CEN/TC 19 2011 Conference

Dealing with test method precision in the daily lab & fuel specification work (EN ISO 4259)

Dr. H. Thomas Feuerhelm,
DIN-FAM, Hamburg (Germany)
(fam@fam-hamburg.de)

CEN, Avenue Marnix 17
1000 Brussels, Belgium
http://www.cen.eu
info@cen.eu

CEN/TC 19, Vlinderweg 6
2623 AX Delft, the Netherlands
energy@nen.nl

Some basic Thoughts

Why do we need specifications ?

→ because we want the same good product quality every day ..

How can product quality be defined ?

→ by a series of characteristics from an applicational model :

.. Product performance criteria
transformed into a list of product properties which can be measured with well defined analytical test methods

.. Limit setting (min, max) for acceptable analytical test results

Requirements for good analytical test methods:

→ must discriminate between “good” and “bad”;

→ standardized, economical and well distributed in the field;

→ suitable measurement range and as accurate and precise as possible;

→ ..., ..., ..., ...,

Some very basic Precision Concepts

EN ISO 4259

→ defines ALL basic issues for round robin tests, precision, quality testing
→ very similar to **ISO 5725**, but much better adapted for petroleum products
→ also gives valuable advice for cases of dispute

Accuracy ←→ True Value

The True Value is, for practical purposes, the value towards which the average of single results obtained by *n* laboratories tends, as *n* tends towards infinity (a theoretical value).

→ most often replaced by "Accepted Reference Value – **ARV** "
→ as opposed to (single) test method results → "Measured Values – **MV** "

Definition of Repeatability " r "

Closeness of agreement between independent results obtained in the normal and correct operation of the same method on identical test material, in a short interval of time, and under the same test conditions (**same operator, same apparatus, same laboratory**) .

Definition of Reproducibility " R "

Closeness of agreement between individual results obtained in the normal and correct operation of the same method on identical test material but under different test conditions **(different operators, different apparatus and different laboratories).**

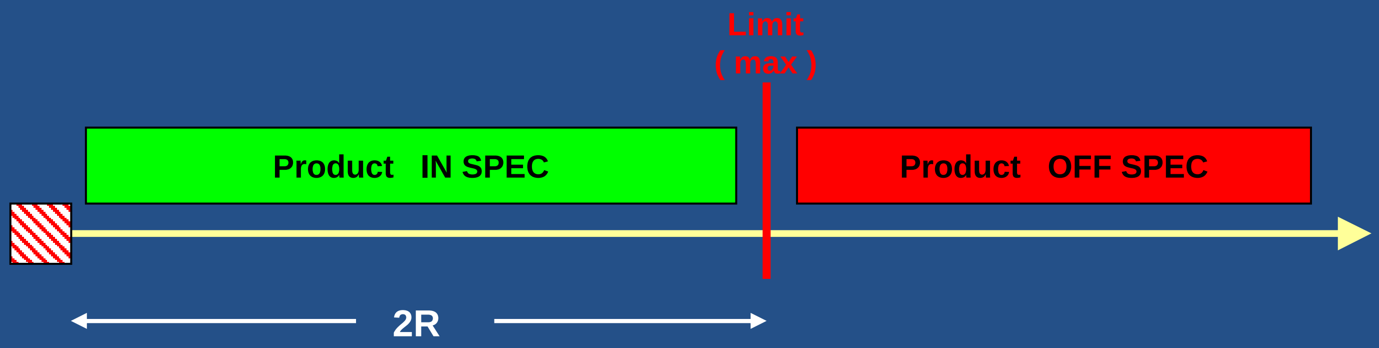
Limit Setting in Specifications (example : MAX)

Closely related "Chain of Custody":

→ Property → Test Method → Precision → Limit
→ Testing is done at the Point of Sale (not only inside e.g. production)
→ Test results may be challenged by customer via his own product analysis

Limit setting Procedure from EN ISO 4259 (MAX – case)

→ (max) Limits shall not be lower than "2R" → 2R rule (see below)
→ Limit is given as True Value / Accepted Reference Value – **ARV**
while test results are expressed as a single Measured Value – **MV**

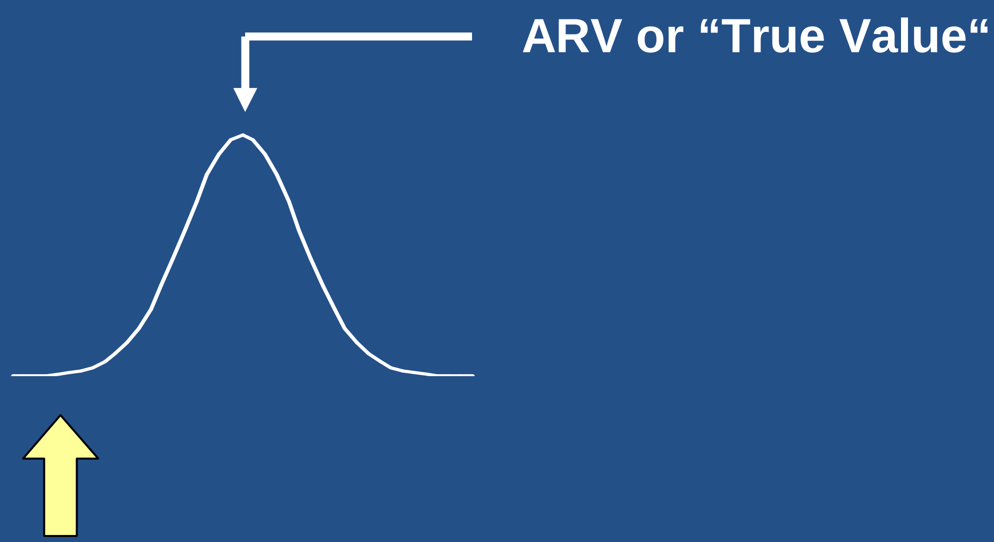


Limit Setting in Specifications (example : MAX)

Respecting (single) Measurement Uncertainty :

→ Two parties (producer & customer) work in a Reproducibility scenario
→ Limit (ARV, multiple measurements) ←→ single test method result
This relation requests special attention:

A single measurement
has some uncertainty :



→ it is analytically incorrect to assume that only one single measurement will “hit” the True Value !!!

