

Methodology and Error in Individual Urine Samples

Tecla Duran Fort

2025-12-11

Contents

Load Data	2
Calibration Curves	2
Reference Polynomial Adjustment to Enforce Origin Constraint	2
Polynomial Fit	2
Harmonization	4
Anisole	4
Final Calibration Curves	4
Heptanone	5
Final Calibration Curves	6
Extrapolation	8
Concentration Prediction	9
Prediction Without Scaling the Curve (Baseline)	9
Prediction Scaling the Curve (Proposed Method)	9
Error Calculation	10
Anisole (Non-Endogenous)	10
RMSE in Pool	10
RMSE in Individual Samples	10
RMSE vs Concentration Level	11
Predicted vs True Concentration	12
2-Heptanone (Endogenous)	12
Endogenous Concentration Estimation	12
Error Calculation	13
RMSE in pool	13
RMSE in Individual Samples	13
RMSE vs Concentration Level	13
Predicted vs True (Total) Concentration	14
Discussion	15

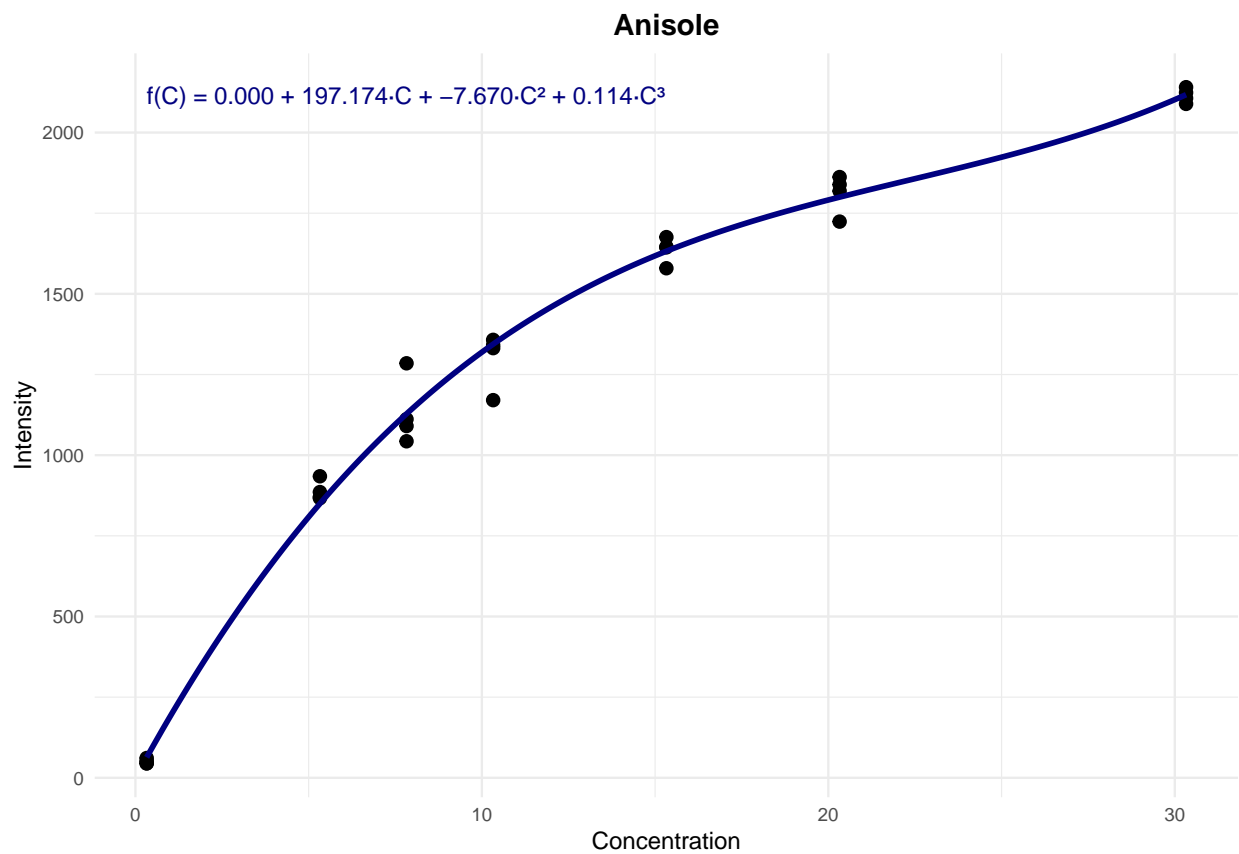
Load Data

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



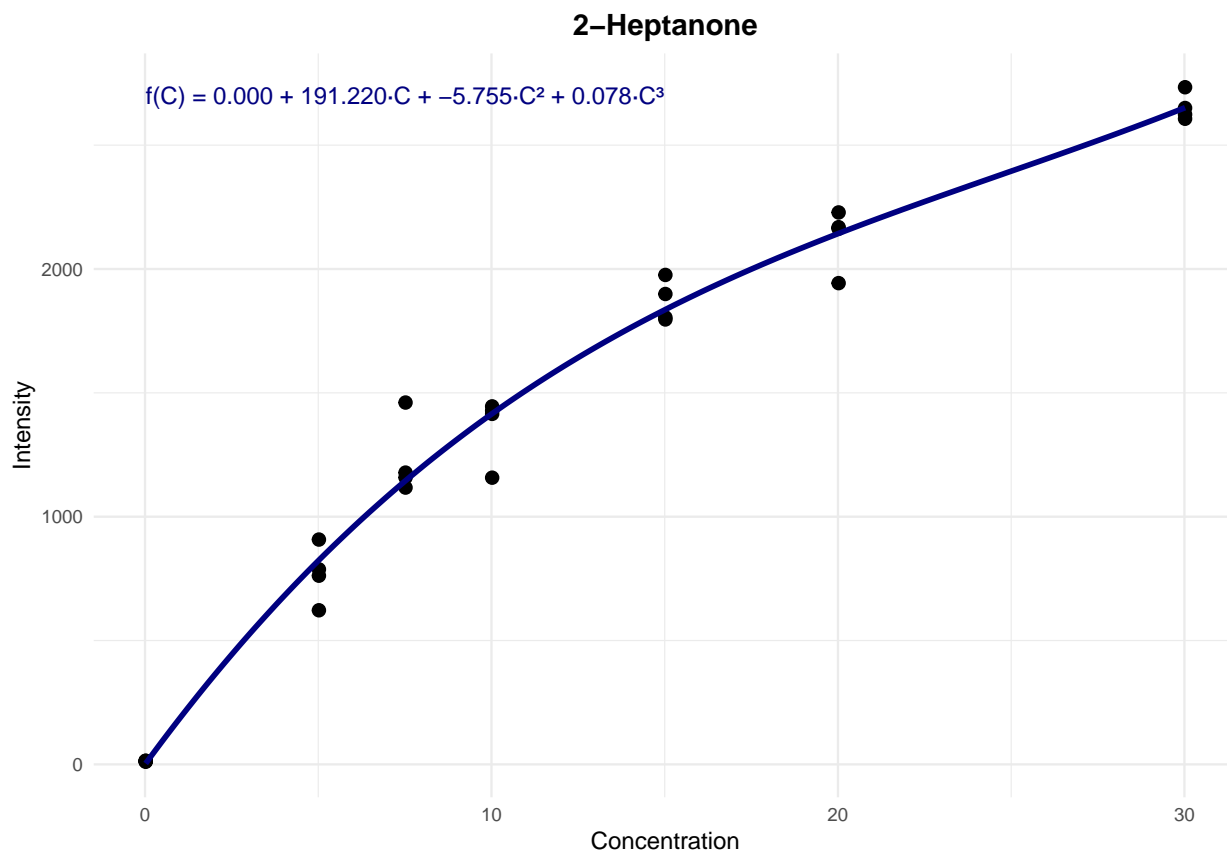


Table 1: Polynomial coefficients (degree 3, no intercept)

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	$f(C) = 191.220 \cdot C + -5.755 \cdot C^2 + 0.078 \cdot C^3$

