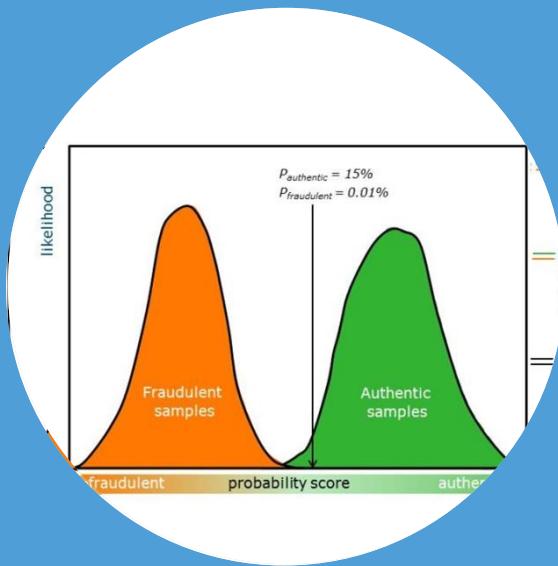


# Validation of multivariate classification methods

Martin Alewijn

RIKILT – Wageningen University and Research (NL)



# RIKILT, Wageningen UR

- Research institute of food safety, Wageningen UR
- > 200 staff
- Clients: government; national/EU scientific funding bodies; industry; NGOs
- Research themes:
  - Natural and chemical contaminants
  - New risks
  - Residues
  - Feed
  - **Product composition / quality / authenticity**



# Multivariate classification methods

A multivariate classifier provides:

Mathematical *translation* of  
(analytical) sample *properties*

Into one of the available  
nominal *classes*

Based on a *database* of  
known samples

**Unknown sample**

Sample	ab	cd	ef	gh	ij	kl	..	xx	yy	zz
Sample	0.47	0.38	-3.23	17.01	11.54	0.01	..	33.78	0.06	7.10

The diagram illustrates the classification process. An orange box labeled "Unknown sample" contains the data row: Sample, ab, cd, ef, gh, ij, kl, .., xx, yy, zz. This row is compared against two reference sets: "Class A reference set" (light blue background) and "Class B reference set" (light green background). Each set contains multiple rows of sample data. A vertical orange bar with a question mark at the top is positioned between the two reference sets, indicating the unknown sample's classification status.

Sample	ab	cd	ef	gh	ij	kl	..	xx	yy	zz
Sample	0.47	2.40	-3.74	36.24	9.59	0.44	..	33.84	0.03	18.62
Sample	0.25	0.17	-3.71	48.68	4.17	0.82	..	33.09	0.13	37.96
Sample	0.64	0.69	-3.42	32.97	24.79	0.00	..	33.09	0.14	52.55
Sample	0.69	1.88	-3.94	58.41	5.46	0.02	..	33.94	0.11	31.79
Sample	0.15	1.51	-3.27	66.07	0.19	0.06	..	33.21	0.01	65.31
Sample	0.43	3.74	-3.90	45.22	0.28	0.08	..	33.49	0.10	16.08
Sample	0.61	0.20	-3.54	45.22	0.28	0.08	..	33.85	0.04	23.21
Sample	0.53	1.12	-3.93	66.62	12.40	0.22	..	33.57	0.22	40.69
Sample	0.44	2.42	-3.47	35.61	10.30	0.00	..	33.82	0.05	50.21
Sample	0.23	2.22	-3.47	51.19	2.79	0.36	..	33.29	0.07	49.97
Sample	0.56	0.72	-3.44	42.53	20.60	0.04	..	33.02	0.02	59.19
Sample	0.7	0.75	-3.05	31.20	2.60	0.00	..	33.06	0.06	33.88
Sample	0.82	0.82	-3.48	81.84	1.64	0.00	..	33.20	0.20	66.80
Sample	0.7	0.7	-3.48	51.61	1.64	0.00	..	33.56	0.24	16.81
Sample	0.41	0.40	-3.3	66.73	1.32	0.02	..	33.23	0.23	0.20
Sample	0.75	0.75	-3.3	53.53	1.32	0.02	..	33.24	0.24	61.03
Sample	0.49	0.49	-3.3	53.53	1.32	0.02	..	33.53	0.09	11.52
Sample	0.35	0.35	-3.3	53.53	1.32	0.02	..	33.74	0.21	64.87
Sample	0.70	0.70	-3.04	51.19	1.32	0.02	..	33.81	0.24	17.58
Sample	0.34	0.34	-3.45	37.42	1.32	0.02	..	33.44	0.17	21.63
Sample	0.94	2.02	-3.37	88.26	25.84	0.00	..	33.75	0.12	54.10
Sample	0.94	2.02	-3.37	88.26	25.84	0.00	..	33.75	0.12	54.10
Sample	0.17	3.65	-3.4	94.74	1.53	0.35	..	33.89	0.24	16.14
Sample	0.51	1.52	-3.90	9.55	2.56	0.55	..	33.15	0.03	68.38
Sample	0.45	1.87	-3.98	85.38	5.66	0.00	..	33.32	0.11	67.49
Sample	0.47	0.32	-3.21	21.99	15.42	0.97	..	33.50	0.07	67.46
Sample	0.81	1.61	-3.97	91.57	11.90	0.48	..	33.28	0.18	28.45
Sample	0.35	0.55	-3.19	23.94	16.24	0.72	..	33.16	0.04	68.13
Sample	0.15	1.15	-3.46	0.58	6.57	0.01	..	33.96	0.21	45.05
Sample	0.58	0.95	-3.34	36.04	18.29	0.01	..	33.54	0.01	32.75
Sample	0.22	3.06	-3.73	37.84	26.16	0.02	..	33.80	0.05	15.30
Sample	0.66	3.99	-3.01	52.44	0.87	0.00	..	33.98	0.13	20.83
Sample	0.47	2.24	-3.68	99.30	21.16	0.52	..	33.98	0.13	20.83

Class A ← or → Class B

# Multivariate classification methods

- Used in areas like food fraud detection, where single analytical markers do not exist
  - Classification using (chemical) properties is indirect
  - New decisions are based on an empirical basis
- **How good is your method?**

# Multivariate performance characteristics

- Correct classification rate
- Accuracy
- Sensitivity
- Specificity
- Kappa-statistic
- Youden's index
- ...
- Predictive Error Sum of Squares (PRESS)
- Total sum of squares (TSS)
- Standard Deviation Error in Calculation (SEC)
- Coefficient of determination ( $R^2$ )
- Cross-validated coefficient of determination ( $Q^2$ )
- ...
- Brier score
- Area Under Receiving Operating Characteristic curve (AU ROC)
- ...

