useful for separating certain noise types and identifying the best averaging time.

`Single-sample MADEV`

The MADEV value at the shortest averaging time, usually one sample interval. This represents short-term sample-to-sample instability.

`Noise floor MADEV`

The minimum MADEV value found across the tau range. This is the lowest stability deviation achieved by averaging.

`at tau`

The averaging time where the MADEV floor occurs.

Example: Noise floor MADEV : 375.41 nV at tau=1183 s This means averaging for around 1183 seconds gives the lowest MADEV in this run.

`95% chi2 CI`

Confidence interval around the MADEV value using a chi-square approximation. Allan-type deviation estimates have finite statistical confidence, especially at long averaging times.

`Degrees of freedom`

Effective degrees of freedom for the MADEV estimate at that tau. Higher degrees of freedom mean a more reliable estimate. At long tau, degrees of freedom usually decrease.

`Optimal m`

Number of samples in the averaging interval:

tau = m × tau0

`Indep. windows`

Approximate number of independent averaging windows at the optimal tau.

Improvement relative to single-sample MADEV.

Example: Averaging gain : 3.81x This means the best averaged stability is 3.81 times better than the shortest-tau stability.

`Floor ppm`

MADEV floor expressed relative to the measured voltage in parts per million.

`SNR vs floor`

Signal-to-noise ratio using the MADEV floor as the noise level.

`SNR single sample`

Signal-to-noise ratio using the single-sample MADEV value.

`TDev at tau_opt`

Time deviation-like integral of voltage noise at the optimal tau. It is calculated from MADEV as:

TDev = tau / sqrt(3) × MADEV

The unit is volt-seconds.

`Dominant noise`

Estimated noise type based on the local slope of the MADEV curve. This should be treated as a guide, not an absolute truth. If slope confidence is low, avoid over-interpreting this label.

ADEV Noise Floor

This section summarizes regular Allan Deviation.

`ADEV floor`

The lowest ADEV value across the tau range.

`at tau`

The averaging time where the ADEV floor occurs.

`Dominant noise`

Noise type estimated from the ADEV slope. ADEV is useful but may not distinguish some phase noise types as well as MADEV.

HDEV Noise Floor

This section summarizes Hadamard Deviation.

`HDEV floor`

The lowest HDEV value across the tau range.

`at tau`

The averaging time where the HDEV floor occurs.

Noise type estimated from the HDEV slope.

`If HDEV << ADEV at long tau, linear drift is significant`

HDEV is less sensitive to linear drift. If HDEV is much lower than ADEV at long averaging time, the ADEV curve may be contaminated by drift. If HDEV and ADEV are similar, linear drift is probably not the dominant issue.

Drift-Aware Statistics

This section compares the original cleaned data with a linearly detrended version.

`Detrended mean`

Mean after removing a fitted linear drift. It should normally be very close to the cleaned mean.

`Detrended std`

Standard deviation after subtracting the fitted linear trend. If this is much lower than the original standard deviation, linear drift is important.

`Detrended MADEV`

MADEV floor after removing linear drift.

`Detrended ADEV`

ADEV floor after removing linear drift.

`Detrended HDEV`

HDEV floor after removing linear drift.

`MADEV drift ratio`

Ratio of original MADEV floor to detrended MADEV floor.

Example: MADEV drift ratio : 1.001x This means linear drift has almost no effect on the MADEV floor.

`ADEV drift ratio`

Ratio of original ADEV floor to detrended ADEV floor.

`HDEV drift ratio`

Ratio of original HDEV floor to detrended HDEV floor.

`Drift influence`

A summary flag:

- **no** — linear drift is not significantly changing the stability metrics.
- **yes** — drift is probably contaminating long-tau stability.

Drift & Systematic Errors

This section reports the fitted linear drift. In V621 and later, drift is preferably reported per hour instead of per second.

Linear slope of voltage versus time.

Example: Drift rate : 21.63 nV/h This means the measured voltage is increasing by about 21.63 nV per hour.

`ppm/h`

Same drift expressed relative to the measured voltage.

`Drift per 24 h`

Expected change over 24 hours based on the fitted linear drift. This is useful for comparing with day/night effects.

`Total drift`

Expected voltage change over the full measurement duration.

Example: Total drift : 1.1267 uV over 1.448e+05 s (1 day 16 hours 13 min) This means the fitted linear drift accounts for about 1.1267 μV over the entire run.

`R2 of linear fit`

Fraction of voltage variance explained by a straight-line drift model. A very small R^2 means linear drift is not a major explanation of the data.

`Drift uncertainty`

Uncertainty of the fitted drift slope. In V621 and later this should be shown per hour.

`Bootstrap drift CI`

Bootstrap confidence interval for the drift slope. This is a more robust uncertainty estimate than the simple regression standard error.

Stationarity / Bootstrap

This section checks whether one global mean is reasonable.

`Bootstrap mean CI`

Block-bootstrap confidence interval for the mean. Block bootstrap is used because samples are correlated.