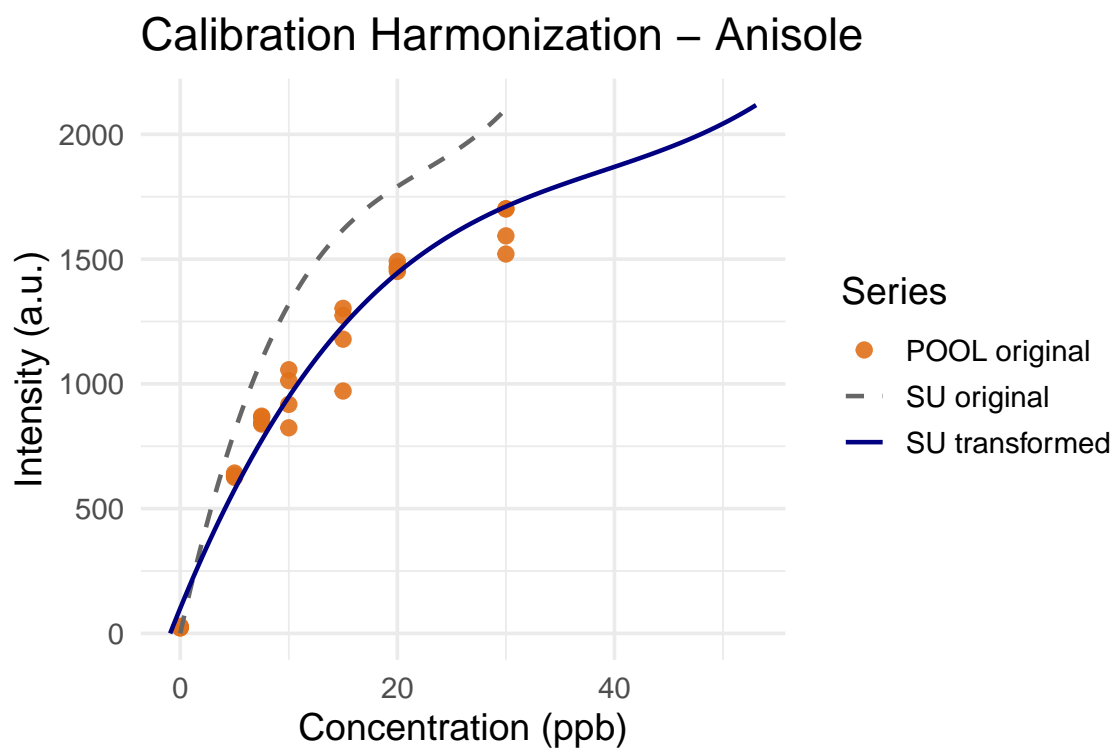
Harmonization

Anisole

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

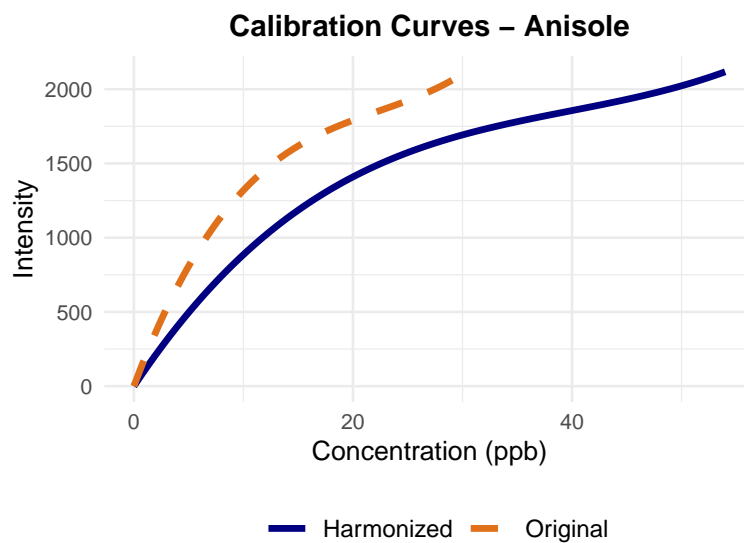
Table 4: Harmonization parameters and uncertainty (Anisole)

Parameter	Value	Error	Units
Scale	0.562	± 0.095	–
Shift	0.930	± 1.619	ppb



Final Calibration Curves

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



Heptanone

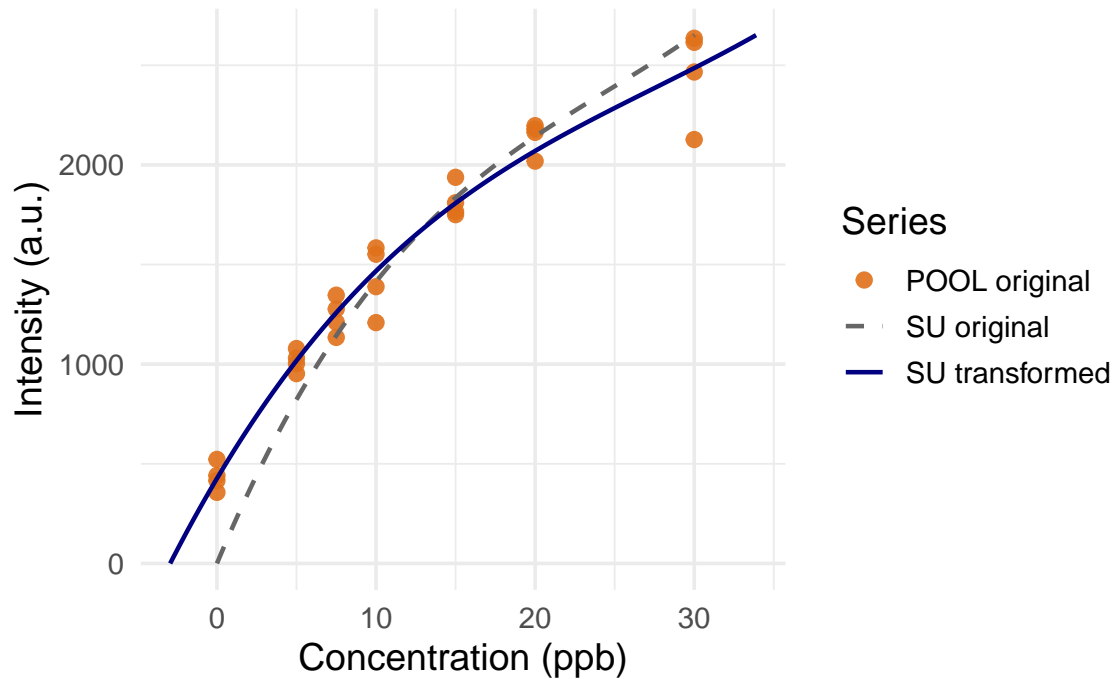
```
res_heptanone <- harmonize(pool_heptanone, su_heptanone)
```

Table 5: Harmonization parameters and uncertainty (Heptanone)