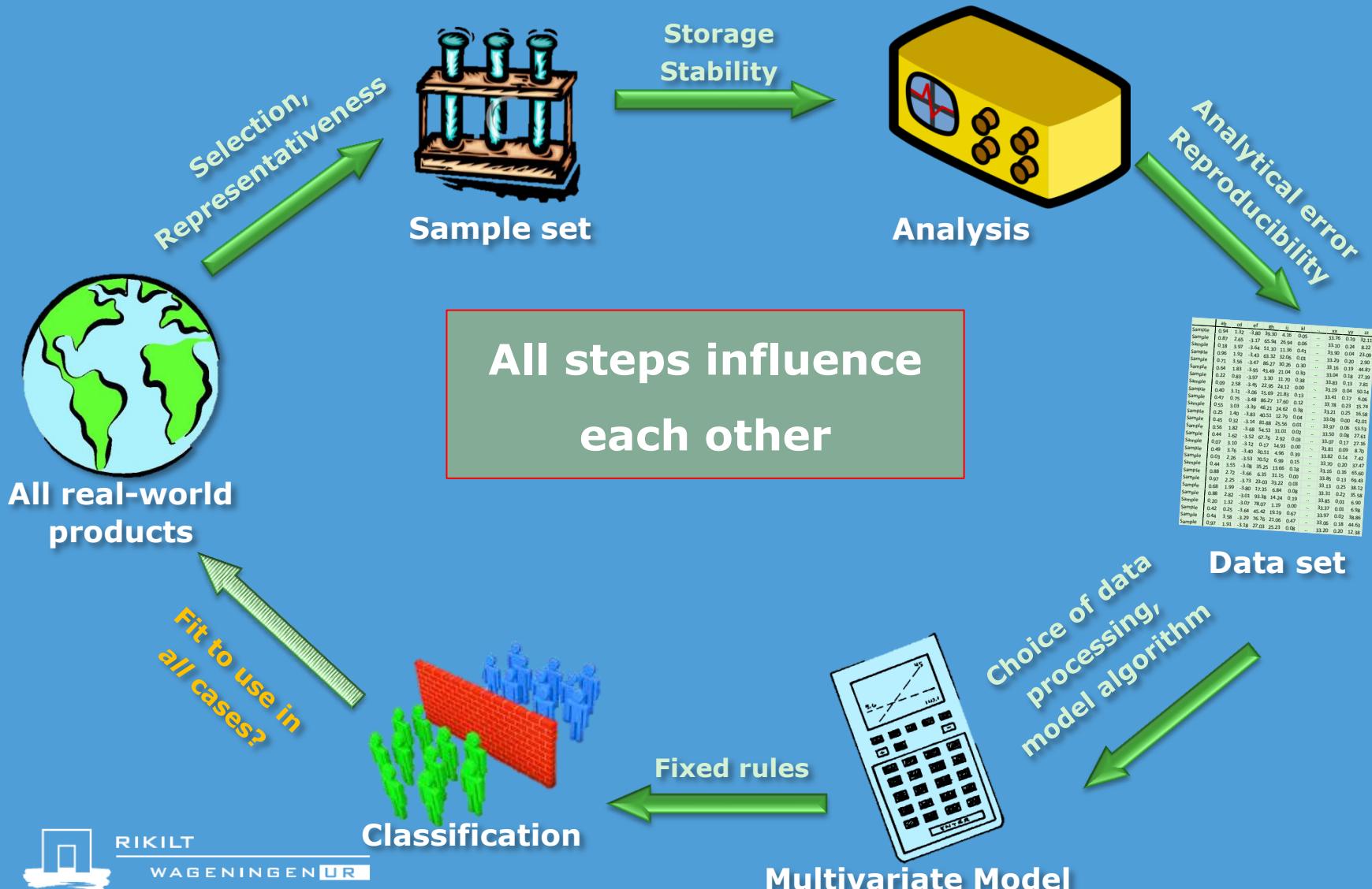
# Analytical method performance evaluation

- Method validation for analytical methods is well-established
- Towards the multivariate (classification) case:
  - Hard to translate all performance characteristics
  - Would this ensure full validation?