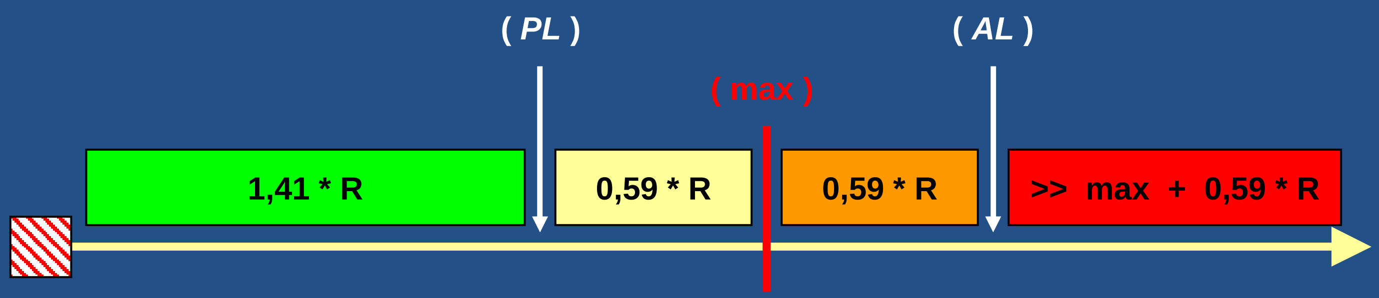
→ Therefore, EN ISO 4259 introduces some safety margin (0,59 * R) to both sides of the limit ….

Limit Setting in Specifications (example : MAX)

Including Measurement Uncertainty :

→ Introduce Production Limit ***"PL"***
below which producers can assume "IN SPEC" with 95% probability

→ Introduce Acceptance Limit ***"AL"***
above which customers can assume "OFF SPEC" with 95% probability



Depending on Company Policy:

→ released product *above* ***"PL"*** may still be "IN SPEC" but with less probability
→ "ON SPEC" blending reduces "IN SPEC" probability to ~ 50% or even less
(think of a Gaussian curve overlaid with its mean at the limit value)

→ For similar reasons, Customer "OFF SPEC" claims make only sense when measured values are higher than ***"AL"*** ..

Some bad Field Examples

Some often found fairy tales :

→ Our Company can measure much more precise than the test method indicates Therefore, we can blend much closer to the specified limit ….

→ *this is where product quality safety competes with quality give-away …*

→ my own *"much better site precision"* does not help when the customer cannot be persuaded to use the same much more precise method *((Producers often confuse* ***REPEATABILITY*** *and* ***REPRODUCIBILTY*** *))*

→ However, producers may have the advantage to have more knowledge about their product from e.g. process control …

→ Similar discussion items are encountered when (trace) limits are requested which fall *below the "2R rule"* or below the test method application range …

→ Some labs claim that they have *"improved"* the test method for better precision, so (only?) they can measure at lower levels. ***This is, of course, incorrect***. These labs have then left the standardized test method and work only under REPEAT conditions using a non-comparable house method ...

How is Precision developed ?

Following EN ISO 4259 :

→ Round Robin testing, using representative numbers of laboratories, sample compositions and measurement ranges yield precision estimators for *Repeatability "r"* and *Reproducibility "R"* which are given in the test method

They take on functional forms like r = const.

or R = [slope] * X + [intercept]

or other functions

→ It is important to recognize that every precision estimation campaign will produce (possibly) similar, but also potentially quite different precision statements because of the stochastic nature of the analytical experiment …

→ which one out of several precision statements / results should be used ??

→ how many more round robin tests are needed before precision is established ??

→ how often do we execute sustainable checks of the precision statements ??

Combining results from two Round Robin Campaigns

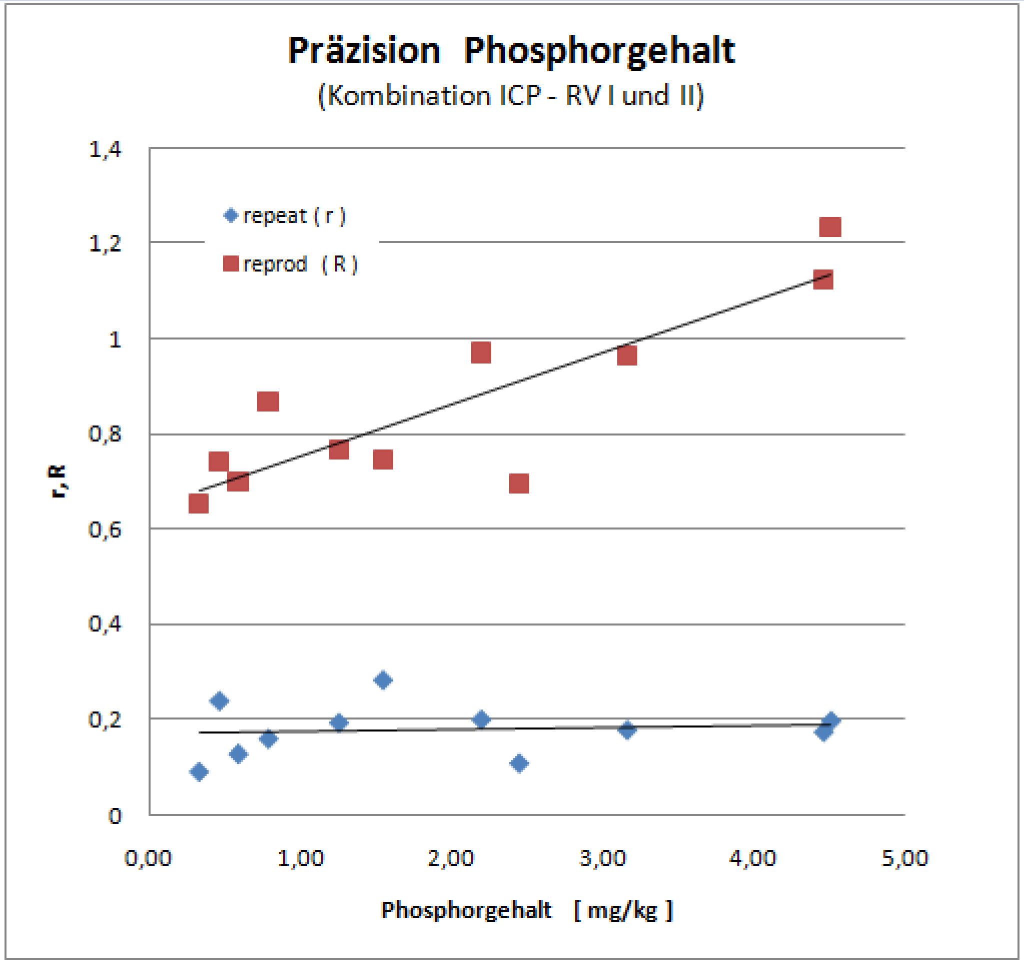
RV Nr.	X	repeat (r)	reprod (R)	(2R)
2	0,331	0,092	0,652	1,363
2	0,464	0,239	0,741	1,392
1	0,588	0,129	0,699	1,419
1	0,787	0,160	0,867	1,462
2	1,254	0,193	0,767	1,563
2	1,545	0,282	0,745	1,626
2	2,194	0,200	0,970	1,767
1	2,445	0,110	0,695	1,822
2	3,159	0,178	0,965	1,977
1	4,457	0,174	1,123	2,258
1	4,505	0,197	1,234	2,269

