VIBRALYTICS · USE CASE

Multimodal Imaging of Pressed Seaweed: RGB, Hyperspectral and XRF Fusion

A pressed seaweed specimen is one object carrying three kinds of information: its visible morphology, its chemistry per pixel and its elemental composition. Each instrument sees that same sheet differently, with its own optics, resolution and point of view. This use case describes a complete computer vision workflow that brings the three views into a single aligned coordinate system and then fuses the spectral and elemental data into better predictive models. The workflow was built and validated end to end on a digital sheet with known ground truth, and each stage was stress tested against real world data.

1. The problem

A herbarium style seaweed sheet is measured with three modalities. **RGB** photography provides the visual and morphological information. **NIR hyperspectral imaging (HSI)** provides a full reflectance spectrum at every spatial location, producing a data cube with two spatial dimensions and one spectral dimension. **XRF** provides elemental information at a set of discrete measurement points distributed over the sheet.

Each instrument produces an image with different spatial coordinates, because the instruments have different optics, sensor positions, resolutions and viewing geometries. The XRF instrument reports its points in sheet millimetres, the RGB camera reports pixels, and the hyperspectral scanner has its own pixel grid and its own distortions. Before any modelling is possible, one question must be answered: **which pixel in one image corresponds to which pixel in the others?** Answering that question is the job of image registration.

Once the three views are aligned, every location on the specimen carries its morphology, its spectrum and its elemental composition. Only then can the spectral and elemental information be fused and modelled together. The acquisition geometry must also be documented carefully, including specimen position, distance, orientation, field of view and spatial resolution, and it matters whether the specimen can move, deform or change between acquisitions. If the seaweed moves between modalities, part of the problem is no longer a simple rigid geometric transformation.

This workflow was designed around four published methods, combined into a single pipeline: automated RGB, hyperspectral and chlorophyll fluorescence registration[1], a thermal and multispectral calibration and analysis package for plants[2], a systematic comparison of multispectral registration strategies[3] and a study of Vis-NIR plus XRF data fusion[4]. On top of those methods we added a practical layer of fiducial markers and a sheet millimetre frame that lets point XRF measurements be joined to hyperspectral spectra.

The complete workflow was validated on a **synthetic sheet with ground truth**: a digital specimen whose true geometry, masks and element values are known exactly, so every step can be scored against the truth before touching real specimens. All accuracy numbers in this document come from that validation run. Sections 9 to 11 describe how each stage is being checked against real data.

2. Step 1: Calibrate the cameras before anything else

Known instrumental errors must be corrected before registration, because registration is more reliable when it works with geometrically corrected images. Cameras can have optical distortion, particularly **radial distortion**, in which straight lines appear slightly curved. Radial distortion occurs when light rays bend more towards the edges of a lens than at its optical centre, and the effect is usually stronger toward the edges of the image than near the centre. In barrel distortion, for example, straight lines appear curved outward, as if the image were being expanded

from the centre. This creates problems for registration, because the same physical point may appear at slightly different pixel positions depending on where it is located in the image.

The intrinsic camera parameters are not entered manually. They are estimated mathematically during **camera calibration**. A checkerboard with known geometry is photographed from different positions and angles. The real coordinates of the checkerboard corners are known because the pattern dimensions are known, while the algorithm detects where those same corners appear in the camera image. The calibration algorithm then estimates the focal lengths f_x and f_y and the optical centre c_x, c_y that best explain how the known 3D points of the checkerboard are projected onto the 2D image. The **distortion coefficients** are estimated in the same process by minimising the **reprojection error**, which is the difference between the detected corner positions and the positions predicted by the mathematical camera model. Once these parameters are estimated, the images are **undistorted**: the pixels are remapped to compensate for the optical deformation, so the image geometry becomes closer to the real geometry of the object.

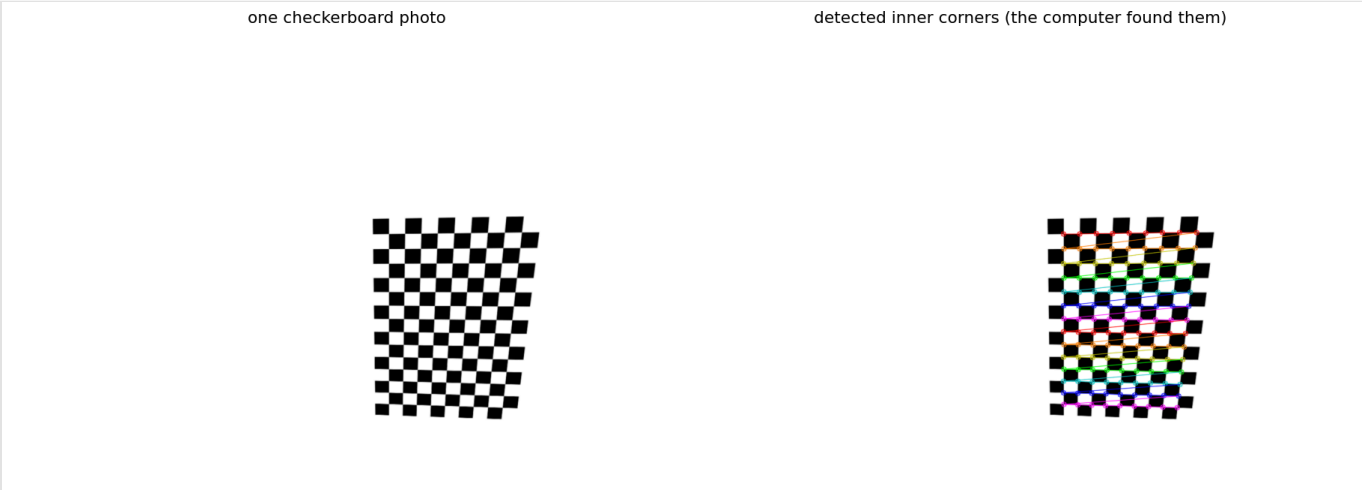

Figure 1. Camera calibration. One photo of the checkerboard (left) and the inner corners detected automatically by the calibration algorithm (right).

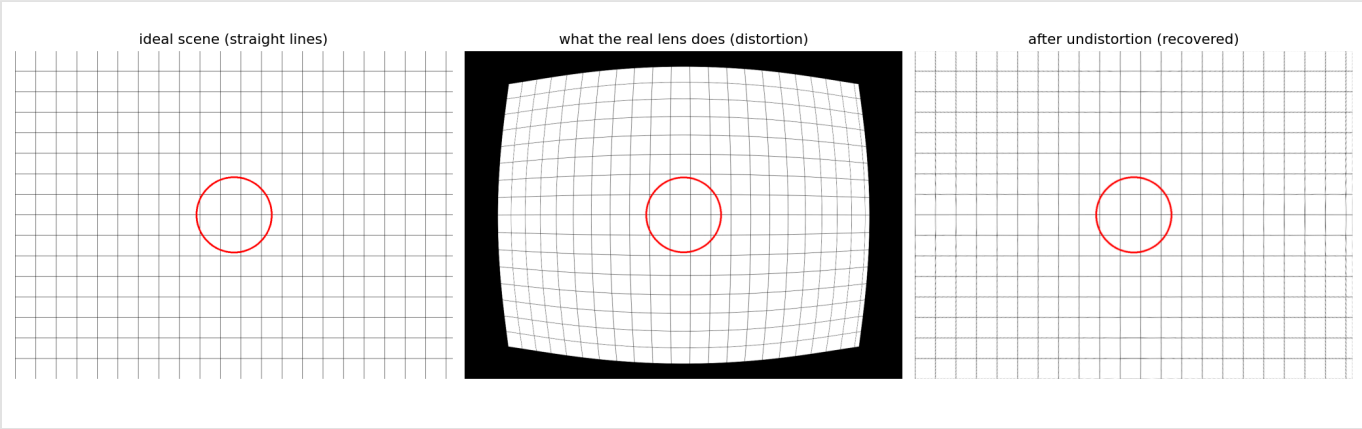

Figure 2. What calibration corrects. An ideal scene with straight lines (left), the same scene as captured by a real lens with barrel distortion (centre) and the recovered image after undistortion (right).

