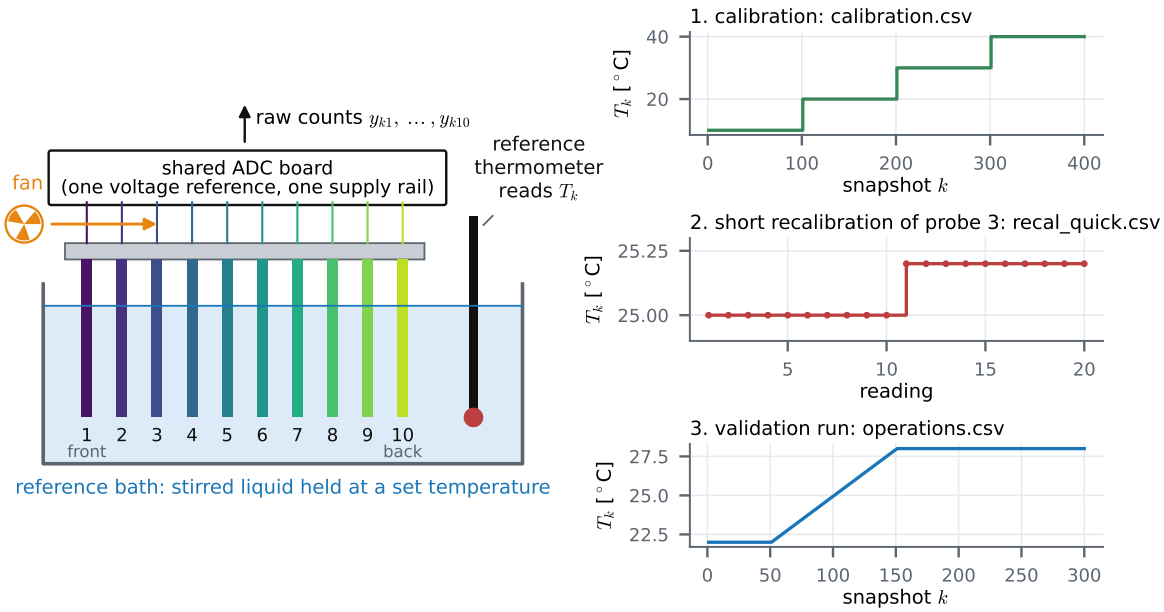
Hands-On 1: Calibrating a Sensor Array

Handout

Signal Processing (ECE-Y523) · University of Patras
Konstantinos Chatzilygeroudis · 29 October 2026

The problem. Ten temperature probes share an ADC, a power supply and airflow along the rack. Their raw counts have different offsets and gains, and their errors are correlated. Calibrate the probes, combine their readings with an uncertainty, and diagnose recurring disagreement.

The measurements. The probes are first calibrated in a stirred reference bath, whose temperature is accurately known. At each of 10, 20, 30 and 40 °C we record 100 snapshots; a snapshot holds the ten simultaneous probe readings. Probe 3 was later recalibrated twice: a short recalibration with the bath at only 25.0 and 25.2 °C, and a full recalibration over the whole range for comparison. Finally, the rack is placed in a temperature-controlled chamber, where the temperature changes over time, and we record 300 more snapshots. This *operational run* plays the role of normal operation. A reference thermometer also records the chamber temperature, but we keep its readings hidden until the final validation.



Left: the rack, shared ADC, fan and reference thermometer in the bath. Right: the temperature programmes of the three experiments.

The route. Follow the same six numbered sections in the notebook. Matrices use bold uppercase symbols ($\boldsymbol{X}$), vectors bold lowercase symbols ($\boldsymbol{w}$), and scalars plain symbols (T_k).

