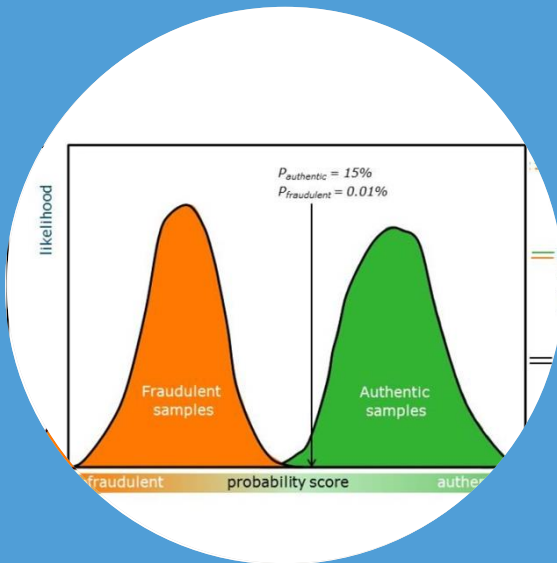


Validation of multivariate classification methods

Martin Alewijn

RIKILT – Wageningen University and Research (NL)



RIKILT

WAGENINGENUR

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

- 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)
- ...

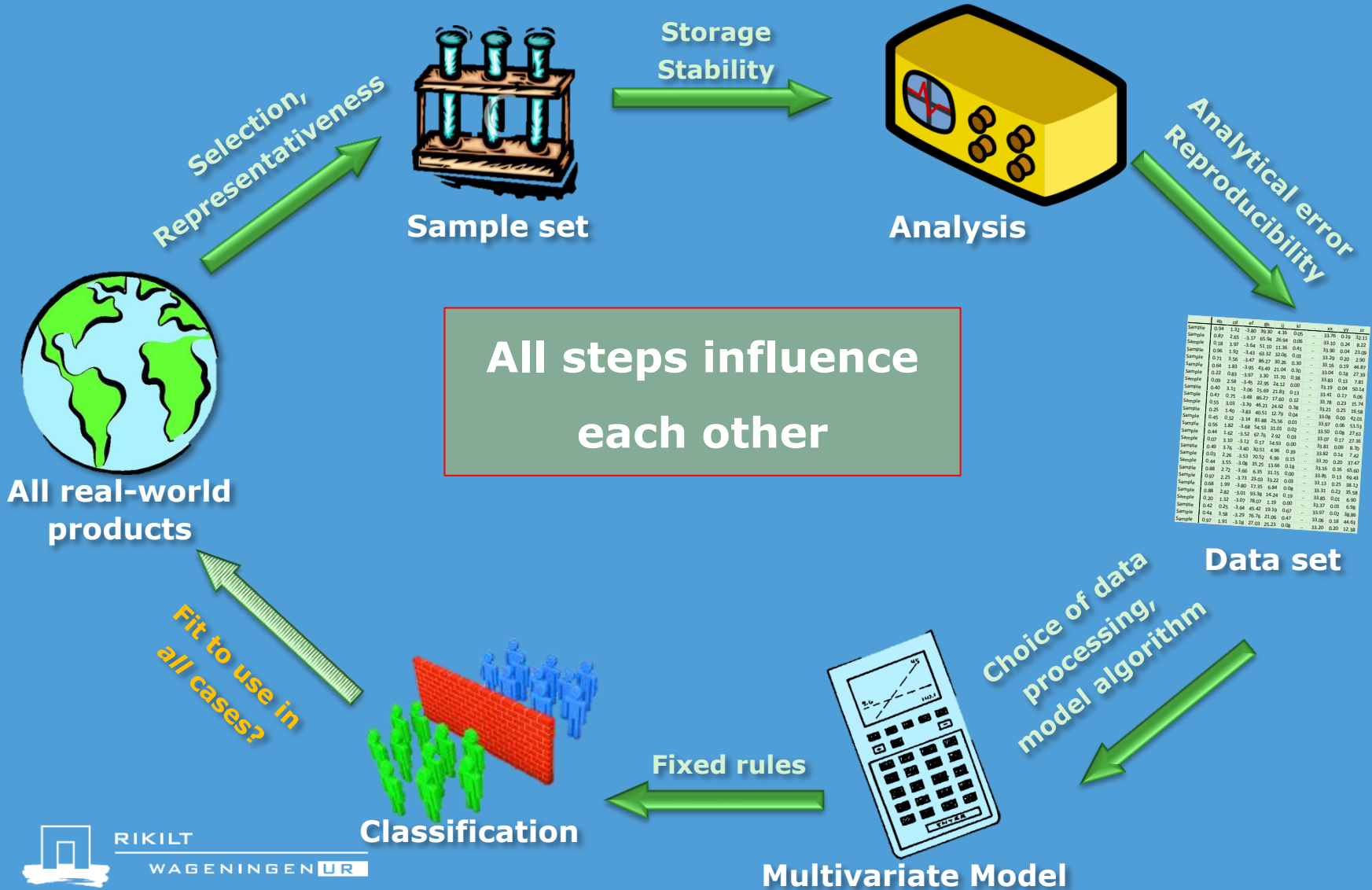
Analytical method performance evaluation