Parameter	Value	Error	Units
Scale	0.815	± 0.11	–
Shift	2.940	± 1.611	ppb

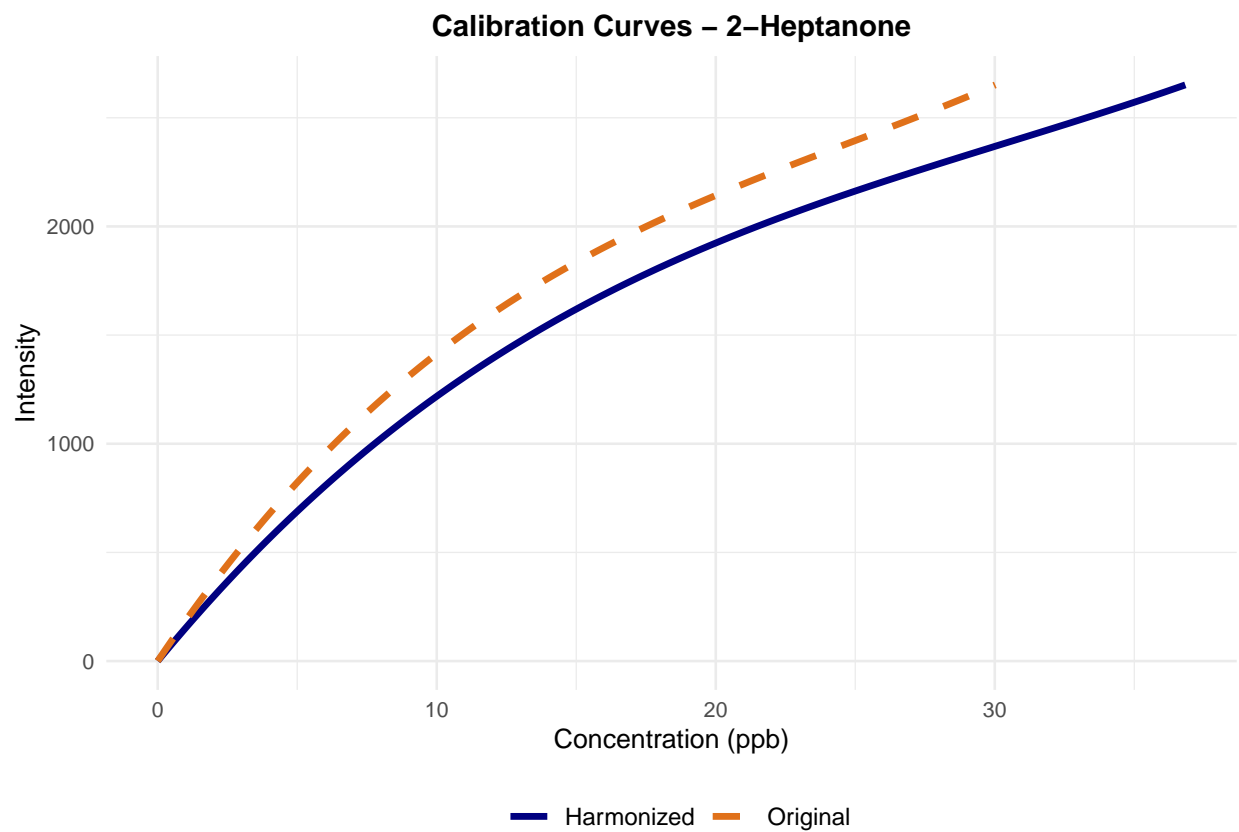
```
plot_harmonization(res_heptanone, title = "Calibration Harmonization - 2- Heptanone")
```