	Stage	What you establish
1	Calibrate	counts become °C
2	Rescue a poor calibration	what ridge and a prior repair
3	Understand the noise	covariance and shared patterns
4	Report a temperature	combination and uncertainty
5	Diagnose faults	persistent evidence and a suspect
6	Validate	compare the report with the reference

1 Calibrate: counts to temperature

Lectures 1 and 5

Question. What line turns each probe's counts into degrees? Let $N = 400$, $C = 10$, k index snapshots and j index probes. Fit the *counts as output*: the reference temperature is accurately known, whereas the probe reading contains the noise.

$$y_{kj} = a_j + b_j T_k + n_{kj}, \qquad \boldsymbol{X}_{k,:} = [1, T_k], \qquad \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\Theta} + \text{noise}.$$

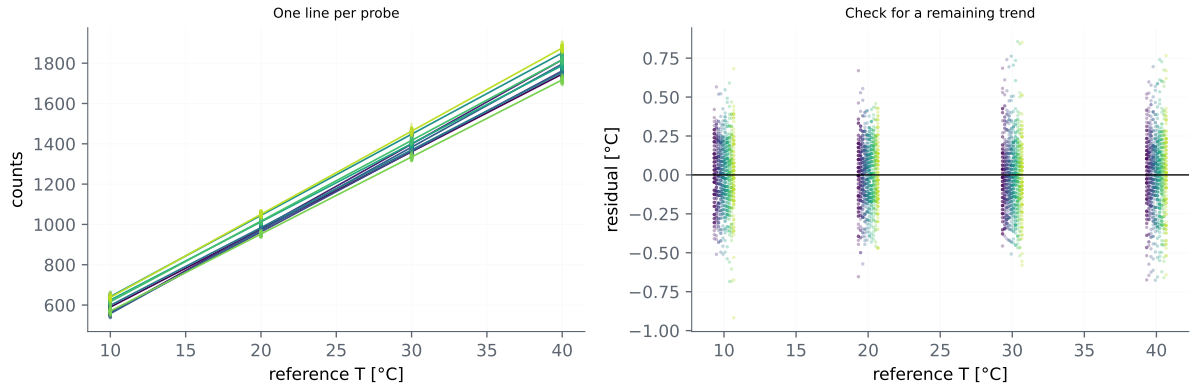
$\boldsymbol{Y}$ is $N \times C$, $\boldsymbol{X}$ is $N \times 2$ and $\boldsymbol{\Theta}$ is $2 \times C$. Offset a_j has units of counts; gain b_j has units of counts/°C. One least-squares call fits all ten columns:

```
X = np.column_stack([np.ones(N), T_cal])
Theta = np.linalg.lstsq(X, Y, rcond=None)[0]
a, b = Theta
R = Y - X @ Theta
E = R / b
```

Invert the fitted lines to obtain temperatures:

$$\hat{T}_{kj} = \frac{y_{kj} - \hat{a}_j}{\hat{b}_j}.$$

The residuals $\boldsymbol{R}$ are in counts, while $\boldsymbol{E}$ is in °C. Dividing each residual column by its fitted gain explains the broadcasting in `R / b`.



Left: fit a separate line to each probe. Right: inspect the remaining errors at each bath temperature.

Read the residuals. Bath-dependent curvature or offsets would suggest that the affine model misses structure. The noise size for each probe is $s_j = \sqrt{\sum_k R_{kj}^2/(N-2)}$ in counts, or $s_j/\hat{b}_j$ in °C. Small residual trends support the model on these data; they do not establish Gaussianity or independence.

Discussion

How do nominal $(a, b) = (200, 40)$ and fitted calibrations differ for one snapshot? Which changes: bias, scatter or both?

Test your understanding. Which errors does calibration remove, and roughly how much noise remains per probe? Why should every later rack calculation use the full calibration? Appendix A gives coefficient uncertainties and the shared-design argument.