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

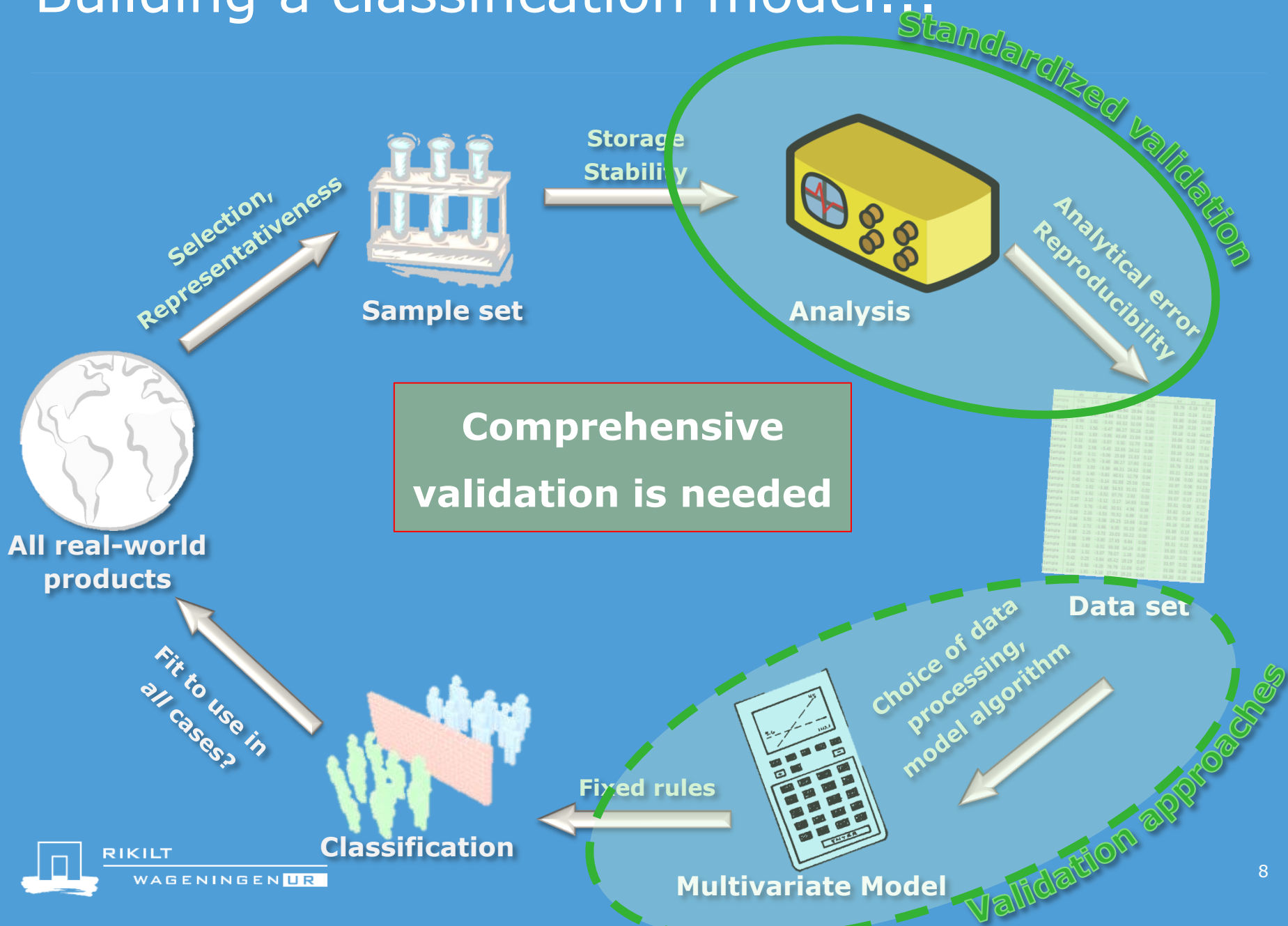
ISO 17025 & Com. Decision 2002/657/EC

Trueness
Accuracy/Recovery
Repeatability
Within-lab reproducibility
Measurement uncertainty
 CC_{α}
 CC_{β}
Limit of detection
Limit of quantification
Linearity
Specificity
Ruggedness
Stability

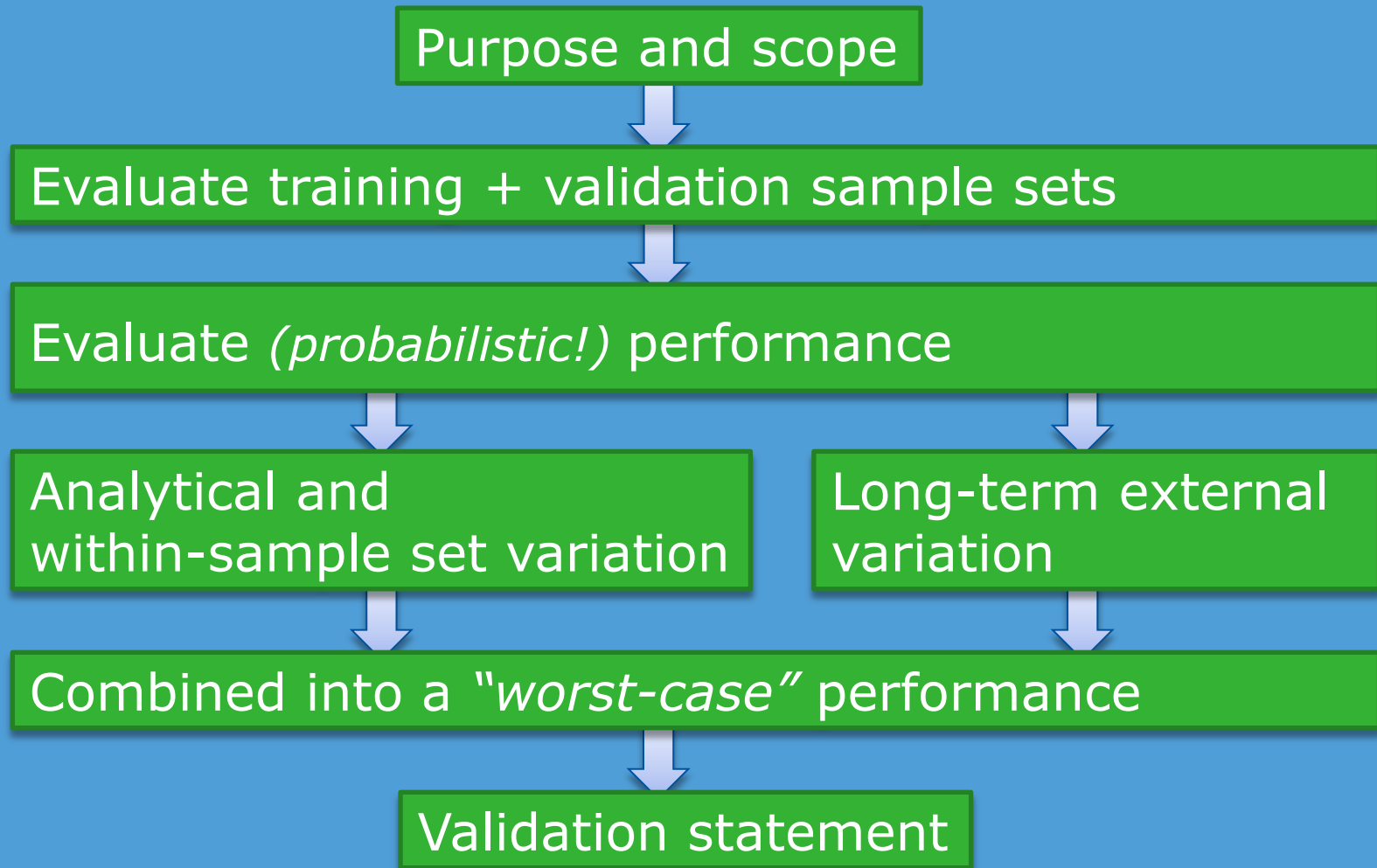
Building a classification model...



