NEXCHROM · TECHNICAL BACKGROUND

Validation of the EIV regression

How the errors-in-variables regression behind NexChrom's calibration was validated: internal-consistency checks and Monte Carlo simulation, with criteria and results.

Qualimetric · Nijmegen, The Netherlands · info@qualimetric.nl · nexchrom.nl · 3 November 2026

CONTENTS

1. Why errors-in-variables regression
2. The regression model
3. Validation approach and its limits
4. Test conditions and reproducibility
5. Checks and results
6. What the results mean for your calibrations
7. Conclusion
8. References
9. Further information

NexChrom calibrates a gas chromatograph against certified reference gases with an errors-in-variables (EIV) regression in conformity with ISO 6143: a Deming-type polynomial fit that accounts for the uncertainty of the certified concentrations as well as the uncertainty of the measured responses. Every mole fraction that NexChrom reports, and every uncertainty that goes with it, passes through this regression. This document summarises how the EIV regression routine of NexChrom was validated, which acceptance criteria were applied, what the results were, and where the limits of the validation lie. The validation consists of internal-consistency checks, Monte Carlo simulation and a comparison with the worked examples of ISO 6143:2001 Annex B. It is written for analysts and quality managers in gas laboratories working to ISO/IEC 17025 who need to judge whether the calculation can be relied upon.

Why errors-in-variables regression

Ordinary least-squares regression assumes that the independent variable is known exactly and that all scatter lies in the response. In gas analysis that assumption does not hold. The reference gases used for calibration carry a certified composition with a stated uncertainty, and for the components that matter most, such as the main components of natural gas, that uncertainty is often of the same order as the repeatability of the chromatograph. Ignoring it has two consequences:

- the calibration function is biased towards the points with the largest concentration uncertainty, because the fit treats them as exact;
- the uncertainty of the calibration function, and therefore of every analysed mole fraction, is understated, because one of the two sources of scatter is left out.

An errors-in-variables regression removes both problems. It adjusts both coordinates of every calibration point, each in proportion to its own uncertainty, and finds the calibration function that requires the smallest total adjustment. This is the approach of ISO 6143 for the comparison methods used to determine and check the composition of gas mixtures, and it builds on the work of Deming and of Bremser and Hässelbarth on uncertainty in calibration.

The price of this approach is numerical: unlike ordinary least squares, an EIV fit has no closed-form solution and must be found iteratively. That makes an independent validation of the implementation essential. A routine that

returns plausible-looking coefficients is not necessarily a routine that has found the correct minimum, reports the correct covariance, or stops for the right reason.

The regression model

Calibration function

The EIV regression routine of NexChrom fits polynomial calibration functions of order 1 to 3, and a straight line through the origin:

$$F(x, b) = b_0 + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 \qquad \text{(orders 1 to 3)}$$
$$F(x, b) = b \cdot x \qquad \text{(slope only, no intercept)}$$

Here x and y are the two coordinates of a calibration point, each with its own uncertainty. In NexChrom x is the measured response and y the certified content: the orientation of the analysis function, which ISO 6143 recommends and which its worked examples use, so that a sample result follows from the curve without an inversion step. The method treats the two variables symmetrically, as shown below.

Estimator

Each calibration point i carries a value x_i with standard uncertainty $u(x_i)$ and a value y_i with standard uncertainty $u(y_i)$. The routine finds adjusted values ξ_i for the x coordinates and coefficients b that minimise the weighted sum of squared deviations

$$S = \Sigma_i \left[\frac{(x_i - \xi_i)^2}{u^2(x_i)} + \frac{(y_i - F(\xi_i, b))^2}{u^2(y_i)} \right]$$

so that every adjusted point $(\xi_i, F(\xi_i, b))$ lies exactly on the calibration curve. The minimisation is carried out by a Gauss-Newton iteration on the Lagrangian of this constrained least-squares problem, with an independent uncertainty per point in each variable.

Reported statistics

For every fit the routine reports:

- the coefficients b and their covariance matrix, computed from the stated uncertainties;
- the standard uncertainty of each coefficient, with t and p values;
- the minimum Smin of the weighted sum of squares, the degrees of freedom $\nu = n - p$ (n calibration points, p coefficients), the Birge ratio BR and the probability of Smin under the χ^2 distribution with ν degrees of freedom;
- the goodness of fit Γ of ISO 6143: the largest weighted deviation of a calibration point from its adjusted point, in either coordinate, together with the signed weighted deviation of every point;
- whether the iteration converged and the reason it stopped: on its convergence tolerance, on its divergence guard, or on the maximum number of iterations.

