Methodology and Error in Individual Urine Samples

Tecla Duran Fort

2025-12-03

Contents

Load Data

Calibration Curves

Reference Polynomial Adjustment to Enforce Origin Constraint

```
## 'data.frame':    28 obs. of  2 variables:
##  $ concentration: num  10.326 30.326 0.326 20.326 5.326 ...
```

```
##  $ intensity    : num  1331.8 2088.8 44.6 1862 885.7 ...
## 'data.frame':    28 obs. of  2 variables:
##  $ concentration: num  10 30 0 20 5 10 30 0 20 5 ...
##  $ intensity    : num  1013 1593.2 21.9 1462.2 642.6 ...
```

Polynomial Fit

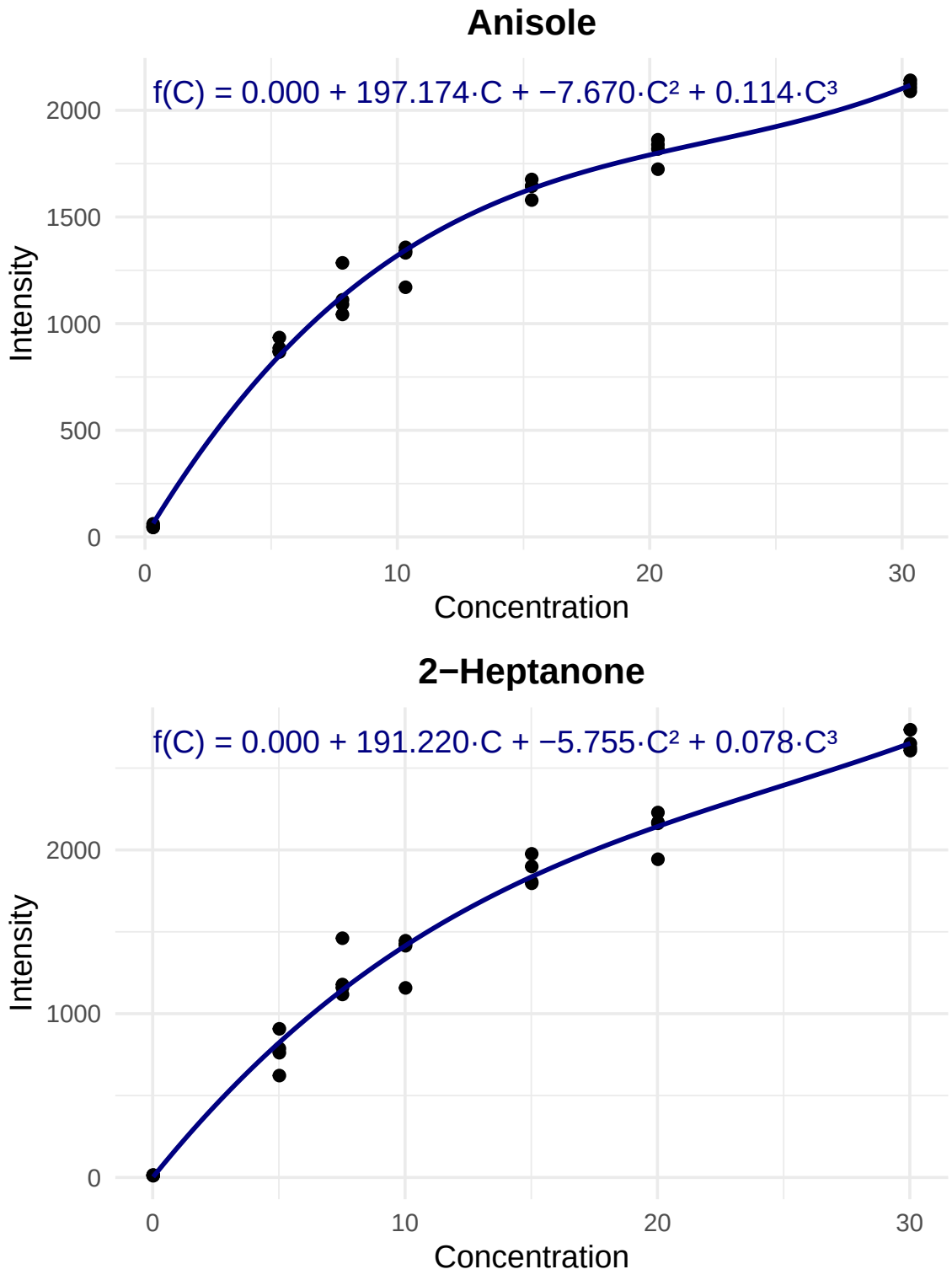


Table 1: Polynomial coefficients (degree 3, no intercept)

Analyte	a0	a1	a2	a3
Anisole	0	197.1742	-7.6705	0.1145
2-Heptanone	0	191.2197	-5.7550	0.0775

Table 2: Model fit metrics

Analyte	R^2	Adjusted R^2	RSE	RSS
anisole	0.9984	0.9982	60.6589	91987.62
heptanone	0.9960	0.9955	110.8239	307048.44