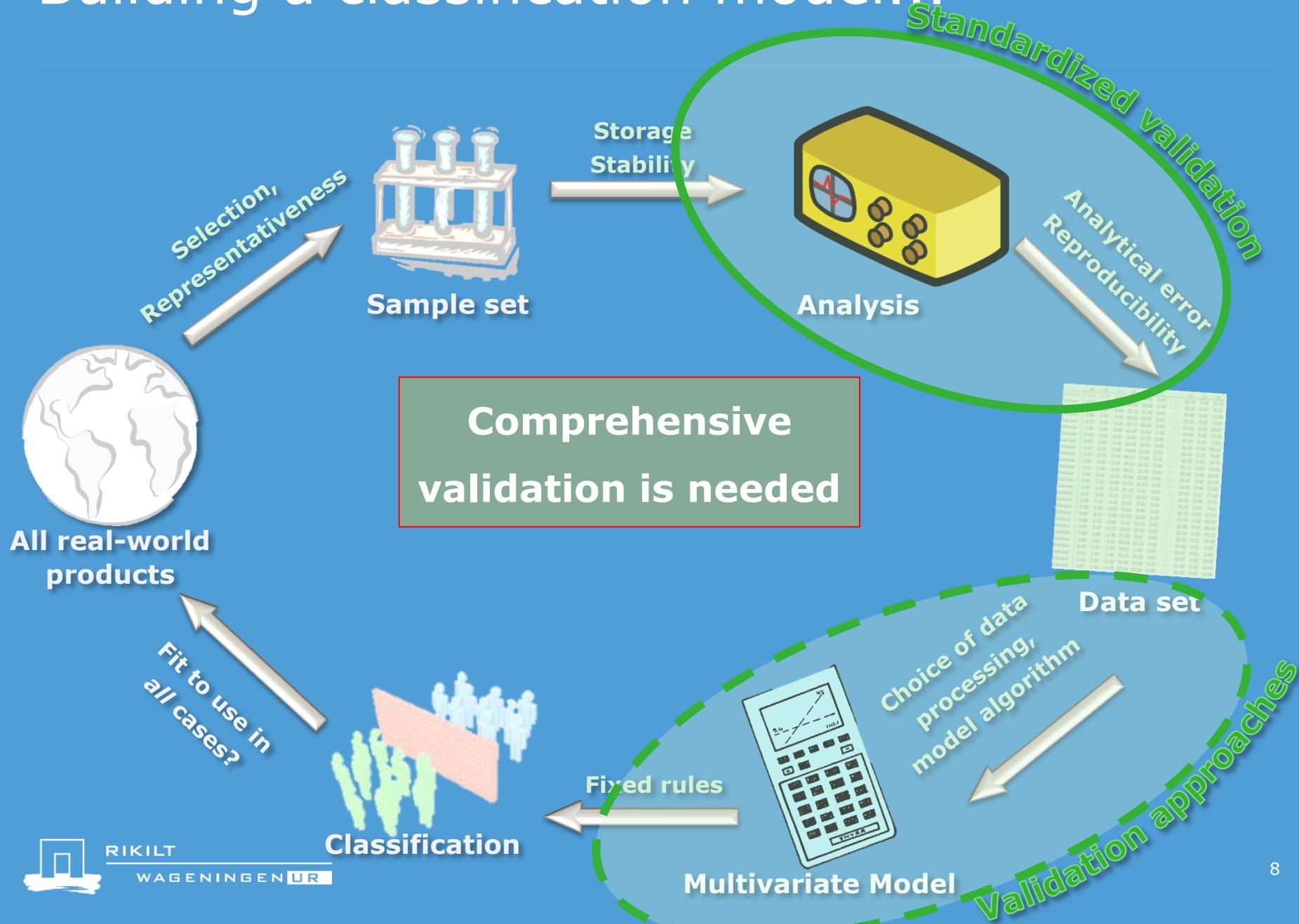
ISO 17025 & Com. Decision 2002/657/EC

Trueness  
Accuracy/Recovery  
Repeatability  
Within-lab reproducibility  
Measurement uncertainty  
 $CC_a$   
 $CC_\beta$   
Limit of detection  
Limit of quantification  
Linearity  
Specificity  
Ruggedness  
Stability

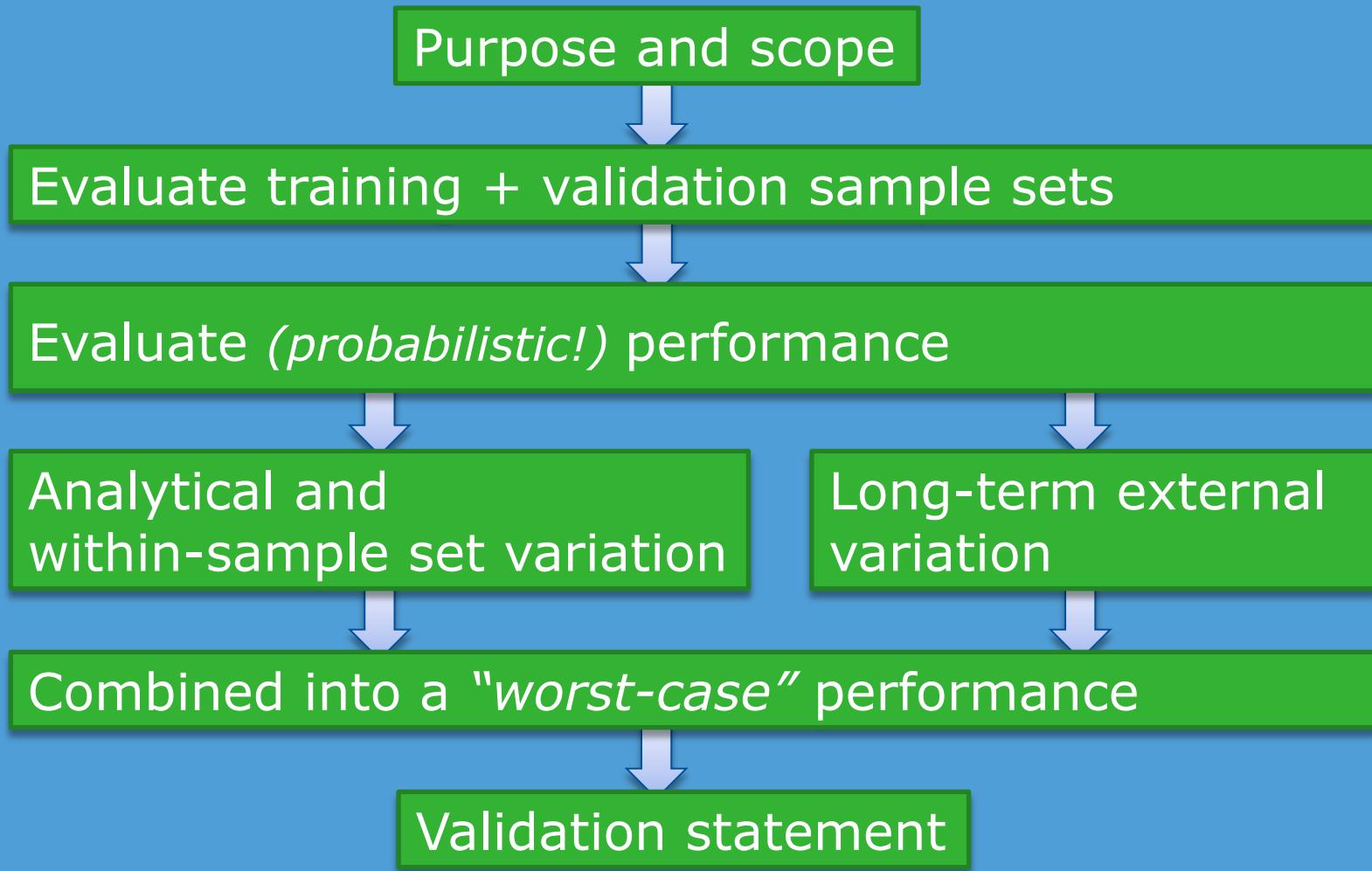
# Building a classification model...