Calibration Harmonization – 2– Heptanone

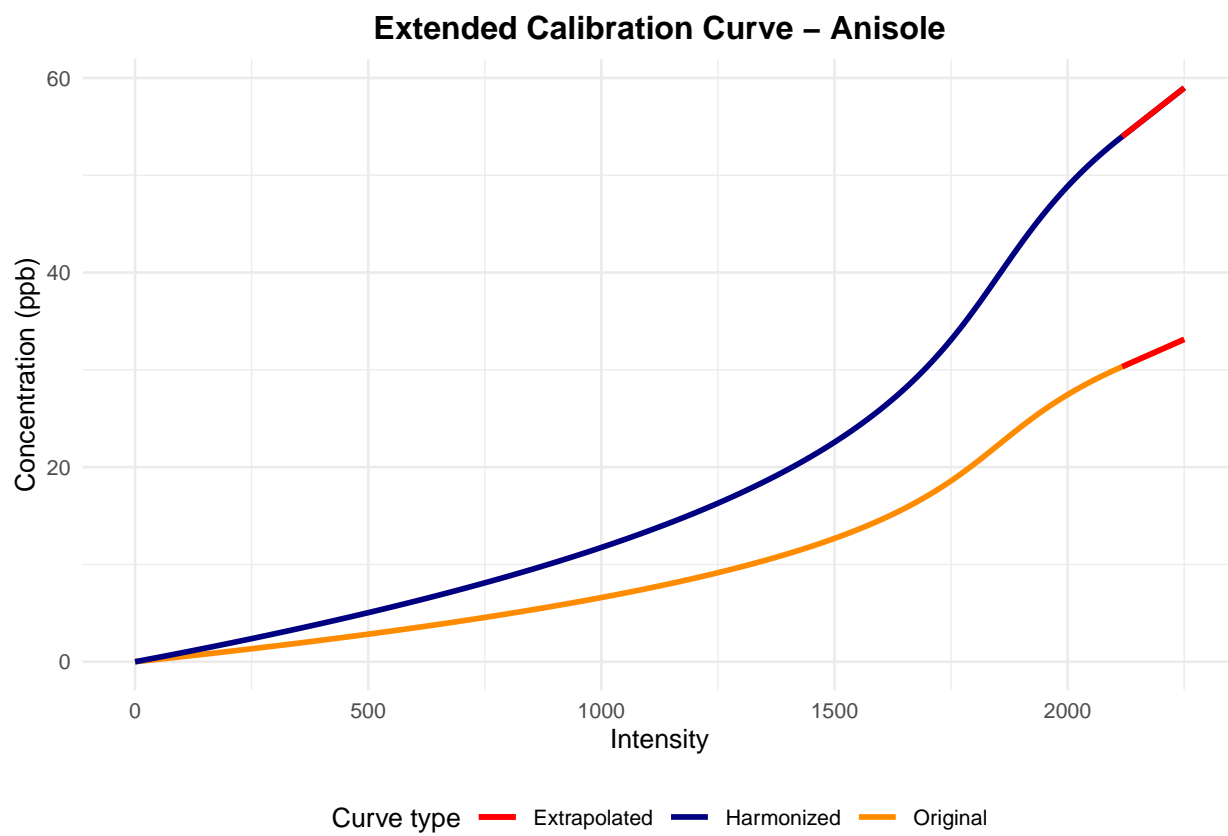


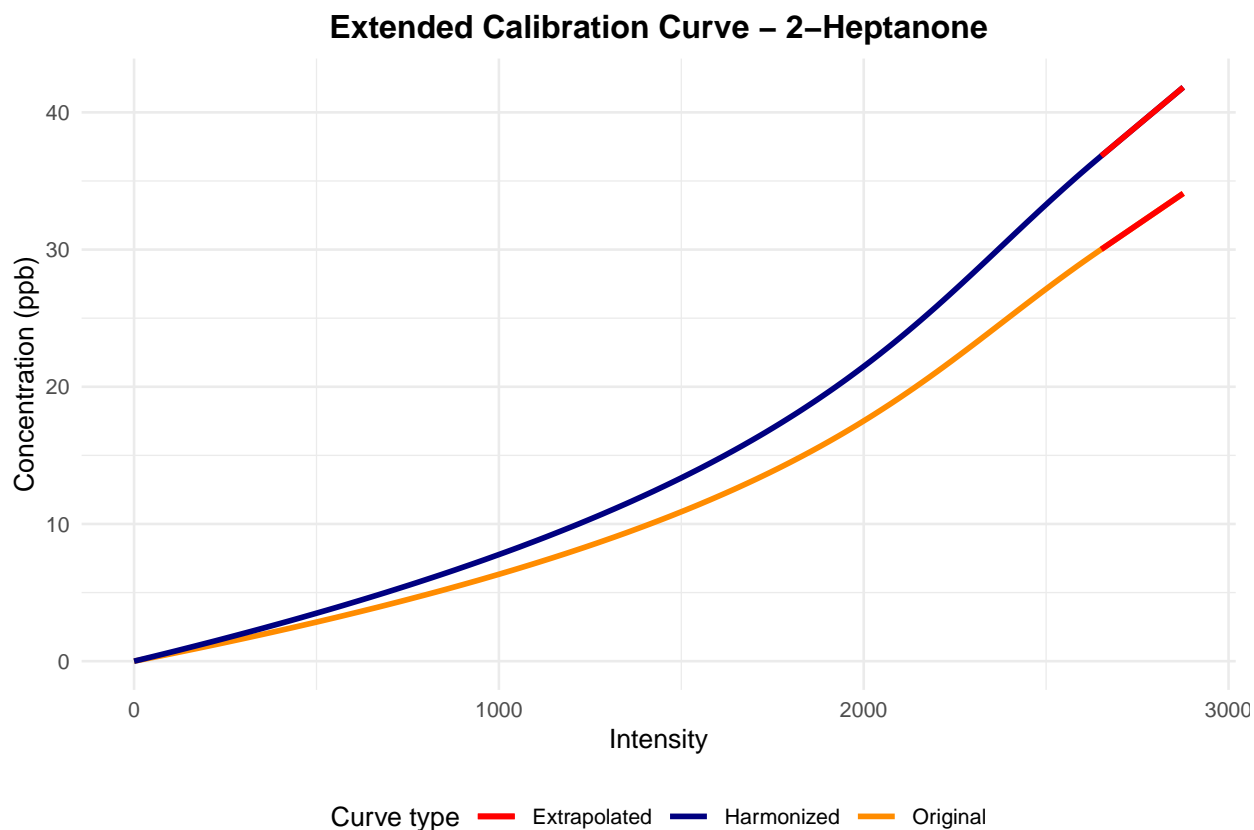
Final Calibration Curves

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



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	11.9554
pool	anisole	30	1593.1510	25.7311
pool	anisole	0	21.8578	0.1985

pool	aniso	20	1462.1706	21.4412
pool	aniso	5	642.6036	6.7372

Error Calculation

Aniso (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:

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

where

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

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

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

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 (aniso)

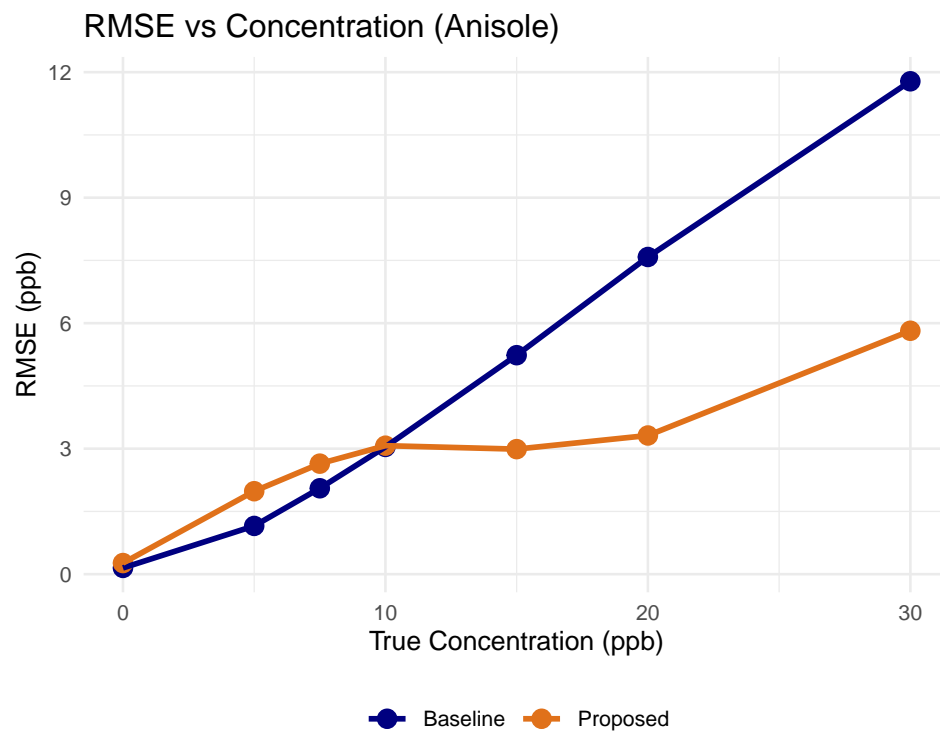
Method	RMSE (ppb)
Baseline	6.993950
Proposed	2.227297

RMSE in Individual Samples

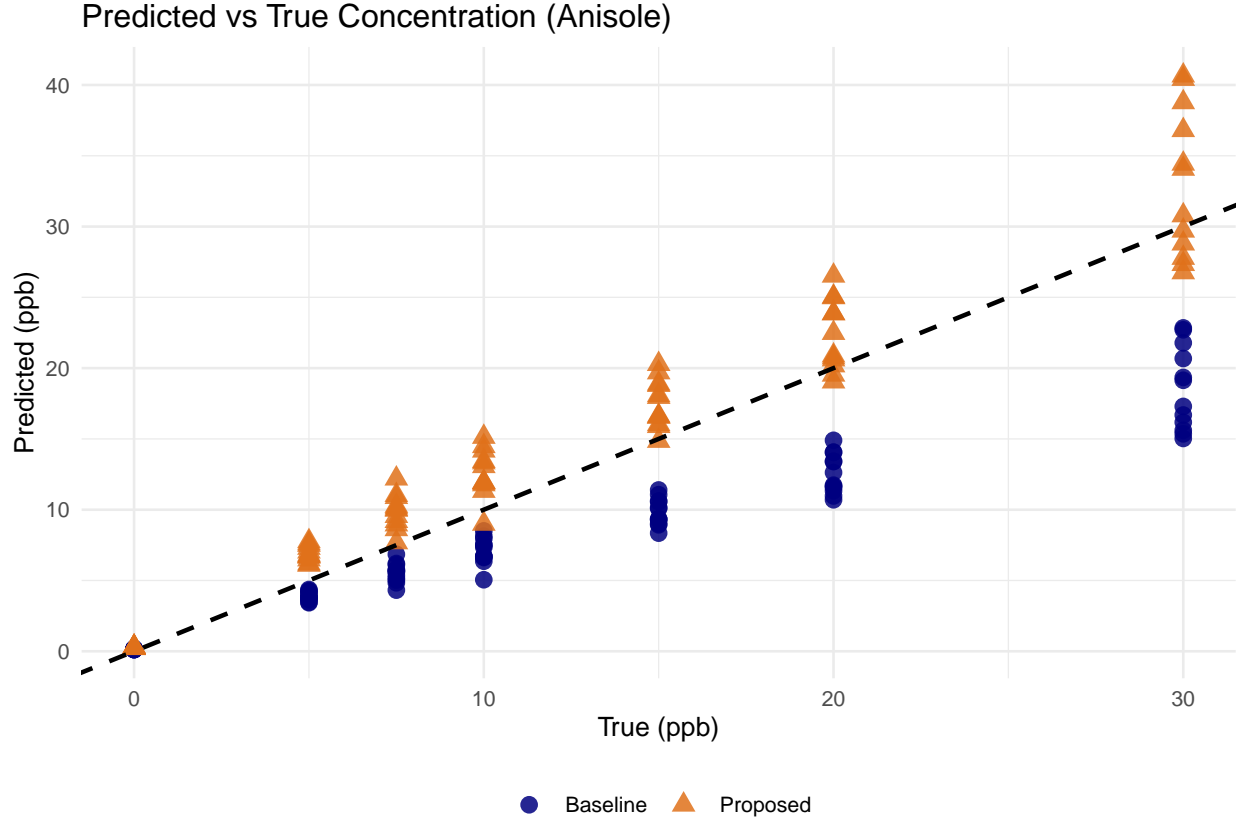
Table 9: Global RMSE (aniso)