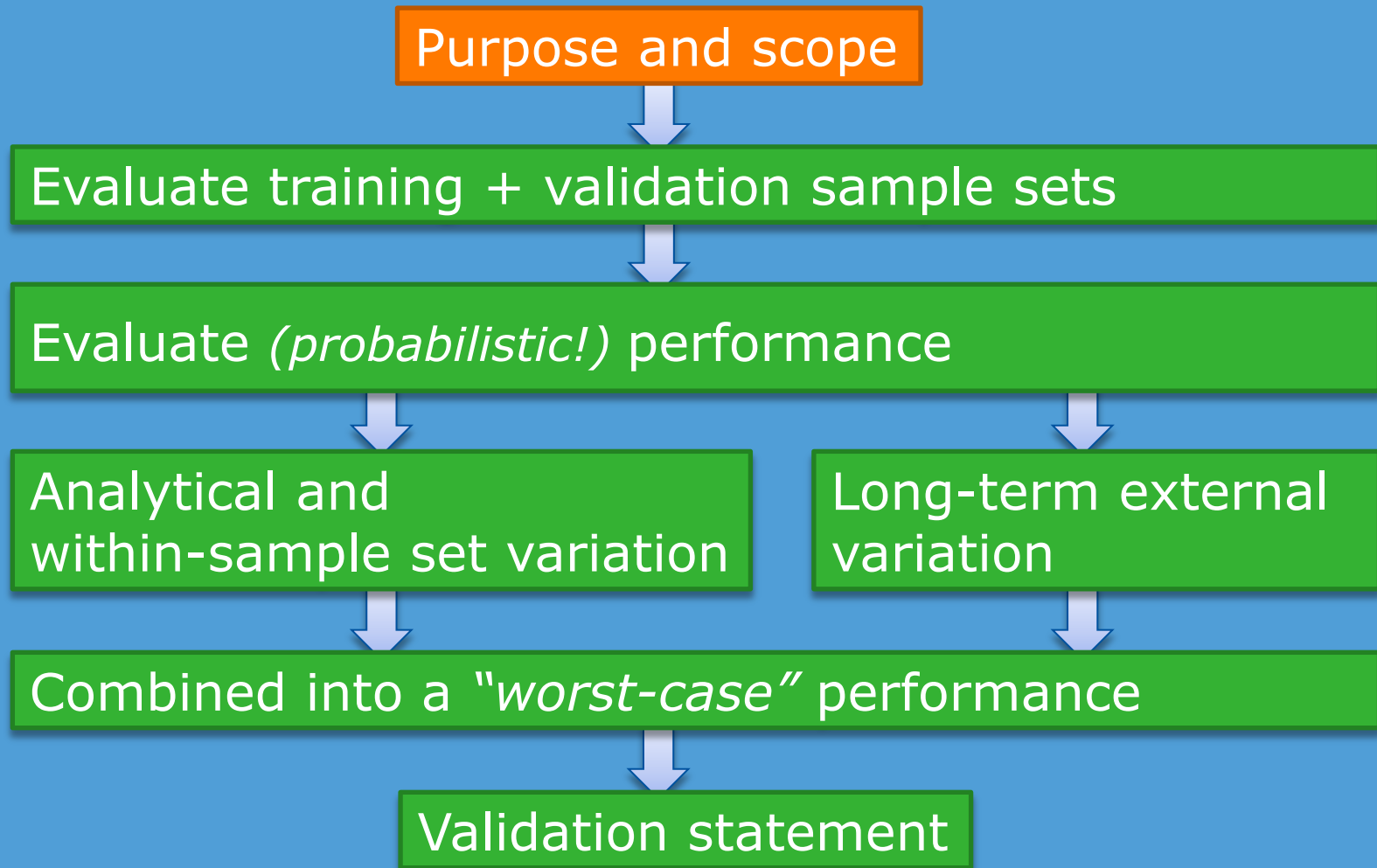
Building a classification model...