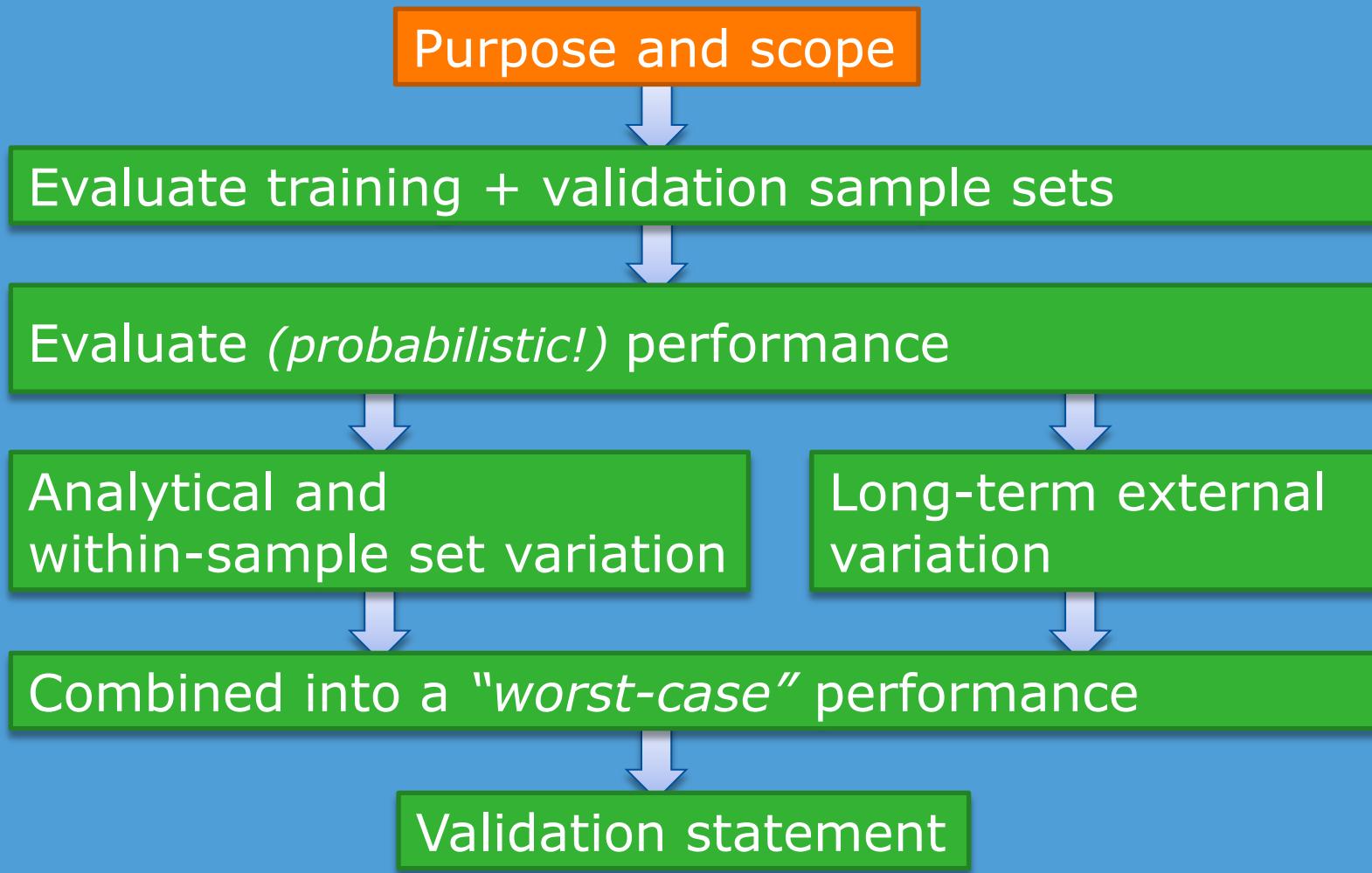
# Building a classification model...



# Proposed validation scheme



# Proposed validation scheme

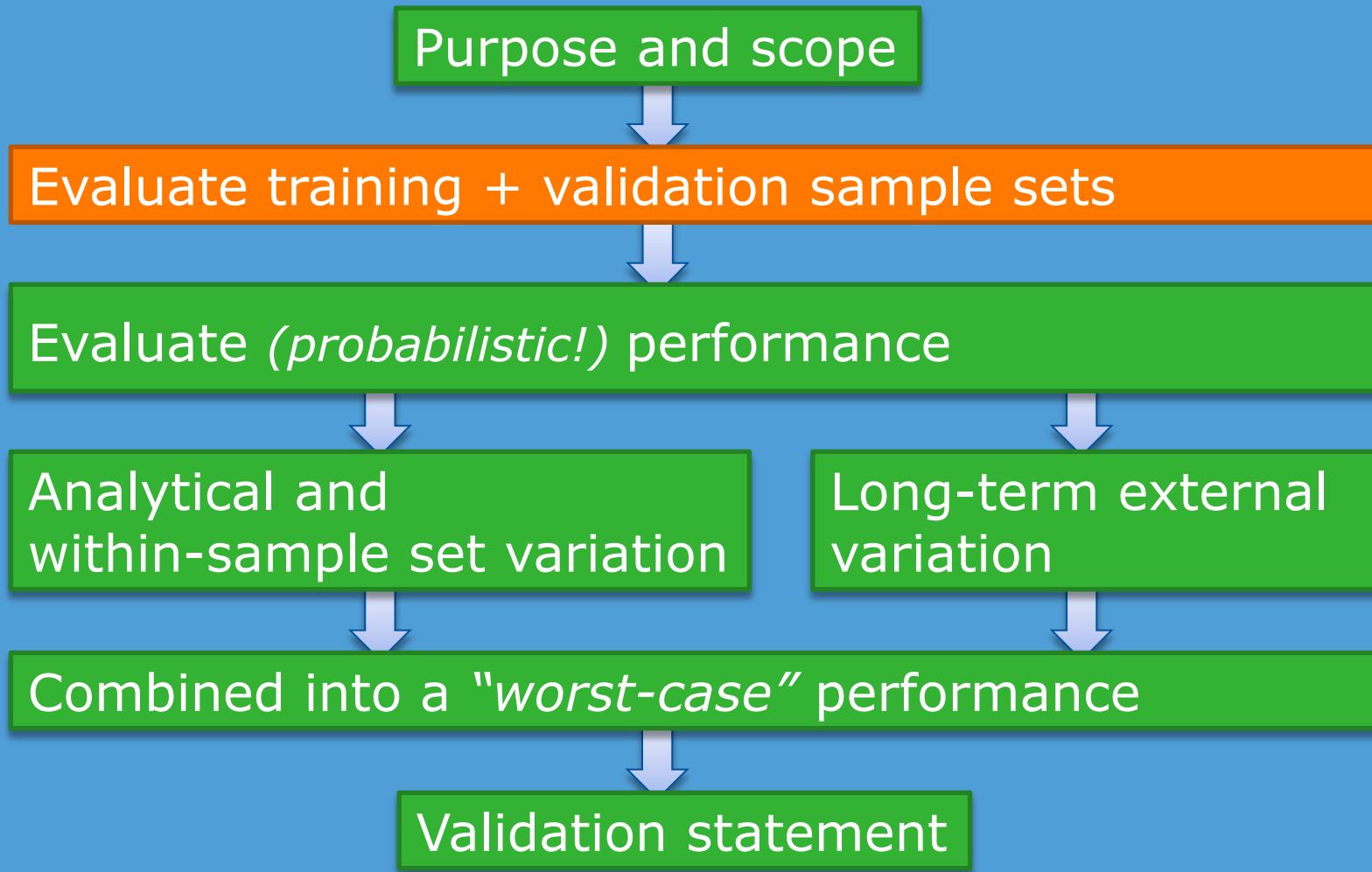


# Purpose and scope

- Define intended use – screening or confirmatory?
- What performance is needed?
- Intended applicability:
  - Matrix/product ranges
  - Origin
  - Year/harvest/season
  - Species/variety/breed
  - Producers
  - ...