2 Rescue a poor calibration

Lectures 2, 5 and 6

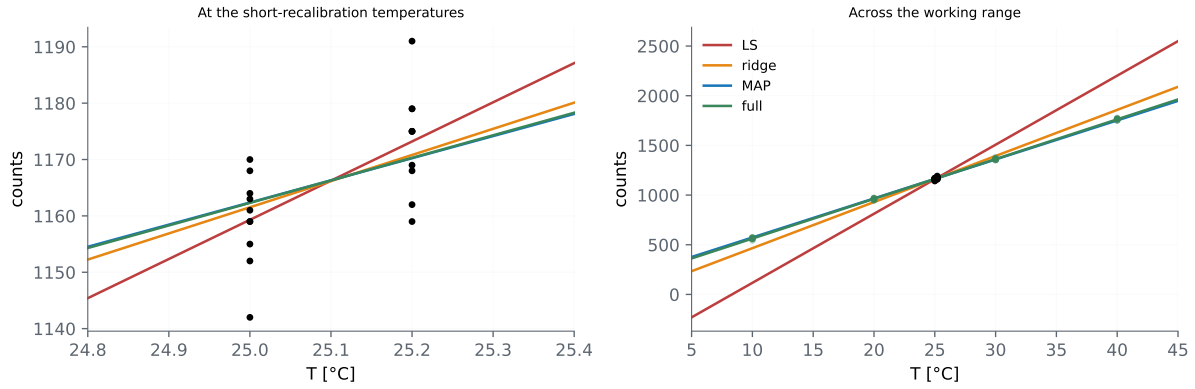
Question. Why can a convincing local fit be a poor calibration? The short recalibration of probe 3 uses ten measurements at 25.0 °C and ten at 25.2 °C. Almost identical reference temperatures barely constrain the gain. The condition number grows from about 67 for the full calibration to 6310 for the short recalibration. This is a controlled comparison; the later rack calculations continue to use the full fit from Section 1.

Compare least squares, plain ridge with $\lambda = 1$, and MAP under a physically motivated prior. For Gaussian noise, least squares is maximum likelihood. Plain ridge is MAP under an isotropic zero-centred Gaussian prior; that prior is a poor physical description of offset and gain in different units.

The other nine probes provide a prior mean $\boldsymbol{\mu}_0$ and separate standard deviations for the two coefficients. With $\boldsymbol{\Lambda}_0 = \mathrm{diag}(1/s_{0,a}^2, 1/s_{0,b}^2)$ and count-noise sd $\sigma = s_3$,

$$\boldsymbol{P} = \frac{\boldsymbol{X}_q^\top \boldsymbol{X}_q}{\sigma^2} + \boldsymbol{\Lambda}_0, \qquad \boldsymbol{P}\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \frac{\boldsymbol{X}_q^\top \boldsymbol{y}_q}{\sigma^2} + \boldsymbol{\Lambda}_0\boldsymbol{\mu}_0.$$

Solve this system; a matrix inverse is unnecessary for computing the estimate.



All fits are plausible locally. Their behaviour across the working range is very different. The full recalibration supplies the comparison.

Fit	Offset [counts]	Gain [counts/°C]	Line error at 40 °C
LS	−578.2	69.50	+10.99 °C
Plain ridge	1.7	46.40	+2.38 °C
Physical MAP	180.5	39.28	−0.27 °C
Full recalibration	162.3	40.00	reference

The line error is the predicted-count difference divided by the full-recalibration gain; it measures disagreement between calibration lines in temperature units.

Discussion

What happens when the prior standard deviations are multiplied by five? How does the balance between measurements and prior information change?

Test your understanding. Ridge stabilises a solution; what determines where it stabilises? Can a stable answer still be wrong? Appendix B and exercises 2–3 of the optional notebook cover SVD, posterior contours and repeated short recalibrations.