Table 3: Polynomial equations

Analyte	Equation
Anisole	$f(C) = 197.174 \cdot C + -7.670 \cdot C^2 + 0.114 \cdot C^3$
2-Heptanone	$f(C) = 191.220 \cdot C + -5.755 \cdot C^2 + 0.078 \cdot C^3$

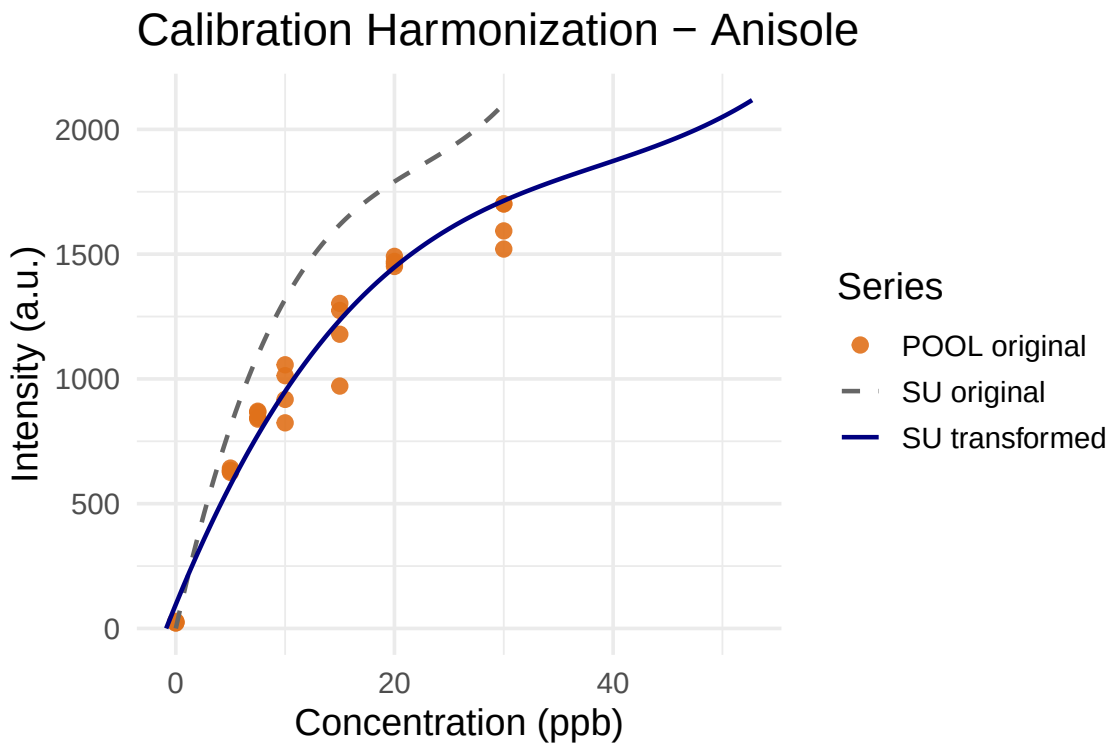
Harmonization

Anisole

```
res_anisole <- harmonize(pool_anisole, su_anisole)
```

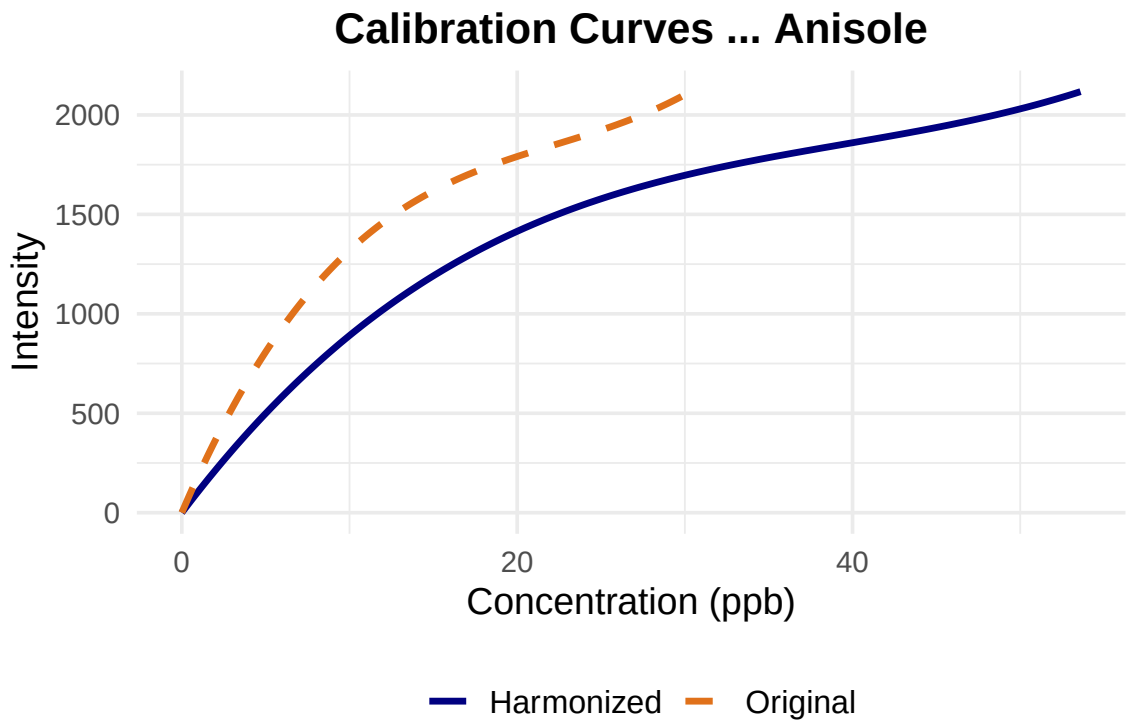
Table 4: Harmonization parameters and uncertainty (Anisole)