Method	RMSE (ppb)
Baseline	5.837247
Proposed	3.255603

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	1.611
s1	heptanone	1.005	0.594
s2	heptanone	0.603	0.435
s3	heptanone	3.568	0.922
s4	heptanone	4.271	0.845

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	2.265698

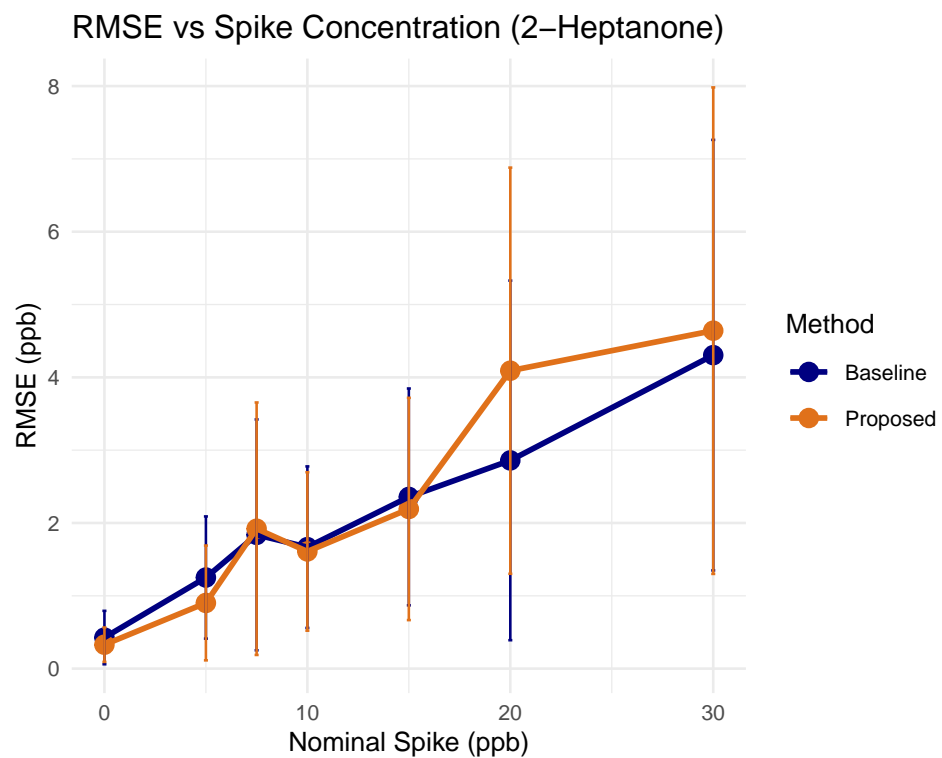
RMSE in Individual Samples

Table 12: Global RMSE (2-Heptanone)