The covariance matrix, the Birge ratio and Γ are defined as

$$\mathrm{Cov}(b) = (J^T \cdot W \cdot J)^{-1}$$
$$BR = \sqrt{(Smin / \nu)}$$
$$\Gamma = \max \text{ over all points } i \text{ of } \left|\frac{x_i - \xi_i}{u(x_i)}\right| \text{ and } \left|\frac{y_i - F(\xi_i, b)}{u(y_i)}\right|$$

where J is the Jacobian of the calibration function with respect to the coefficients at the solution and W the corresponding weight matrix.

The Birge ratio, the root of the reduced χ^2, describes the fit as a whole. Under a correctly specified uncertainty model, Smin follows a χ^2 distribution with ν degrees of freedom, so BR^2 has an expectation of 1. In earlier versions of this document the Birge ratio was called the goodness of fit (GOF). Γ examines every calibration point separately and is the goodness-of-fit criterion of ISO 6143: a calibration function is admissible when $\Gamma \leq 2$, which means that the curve passes through every calibration rectangle $x \pm 2u(x)$, $y \pm 2u(y)$. How NexChrom uses the two statistics to accept a fit and to choose its order is described in the companion document *Calculation and uncertainty*.

Validation approach and its limits

Internal consistency, Monte Carlo simulation and the worked examples of the standard

The validation tests the routine against properties that any correct implementation of the estimator must have, against simulated data for which the true answer is known, and against the numbers that ISO 6143:2001 prints for its own worked examples:

- **exact recovery**: data lying exactly on a known polynomial must return that polynomial;
- **a known limiting case**: when the x-uncertainty tends to zero, the EIV estimator must reduce to ordinary weighted least squares, both in its coefficients and in its covariance;
- **symmetry**: exchanging the roles of x and y, with their uncertainties, must lead to the same minimum;
- **Monte Carlo coverage**: over many simulated datasets with known noise, the scatter of the estimated coefficients must match the covariance the routine reports, the estimates must be unbiased and BR^2 must average 1;
- **Monte Carlo sensitivity**: when the true noise is larger than the stated uncertainty, the Birge ratio must rise accordingly;
- **worked examples of the standard**: on the data of the worked examples of ISO 6143:2001 Annex B, the routine must reproduce the weighted sum of squares and the goodness of fit Γ that the standard prints.

All reference values are computed independently of the routine under test, from the normal equations of weighted least squares, from analytical identities or from the empirical statistics of the simulation, or they are taken from the standard. The routine is never used to produce its own reference.

What the validation does not cover

The validation does not compare the routine with another implementation of the same estimator on a shared dataset. The correctness of the estimator rests on the properties above and on the agreement with the numbers of the standard. The comparison with ISO 6143:2001 Annex B covers the straight-line fits of the worked examples; the exponential and power functions of the third example lie outside the routine, which fits polynomials, and the second example is compared as information only, because the standard prints its data to four significant figures.

The validation concerns the numerical behaviour of the regression routine itself. How NexChrom derives the uncertainties $u(x_i)$ and $u(y_i)$ from certificates and replicate injections, and how it uses Γ to accept a model and choose its order, is described in the companion document *Calculation and uncertainty* and is outside the scope of this document.

Review and current version

During the validation the routine was reviewed in detail, including its convergence control, and refined where the review showed room for improvement. It was later extended with the goodness of fit Γ of ISO 6143, the χ^2 probability and the weighted deviations per point; the estimator itself is unchanged, and the earlier checks

reproduce their results exactly. All results in this document were obtained with the current version of the routine, which is the version used in NexChrom.

Test conditions and reproducibility

Item	Value
Checks	Six groups, 19 recorded acceptance checks
Monte Carlo size	2000 simulated datasets per Monte Carlo check
Random numbers	Fixed seed, with a fixed offset per check
Reference data	Computed independently inside the validation run, or taken from ISO 6143:2001 Annex B as printed
Computing environment	Python with NumPy and SciPy; the library versions are recorded with every run
Outcome reporting	One pass or fail line per check, with the actual numbers; the run succeeds only if every check passes

All randomness in the validation derives from a fixed random seed. With the same library versions the run is bit-reproducible: repeating it yields exactly the same numbers as reported here. This makes the validation repeatable after any future change to the routine or its computing environment.