Parameter	Value	Error	Units
Scale	0.566	±0.125	–
Shift	0.879	±2.015	ppb



Final Calibration Curves

Calibration curves without harmonization (baseline model) and with pool harmonization (proposed model).



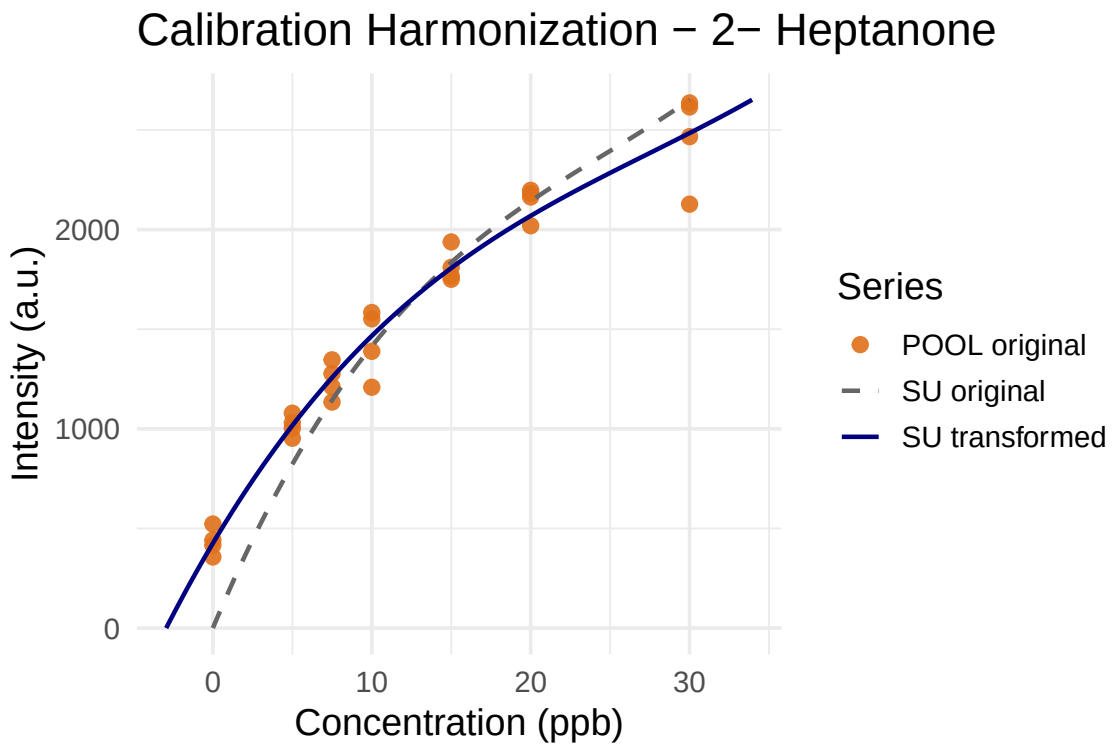
Heptanone

```
res_heptanone <- harmonize(pool_heptanone, su_heptanone)
```

Table 5: Harmonization parameters and uncertainty (Heptanone)