For XRF, appropriate elemental calibration is also required, together with consideration of background, sensitivity, measurement geometry and matrix effects when signal intensity is converted into concentration. The XRF instrument has no lens, so there is no lens distortion; its geometric treatment is the registration of its measurement points to the sheet frame, described in Step 7.

Illumination correction for the hyperspectral images

Illumination variation is caused by (i) non uniformness in the spatial distribution of radiation on the object plane due to the direction of the incident radiation, and (ii) changes in the intensity of the incident radiation through time, for example between images taken at different moments. Inaccurate retrieval of reflectance images and inaccuracies in image analysis and segmentation are some of the main problems associated with illumination variation.

Two corrections are applied. The first removes the **illumination gradient** across the image. A gradient reference image (G_{ref}) is estimated from small 4 by 4 pixel regions taken at the four corners of the original image: from the average intensity of those four regions, a smooth interpolated surface is created that represents how illumination changes across the field of view. G_{ref} is not a photograph of the scene; it is an estimated map of the unwanted illumination variation. The difference between G_{ref} and its minimum value (G_{dif}) quantifies how much brighter each position is compared with the least illuminated position, and this pattern is subtracted from the original image to obtain the gradient corrected image (G_{cor}).

The second correction normalises the overall illumination level and the sensor response. A **dark reference** image is taken by covering the lens, which measures the signal the sensor produces in the absence of light. A **white reference** is interpolated from the gradient corrected image. The corrected image is then computed as the gradient corrected data minus the dark reference, divided by the difference between the white and dark references, and multiplied by the known reflectance factor of the reference material, which in this workflow is 80 percent. The goal of both corrections is that differences between pixels mainly reflect real differences in the specimen rather than differences caused by uneven illumination or by the camera itself.

3. Step 2: One coordinate system for all modalities

One modality needs to define the reference coordinate system. **RGB is a reasonable candidate** when it provides the highest spatial resolution and the clearest morphology. The hyperspectral cube and the XRF points are then transformed into the RGB coordinate system. Another modality could be selected if the actual experimental setup makes it more appropriate, for example when the reference sensor does not cover the full specimen or when a specific instrument defines the geometry of the experiment.

To make the mapping precise, the physical sheet gets its own frame: **four fiducial markers** (ArUco codes) are placed at the corners of the sheet. Fiducial 0 defines the origin, fiducial 1 defines the X axis and fiducial 3 defines the Y axis. Because the positions of the markers on the physical sheet are known in millimetres, they link the real sheet to every image in which the markers are visible, and they give the XRF points (reported in millimetres) a path into the pixel world of the cameras.

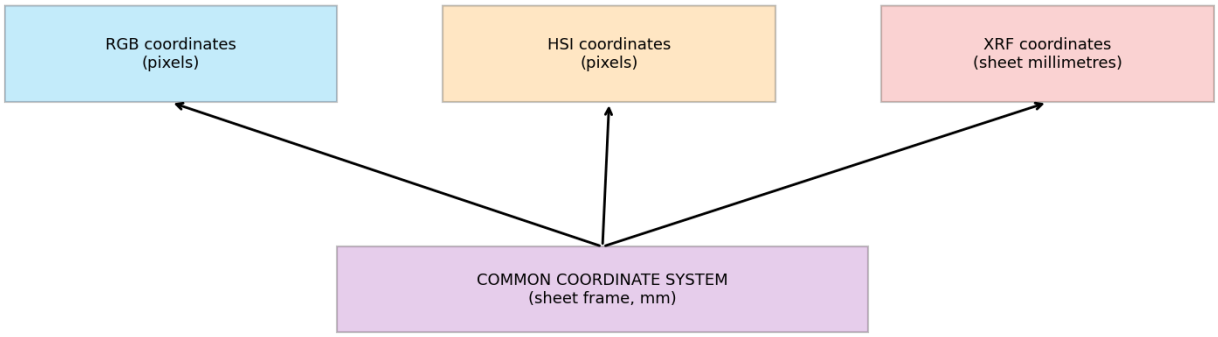

Figure 3. The three modalities live in different coordinate systems: RGB and HSI in pixels, XRF in sheet millimetres. Registration maps all of them into one common sheet frame.

Figure 4. The sheet frame. Four ArUco fiducials at the corners of the sheet define a millimetre coordinate system: fiducial 0 is the origin, fiducial 1 defines the X axis and fiducial 3 defines the Y axis.

A golden rule applies here: registration needs images that show **the same object**. One sheet is never registered against another sheet; every sheet has its own transformation matrix, obtained from its own fiducials.

4. Step 3: Coarse registration with image features

Registration means finding the transformation that makes the same physical point fall on the same pixel in two images. Image registration matches two or more images acquired from different viewpoints, sensors, times or fields of view, and it is what makes a combined analysis possible at all.

Several families of registration techniques exist. Frequency based methods such as phase correlation transform both images into the Fourier domain and estimate the transform from the global correlation peak, and the phase only correlation variant focuses exclusively on phase information, which makes it robust to intensity differences and noise. Feature based methods identify key points such as edges, corners or gradients in both images and compute the transformation from feature matches, typically filtered with the random sample consensus (RANSAC) algorithm. Mutual information based methods maximise a statistical dependency measure between the images. For this workflow we use the **feature based approach with SIFT and RANSAC** for the coarse step[3]:

SIFT finds distinctive points (corners and blobs) in both images. A matcher pairs the points that look alike. RANSAC then finds the transformation that fits the majority of pairs and discards the wrong ones, called outliers. This produces a **coarse alignment**: good, but not pixel perfect, because features are sparse and sometimes noisy.

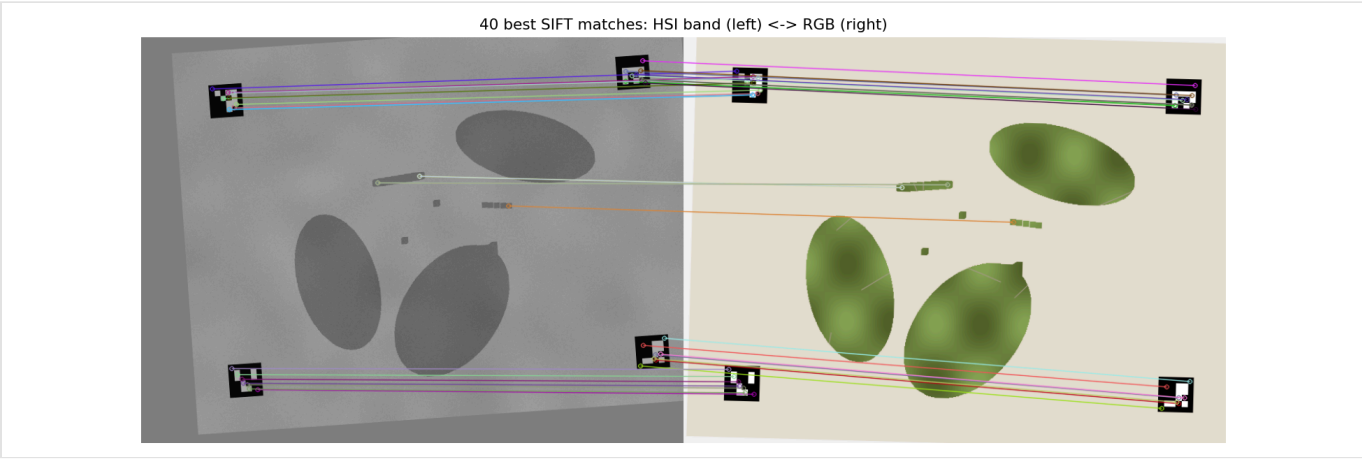

Figure 5. Coarse registration. SIFT keypoints are detected in the RGB reference and in the HSI band image, matched across the two views and filtered with RANSAC to fit the transformation.

Affine or homography?

When several spectral images of the same object are acquired, the object does not necessarily appear in exactly the same position or with exactly the same shape in every image. One reason is **foreshortening**: an object appears shorter or differently shaped when viewed from a particular angle. Another reason is **inter sensor separation**: the sensors are physically apart, so each observes the object from a slightly different viewpoint.