Checks and results

Summary

Check	What it demonstrates	Checks	Result
Exact recovery on noise-free data	The iteration finds the exact solution, for every model order	4	4 pass
Limiting case: weighted least squares	Coefficients and covariance are correct where an independent closed-form reference exists	2	2 pass
Role-swap symmetry	Both error directions are treated symmetrically; the global minimum is found	3	3 pass
Monte Carlo coverage	The reported covariance describes the true scatter; the estimates are unbiased; BR^2 averages 1	3	3 pass
Monte Carlo sensitivity of the Birge ratio	The Birge ratio detects understated uncertainties	1	1 pass
Worked examples of ISO 6143:2001 Annex B	The weighted sum of squares and Γ agree with the numbers printed in the standard	6	6 pass
Total		**19**	**19 pass, 0 fail**

Exact recovery on noise-free data

Set-up. Twelve points lying exactly on a known monotonic polynomial, with x between 0.5 and 10, u(x) = 0.01 and u(y) = 0.05, for the slope-only model and for polynomial orders 1, 2 and 3.

Acceptance criterion. The coefficients are recovered to within 10^{-8} relative, and the iteration stops on its own convergence tolerance, not on the divergence guard or the iteration limit.

Result.

Model	Largest relative coefficient error	Smin	Iterations	Stopped on	Verdict
Slope only	0	0	1	tolerance	pass
Order 1	0	3.2×10^{-28}	1	tolerance	pass
Order 2	2.1×10^{-15}	7.7×10^{-27}	3	tolerance	pass
Order 3	1.3×10^{-15}	2.7×10^{-27}	4	tolerance	pass

The coefficients are recovered to machine precision for every model, in at most four iterations. As additional information, outside the formal checks, the same test was run on a quadratic that is not monotonic over the fitted range (true coefficients 0.7, −2.1 and 0.35), a case outside the intended domain of a calibration function. The routine also converged there, in five iterations, to coefficients within 5.3×10^{-14} standard uncertainties of the truth.

Limiting case: weighted least squares

Set-up. A noisy quadratic dataset of 15 points with a y-uncertainty that grows with x, u(y) = 0.2 + 0.02·x, and a negligible x-uncertainty, u(x) = 10^{-12}. In this limit the EIV estimator must reduce to weighted least squares with weights $1/u^2(y_i)$. The reference, coefficients and covariance matrix, was computed independently from the normal equations.

Acceptance criterion. Coefficients agree to within 10^{-8} relative; every element of the covariance matrix agrees to within 10^{-6} relative.

Result.

Quantity	EIV regression	Weighted least squares	Largest relative difference	Verdict
b_0	1.7692685149	1.7692685149		
b_1	1.5821644407	1.5821644407		
b_2	−0.0345398135	−0.0345398135		
Coefficients			5.8×10^{-15}	pass
Covariance matrix			0 (identical)	pass

This check is important because it compares not only the coefficients but also the covariance matrix against an independent, closed-form reference. The covariance is what carries the calibration uncertainty into the analysed mole fractions.

Role-swap symmetry

Set-up. A noisy straight-line dataset of 12 points with u(x) = 0.30 and u(y) = 0.50, so that both error directions contribute substantially. The forward fit y = f(x) uses the uncertainties as given; the reverse fit x = g(y) swaps the data and the uncertainties. Because the EIV objective treats both variables alike, both fits must find the same minimum:

```
b₁(reverse) = 1 / b₁(forward)
b₀(reverse) = −b₀(forward) / b₁(forward)
Smin(reverse) = Smin(forward)
```

Acceptance criterion. Slopes and intercepts consistent to within 10^{-6} relative; minimum values identical to within 10^{-8} relative.

Result.

Quantity	Forward fit	Reverse fit	Relative difference	Verdict
Slope	$1 / b_1 = 0.5586261986$ ($b_1 = 1.7901058034$)	0.5586261986	4.9×10^{-13}	pass
Intercept	$-b_0 / b_1 = -2.5701357749$	−2.5701357749	3.5×10^{-12}	pass
Smin	13.099809593425	13.099809593425	8.1×10^{-16}	pass

Both fits stop on their convergence tolerance, in seven iterations. The symmetry distinguishes a genuine errors-in-variables fit from ordinary regression, where regressing y on x and x on y give different lines.