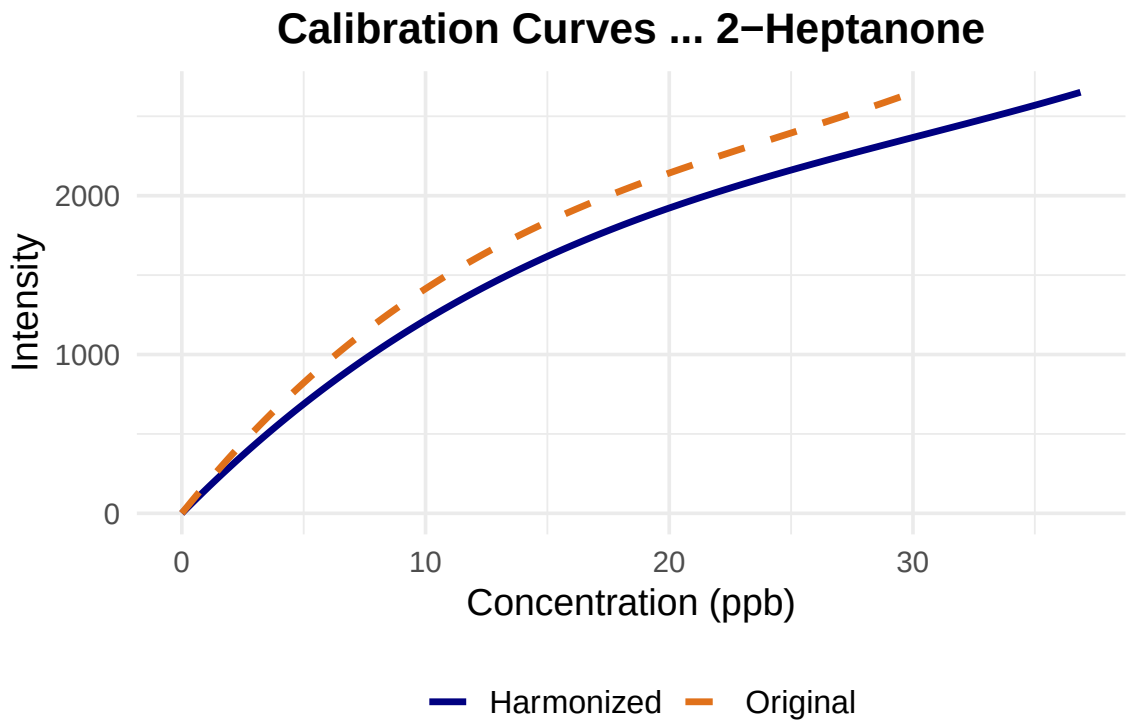
Parameter	Value	Error	Units
Scale	0.814	±NA	–
Shift	2.940	±NA	ppb

```
plot_harmonization(res_heptanone, title = "Calibration Harmonization - 2- Heptanone")
```