Grand Mean: 1,975

Wiederholbarkeit: r = 0,0040 X + 0,1696

Vergleichbarkeit: R = 0,1085 X + 0,6455

Grenzwertsetzung: " max. 3,0 mg/kg"



Some Issues with Precision Function "R"

Which function type should be used?

.. tabulated values;

.. Line equations like $R = 0{,}123 * X + 1{,}200$

.. Power functions like $R = 0{,}345 * X^{-0{,}8677}$ and so on …

→ We can see that the *actual format* of the precision function is not of utmost importance. Programs like "TableCurve" calculate more than 1100 of them….

Rounding

→ How many digits to use for the regression coefficients (rounding) ?
How many digits do we use e.g. for ***SQRT(2)*** or ***SIN(45°)*** or ***PI*** **?? ?? ?? ??**
What about e.g. the rounding in Excel Worksheets **?? ?? ?? ??**

→ Rounding should respect the Standard Errors **"*SE*"** for the regression coefficients and should be not *less than three* and not *more than four* (significant) digits …

→ Rounding must be compatible for precision, test results, limits, ***PL*** and ***AL*** !

Standard Errors "SE": reflect the uncertainties for the regression coefficients:

so $Y = a * X + b$

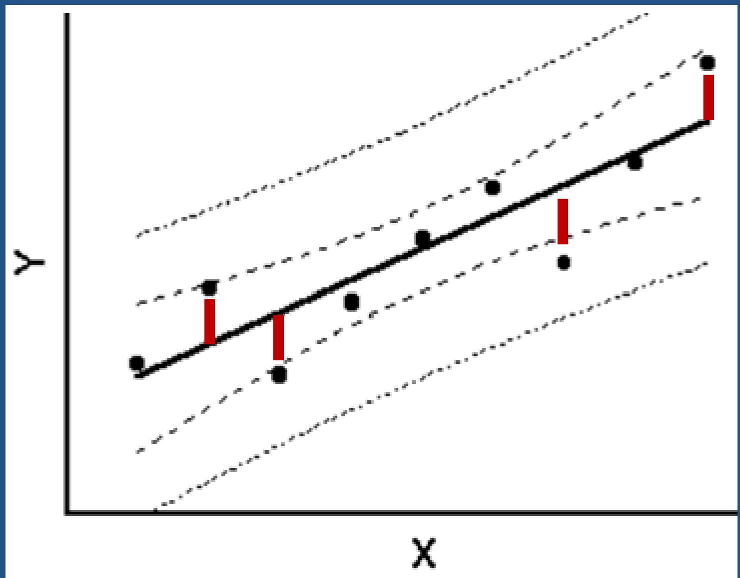
becomes $Y = (a \pm 1{,}96 * SE(a)) * X + (b \pm 1{,}96 * SE(b))$

Confidence - & Prediction Intervals

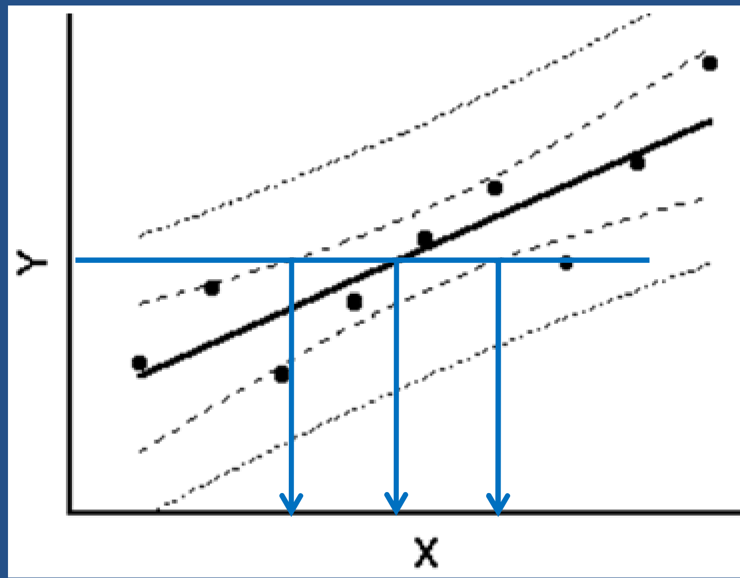
Dispersion

→ More interesting than the functions alone are their corresponding confidence- and prediction intervals. These intervals constitute corridors in which many different regression / approximation functions could be used…

Calibration …

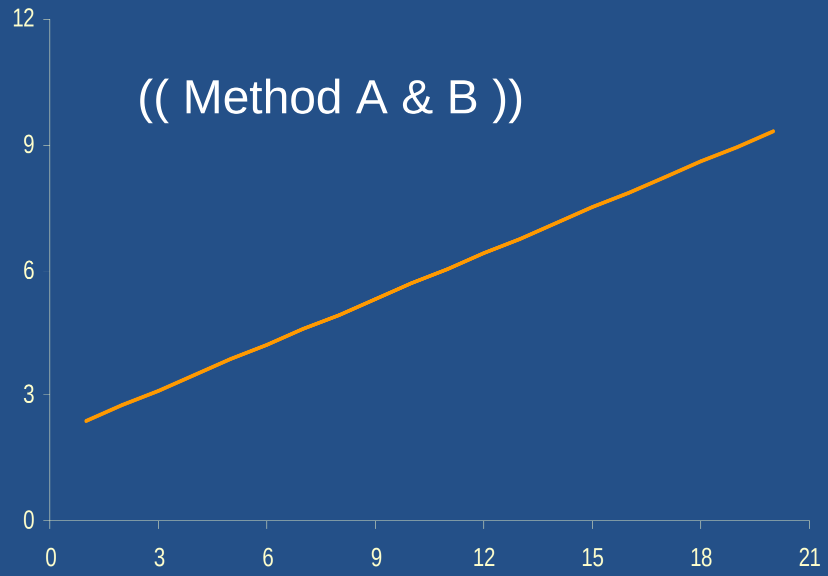


Prediction …



→ Interpretation of rules like the “2R – rule” for limit setting can sometimes become difficult when confidence / prediction corridors are not very small ..