`Bootstrap block`

Number of samples in each bootstrap block. The block length is usually chosen based on the autocorrelation structure.

`Split mean shift`

Difference between the first half and second half means, expressed in units of standard deviation. Small values mean the mean is stable over time.

`Split std ratio`

Ratio of standard deviation between halves of the run. A value near 1 means the noise level is similar in both halves.

`Nonstationary flag`

Summary flag indicating whether the record appears nonstationary:

- **no** — one global mean is probably acceptable.

- **yes** — the run may contain changing conditions and should be analyzed in sections.

Advanced Diagnostics

This section contains deeper diagnostics added in later versions.

`Fitted MADEV slope`

Slope of the MADEV curve near the noise floor, fitted over a local log-log window. This is more stable than using a point-by-point gradient.

`Flicker freq./1f`

Estimated noise type from the fitted slope.

`confidence`

Confidence in the slope-based noise classification. Low confidence means the slope fit is weak or the curve is nearly flat/noisy.

`R2`

Goodness of fit of the local slope line. Low R^2 means the local curve does not follow a clean power law.

`Noise-floor plateau`

Range of tau values whose confidence intervals overlap the minimum deviation.

Example: Noise-floor plateau : 448.5 .. 3120 s This means many neighboring tau values are statistically similar to the minimum. In this case it is better to say the best averaging time is a range, not one exact tau.

`11 tau points overlap CI`

Number of tau points statistically consistent with the noise floor.

`Residual std`

Standard deviation after removing the fitted linear drift.

`Residual lag-1 rho`

Autocorrelation between adjacent residual samples. A high value means the noise is strongly correlated from one sample to the next.

`Residual MADEV`

MADEV floor calculated on residuals after linear drift removal. If this is similar to the original MADEV floor, linear drift removal did not change stability much.

`Robust drift fit`

Drift slope estimated using a robust regression method. For large datasets, the script may use Huber IRLS on a deterministic subsample.

- **Huber IRLS** — a regression method that reduces the influence of outliers.
- **deterministic subsample** — a repeatable subset of the data used to make robust regression faster for large files.

`Robust-vs-OLS drift`

Difference between robust drift and ordinary least-squares drift. If this is small, the drift estimate is not strongly affected by outliers.

24h Periodic / Environmental Analysis

This section checks for a repeating 24-hour component. This may be due to temperature, HVAC, sunlight, lab activity, or other daily effects. Without an actual temperature log, it should be called environmental or periodic, not definitely thermal.

`24h sine amplitude`

Amplitude of a fitted 24-hour sine wave.

Example: 24h sine amplitude : 1.9926 uV peak This means the best-fit daily sinusoidal component has a peak amplitude of about 1.99 µV.

`RMS`

Root-mean-square value of that sine component. For a sine wave:

RMS = peak amplitude / sqrt(2)

`24h sine p-p`

Peak-to-peak size of the fitted 24-hour sine wave. For a sine wave:

peak-to-peak = 2 × peak amplitude

`R2`

Fraction of total variance explained by the 24-hour sine model.

Example: R2=0.1891 This means the 24-hour component explains about 18.9% of the voltage variance.

`variance`

Same idea as R^2, expressed as a percentage.

`24h hour-bin p-p`

Peak-to-peak range of average voltage after grouping samples by hour of day. This is a non-sinusoidal check. If hour-bin p-p is larger than sine p-p, the daily pattern may not be a pure sine wave.

`24h peak phase`

Time of the fitted daily sine peak, expressed as hours after the run started. To convert this to clock time, add it to the start time of the measurement.

`24h equiv drift`

Daily peak-to-peak change divided by 24 hours. This is not a true linear drift; it is an easy way to express daily-cycle size as an equivalent rate.

`24h MADEV impact`

Ratio showing how much the 24-hour component affects long-tau MADEV.

Example: 24h MADEV impact : 8.732x at tau=3.882e+04 s This means the daily periodic component is strongly affecting the long-averaging-time stability estimate.

`24h influence`

Summary label. Typical interpretation:

- **low** — daily cycle is weak.

- **moderate** — daily cycle is noticeable.
- **high** — daily cycle is significant and should be reported.

`flag=YES`

The script detected a significant 24-hour periodic component.

`24h confidence`

In V62 and later, the report may include a confidence label based on the number of observed 24-hour cycles. If the run is shorter than two full days, the 24-hour fit is useful but not highly robust.

24h-Corrected Statistics

This section appears in V62 and later. It shows statistics after subtracting the fitted 24-hour periodic component.

`24h-corrected mean`

Mean after removing the fitted 24-hour component. This should usually remain close to the original mean.

`24h-corrected std`

Standard deviation after removing the fitted 24-hour component. If this is much lower than the original standard deviation, the daily cycle is a significant contributor to noise.

`24h-corrected MADEV`

MADEV floor after subtracting the fitted 24-hour sine component.