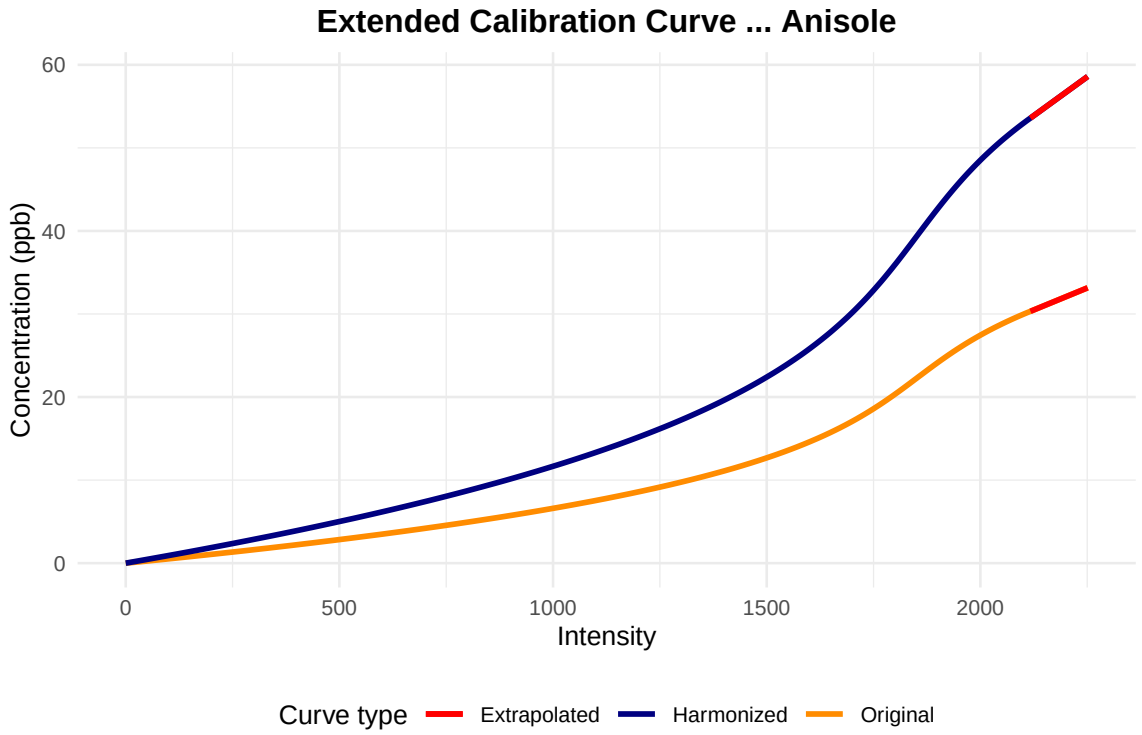


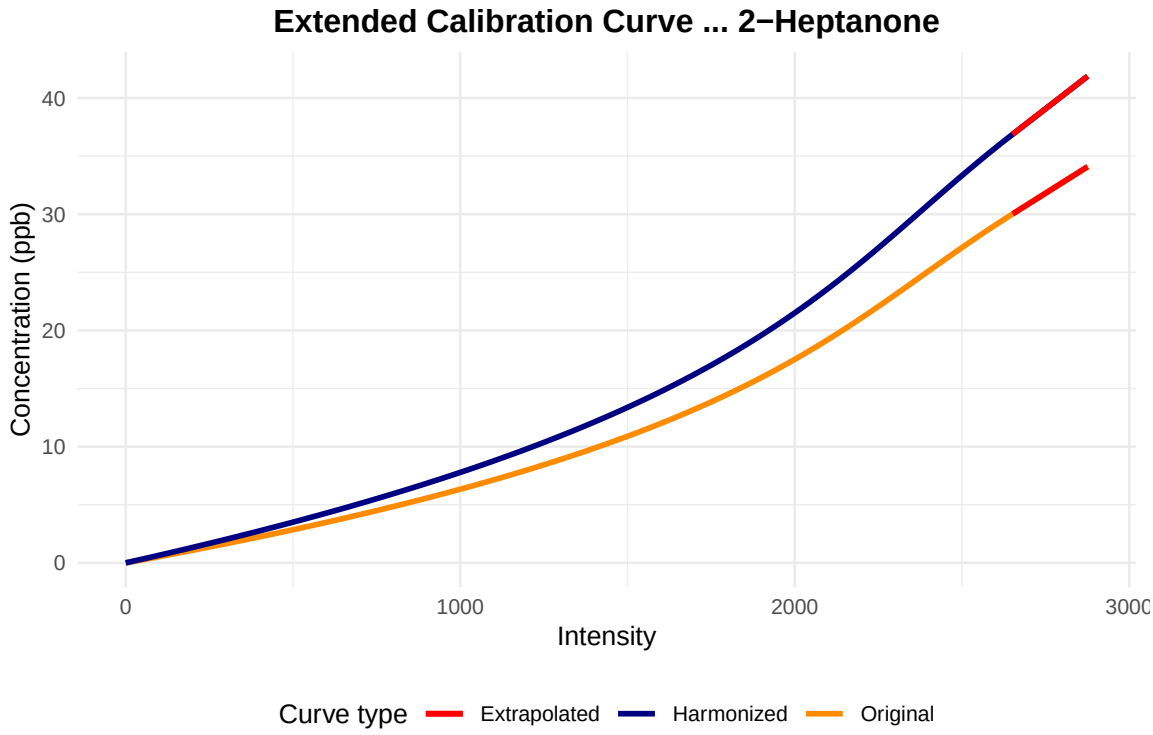
Final Calibration Curves

Calibration curves without harmonization (baseline model) and with pool harmonization (proposed model).


Calibration Curves ... 2−Heptanone
Intensity
2000
1000
0
0
10
20
30
Concentration (ppb)
Harmonized
Original

Extrapolation





Concentration Prediction

Prediction Without Scaling the Curve (Baseline)

Table 6: Estimated concentrations across datasets (baseline calibration)

Sample	Analyte	Spiked Conc. (ppb)	Intensity (a.u.)	Estimated Conc. (ppb)
pool	anisole	10	1012.9812	6.7167
pool	anisole	30	1593.1510	14.4560
pool	anisole	0	21.8578	0.1115
pool	anisole	20	1462.1706	12.0458
pool	anisole	5	642.6036	3.7850

Prediction Scaling the Curve (Proposed Method)

Table 7: Estimated concentrations across datasets (proposed calibration)

Sample	Analyte	Spiked Conc. (ppb)	Intensity (a.u.)	Estimated Conc. (ppb)
pool	anisole	10	1012.9812	11.8705
pool	anisole	30	1593.1510	25.5483
pool	anisole	0	21.8578	0.1971
pool	anisole	20	1462.1706	21.2888
pool	anisole	5	642.6036	6.6893

Error Calculation

Anisole (Non-Endogenous)

In this section we evaluate the accuracy of the estimated concentrations obtained from the calibration models. For each measurement we define the **error** as:

$$\varepsilon_i = \widehat{C}_i - C_i,$$

where
- C_i = true (spiked) concentration, since anisole is non endogenous analyte
- $\widehat{C}_i$ = estimated concentration from the calibration model.

To summarise the error magnitude we use the **root-mean-square error (RMSE)**:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_i (\,\widehat{C}_i - C_i\,)^2}.$$

RMSE gives a single value (in ppb) that increases when predictions deviate from the true concentration. We compare RMSE between:

- **Baseline model** (using the original SU calibration curve)
- **Proposed model** (using the harmonised and scaled calibration curve)

for: - Pool sample
- Individual samples (s1–s3, s4)
- Global error across all samples
- Each concentration level

The following tables and plots present these comparisons.