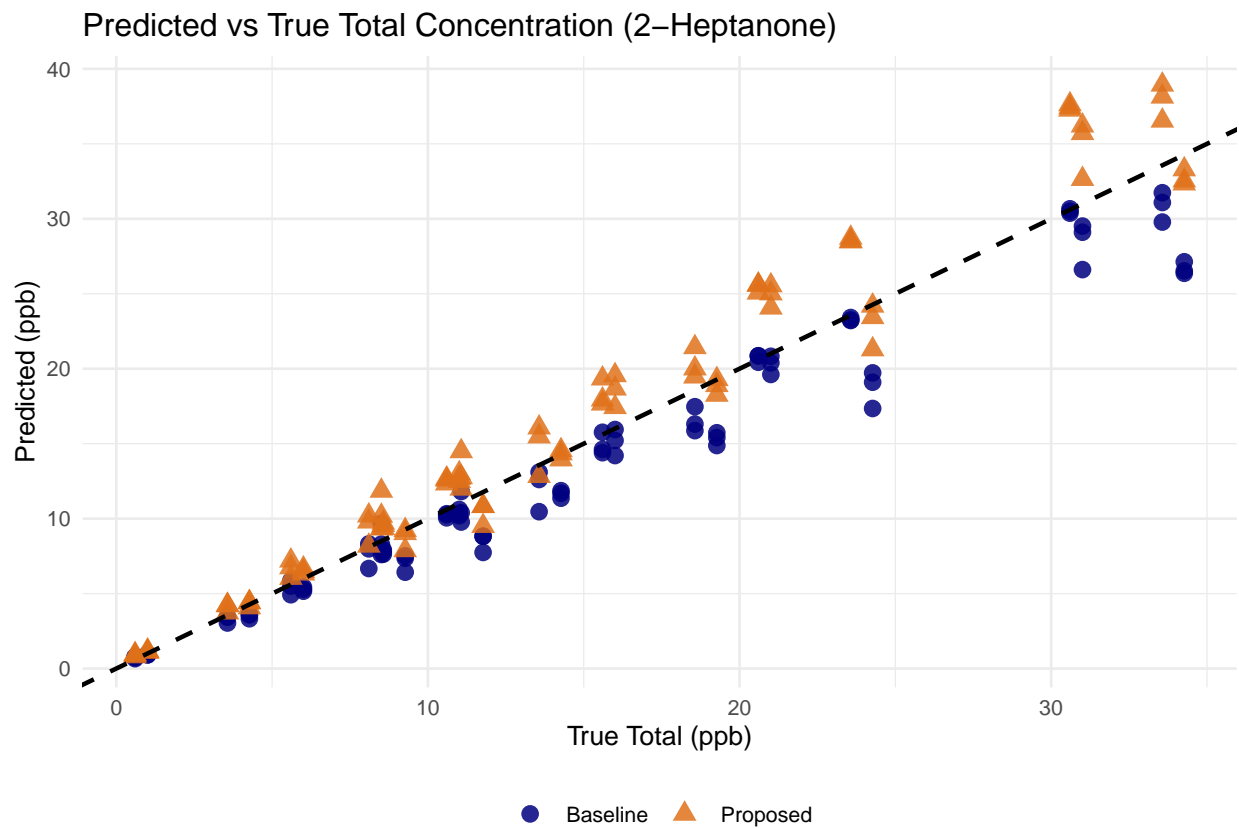
Method	RMSE (ppb)
Baseline	2.395582
Proposed	2.680237

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.

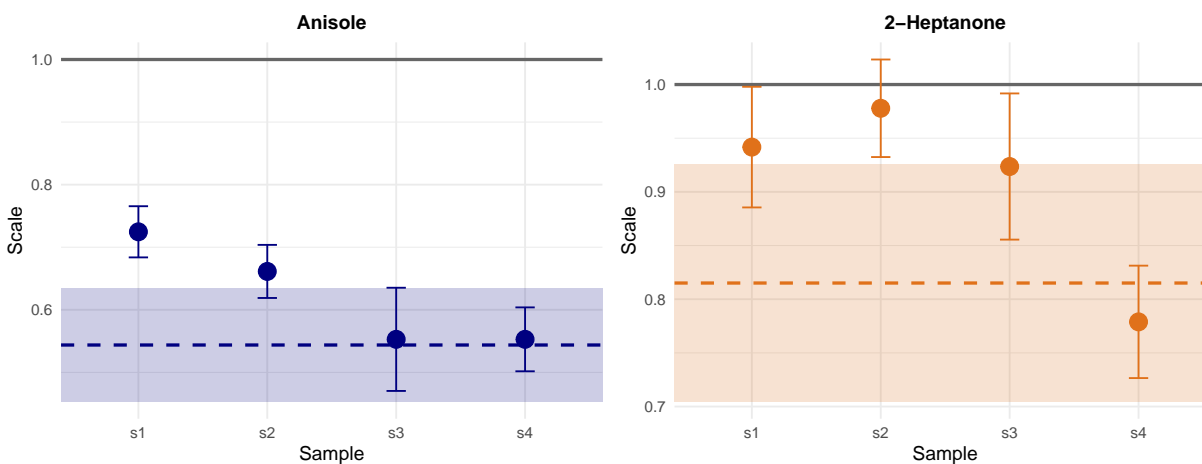
Since we have a whole calibration for each individual sample, we can estimate the scale wrt the synthetic urine.

Table 13: Estimated scaling factors with uncertainty

Sample	Analyte	Scale \pm Error
pool	anisole	0.544 ± 0.090
pool	heptanone	0.815 ± 0.110
s1	anisole	0.725 ± 0.041
s1	heptanone	0.942 ± 0.056
s2	anisole	0.661 ± 0.042
s2	heptanone	0.978 ± 0.045
s3	anisole	0.553 ± 0.082
s3	heptanone	0.924 ± 0.068
s4	anisole	0.553 ± 0.051
s4	heptanone	0.779 ± 0.052

Scaling Factors for All Samples

Dashed line = Pool scale



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.