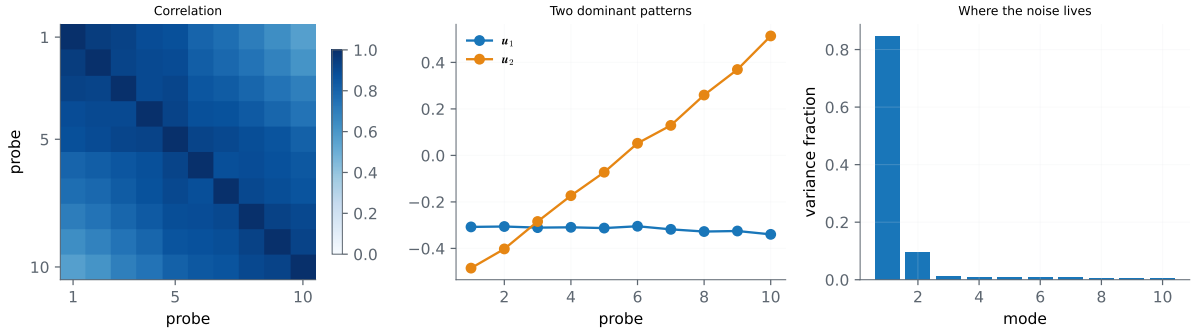
3 Understand the noise

Lectures 4 and 5

Question. How do the probes' errors move together? Each row of $\boldsymbol{E}$ is a residual across all ten probes in °C. The calibration fit used two parameters per probe, so estimate covariance as

$$\boldsymbol{S} = \frac{\boldsymbol{E}^\top \boldsymbol{E}}{N-2}, \qquad \rho_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} S_{jj}}}.$$

$\boldsymbol{S}$ is $C \times C$, with units of °C^2. Off-diagonal entries matter: these measurements are simultaneous, and the ADC and airflow are shared.



Read the correlation matrix, the two dominant eigenvectors, and the fraction of variance carried by each mode. Eigenvector signs are arbitrary.

Read the patterns. For $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$, eigenvector $\boldsymbol{u}_i$ is a pattern across probes and λ_i is the variance along that pattern. A nearly uniform pattern suggests a shared electronic disturbance. A front-to-back pattern is consistent with airflow along the rack. These two modes carry about 94% of the variance; covariance alone does not prove their physical causes.

Whiten. Cholesky factorisation gives $\boldsymbol{S} = \boldsymbol{L}\boldsymbol{L}^\top$. For a column error vector, solve $\boldsymbol{L}\boldsymbol{q} = \boldsymbol{e}$. For all rows at once:

```
L = np.linalg.cholesky(S)
Q = np.linalg.solve(L, E.T).T
```

Whitening rescales and mixes coordinates. Under the Gaussian model with the true covariance, the whitened coordinates are independent standard normals. With this estimated covariance, $Q^\top Q/(N-2) = \boldsymbol{I}$ on the calibration data by construction. It checks the computation. Independent validation comes later.

Discussion

Which patterns are uniform and front-to-back? What does a negative eigenvector entry mean?

Test your understanding. The same covariance model is now used for both combination and diagnosis. Which two assumptions about the errors does it rely on? Appendix C derives the estimate; exercise 4 of the optional notebook compares whitening matrices.

4 Report one temperature with uncertainty

Lecture 5

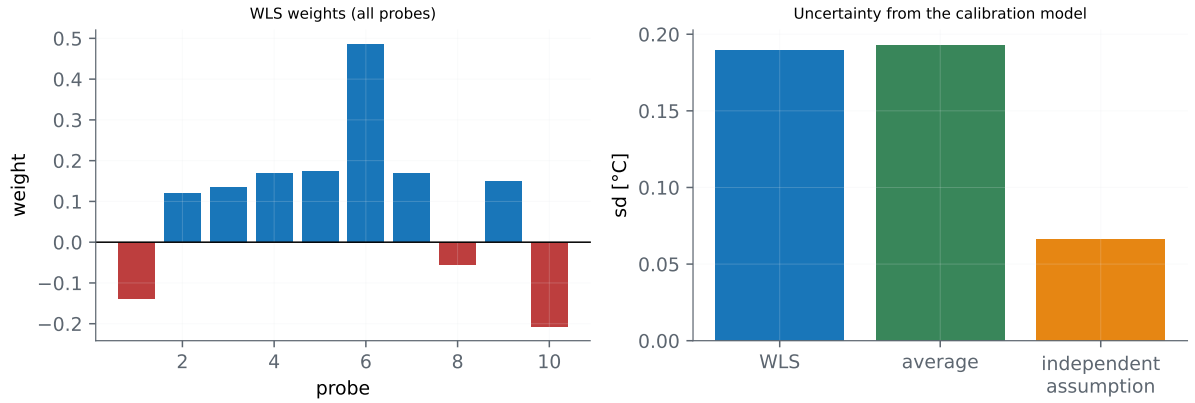
Question. How much does combining ten correlated probes help? For one snapshot, let $\boldsymbol{t}$ contain the calibrated probe temperatures: $\boldsymbol{t} = T\mathbf{1} + \boldsymbol{e}$, with $\mathrm{Cov}(\boldsymbol{e}) = \boldsymbol{S}$. Weighted least squares gives