# Proposed validation scheme



# Sample sets: training and validation set

Both sets:

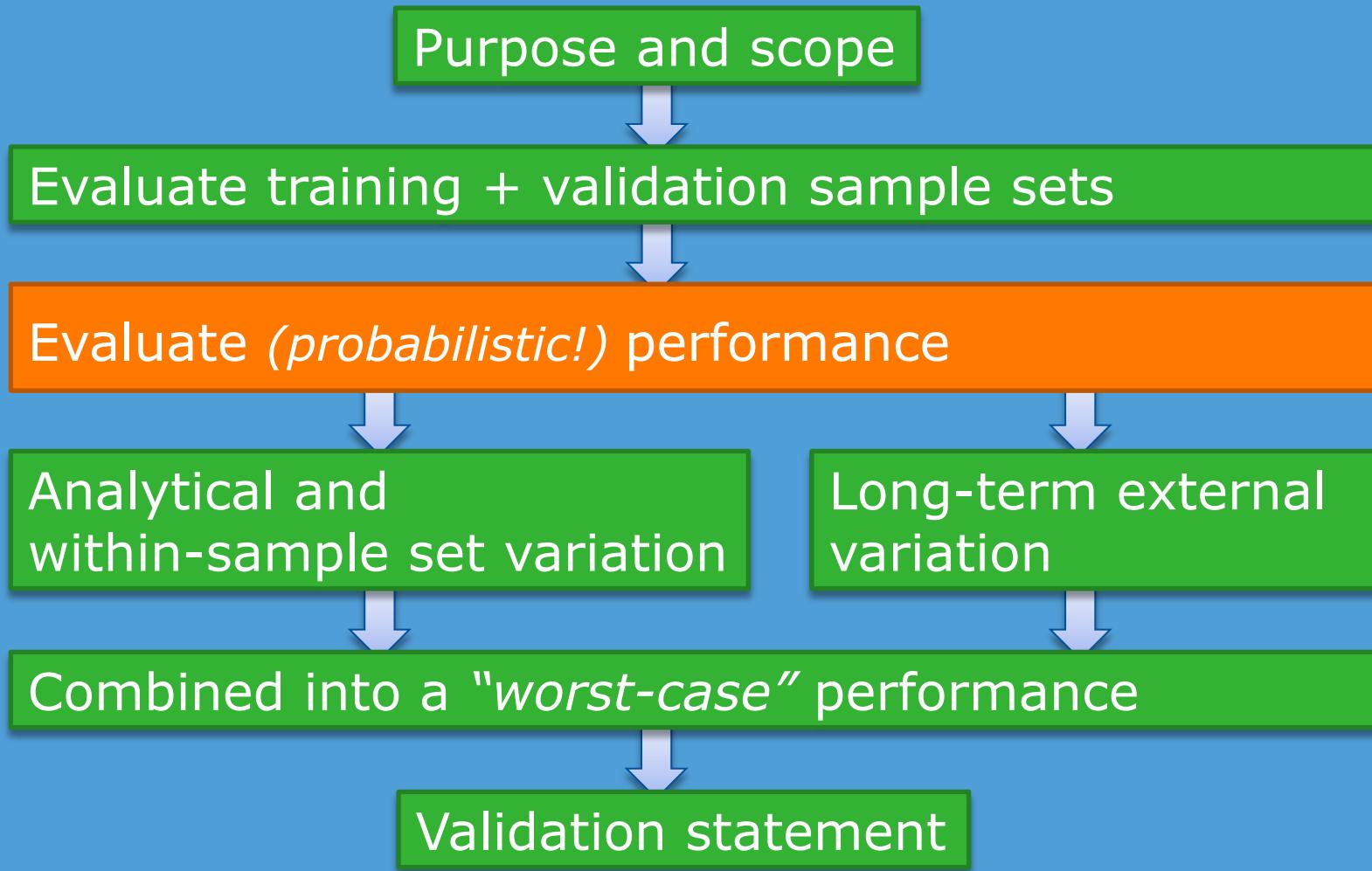
- Is sampling representative, true, and balanced for:
  - Target class
  - Relevant subgroups (season, variety, storage time, etc.)
- Is the sampling quantitatively sufficient?

Validation set: (*within scope*)

- Is all *additional* variation included?
  - Separate sample source, different time period, use of different equipment/technicians, etc.

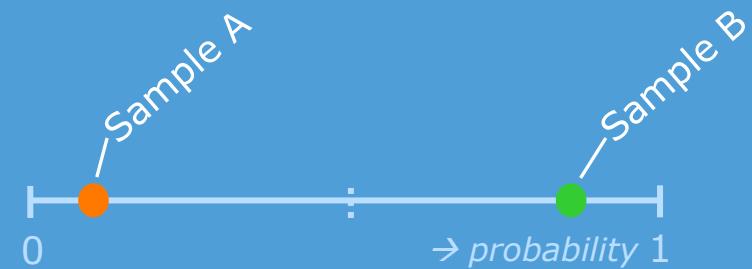
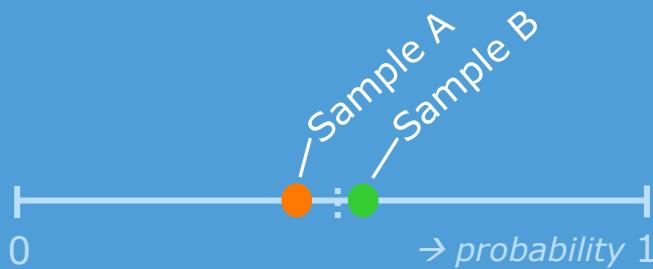


# Proposed validation scheme

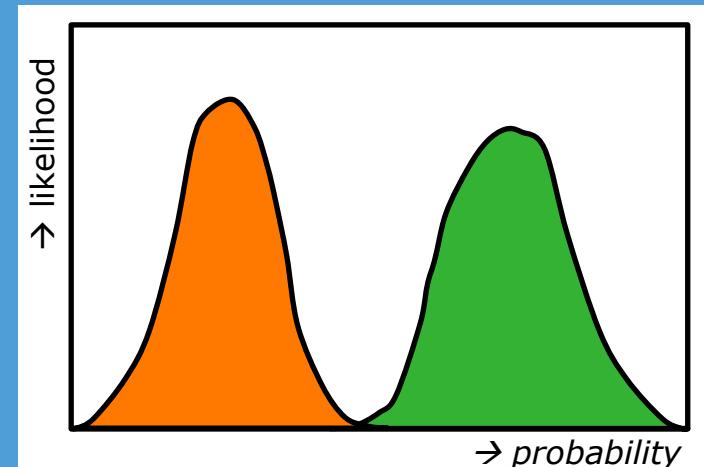


# Performance evaluation

- Two examples, the same classification result:

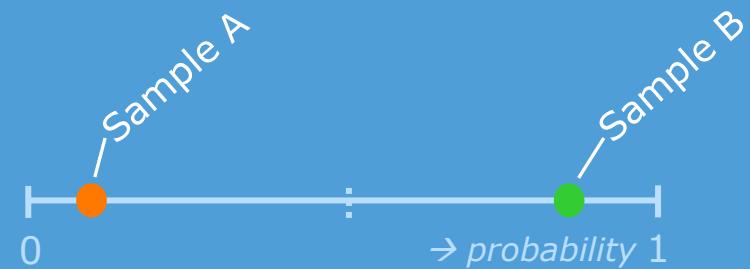
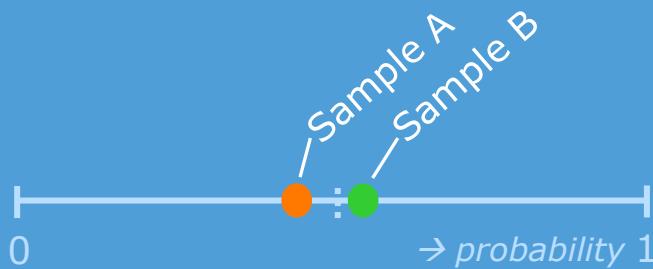


- During evaluation: **go probabilistic!**
- Every algorithm can be made probabilistic
- Obtain model probability distributions



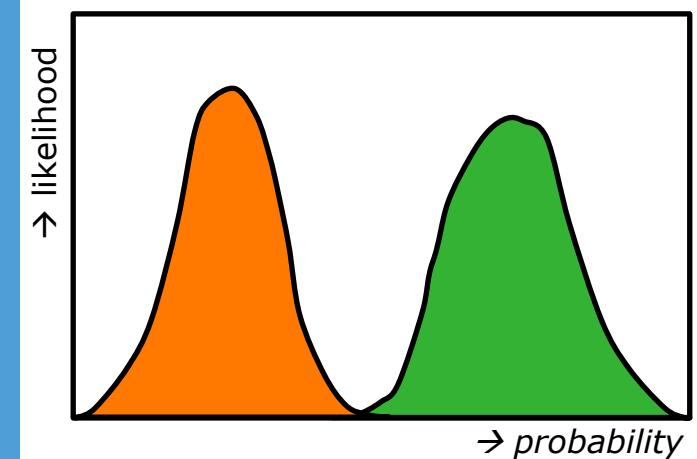
# Performance evaluation

- Two examples, the same classification result:

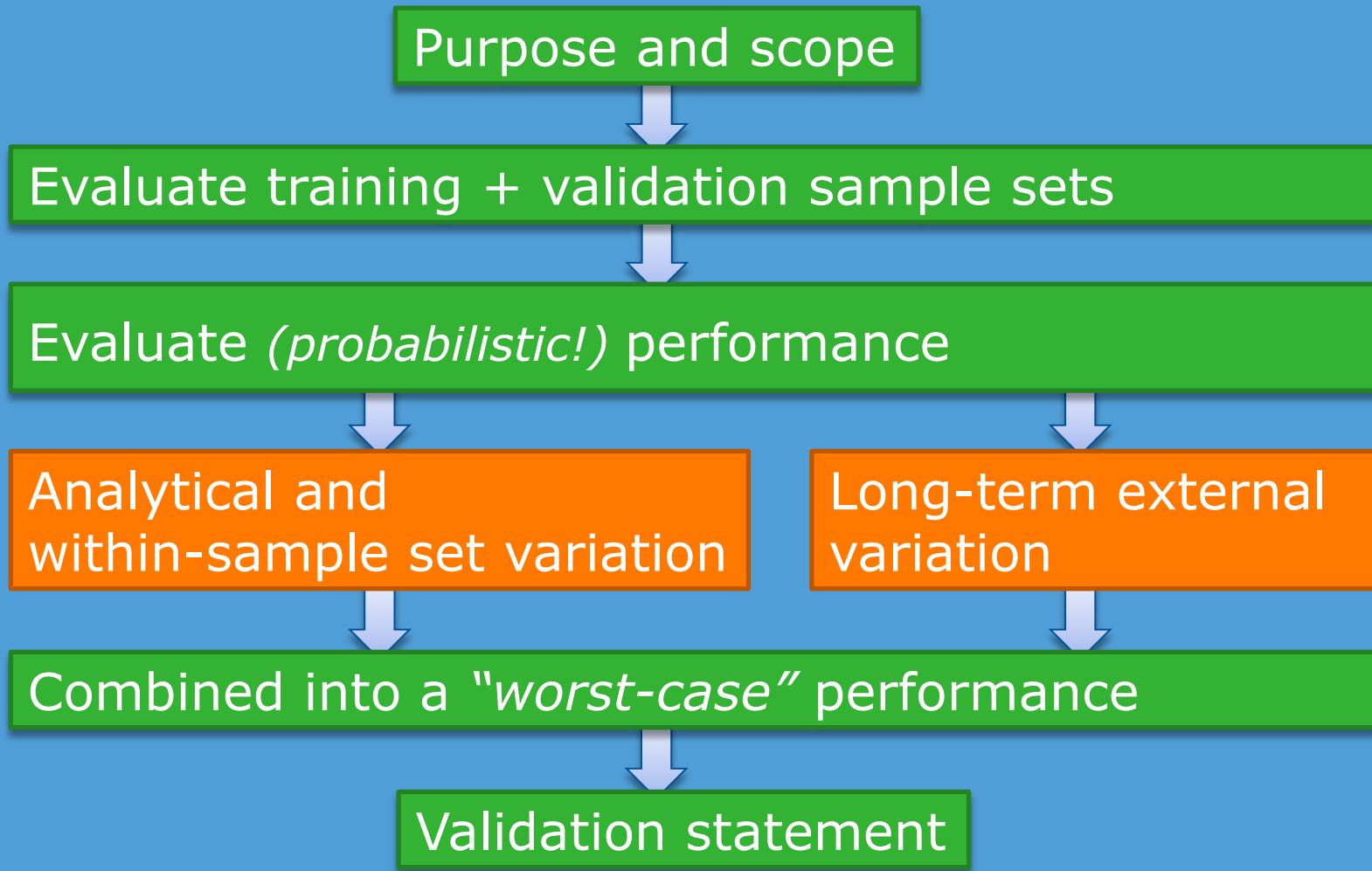


Probability score:

- a candidate for *multivariate* repeatability and reproducibility?