An **affine transformation** can translate, rotate, scale and skew an image. It has six degrees of freedom, and an important property is that it preserves parallel lines: two lines that are parallel before the transformation remain parallel after it. This model is appropriate when the cameras look roughly straight at a flat object.

A **perspective transformation**, also called a projective transformation or homography, has eight degrees of freedom. That does not make it automatically better in every situation, but the extra flexibility allows it to model perspective effects that an affine transformation cannot represent, such as different parts of the object appearing displaced by different amounts because of the viewing geometry.

In this pipeline both models are estimated and compared on real numbers before choosing. On the validation sheet the affine model reached a landmark error of 0.80 px with NCC 0.927 and Mutual Information 0.224, while the homography reached 0.79 px with NCC 0.927 and Mutual Information 0.226. The **homography modelled the viewpoint difference slightly better** and was selected.

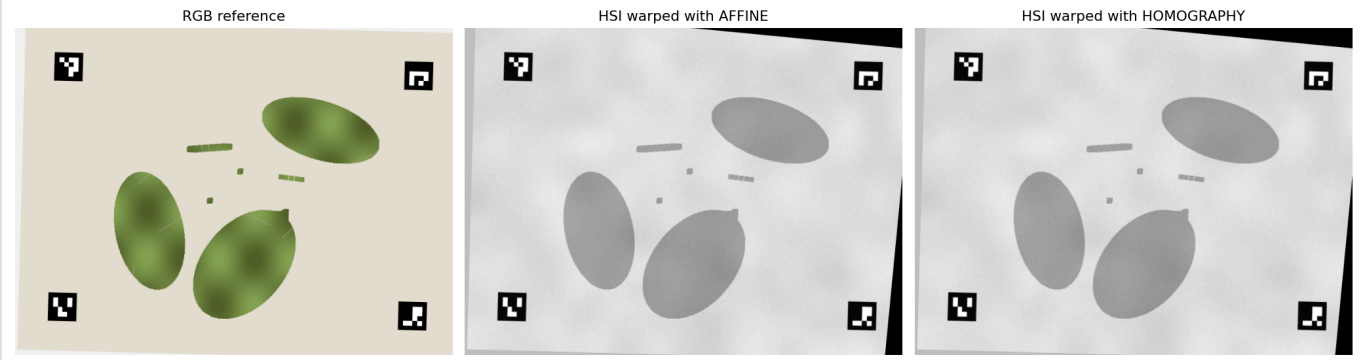

Figure 6. The HSI band warped onto the RGB frame with an affine transformation (centre) and with a homography (right), compared against the RGB reference (left).

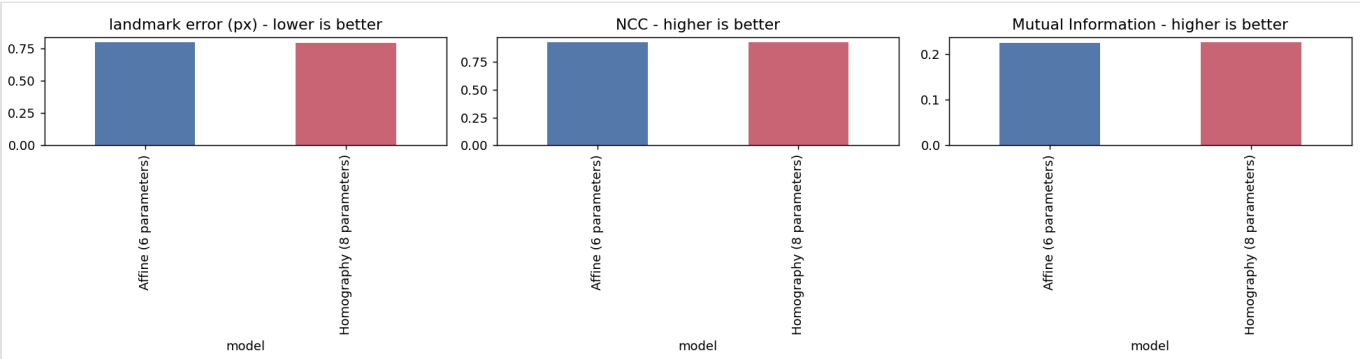

Figure 7. Affine versus homography scored by landmark error (lower is better), NCC and Mutual Information (higher is better). Both models perform similarly, with a small advantage for the homography.

Three ways to register: full image, masks or masked pixels

There is more than one way to feed an image pair into the registration step. Following the systematic comparison of Rana and colleagues[3], three strategies were implemented and measured: **(A) register the complete images** and later move the annotation points with the found transformation, **(B) register the binary specimen masks**, and **(C) register only the pixels inside the mask**. On the validation sheet, strategy A gave a landmark error of 0.79 px, strategy B 8.94 px and strategy C 12.58 px, so the full image strategy won here. In the original paper the mask based strategy performed best, which is exactly why the comparison must always be repeated on your own images instead of assumed.

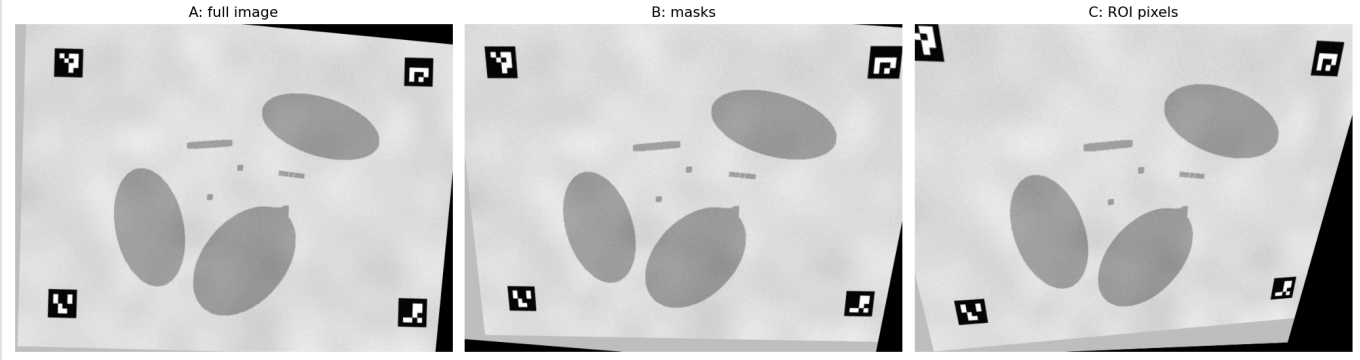

Figure 8. The three registration strategies of Rana and colleagues: A full image, B binary masks, C masked pixels.