RMSE in Pool

Table 8: RMSE in pool (anisole)

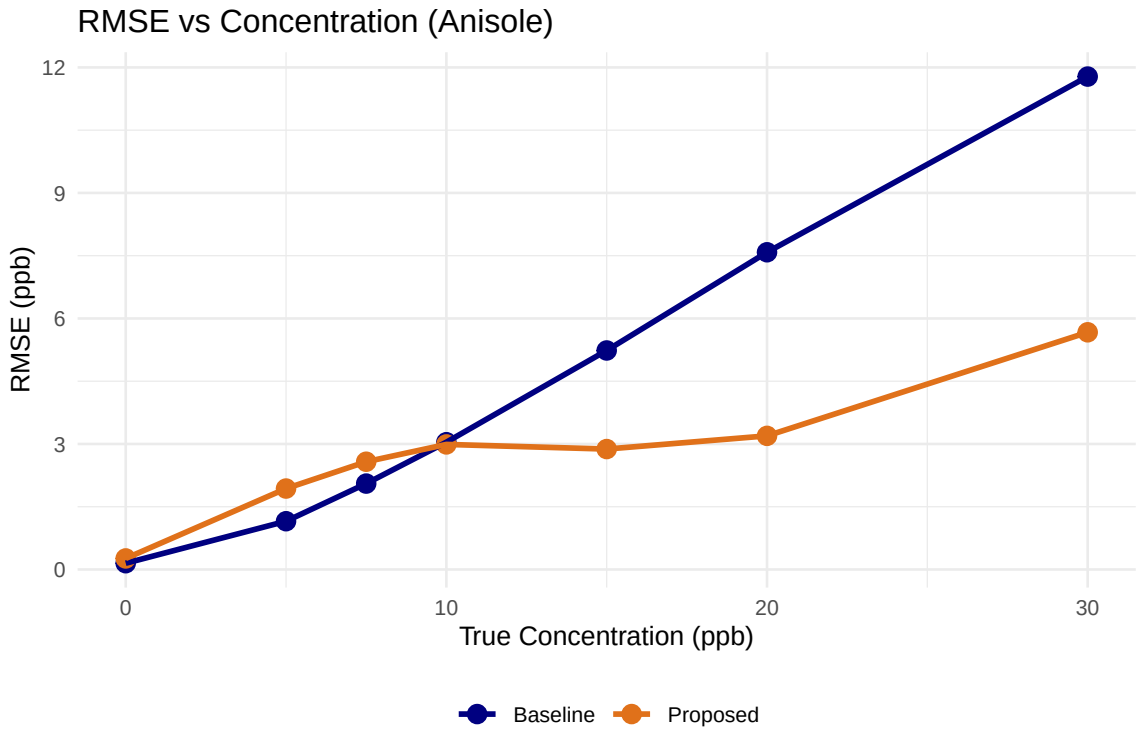
Method	RMSE (ppb)
Baseline	6.993950
Proposed	2.217788

RMSE in Individual Samples

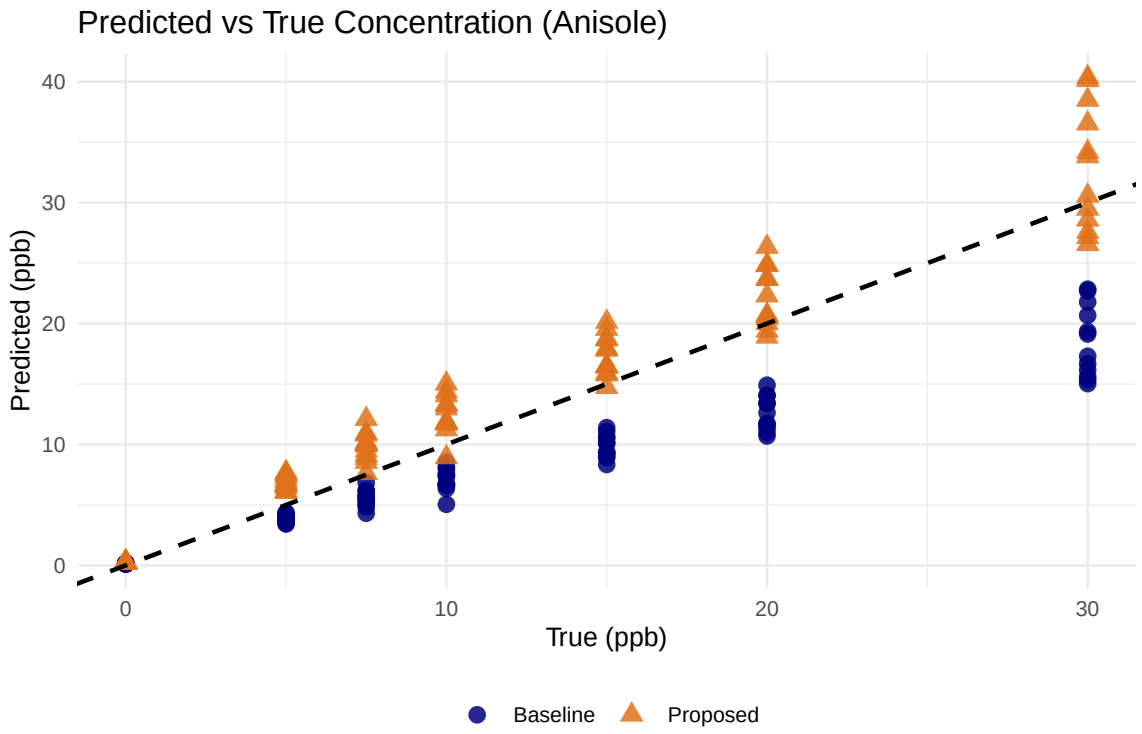
Table 9: Global RMSE (anisole)

Method	RMSE (ppb)
Baseline	5.837247
Proposed	3.162367

RMSE vs Concentration Level



Predicted vs True Concentration



2-Heptanone (Endogenous)

Unlike anisole, **2-heptanone is naturally present** in urine without spiking.
Therefore, to compute the prediction error, we must estimate the **endogenous (baseline) concentration** C_0 for each sample.