Monte Carlo coverage of the reported covariance

Set-up. A fixed true quadratic, b = (1.0, 2.0, 0.05), with 15 points and Gaussian noise of known size, u(x) = 0.05 and u(y) = 0.30. Two thousand datasets were simulated and each was fitted with the correct stated uncertainties (ν = 12).

Acceptance criteria.

- For every coefficient, the ratio of the empirical variance over the 2000 fits to the mean variance reported by the routine lies between 0.90 and 1.10. The Monte Carlo standard uncertainty of such a ratio at N = 2000 is about √(2/2000) ≈ 3.2 %, so this tolerance leaves roughly three standard deviations of headroom.
- No coefficient shows a bias larger than four Monte Carlo standard deviations of the mean.
- The mean of BR^2 lies within 1.00 ± 0.05. BR^2 is the unbiased quantity; BR itself has a finite-sample expectation slightly below 1, about 0.979 for ν = 12.

Result.

Quantity	b_0	b_1	b_2	Criterion	Verdict
Empirical / reported variance	1.022	1.010	1.016	0.90 to 1.10	pass
Mean estimate minus true value	1.1×10^{-3}	-3.0×10^{-4}	1.4×10^{-5}	within 4σ of the mean	pass
4σ limit of the mean	0.026	0.008	$< 5 \times 10^{-4}$		

Quantity	Result	Expected	Verdict
Mean of BR^2	1.007	1.00 ± 0.05	pass
Mean of BR	0.983	about 0.979 for ν = 12	(information)
Fits not converged	0 of 2000	0	

The covariance matrix reported by the routine agrees with the true sampling spread of the coefficients to within 1 to 2 %, which is inside the Monte Carlo resolution. The estimates are unbiased within that resolution, and the Birge ratio behaves as a χ^2-based statistic with expectation 1 when the uncertainty model is correct. All 2000 fits stopped on the convergence tolerance.

Monte Carlo sensitivity of the Birge ratio

Set-up. The same simulation as above, but the actual noise is twice the stated uncertainty in both variables, while the fit is still given the original values of u(x) and u(y).

Acceptance criterion. The mean of BR lies between 1.85 and 2.05 (expected about 1.96 for ν = 12), and the mean of BR^2 within 4.00 ± 0.40.

Result.

Quantity	Result	Expected	Verdict
Mean of BR	1.965	about 1.96 (1.85 to 2.05)	pass
Mean of BR^2	4.02	4.00 ± 0.40	pass
Fits not converged	0 of 2000	0	

The Birge ratio scales in direct proportion to the ratio of true to stated noise. This confirms that the statistic does what is required of an overall measure of the fit.

Worked examples of ISO 6143:2001 Annex B

Set-up. Annex B of ISO 6143:2001 works three examples of the comparison method, with the content as a function of the response, the orientation NexChrom uses. The data were taken as printed in the standard. The routine was run on the straight-line fits of the first example (three calibration points) and the third example (twelve points), and its weighted sum of squares Smin and goodness of fit Γ were compared with the values the standard prints for them (Sres and Γ in the notation of the standard).

Acceptance criteria. For the first example, which prints Sres to four and Γ to three significant figures, the difference must stay within half a unit of the last printed digit. For the third example, printed to more digits, Smin must agree within 10^{-6} relative and Γ within 2×10^{-4} relative. In both examples Γ must equal the largest absolute weighted deviation that the routine reports for the individual points.

Result.

Quantity	EIV regression	ISO 6143:2001	Difference	Verdict
Example 1, Smin	0.6743049	0.6743	4.9×10^{-6}	pass
Example 1, Γ	0.567950	0.568	5.0×10^{-5}	pass
Example 3, Smin	272.6391499	272.6392	1.8×10^{-7} relative	pass
Example 3, Γ	6.836153	6.8352	1.4×10^{-4} relative	pass
Γ equals the largest weighted deviation of a point, examples 1 and 3	identical in both			2 pass

As information, outside the formal checks: for the first example the coefficients agree with the standard to five digits (b_0 −0.357468 against −0.35747, b_1 24.6115 against 24.612) and their uncertainties to three or four ($u(b_0)$ 0.157131 against 0.15716, $u(b_1)$ 0.480355 against 0.48048). The straight line of the third example is, as in the standard, not admissible (Γ > 2); the standard then turns to exponential and power functions, which lie outside the routine. For the second example, a straight line and a second-order polynomial, the routine gives Smin 6.04 and 1.40 against the printed 6.17 and 1.47, and Γ 1.627 and 0.866 against 1.632 and 0.868, while the coefficient uncertainties agree to four digits; the standard prints the data of this example to four significant figures, which most likely explains the difference, so the example is recorded as information and not as a pass or fail check.