Proposed validation scheme



Proposed validation scheme

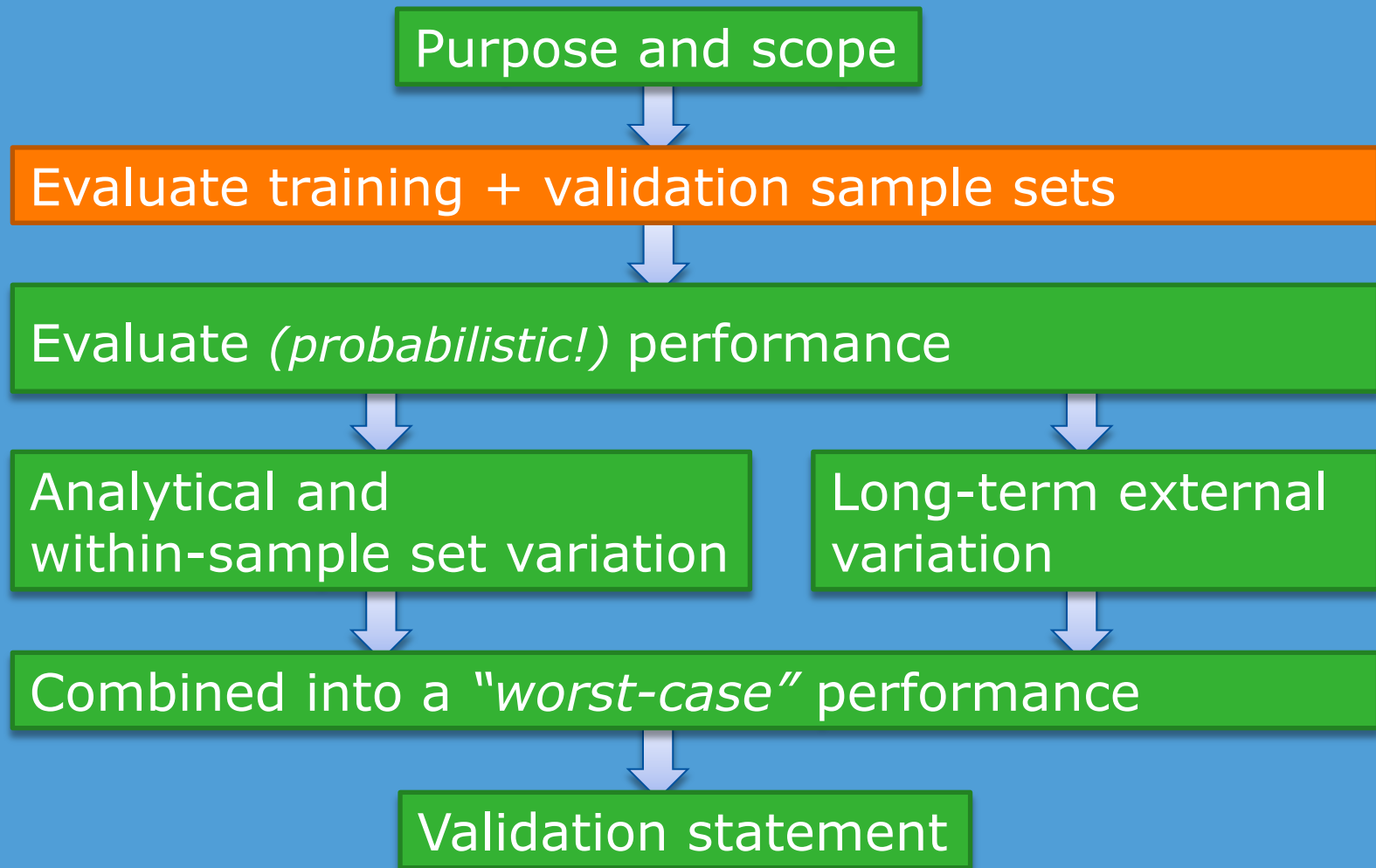


Purpose and scope

- 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:

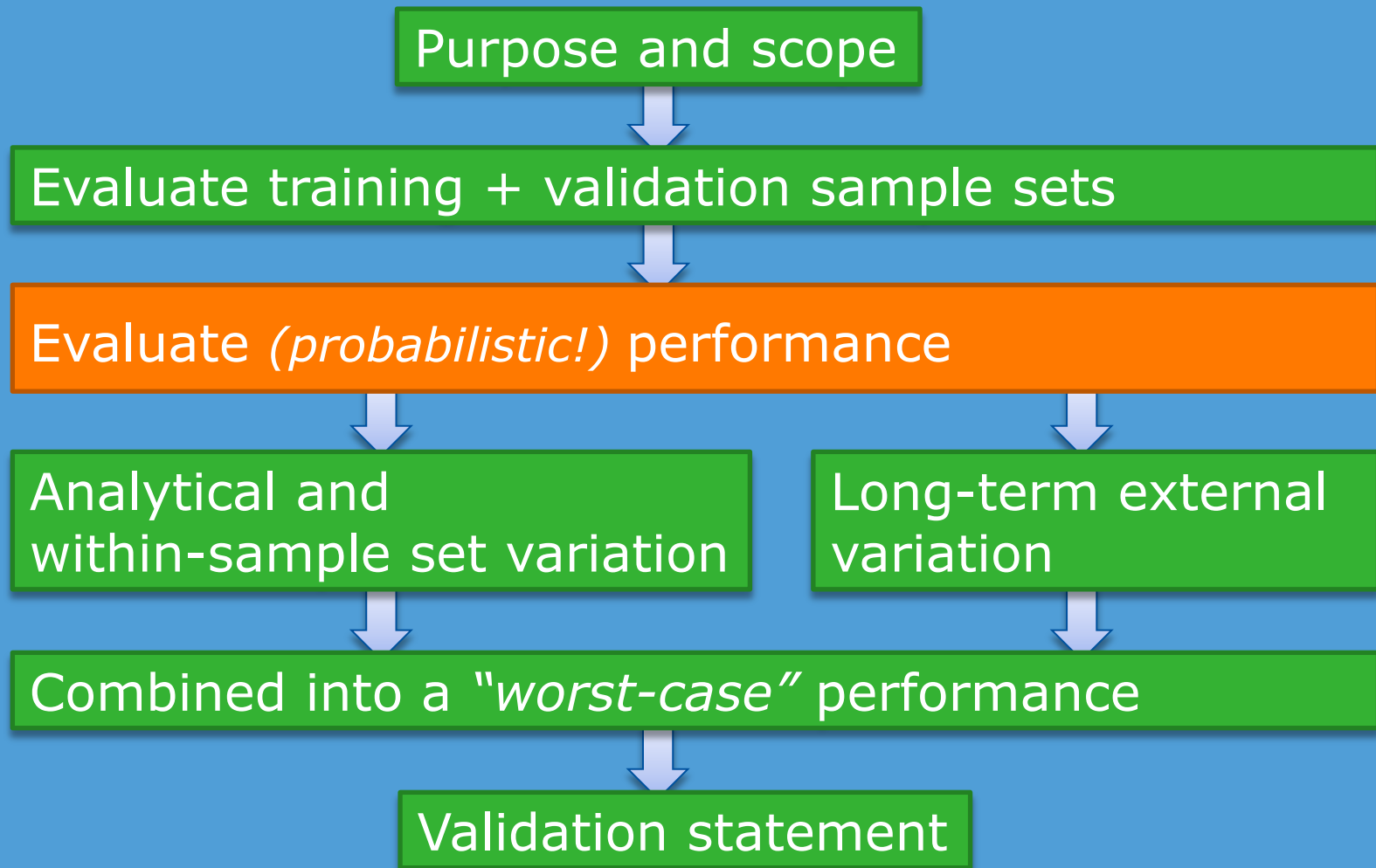
- Is sampling representative, true, and balanced for:
 - Target class
 - Relevant subgroups (season, variety, storage time, etc.)
- Is the sampling quantitatively sufficient?