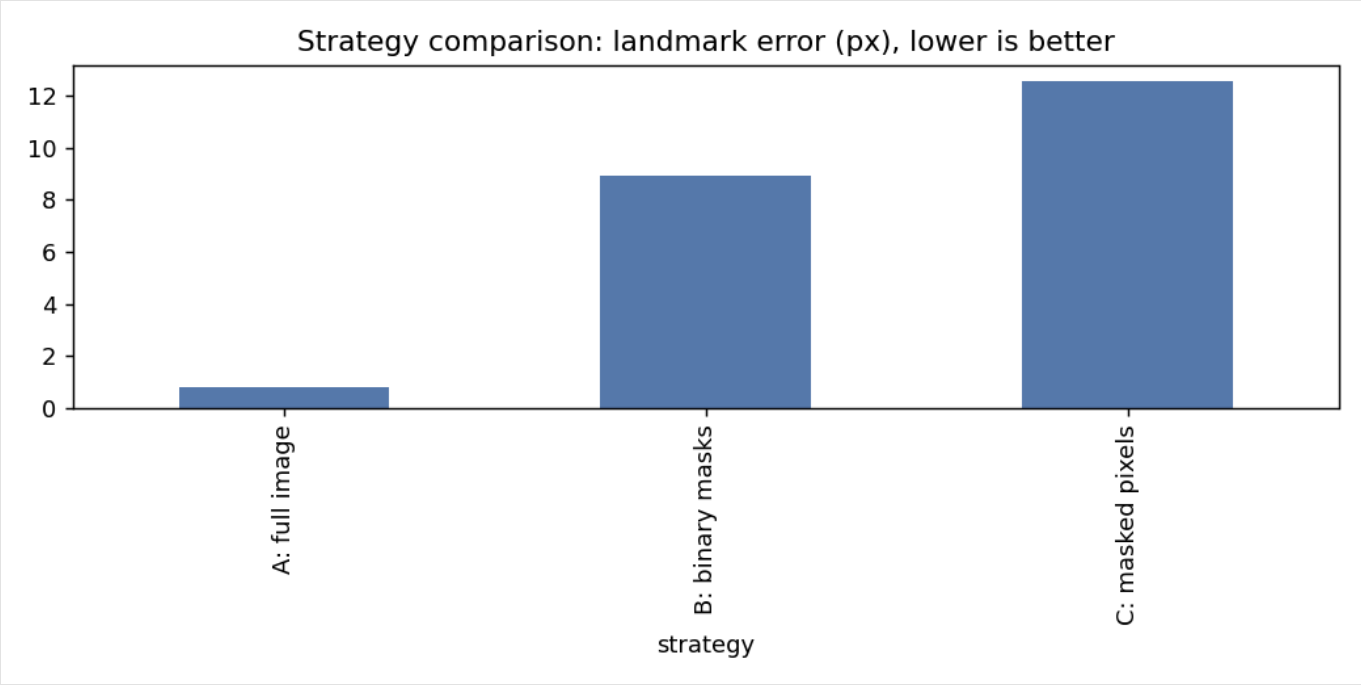

Figure 9. Strategy comparison by landmark error in pixels (lower is better). On this sheet, strategy A (full image) is clearly the best.

5. Step 4: Fine registration with intensity information

The feature based transformation is a coarse alignment: good, but not pixel perfect, because features are sparse and noisy. The fine step refines it with the actual image intensities.

Mutual Information (MI) is particularly useful for multimodal registration, because it does not require two images to have similar intensity values. RGB, HSI and XRF measure different physical properties, so the same location on the specimen can have completely different intensity values in each image. Instead of comparing intensities directly, MI looks for a statistical relationship between the measurements in the two images. Starting from the coarse alignment, the algorithm makes small adjustments to the transformation, and for each adjustment it examines the values that fall at corresponding pixel locations and calculates how strongly the measurements are statistically related. If the images are poorly aligned, pixels from unrelated physical regions are compared and the relationship is weaker. If the images are correctly aligned, the corresponding pixels represent the same physical locations, so the relationship becomes stronger. The algorithm keeps searching for the transformation that maximises Mutual Information, resulting in a more precise alignment.

In this pipeline the fine step maximises the **normalised cross correlation (NCC)** between the RGB reference and the warped HSI band, optimised numerically on top of the coarse homography, and MI is reported as the multimodal quality metric. On the validation sheet the refinement improved NCC from 0.927 to 0.928 and left the landmark error at 0.87 px, confirming that the coarse step was already close and that the fine step corrected the remaining small misalignment.

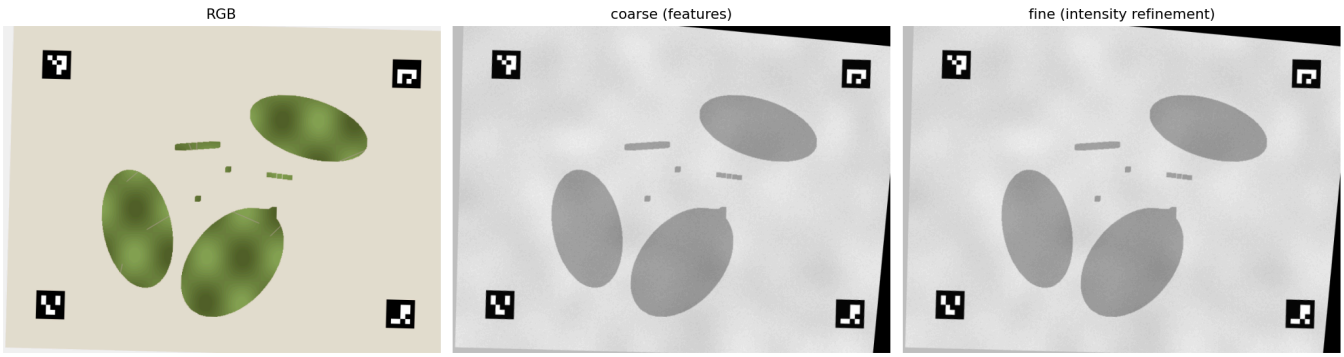

Figure 10. Fine registration. The RGB reference (left), the HSI band after the coarse feature based alignment (centre) and after the intensity based refinement (right).

6. Step 5: Segmentation, separating the specimen from the background

For RGB segmentation the objective is to separate the specimen from the background, such as paper, labels or shadows. A useful method is the **Excess Green (ExG) index**, which emphasises green areas using the RGB channels (ExG = 2G minus R minus B). This is particularly useful for plant or seaweed specimens because they are often greener than the surrounding background. Once the ExG image has been calculated, **Otsu thresholding** automatically determines a threshold that separates specimen pixels from background pixels. The result is a binary mask, which is then cleaned with **morphological operations** (erosion and dilation, implemented as opening and closing) to remove small isolated noise, eliminate unwanted small regions and fill small holes inside the specimen, followed by a minimum component size filter.

For HSI segmentation the approach can exploit information outside the visible range. One possibility is the **NDVI** (Normalised Difference Vegetation Index) computed from a near infrared band and a red band, since vegetation reflects near infrared radiation strongly. In this workflow, however, the final specimen mask is created from the RGB image, because RGB provides the best spatial quality. Once the mask exists in the RGB frame, it is **transferred to the registered HSI and XRF layers**: the same spatial region identified as specimen in RGB is kept as specimen everywhere, while background pixels are excluded from the analysis.

If standard methods are not accurate enough, an automatic computer vision model such as **YOLOv8-seg** can be investigated: the model would detect the specimen and produce the segmentation mask directly.

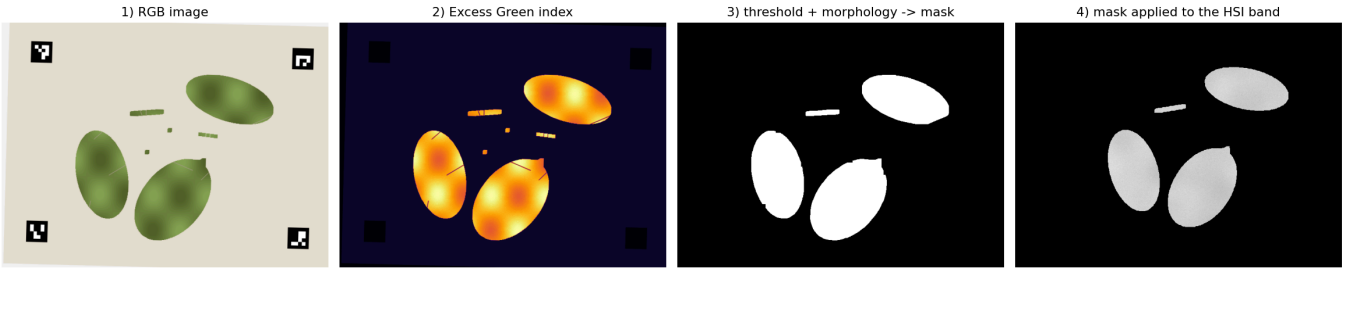

Figure 11. Segmentation. The RGB image (1), the Excess Green index (2), the binary mask after Otsu thresholding and morphological cleaning (3) and the mask applied to the registered HSI band (4).

Vegetation indices were also evaluated beyond ExG and NDVI, and their behaviour was checked on real spectra as well (Figure 12).