$$\boldsymbol{w} = \frac{\boldsymbol{S}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{S}^{-1}\mathbf{1}}, \qquad \hat{T} = \boldsymbol{w}^\top \boldsymbol{t}, \qquad \mathrm{sd}(\hat{T}) = \frac{1}{\sqrt{\mathbf{1}^\top \boldsymbol{S}^{-1}\mathbf{1}}}.$$

The weights sum to one. Among unbiased linear combinations under this model, these weights have minimum variance. Compute them with a solve:

```
Si1 = np.linalg.solve(S, np.ones(C))
w = Si1 / Si1.sum()
T_all = T_probes @ w
sd_all = 1 / np.sqrt(Si1.sum())
```

The notebook's `combine` function also returns the residual distance needed in the next section. The same function is reused after diagnosis.



Some WLS weights are negative. Both WLS and a plain mean have much larger uncertainty than an independent-probe calculation suggests. All probes are still included here.

Compare uncertainties. The plain mean has variance $\mathbf{1}^\top \boldsymbol{S}\mathbf{1}/C^2$. If the probes were independent, that variance would instead be $\sum_j S_{jj}/C^2$. Here these give standard deviations near 0.193 °C and 0.066 °C, respectively; WLS gives 0.190 °C. The shared component survives averaging. Negative weights can cancel some correlated disturbances, but cannot remove a disturbance identical to a true change in temperature.

These are model-based uncertainties for healthy probes. A faulty probe can bias a combined reading. The full calibration's coefficient uncertainty is small here and is omitted (appendix F); the short recalibration would require more care.

Discussion

How does uncertainty change without the off-diagonal entries of $\boldsymbol{S}$? Which assumption changed?

Test your understanding. Why must a reported uncertainty include the cross-probe covariance? Why does the operational reference stay hidden until Section 6? Appendix F proves the combination formula and the common-noise floor.

5 Diagnose recurring disagreement

Lecture 4, with a short introduction to χ^2

Question. Is the disagreement too large, and which probe explains it? Use the residual about the estimated temperature, $\boldsymbol{r} = \boldsymbol{t} - \hat{T}\mathbf{1}$. The unknown true temperature cancels. Under the Gaussian model with known covariance,

$$d^2 = \boldsymbol{r}^\top \boldsymbol{S}^{-1} \boldsymbol{r} \sim \chi^2_{C-1}.$$