Precision Functions and Dispersion

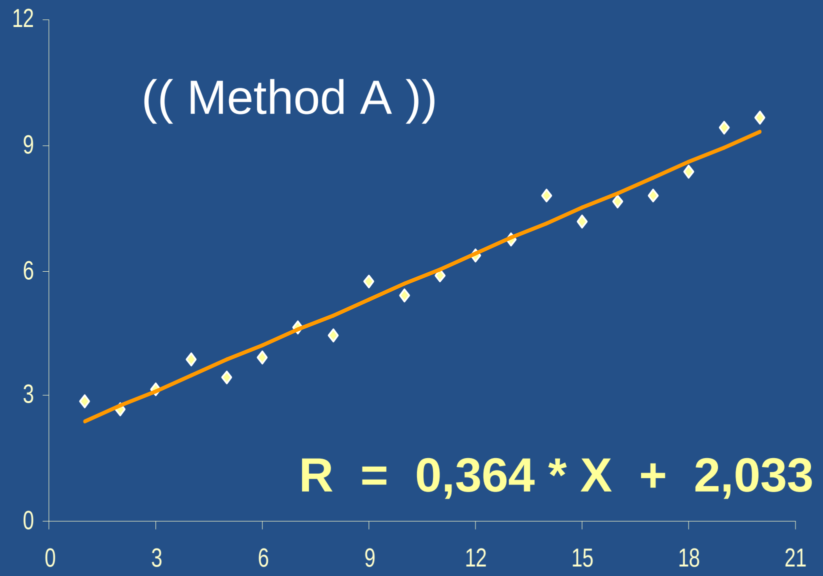


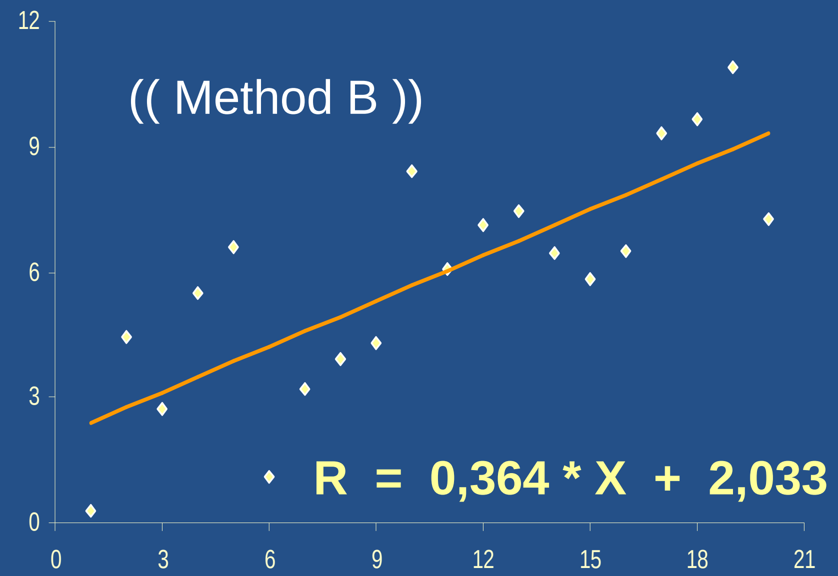
Assume Test Method A & B :

.. same measurement range;
.. same Grand Means;
.. same precision function :

R = 0,364 * X + 2,033

→ same precision ????