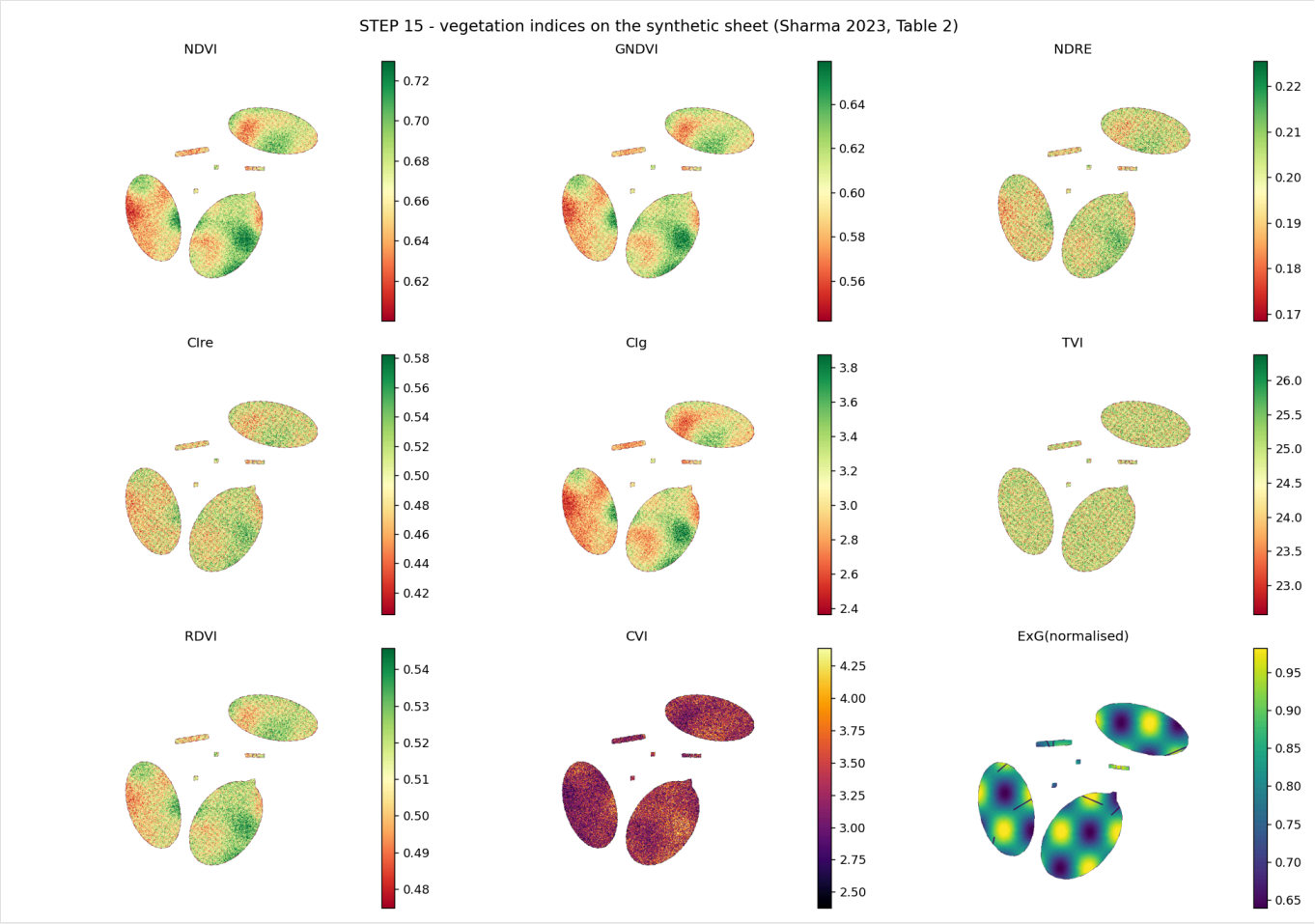

Figure 12. Vegetation indices computed on the synthetic sheet (top rows) and the distribution of index values on 500 real GLORIA spectra, which are mostly water and coastal samples and therefore sit close to zero.

7. Step 6: Did the registration actually work?

It is not enough to assume that the transformation succeeded; the alignment must be verified. The evaluation considers three aspects.

Geometric quality tells us whether the specimen is in the same position in the different images. The **Overlap Ratio (OR)** measures how much of the specimen area coincides between the registered images, **centroid displacement** measures the distance between the centres of the specimen in the two images, and **landmark error** measures the distance between identifiable corresponding points. **Multimodal quality** is evaluated with **Mutual Information**, which indicates whether the modalities contain statistically related information at corresponding spatial locations even when the images look very different. Finally, the **failure rate** measures how often the registration algorithm produces an incorrect result. Considering both geometric and multimodal metrics gives a much stronger evaluation, because a single metric may indicate good alignment while missing another type of error.

On the validation sheet the aligned masks reached an overlap ratio of 0.989 and a Dice coefficient of 0.994, the centroid displacement was 0.70 px, the landmark error 0.87 px, and the image pair showed NCC 0.928 with Mutual Information 0.225. A simpler coefficient based check (a correlation between corresponding regions) can serve as a quick additional sanity check, but it should be treated as an option rather than the main evaluation, especially for strongly multimodal data, because different modalities may not have a direct intensity relationship.

The exact order of registration, segmentation and registration evaluation may vary with the workflow, since segmentation can be performed before or after registration, either to support the registration or for the subsequent scientific analysis.

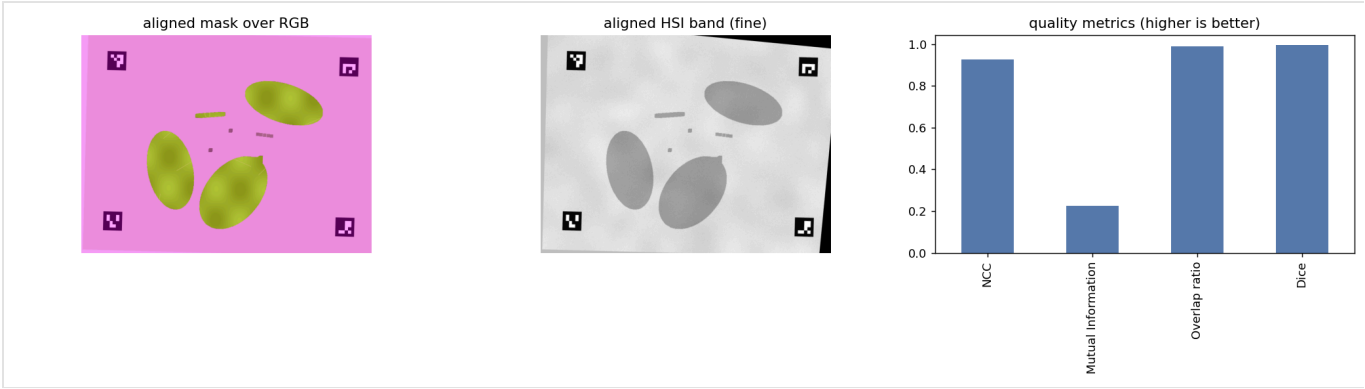

Figure 13. Registration evaluation. The aligned specimen mask over the RGB image, the aligned HSI band after fine registration, and the resulting quality metrics.

8. Step 7: Bringing the XRF data in

The XRF instrument reports elemental concentrations at discrete points on the sheet, expressed in millimetres, together with the beam footprint and the positioning uncertainty. The XRF modality has no lens and therefore no lens distortion; its calibration is the energy calibration described in Step 1 plus the geometric registration of its measurement points to the sheet frame.

The join works through the fiducials. The four corner markers, known in millimetres on the sheet and detected in pixels in each image, define an affine mapping from sheet millimetres to image pixels. Each XRF point is mapped into the registered images, and around every point a circular region of interest is extracted with a radius that covers the beam diameter plus the positioning uncertainty. Within that region the mean NIR reflectance spectrum is extracted from the hyperspectral cube.

The result is a **joined table with one row per XRF point**: sheet coordinates, pixel coordinates, elemental concentrations and the full NIR spectrum at that location. That table is the bridge between imaging and chemistry, and it is the input of every fusion model in the next section. On the validation sheet, the mean position error of the mapped XRF points was 0.84 px with a maximum of 1.30 px, and the spectra extracted at the mapped positions agreed with the ground truth with a correlation of r = 1.000.