We treat each sample as a shifted version of the SU calibration curve.
Using the affine alignment model:

$$f_{\text{sample}}(C_{\text{total}}) \approx f_{\text{SU}}(\alpha \cdot C_{\text{total}}) \,,$$

we obtain:

- C_0: endogenous concentration (offset)
- α: scale factor due to matrix effects

The **true concentration** for each point becomes:

$$C_{\text{total}} = C + C_0$$

All RMSE calculations use this corrected concentration.

Endogenous Concentration Estimation

Table 10: Estimated endogenous concentrations (2-Heptanone)

Sample	Analyte	C0 (ppb)	Error

pool	heptanone	2.940	2.255
s1	heptanone	1.030	0.816
s2	heptanone	0.553	0.664
s3	heptanone	3.543	1.274
s4	heptanone	4.322	1.320

Error Calculation

Same RMSE approach as anisole, but using total concentration.

RMSE in pool

Table 11: RMSE in pool (2-Heptanone)

Method	RMSE (ppb)
Baseline	3.747469
Proposed	2.269529

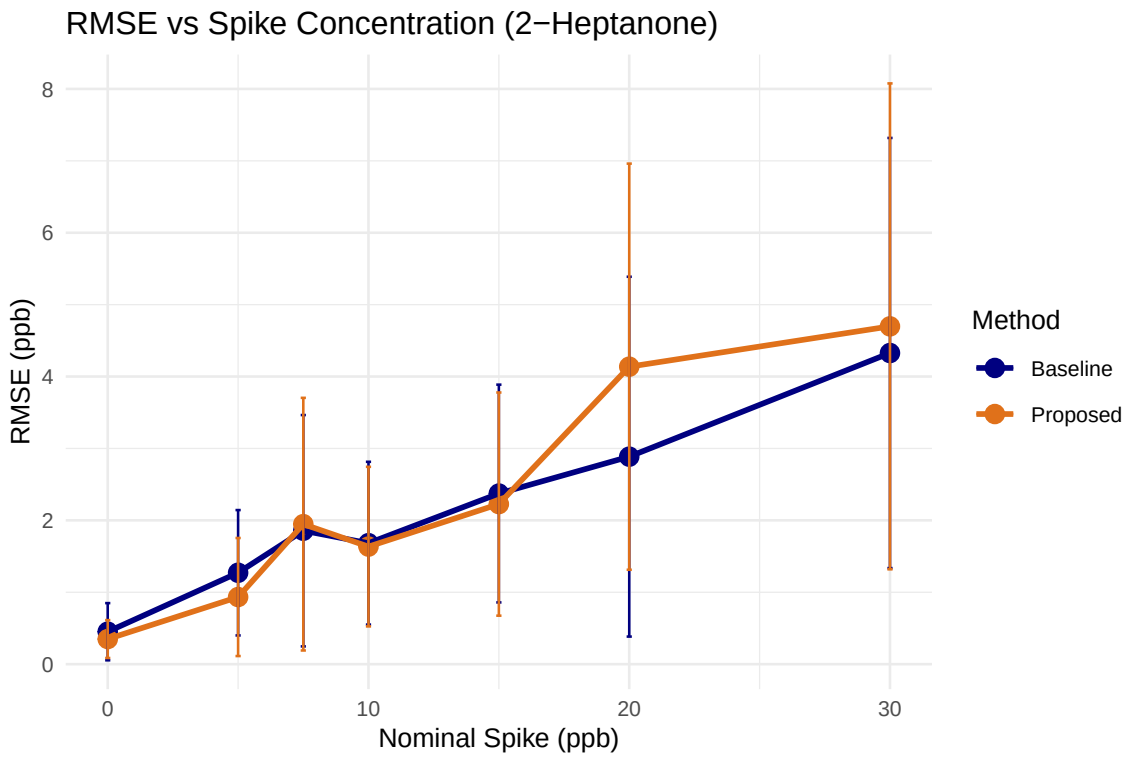
RMSE in Individual Samples

Table 12: Global RMSE (2-Heptanone)