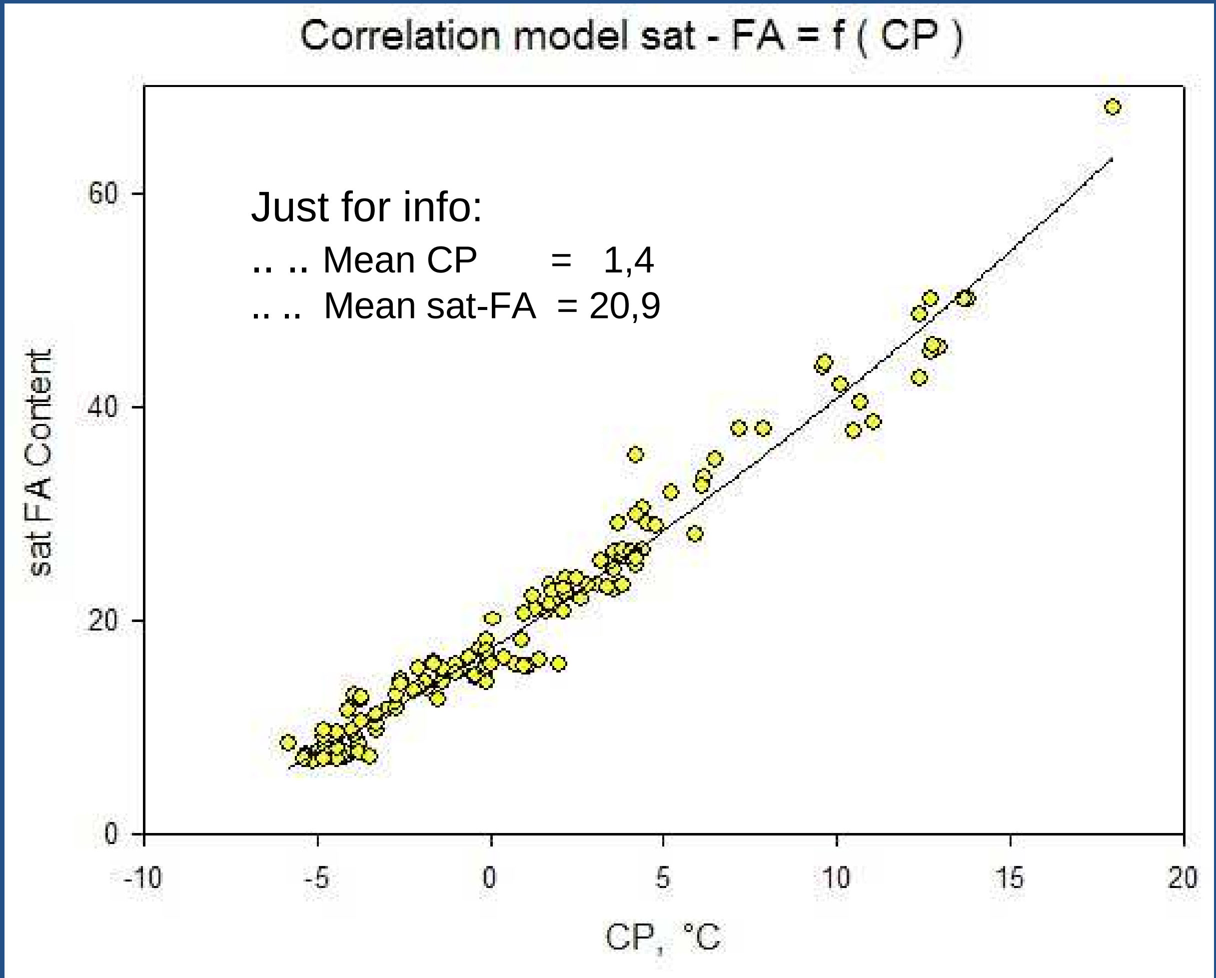




Problematic Model Interpretation without Dispersion ?

Example: CONCAWE prediction model for ***saturated Fatty Acids*** from CP

→ 1st impression: acceptable .. good correlation → sat. FA predicable
→ 2nd impression: uneven point distribution, not many points in interesting region



Unanswered Questions:

→How much dispersion?

→Influence on prediction quality ?

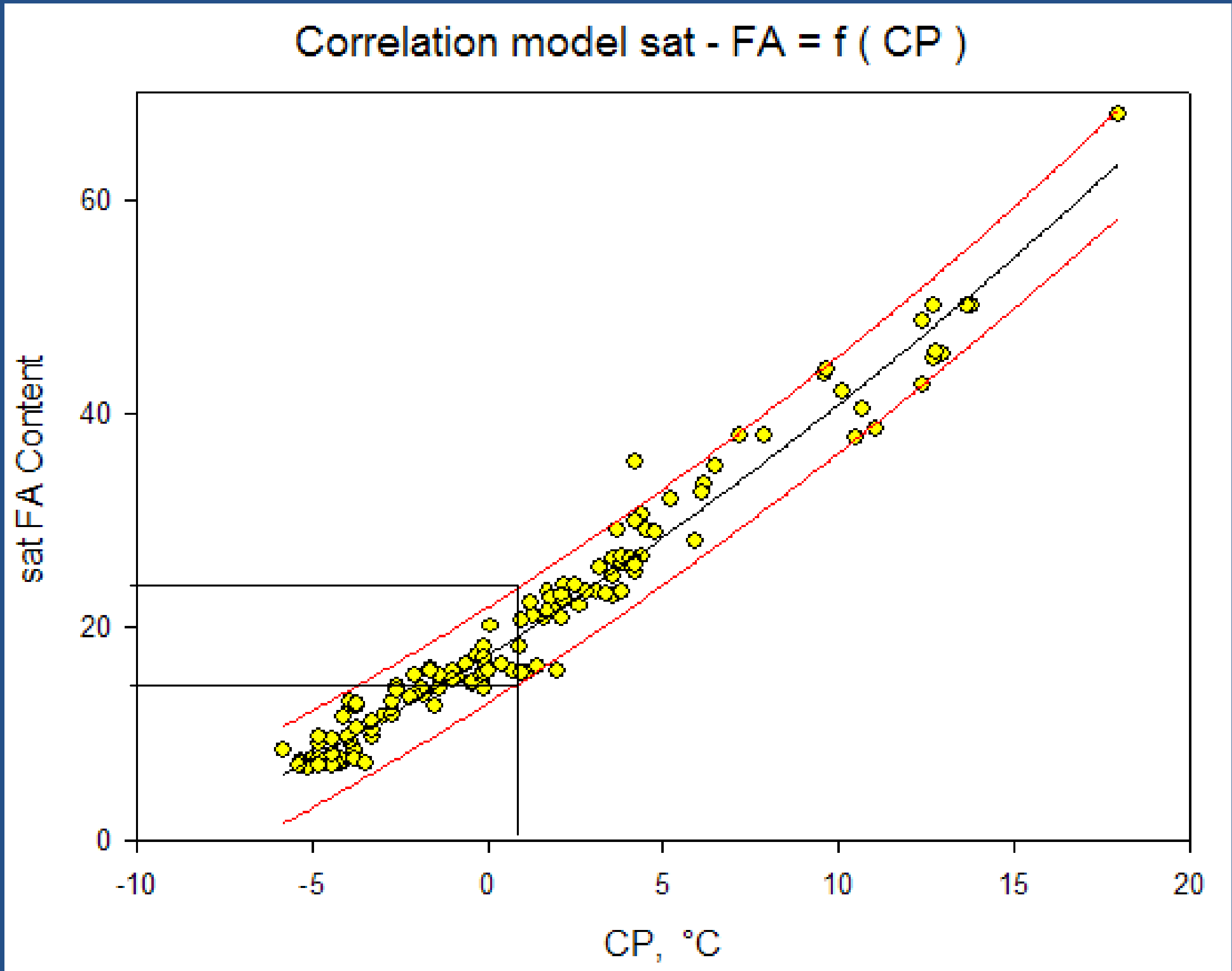
→How big is the prediction error anyway ?

→Uncertainty for CP measurement requires orthogonal regression ?

Problematic Model Interpretation with Dispersion

Example: CONCAWE prediction model for *saturated Fatty Acids* from CP

→ 1st amendment: 95% prediction corridor introduced
→ assumption: assume negligible error in CP measurement



Inspection Results:
(also confirmed by PEG)

Instead of using the regression „line“ alone, the „prediction uncertainty“ must be kept in consideration when the regression model is applied.