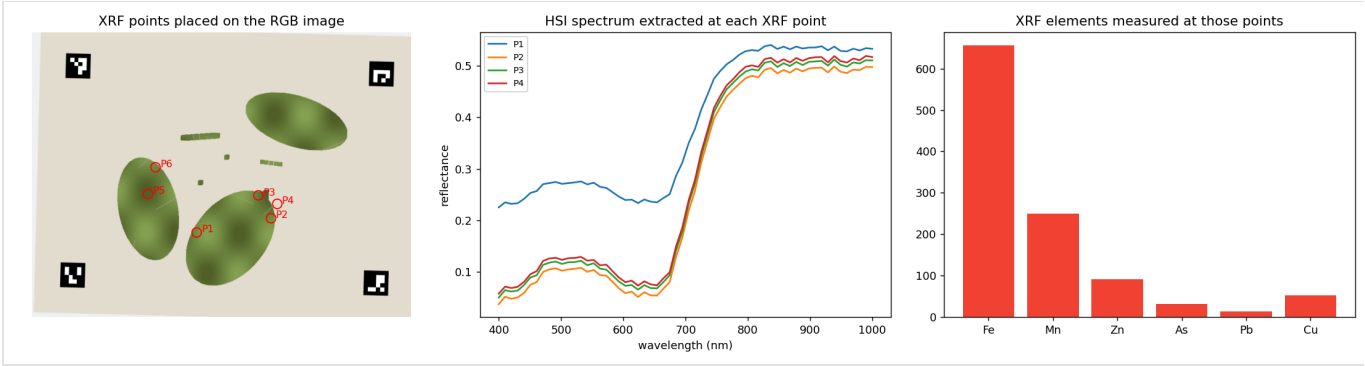

Figure 13. Joining XRF and HSI. The XRF measurement points placed on the RGB image (left), the NIR reflectance spectrum extracted at each point (centre) and the elemental concentrations measured at those same points (right).

9. Step 8: Fusing the data and evaluating the models

Once the spectra and the elemental concentrations are aligned, the data can be fused. Several strategies exist, and they were compared systematically rather than assumed[4].

In **early fusion**, the spectral variables and the elemental concentrations are concatenated into one combined dataset that feeds a single predictive model. Other strategies perform the fusion at the **model level**: an individual model is developed for each sensor, producing separate predictions, and those predictions are then combined by **model averaging** (GR2) or by **stacking** (LS3). The least squares combination is particularly interesting because it considers not only the individual predictions but also the covariance between their prediction residuals: if the individual sensor models tend to make correlated errors, this shared error structure can be incorporated into the final combination, allowing the contribution of each sensor to be determined while accounting for the relationship between their errors. As an extension, a **residual structure meta model** was investigated: it feeds the difference between the two model residuals, its absolute value and a local residual correlation into a meta learner, exploring whether the spatial pattern of disagreement between sensors is itself informative.

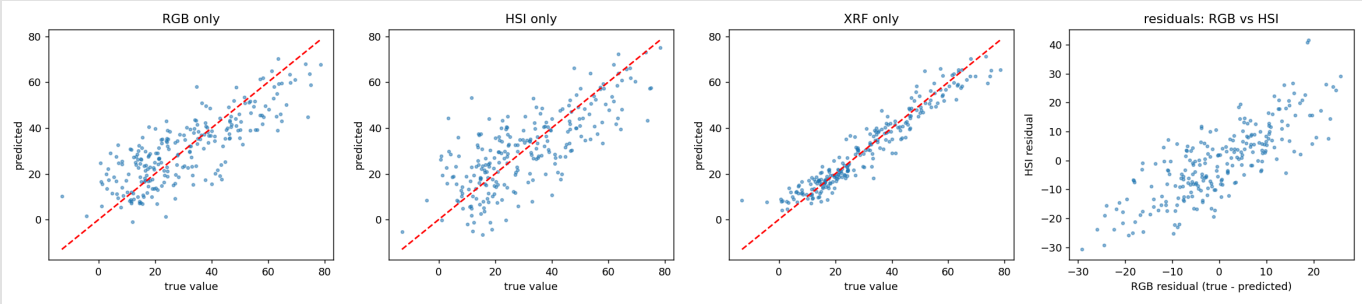

Figure 14. Single sensor models. Predictions versus true values for each individual model, and the relationship between the residuals of the different models, which is the basis of the residual aware combinations.

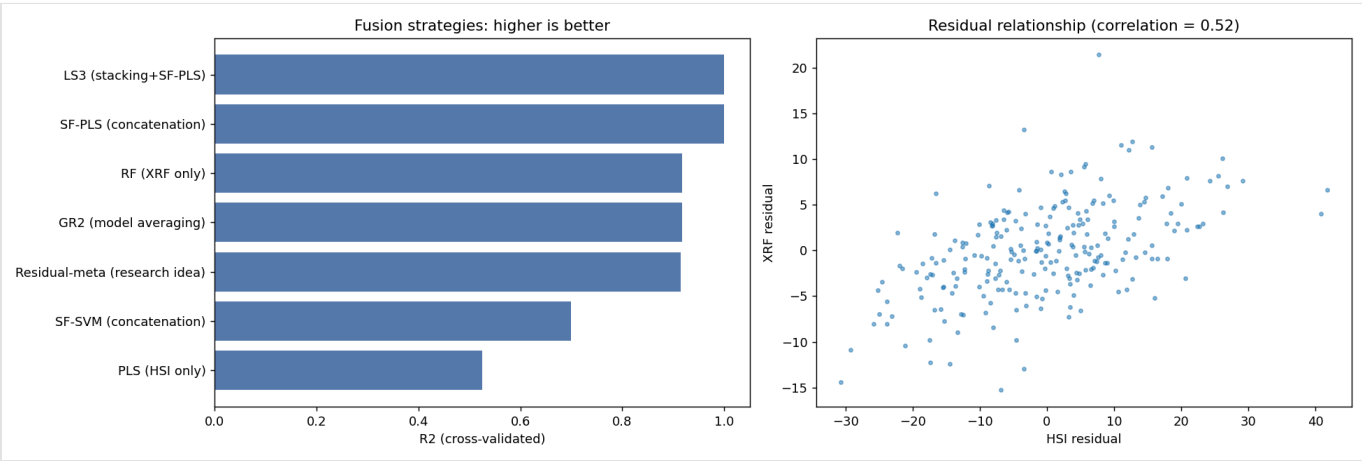

Figure 15. Data fusion. Cross validated R2 of the fusion strategies (left, higher is better) and the residual relationship between the individual HSI and XRF models (right).

On the validation cohort (cross validated), the single sensor models reached R^2 = 0.525 for PLS on the spectra alone and R^2 = 0.918 for a random forest on the elements alone. The fusion strategies clearly outperformed both individual models: early fusion by concatenation (SF-PLS) reached R^2 = 0.9998, and the stacking combination (LS3) reached R^2 = 0.9998, with the residual meta model at 0.913. The full comparison is shown in the next figure.