# Proposed validation scheme

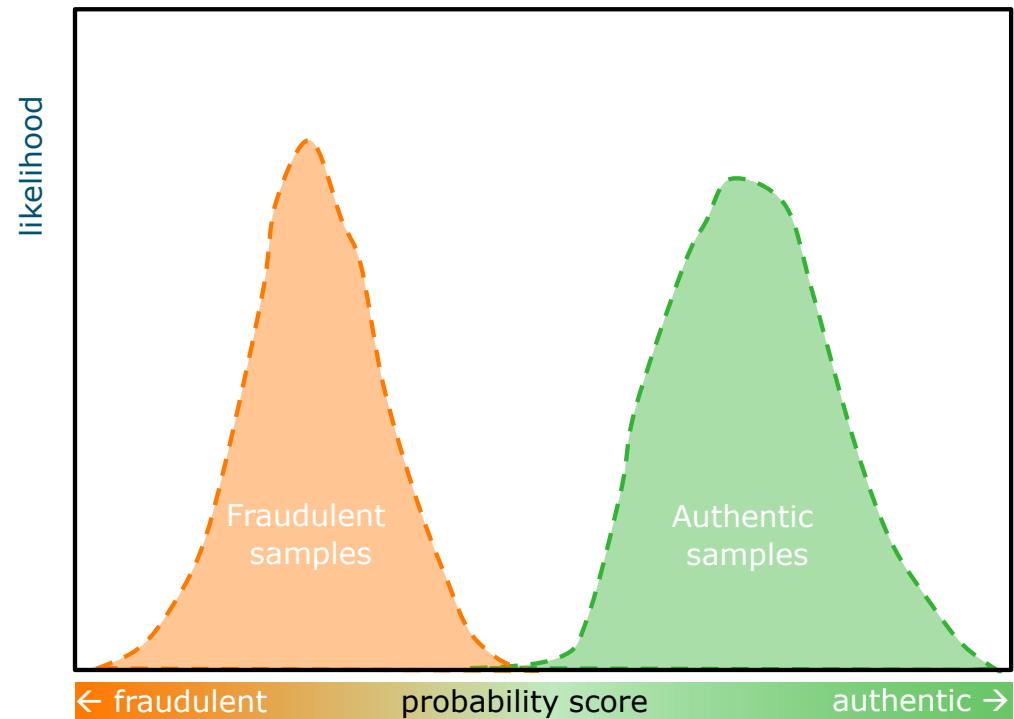


# Sources of error/variation: training set

## Error & Variation:

- Analytical variation:  
→ all usual analytical variation (for each of the variables!)
- Natural variability
- Storage
- Sample handling
- Species
- Regions
- ....

Probability distribution obtained using resampling techniques (e.g.  $rCV$ ) on training set