Validation set: *(within scope)*

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

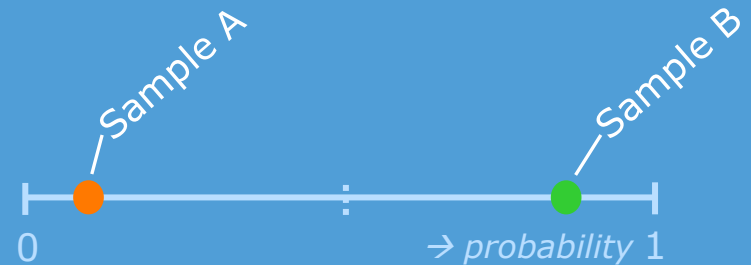
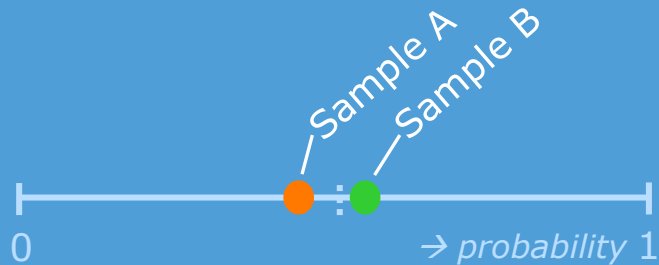


Proposed validation scheme

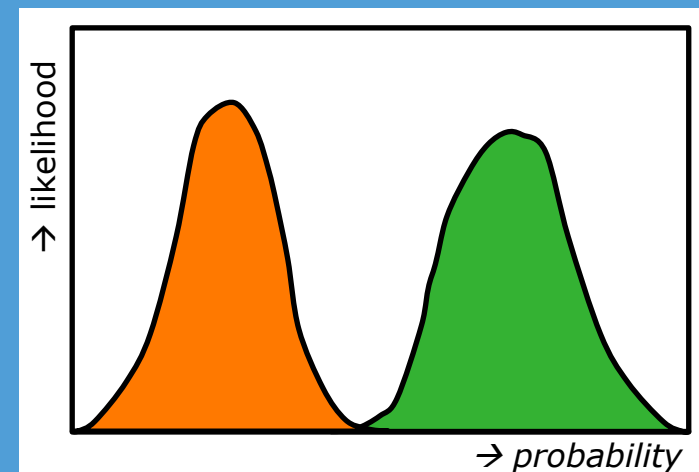


Performance evaluation

- Two examples, the same classification result:

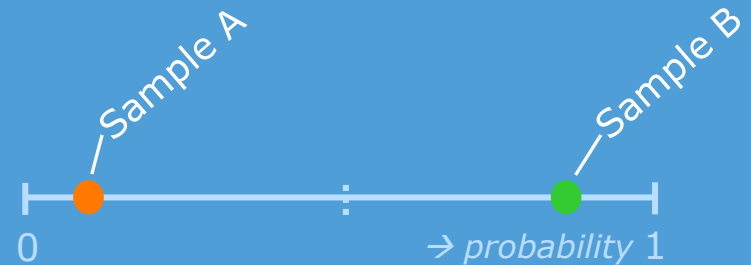
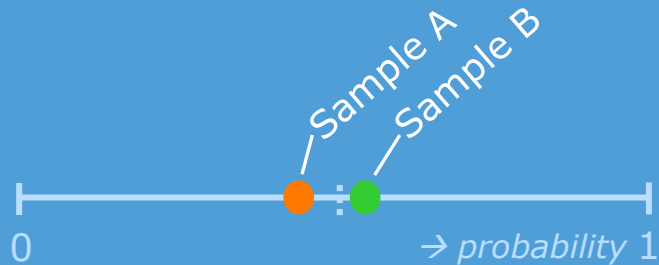


- During evaluation: **go probabilistic!**
- Every algorithm can be made probabilistic
- Obtain model probability distributions



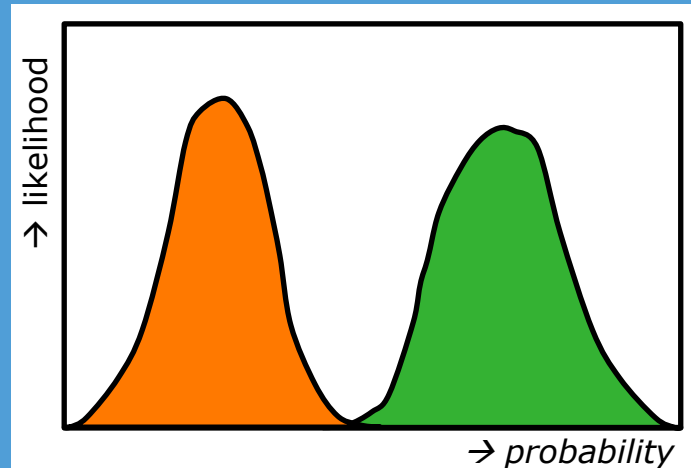
Performance evaluation

- Two examples, the same classification result:

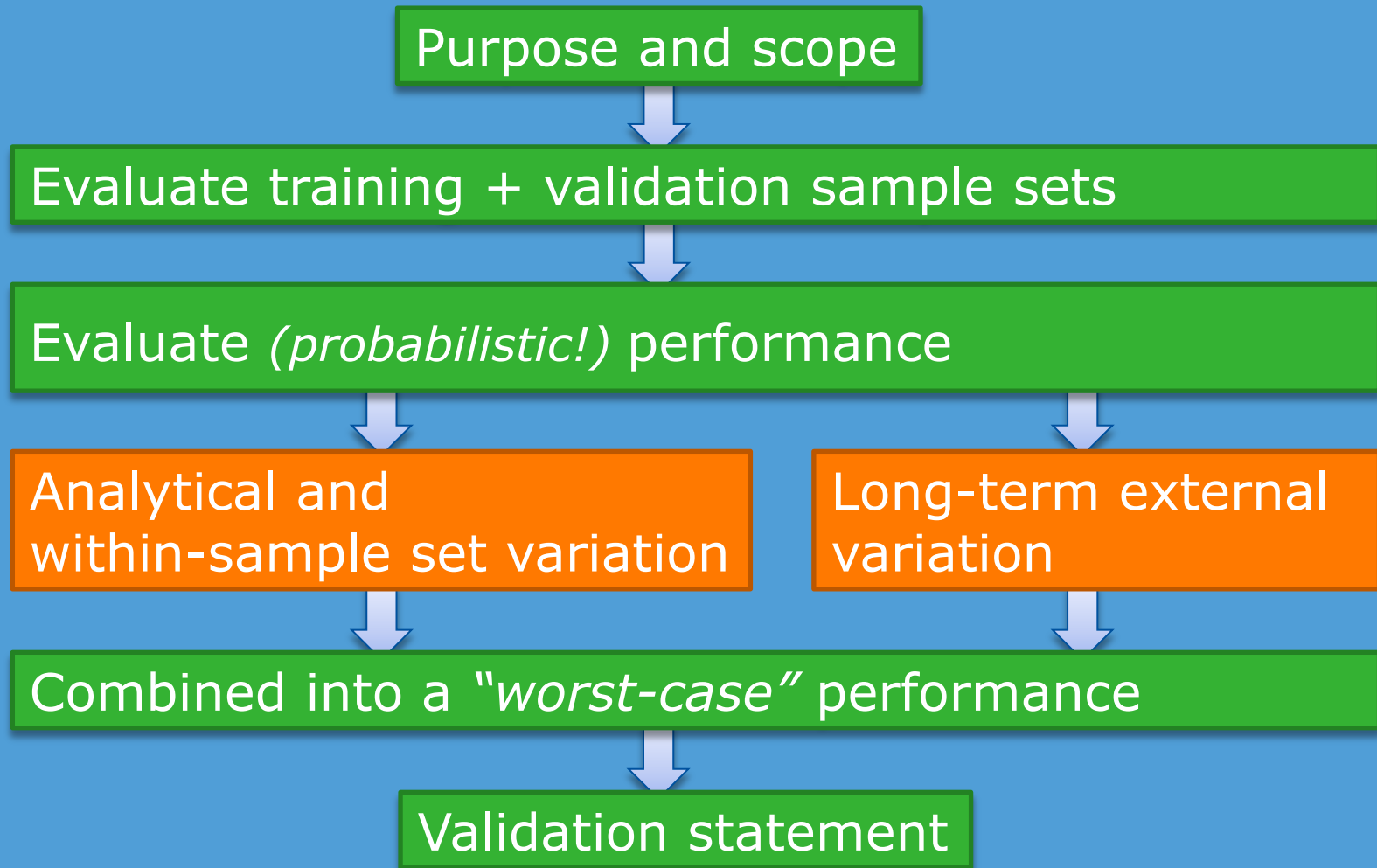


Probability score:

- a candidate for *multivariate* repeatability and reproducibility?