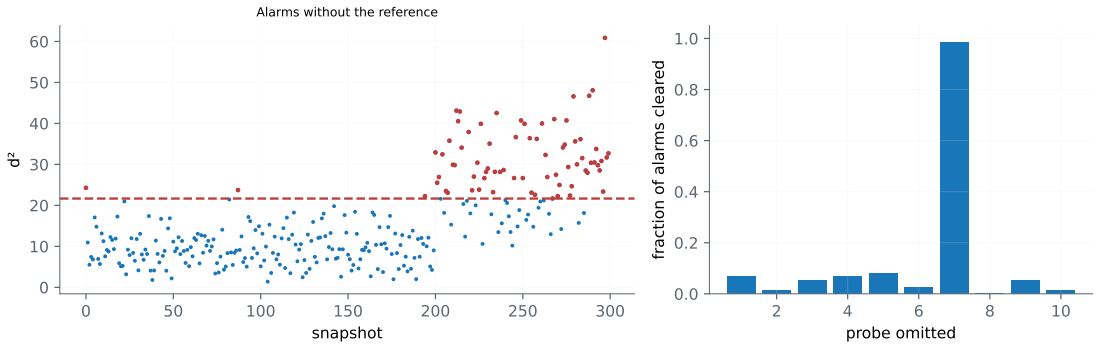
A χ^2_ν variable is a sum of ν squared independent standard normals. Fitting one temperature removes one direction, hence $C - 1$ degrees of freedom. For ten probes the 99% quantile is 21.67. Estimated $\boldsymbol{S}$ makes the nominal 1% per-snapshot false-alarm rate approximate.

An alarm is evidence about a snapshot. Even a healthy operational run of 300 snapshots should produce about three alarms. Before excluding a probe, check whether there are unusually many alarms. If m of n snapshots alarm, `stats.binom.sf(m-1, n, 0.01)` gives the probability of at least m alarms under an independent healthy-snapshot model with rate 1%. We call the run's alarms excessive when this probability is below 0.01.

A conservative single-fault rule.

1. If the run has no excess alarms, retain all probes.
2. Otherwise omit each probe, refit T, and retest with $C - 2$ degrees of freedom.
3. Exclude one probe only if it uniquely clears more than half the original alarms and the remaining rack no longer has excess alarms.
4. If the evidence is ambiguous, retain the probes and report an unresolved cause.

The notebook implements this rule in `diagnose`. It is an offline investigation under a single recurring fault assumption. The search and estimated covariance prevent interpreting it as a guaranteed fault-classification test.



The code uses no supplied fault time, probe number or operational reference temperature. Investigate the timeline before consulting the validation labels.

Discussion

When does disagreement become persistent? Do the alarms in the first 100 snapshots justify exclusion?

Test your understanding. Why does a single snapshot alarm not identify a faulty probe? This code does not estimate a change point; see appendix G for the degrees of freedom.

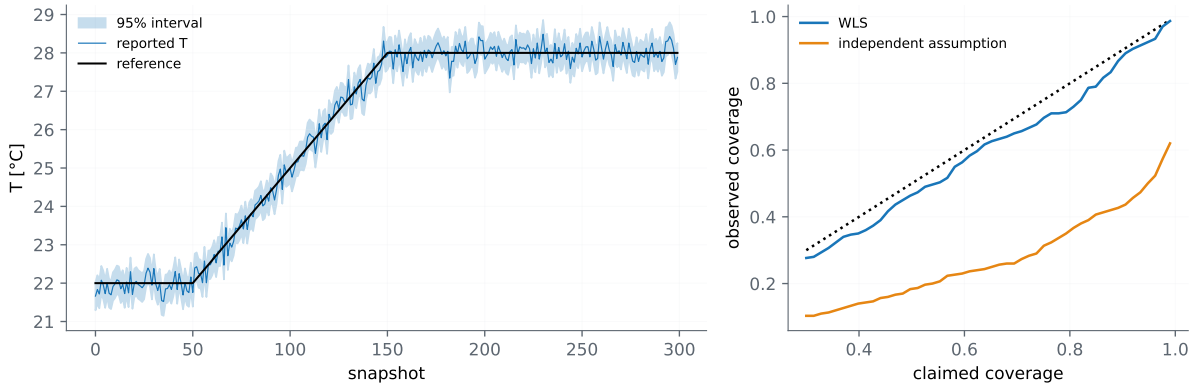
6 Validate the final report

Lectures 4 and 5

Question. Does the output agree with information withheld from diagnosis? Recombine the retained probes using their covariance submatrix. Now reveal `T_ref`. Check the temperature errors and their coverage; inspect whitening on the independent healthy part of this run.

```
keep = np.ones(C, dtype=bool)
if suspect is not None:
    keep[suspect] = False
T_final, sd_final, _ = combine(
    T_probes[:, keep], S[np.ix_(keep, keep)])
T_ref = ops[:, 1]
coverage = np.mean(abs(T_final - T_ref) <= 1.96*sd_final)
```