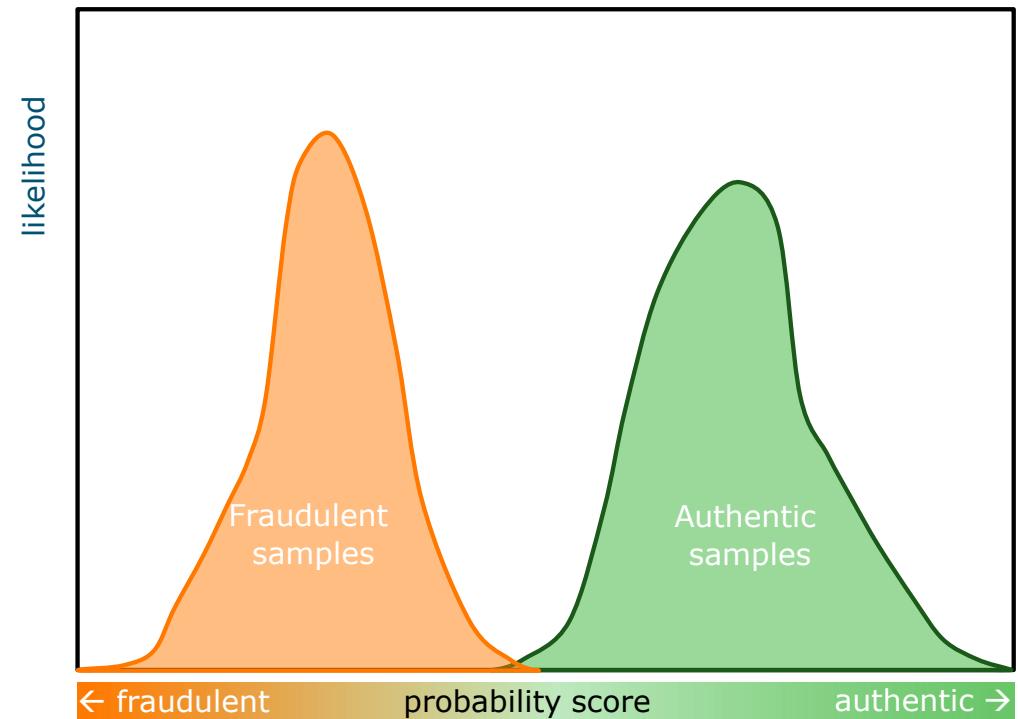


# Sources of error/variation: validation set

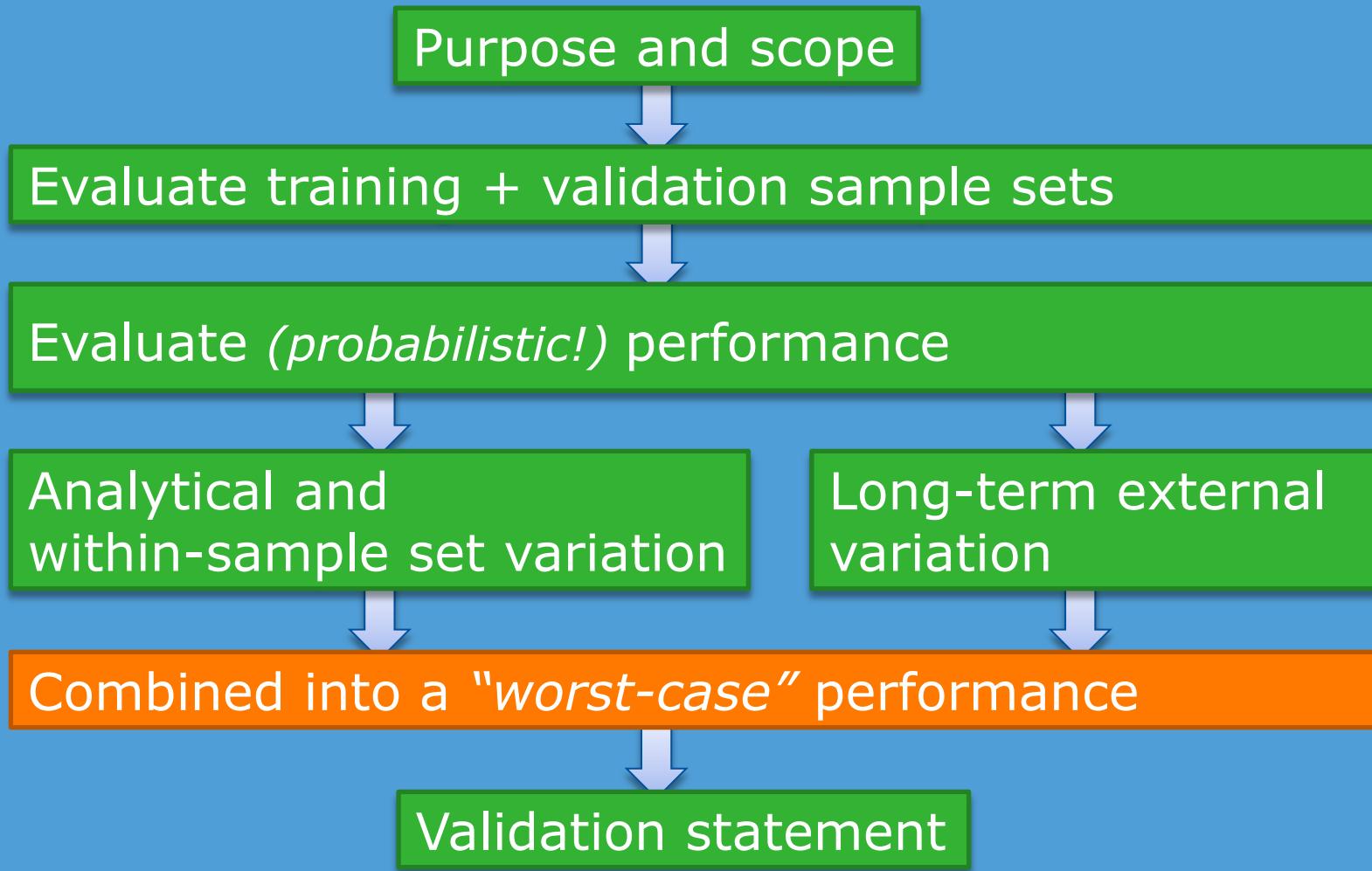
*Additional* variation  
(within scope!) due to  
obtained differences in:

- Harvest
- Producers
- Storage conditions
- ....
- Analytical equipment
- Technician
- Preprocessing protocol
- Solvents
- ....

Probability distribution obtained  
predicting the validation set  
samples by the model

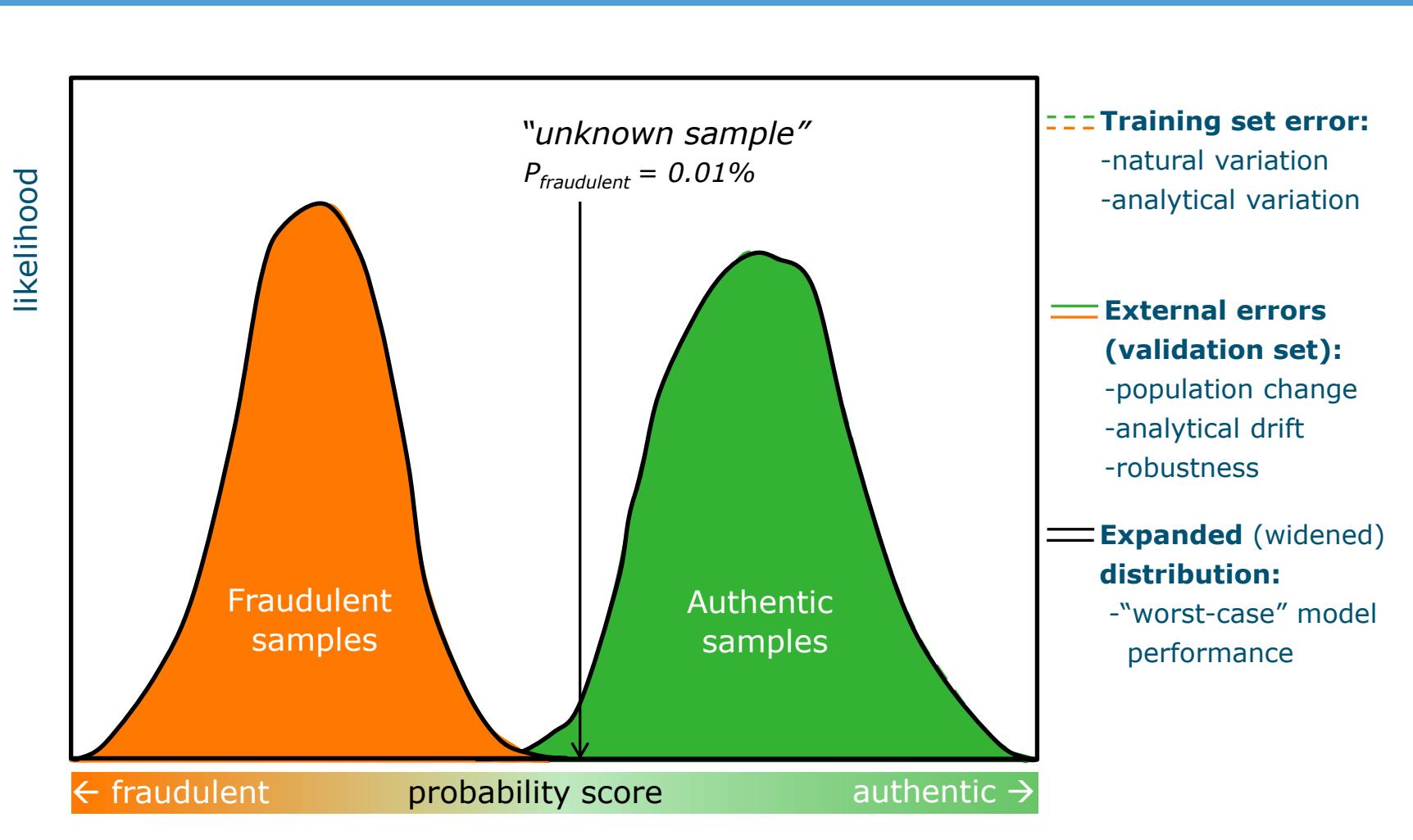


# Proposed validation scheme



# "Worst-case" probability distribution

Combining the sources of error: widening effect



# Proposed validation approach:

- Evaluates the sum of sample set + analysis + model
- Bases performance on probabilities rather than binary results
- Combines different sources of error into the “worst-case” overall (un)certainty profile
- Allows adding certainty statements to future samples

# Further steps:

- International input
- CEN workgroup envisaged
  
- Approach for quantitative validation of sample size and composition
- Add an “expanded measurement uncertainty” to the final distribution?
  
- Get a (test) method accredited – *in progress*
  
- Hopefully leads to a robust, useable and accepted system to validate non-targeted classification methods



# Thank you for your attention



[martin.alewijn@wur.nl](mailto:martin.alewijn@wur.nl)