`Raw/corrected std ratio`

Ratio of original cleaned standard deviation to 24h-corrected standard deviation. A value greater than 1 means removing the daily component reduced the spread.

Stability Summary

This section condenses the most important results.

`Mean +/- corr SEM`

Best voltage estimate with autocorrelation-corrected standard error. This is often the most practical line for quoting the measured value.

`Bootstrap mean CI`

Block-bootstrap confidence interval for the mean. Use this as a realistic confidence interval when samples are correlated.

`MADEV floor`

Best stability floor and corresponding tau.

`Drift / 24 h`

Linear drift projected over 24 hours.

`24h periodic/environmental effect`

Size and strength of daily cycle. This is especially important if the effect is moderate or high.

`Quantization flag`

Whether measurement resolution appears to limit the results. **no** means analog/statistical noise is much larger than quantization noise.

Quantization

This section checks whether the instrument resolution or ADC step size is limiting the analysis.

`Detected LSD`

Estimated least significant digit or code step.

Example: Detected LSD : 100 nV This means the voltage readings appear quantized in 100 nV increments.

`mode-of-positive-differences`

The LSD was estimated from the most common positive difference between unique voltage values. This is usually more reliable than using the smallest difference, because floating-point artifacts can create tiny fake differences.

`LSD percentile`

A diagnostic low-percentile difference estimate. If it is reported as ignored, it means it was contaminated by floating-point precision artifacts.

`LSD mode estimate`

The physical LSD estimate actually used for quantization analysis.

`Levels hit`

Number of quantization codes that appeared in the cleaned data.

`of possible`

Number of possible quantization codes between cleaned minimum and maximum voltage.

`Quantized range`

Minimum and maximum voltage covered by the common quantization grid.

`Quant. origin`

Reference voltage used for assigning samples to quantization codes. Using one common origin avoids impossible results like "levels hit greater than possible levels."

`Missing levels`

Possible quantization codes in the range that were not observed.

`Occupancy`

Percentage of possible quantization codes that were observed. High occupancy means the distribution fills the available voltage range smoothly.

`Occupancy entropy`

Measures how evenly the occupied quantization codes are used:

- **1** — evenly used.

- **lower values** — distribution is concentrated in fewer codes.

`Quantization RMS`

The theoretical RMS noise from uniform quantization:

Quantization RMS = LSD / sqrt(12)

`Sigma / LSD`

Ratio of measured standard deviation to quantization step size. Large values mean the measurement noise spans many codes.

`Quantization limit`

Summary flag:

- **no** — quantization is not limiting the result.
- **yes** — the ADC/code width may limit further averaging or interpretation.

Recommendations

This section gives conditional advice based on the measured data.

`Use corrected SEM or bootstrap CI instead of naive SEM`

This appears when the data is correlated. It means the ordinary SEM is too optimistic.

`Stationarity checks passed`

The record appears stable enough for one global mean.

`LSD percentile estimate was ignored`

The percentile-based LSD estimate was affected by floating-point artifacts. The mode-based LSD estimate was used instead.

`24h cycle detected`

The data contains a measurable 24-hour periodic component. The script recommends correlating it with temperature, HVAC, sunlight, or lab operation schedule.

`Quote the MADEV floor as a plateau range`

Several tau values are statistically similar to the minimum. Report the noise floor over a tau range rather than pretending one exact tau is uniquely optimal.

`Noise-type slope confidence is low`

The slope-based noise classification should not be over-interpreted. The label is useful, but not definitive.

`Residuals remain strongly autocorrelated`

Even after removing linear drift, the data still has time correlation. This suggests environmental effects, low-frequency noise, instrument settling, or periodic behavior.

How to Interpret the Whole Report

A good concise human interpretation might be:

> The cleaned mean voltage is stable and well estimated. The raw file contained isolated spikes, but spike removal had only a small effect on the mean. The sample noise is strongly correlated, so the corrected SEM and bootstrap CI are more realistic than the naive SEM. The MADEV floor occurs around the 20-minute averaging range, but the statistically equivalent plateau is broader. Linear drift is small, but a 24-hour environmental cycle is significant and affects long-tau stability. Quantization is not limiting the measurement because the noise spans many ADC codes.

What to Quote in a Summary Report

For a short measurement report, quote the following:

- Mean voltage
- Corrected SEM
- Bootstrap mean CI
- MADEV floor and plateau range
- Linear drift per 24 h
- 24h periodic/environmental amplitude and R^2
- Quantization status
- Sanitization count

Example format:

```text
Mean voltage:          7.05652096 V
Corrected SEM:         144 nV
Bootstrap CI:          7.05652082 V to 7.05652112 V
MADEV floor:           375 nV near tau = 1180 s
Noise-floor plateau: approximately 450 s to 3100 s
Linear drift:          approximately 0.52 µV / 24 h
24h periodic effect: approximately 4.0 µV peak-to-peak, R² ≈ 0.19
Quantization:          100 nV LSD, not quantization-limited
Sanitization:          about 0.39% isolated spikes replaced
```