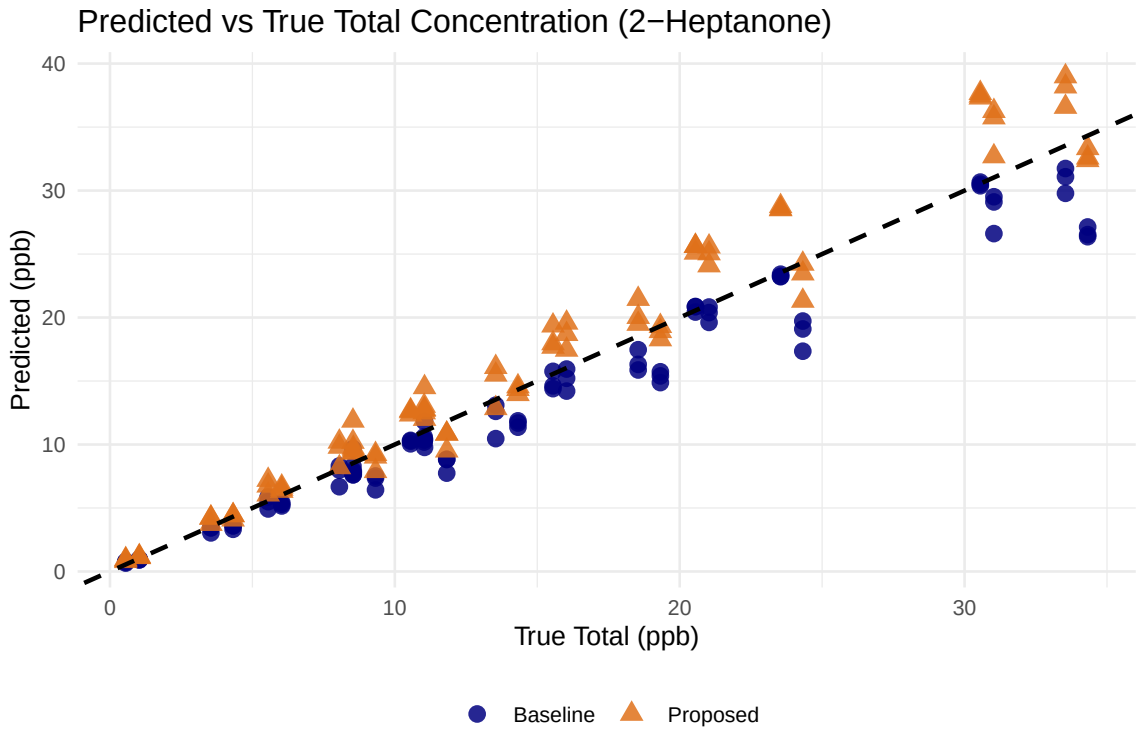
Method	RMSE (ppb)
Baseline	2.413547
Proposed	2.715028

RMSE vs Concentration Level

Grouped by spiked concentration



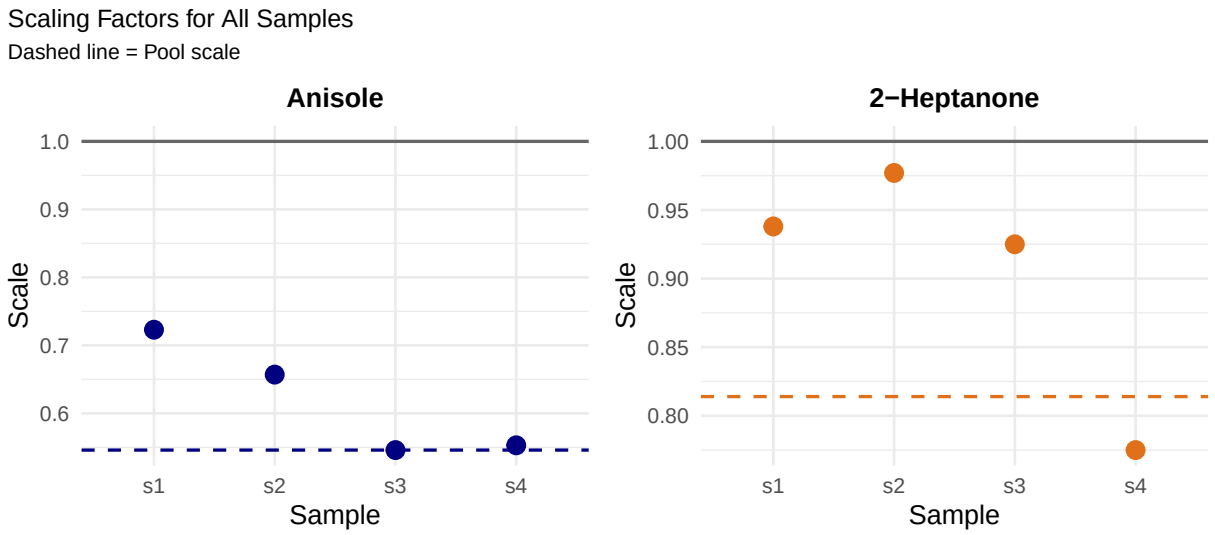
Predicted vs True (Total) Concentration



Discussion

The error is lower with our methodology in anisole but not in heptanone. This is probably due to the fact that the effect of the matrix is lower for 2-heptanone.

Since we have a whole calibration for each individual sample, we can estimate the scale wrt the synthetic urine.



For anisole, the individual samples deviate more from the pool reference, while for 2-heptanone the scaling factors stay much closer to 1 (synthetic urine), indicating that matrix effects are smaller and the harmonization may not be necessary or helpful in this case.