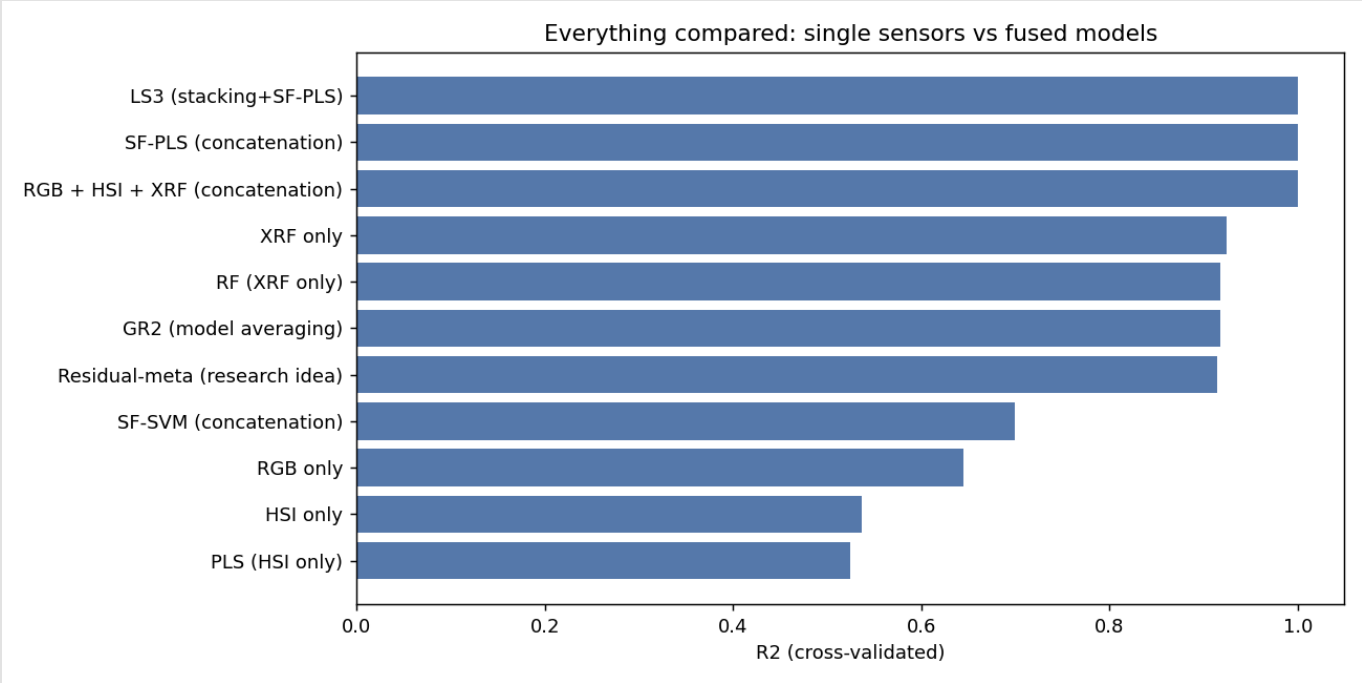

Figure 16. Full model comparison by cross validated R2. The fused models sit at the top of the ranking.

The final question is not simply which model has the lowest error. It is equally important to understand **what information each modality contributes**, when multimodal integration improves the result, and whether the improvement has a biological interpretation. In this synthetic cohort most of the predictable signal was carried by the elemental data, and the fusion strategies that kept the spectral and elemental blocks separated and then combined their predictions performed best. On real specimens the ranking must be re measured with nested, per specimen cross validation, and the fusion cohort used here must be replaced by real measurements.

10. Real data behind every stage

No public dataset combines RGB, hyperspectral and XRF measurements of the same object, so each stage of the pipeline was validated against a real dataset of its own while the end to end behaviour was proven on the synthetic ground truth sheet.

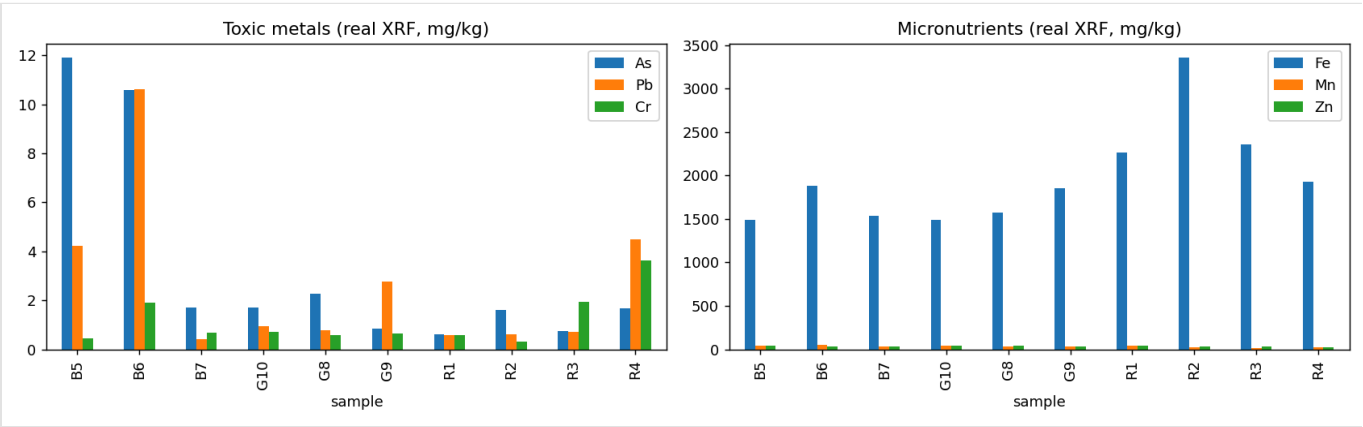

Figure 17. Real XRF measurements in ten macroalgae samples measured by portable EDXRF: toxic metals (left) and micronutrients (right), in mg per kg.

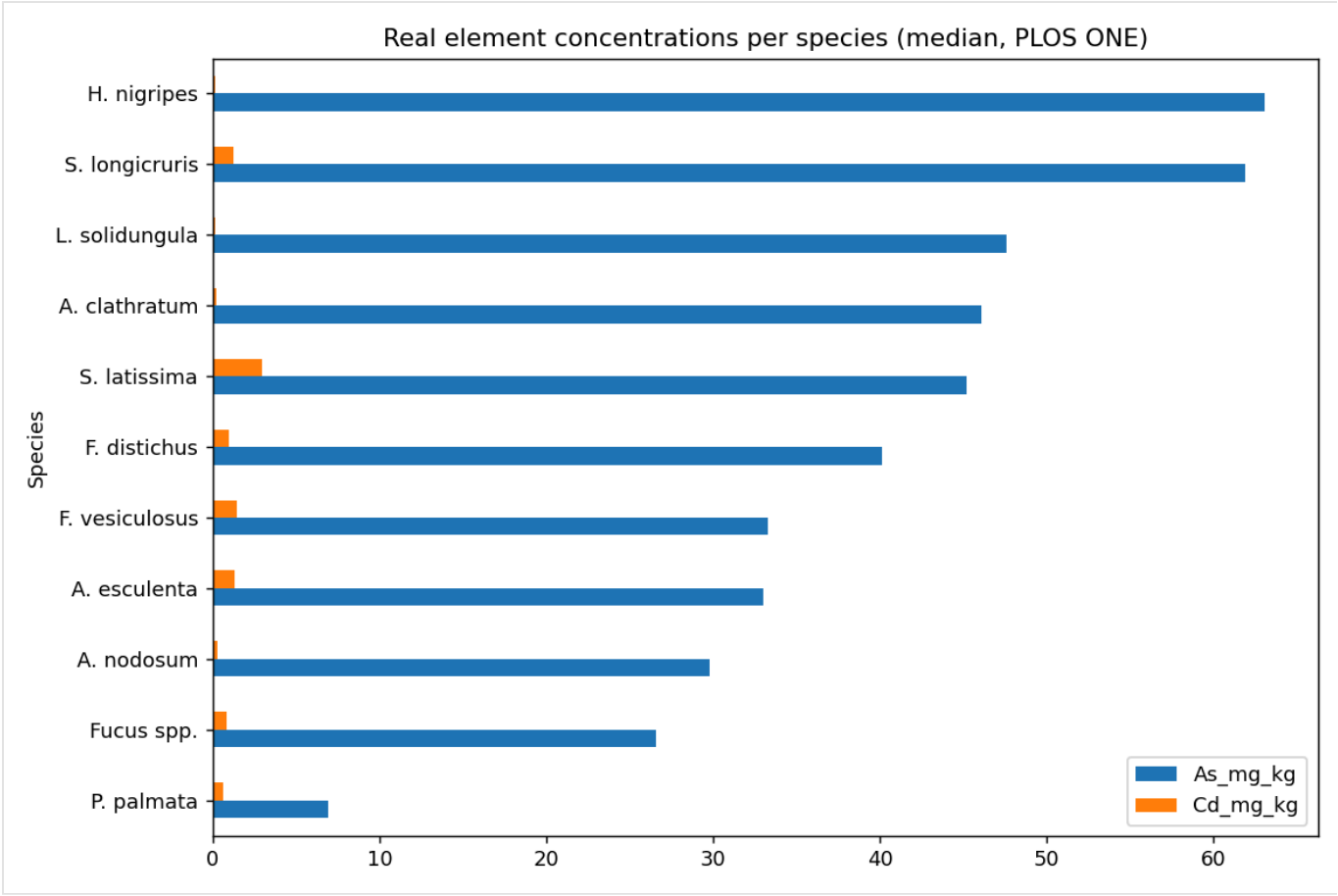

Figure 18. Real element concentrations per seaweed species, median values for ten species from Greenland across 17 elements.