Proposed validation scheme

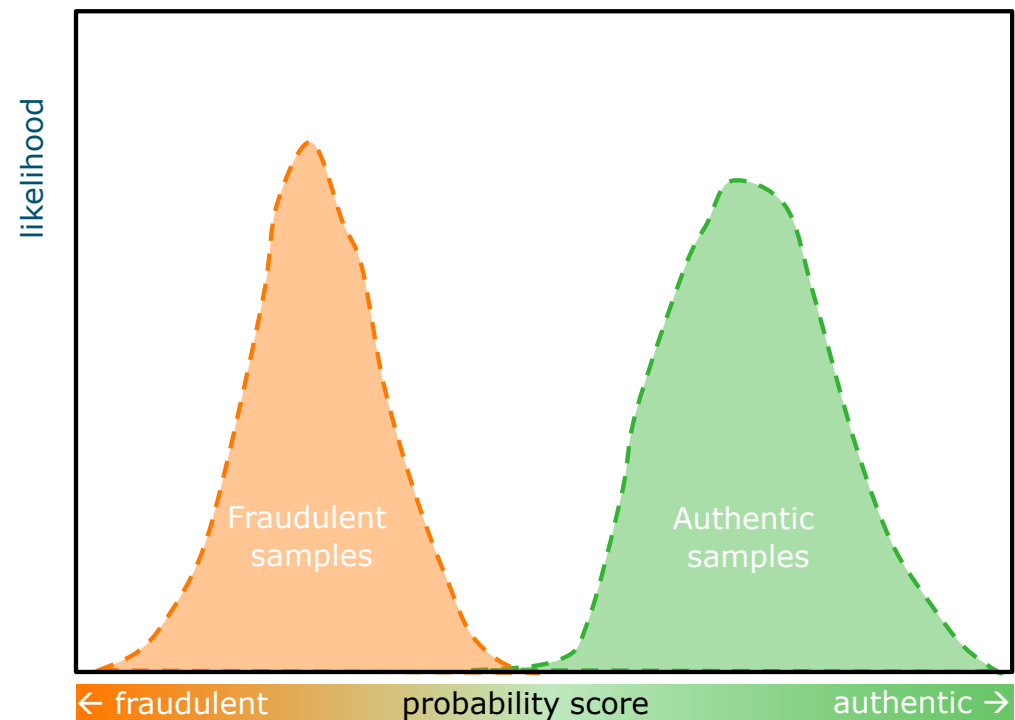


Sources of error/variation: training set

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

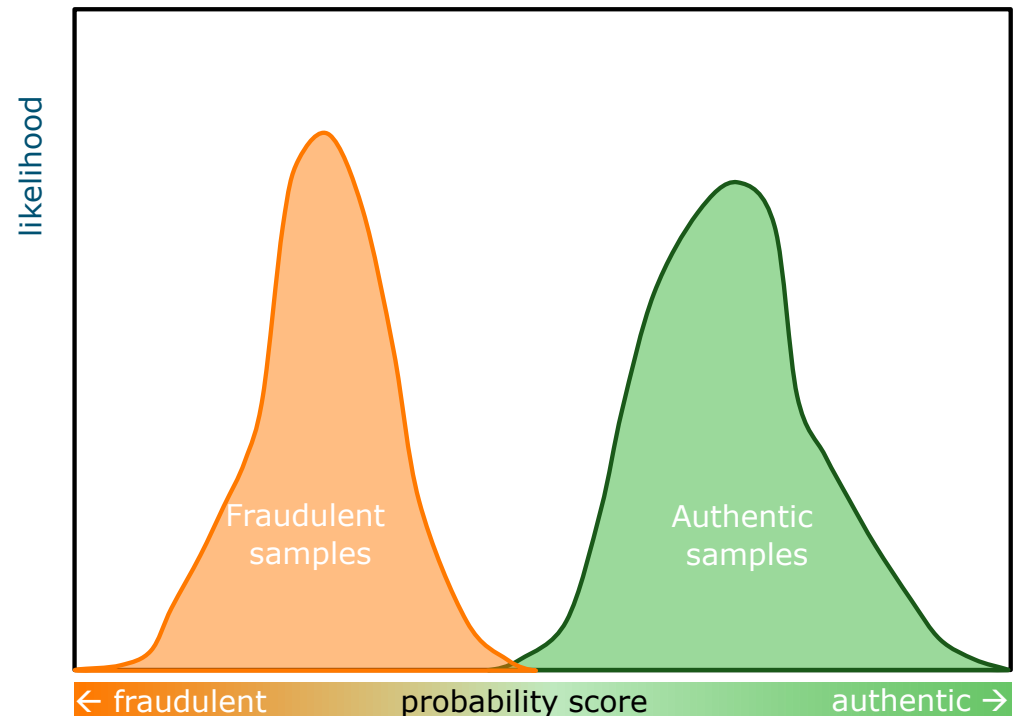


Sources of error/variation: validation set

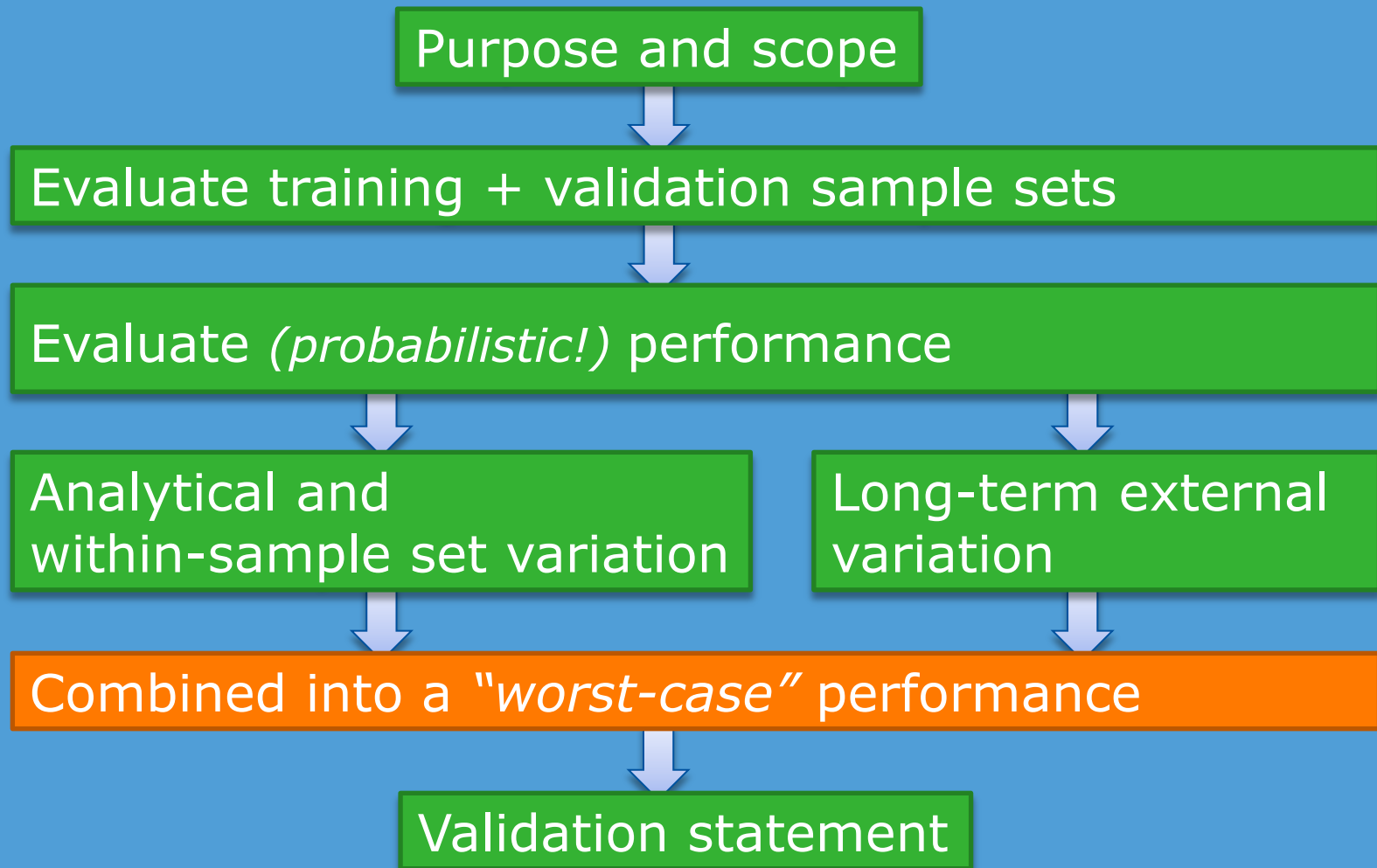
Additional variation (within scope!) due to obtained differences in:

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

Probability distribution obtained predicting the validation set samples by the model

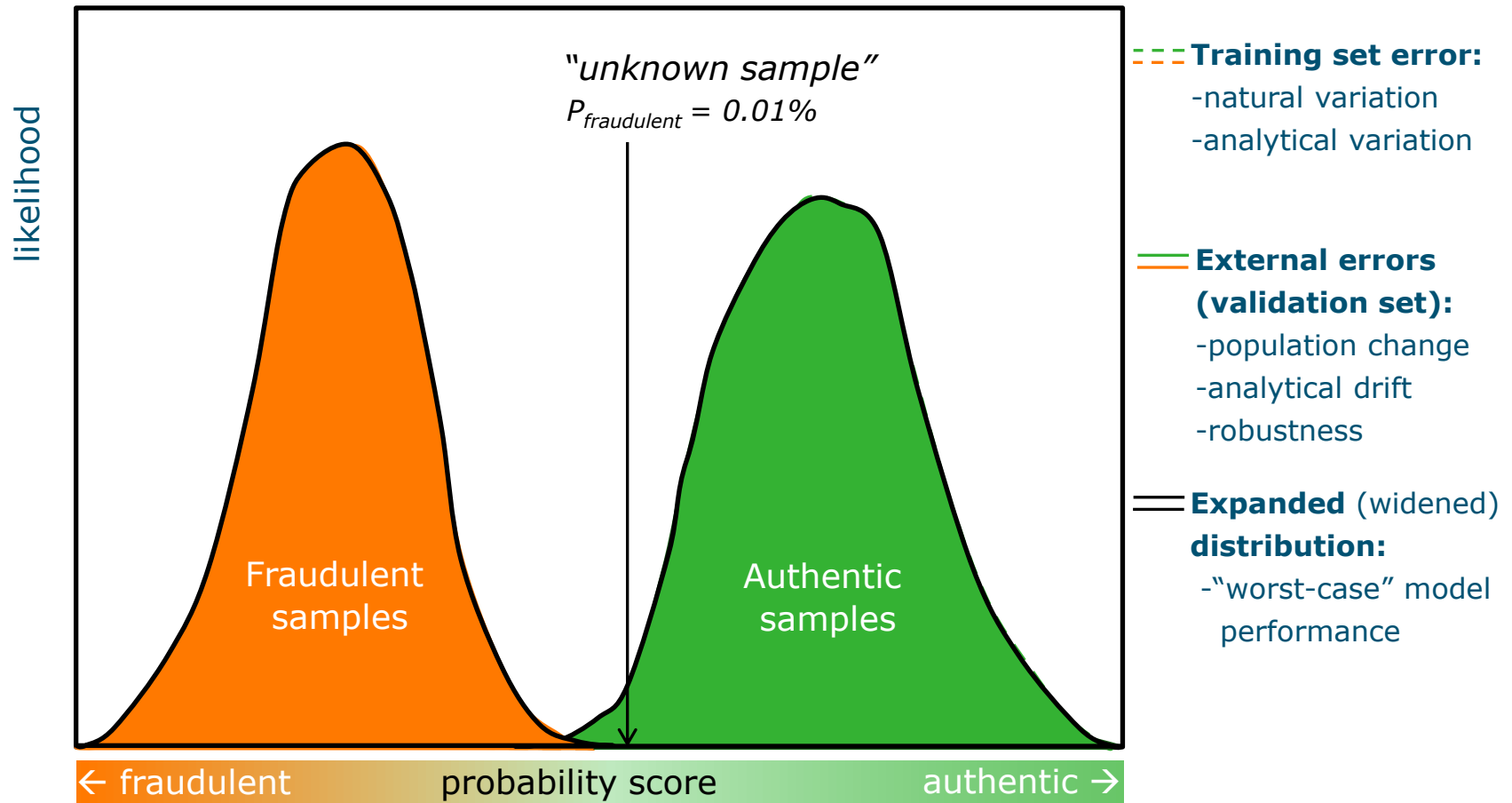


Proposed validation scheme



"Worst-case" probability distribution

Combining the sources of error: widening effect

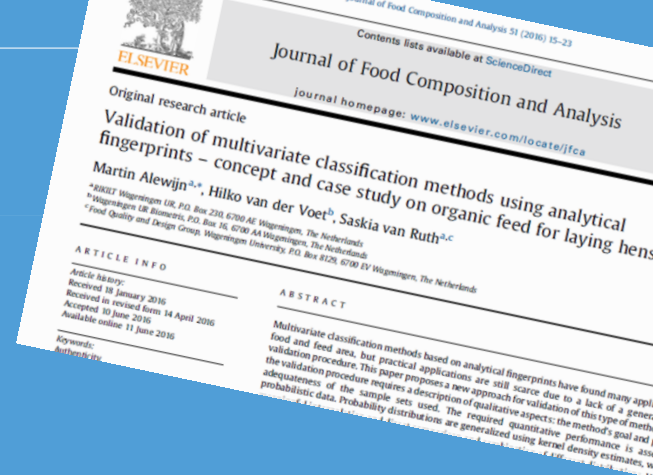


Proposed validation approach:

- 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



RIKILT

WAGENINGENUR