Example:

.. For CP = 1,4 and
.. avg (sat FA) = 20,9

The prediction span is from

avg (sat FA) = 15,6 .. 24,7

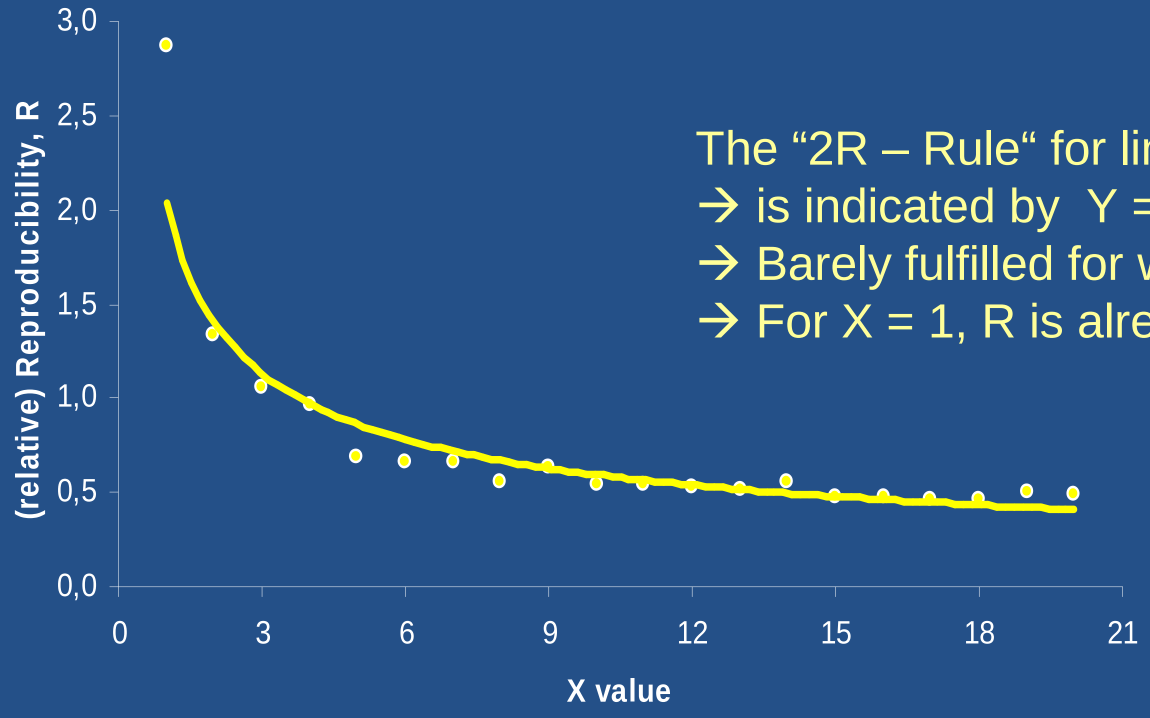
Thereby indicating an error of about 54% of the sat.FA value obtained from the regression

Useful (relative) Precision

For any Test Method Precision Footprint :

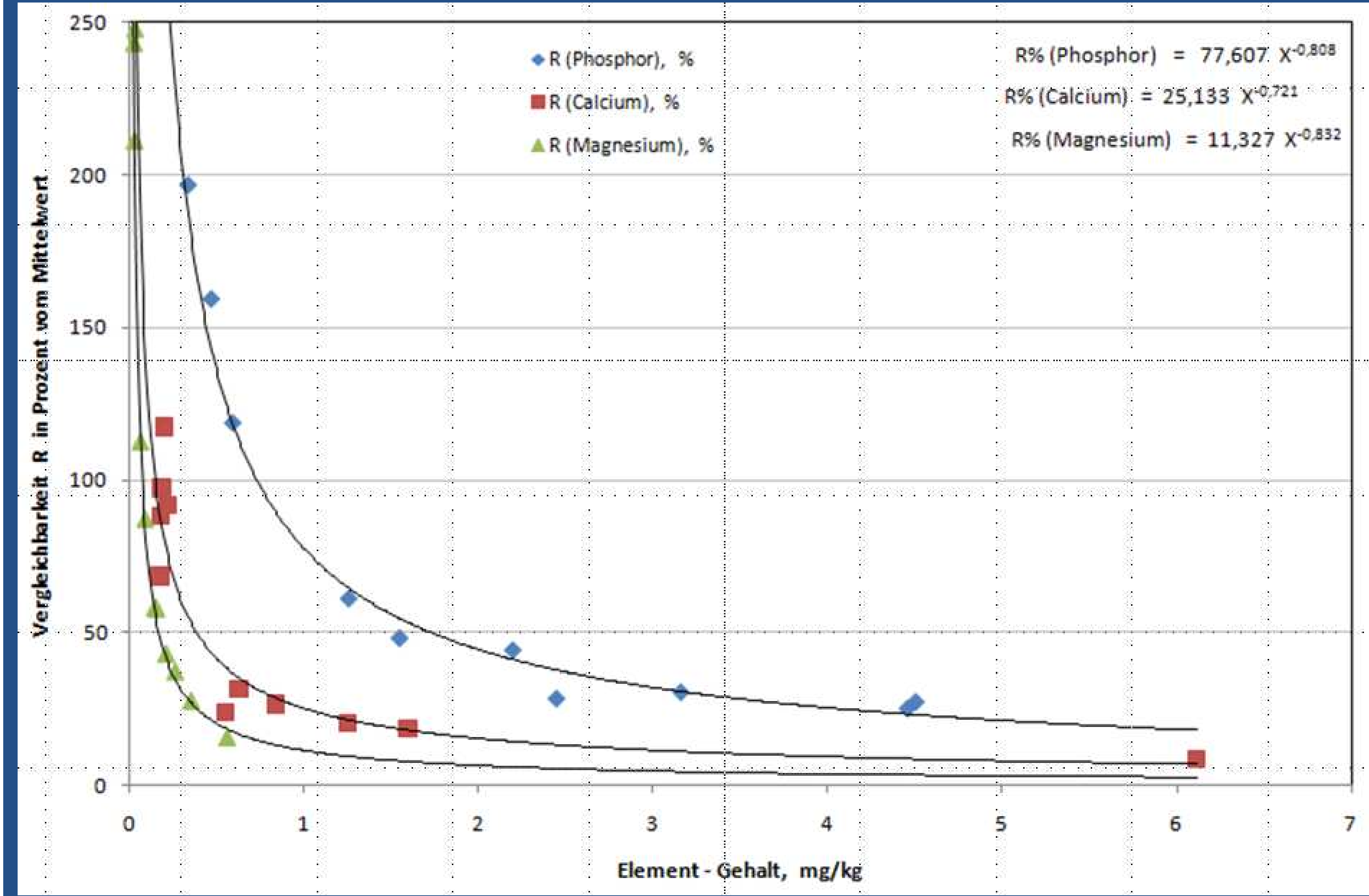
.. Instead of *absolute* values we now use *relative* values, i.e. the proportion of R relative to X with a clear and well – defined functional model:

$$R = a * (1/X)^{b} \quad \text{(same as)} \quad R = a * X^{-b}$$

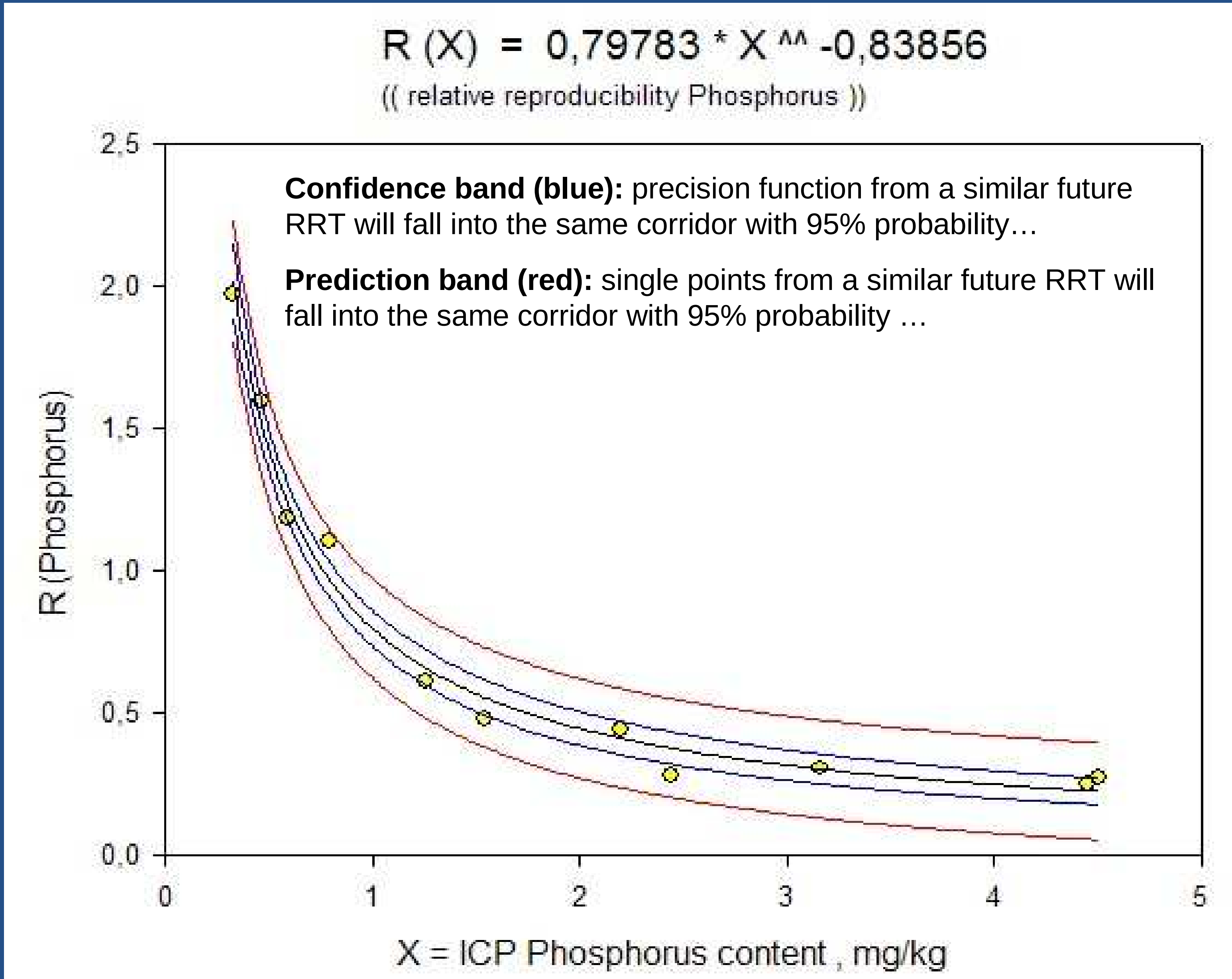


The “2R – Rule“ for limit setting
→ is indicated by Y = 0,5
→ Barely fulfilled for whole measurement range ..
→ For X = 1, R is already three times that value ..