Validation labels, revealed here. The synthetic data contain a +0.30 °C offset on probe 7 from snapshot 200. The preceding 200 snapshots are healthy. These labels are used only now, to compare healthy and faulty rates and to check the calibration whitener on new healthy errors. A 0.30 °C fault is about 1.5 times probe 7's own standard deviation; shared noise makes the individual readings less informative than their joint pattern.



Final temperature and nominal 95% interval, followed by observed coverage against claimed coverage. The reference is used for validation only.

Validation result	Value
Suspect supported by the full run	probe 7
Healthy-snapshot alarm rate	1.5%
Faulty-snapshot alarm rate	70%
Final temperature sd	0.190 °C
Observed coverage of nominal 95% interval	92.3%
Excluded probes when only healthy data are processed	none

Interpret the limits. The held-out whitened covariance is near, rather than exactly equal to, the identity. The covariance was estimated from 400 calibration snapshots; coverage need not equal its nominal value. The suspect was also selected from this same run. This is a worked validation, not an independent assessment of fault-classification performance. A changed fan, time-correlated errors or several faulty probes would require revisiting the assumptions and testing on new data.

Discussion and further reading

At the end of the session. The final temperature plot and diagnostic timeline bring together calibration, correlated noise, estimation and fault detection. The questions below connect these results to the mathematical tools used throughout the session.

Questions.

1. Why fit counts against reference temperature before inverting the line?
2. Why does the spread of reference temperatures matter more than their absolute level?
3. Why does averaging fail to remove shared noise?
4. Why is the operational threshold based on $C - 1$, and the leave-one-out threshold on $C - 2$?
5. Why are three alarms in 300 snapshots insufficient to condemn a probe?
6. Why is identity covariance on the calibration data a weak check of whitening?

Short answers. (1) Noise is predominantly in counts, while the reference is accurately known. (2) Temperature spread determines the gain's lever arm. (3) A shared disturbance appears in all readings together. (4) Estimating the common temperature uses one degree of freedom, including after dropping a probe. (5) A 1% per-snapshot test produces about three alarms in a healthy run of that length. (6) The same data were used to construct the whitener.

Data files. All files are in `data/`.

File in `data/`	Contents
`calibration.csv`	400 snapshots; columns: snapshot, reference temperature, ten probes
`recal_quick.csv`	Short recalibration of probe 3 at 25.0 and 25.2 °C; reference temperature and counts
`recal_full.csv`	Full recalibration of probe 3 over the whole range; answer key for the short recalibration
`operations.csv`	300 new snapshots; same columns as calibration

Additional material. After the session, you can read the mathematical appendices, which give the full derivations and further analysis, or run the exercises of the optional notebook, which explore the same ideas numerically.

Mathematical appendix	Topic
A	Affine coefficient uncertainty; residuals; shared-design LS and WLS
B	Ill-conditioning; ridge through SVD; MAP and bias–variance
C	Covariance estimation and inverse-covariance inflation
D	χ^2 properties, quantiles and dimension
E	Ellipses, whitening and reference-based detector power
F	Combining correlated probes and calibration uncertainty
G	Reference-free residuals and the lost degree of freedom

Notebook exercise	Question
1	Why does one matrix solve equal ten separate fits?
2	How can the local fit look convincing while the gain is unreliable?
3	What do the prior and the likelihood each contribute?
4	Why does a two-dimensional "2-sigma ellipse" contain only 86%?
5	Why can correlation make a small single-probe fault easier to detect?