The routine therefore reproduces, on the standard's own data, the two quantities that ISO 6143 derives from a comparison: the overall measure Smin and the acceptance statistic Γ.

What the results mean for your calibrations

The reported uncertainties can be trusted, provided the inputs are right

The limiting-case and Monte Carlo checks show that the covariance matrix of the calibration coefficients is computed correctly and describes the real spread of the coefficients. That covariance is the route by which the calibration contributes to the uncertainty of every analysed mole fraction.

The covariance is built from the uncertainties that are stated for the calibration points. It is deliberately not rescaled by the observed scatter. If the stated uncertainties are too small, the reported covariance is too small as well. The safeguard against that situation is the fit test.

What the two fit statistics tell you

The sensitivity check shows how to read the Birge ratio. A Birge ratio of 1 means that the scatter of the calibration points around the curve is as large as the stated uncertainties predict; a value of 2 means twice as large. The statistic bounds the extent to which uncertainties may be understated; it does not by itself prove that the chosen polynomial has the right shape. A value approaching 2 is a signal to review the uncertainty budget of the reference gases or the repeatability of the instrument.

Γ answers a different question: how far the worst calibration point lies from the curve, in standard uncertainties. The comparison with the worked examples shows that the routine computes it as the standard does. ISO 6143 accepts a calibration function when $\Gamma \leq 2$ and, of several admissible functions, prefers the one with the fewest coefficients. The two statistics do not map one to one: Γ takes the worst point, the Birge ratio averages over all points, so a fit can pass one of them and fail the other.

NexChrom tests every fit on Γ against the fixed limit 2 of the standard, which ISO 6143:2001 names as the definitive acceptance test. The Birge ratio decides nothing: the software shows it for information, with its χ^2 test. The companion document *Calculation and uncertainty* describes both.

Convergence is reported, not assumed

Every fit reports whether the iteration converged and why it stopped. In all checks of this validation, including the 4000 Monte Carlo fits, the iteration stopped on its convergence tolerance. In supplementary tests with data deliberately constructed far outside the intended domain, such as exponential or sigmoidal data forced into a cubic with a dominant x-uncertainty, the routine stops and reports that it has not converged, with a warning, rather than returning a result that appears valid.

Conclusion

All 19 acceptance checks passed and none failed. The checks against an independent reference (the weighted least-squares limit, for coefficients and covariance), the symmetry check, the Monte Carlo checks (covariance coverage, unbiasedness and the sensitivity of the Birge ratio) and the comparison with the worked examples of ISO 6143:2001 Annex B (weighted sum of squares and Γ) together cover the defining properties of the estimator and its agreement with the standard.

Within the limits stated in this document, the EIV regression routine of NexChrom is validated as a correct errors-in-variables estimator for monotonic polynomial calibration functions of order 1 to 3 and for the straight line through the origin, with trustworthy reported coefficients, covariance and Birge ratio, and with a weighted sum of squares and goodness of fit Γ that agree with the worked examples of ISO 6143:2001. The validation is fully reproducible from a fixed random seed and the printed data of the standard, and can be repeated after any change to the routine.

References

- Deming, W.E. (1964). *Statistical Adjustment of Data*. Wiley, New York.
- Bremser, W. and Hässelbarth, W. (1997). Controlling uncertainty in calibration. *Analytica Chimica Acta* 348, 61–69.
- ISO 6143:2001, *Gas analysis — Comparison methods for determining and checking the composition of calibration gas mixtures*.
- JCGM 100:2008, *Evaluation of measurement data — Guide to the expression of uncertainty in measurement* (GUM).
- ISO/IEC 17025, *General requirements for the competence of testing and calibration laboratories*.

Further information

The full validation documentation, including the complete output of the validation run, is available to customers with the software, for use in their own software validation and audits.

NexChrom is in development; the release is expected in early 2027. For questions about the EIV regression or its validation, or to register your interest: info@qualimetric.nl · nexchrom.nl