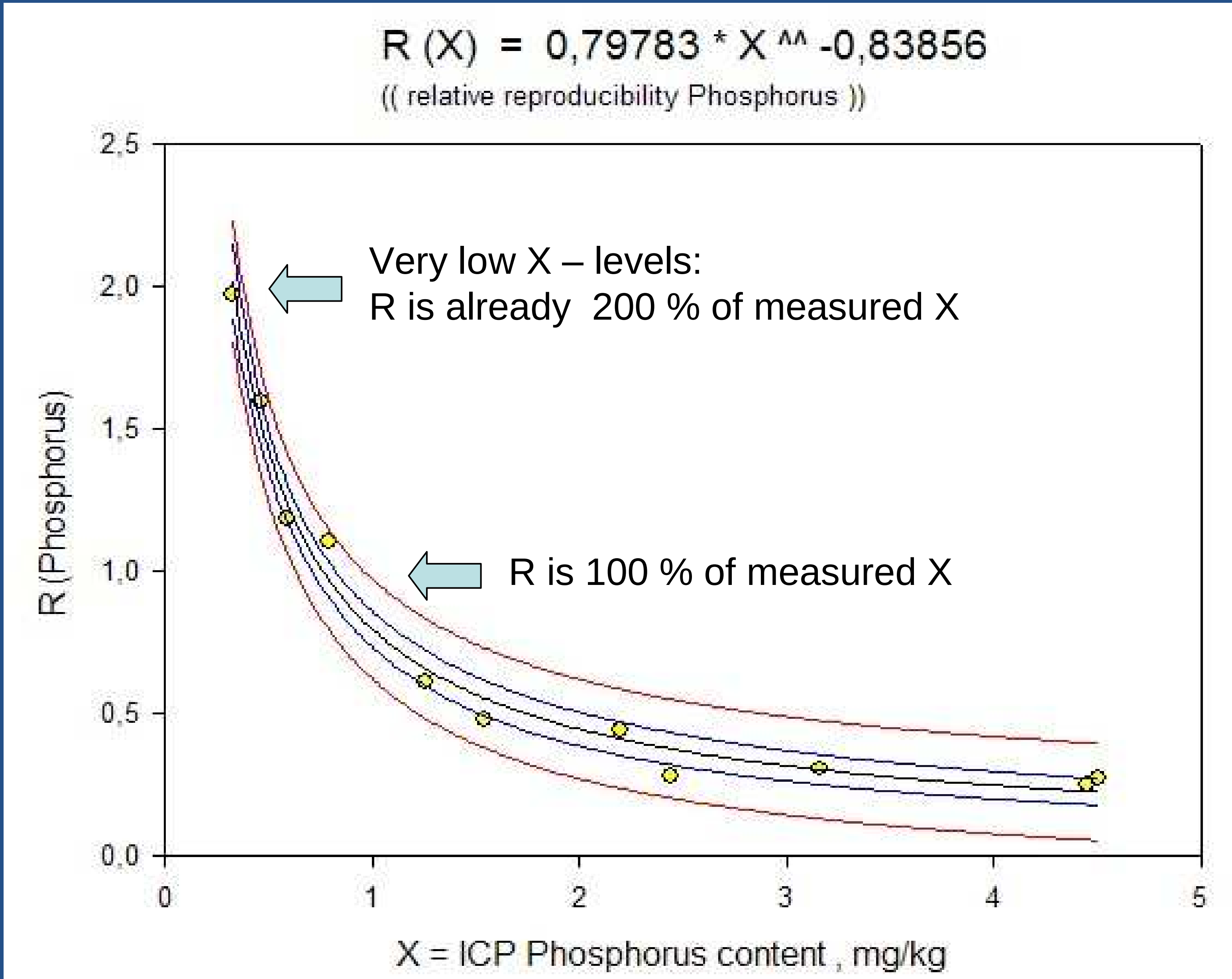
Comparing Test Methods with (relative) Precision



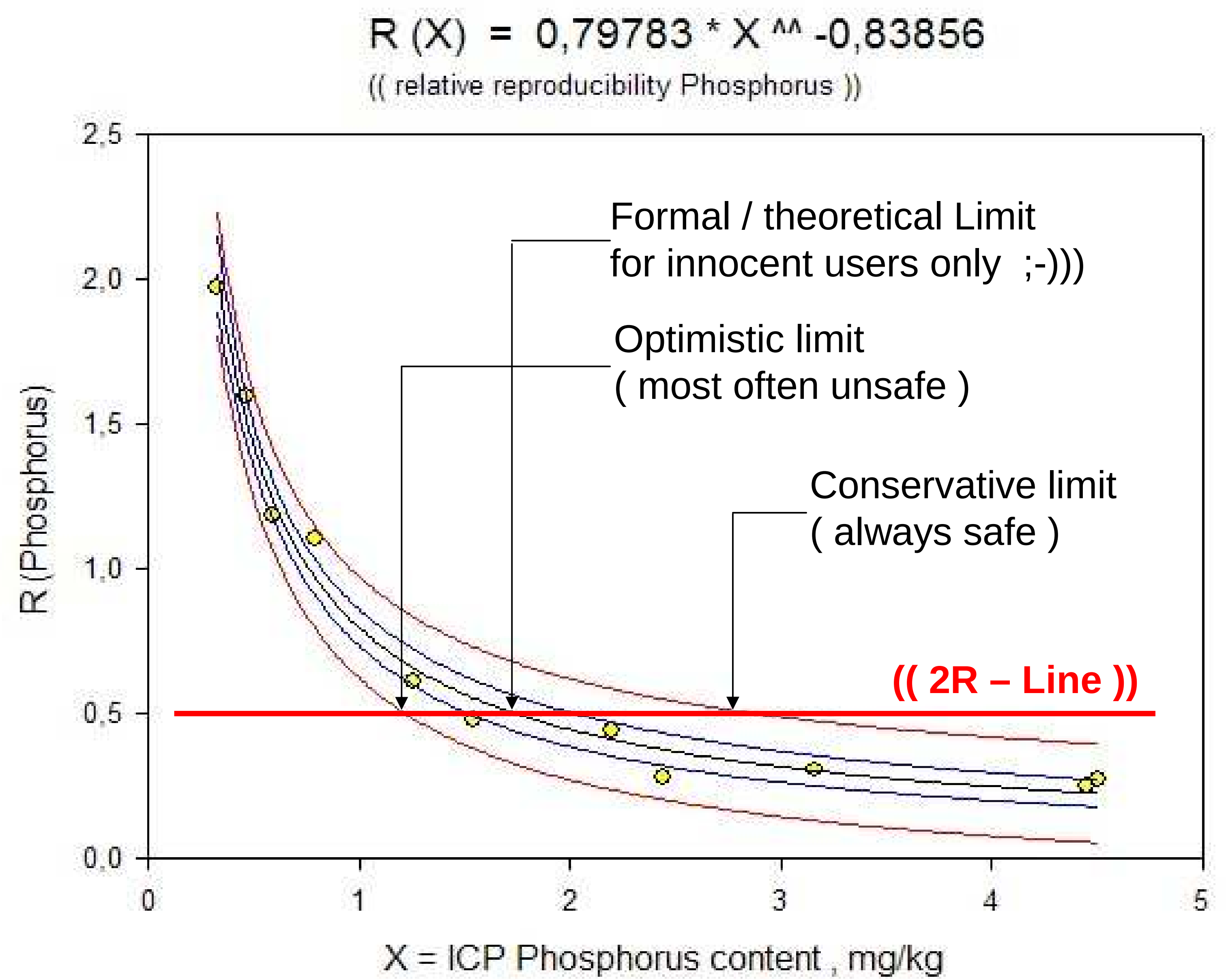
(rel) Precision & Confidence Bands in Limit Setting (I)



Test Method Judgements with (relative) Precision



Limit Setting with (relative) Precision



Some Conclusions

→ The presentation indicates the importance of dispersion and confidence band concepts for the judgement of test method result quality.

→ Although from a superficial viewpoint being more complicated, the use of such extended analytical tools will give us much more robust and trustworthy test methods which will also save time and effort in test method maintenance.

→ Test method robustness and functional reliability of limits are getting more important as requirements and test methods as well are pushed more and more to their theoretical and practical limits …

This is all for today.

Thank you for patient listening !