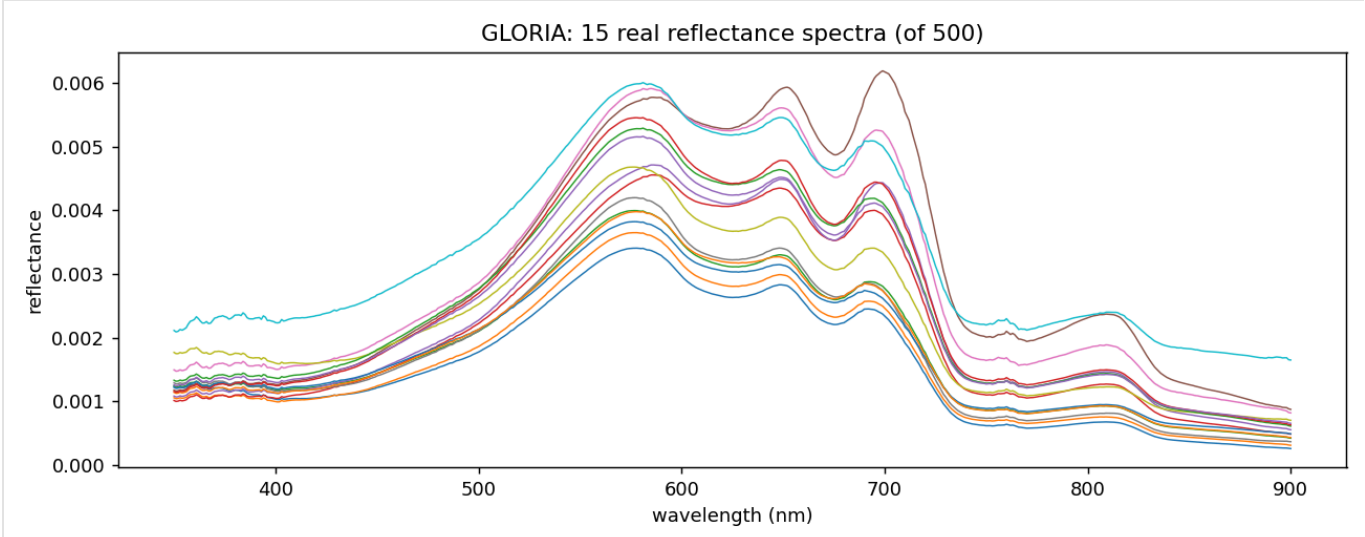

Figure 19. Real hyperspectral reflectance spectra from the GLORIA dataset of inland and coastal waters, used to test the spectral stages with measured data.

Figure 20. Real seaweed photographs, used to test segmentation and feature detection on non synthetic images.

The XRF side follows the published protocol for elemental analysis of herbarium sheets with portable EDXRF, including the cardboard backing controls, the adhesive tape considerations and the treatment of overlapping emission peaks. Real XRF tables from macroalgae (13 elements per sample) and published element concentrations across ten seaweed species (17 elements) are already integrated into the pipeline loaders, together with real reflectance spectra from the GLORIA water quality dataset and real seaweed photographs. A real hyperspectral dataset of plant root images (HyperPRI) was additionally used to test the undistortion and registration path on measured camera data.

11. The pipeline at a glance

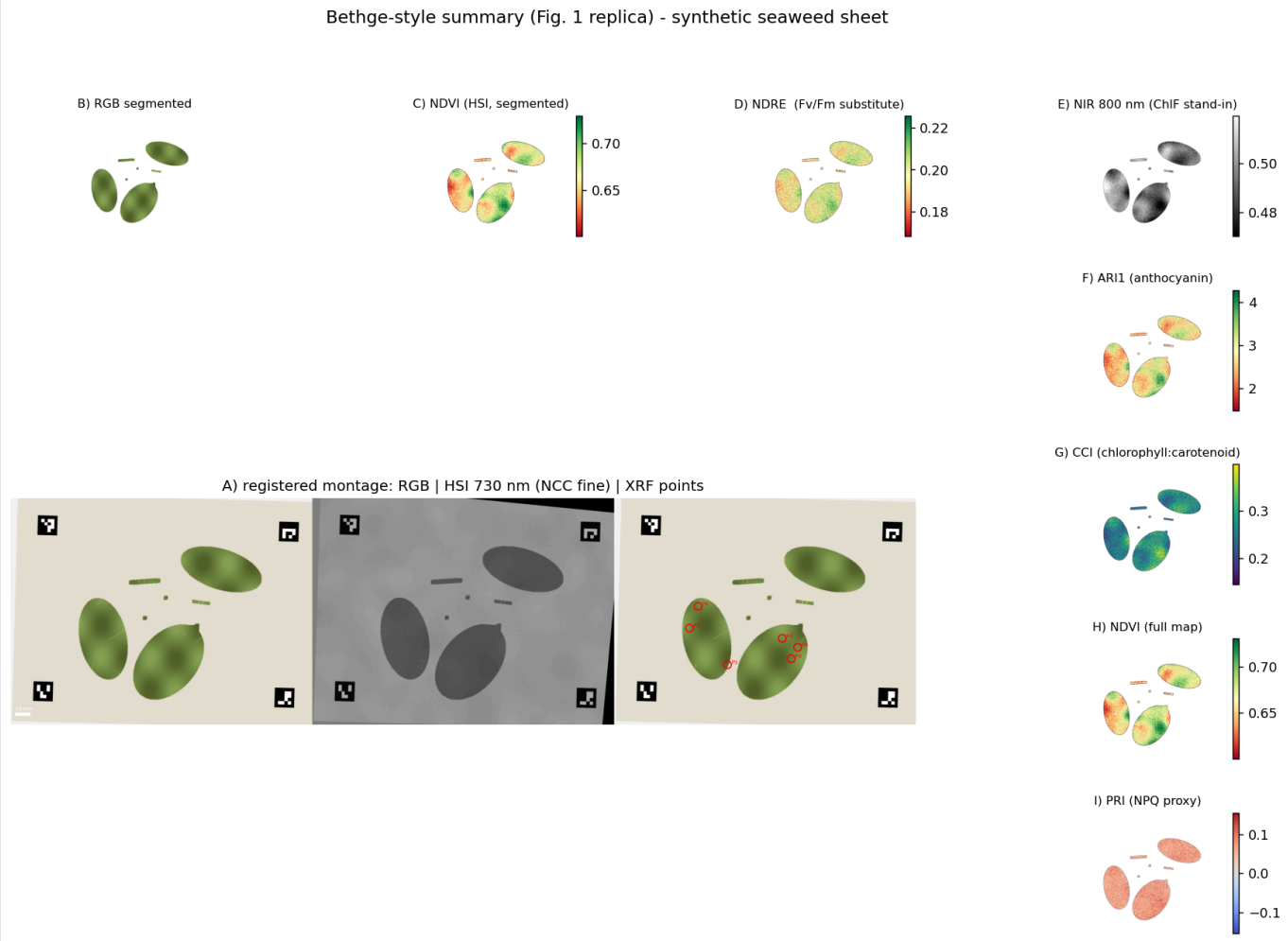

Figure 21. The whole workflow at a glance: the registered montage of the RGB image, the HSI band at 730 nm after NCC fine registration and the XRF points, all expressed in the same sheet frame.

The three Bethge style panels below are additional demonstrations of the same machinery on simulated experiments: a registered time series of a drying process, the intersection of registered binary masks as an evaluation view, and a control versus treated comparison at different resolutions. They show that the registered stack supports time series studies, mask based evaluations and group comparisons once the geometry is solved.

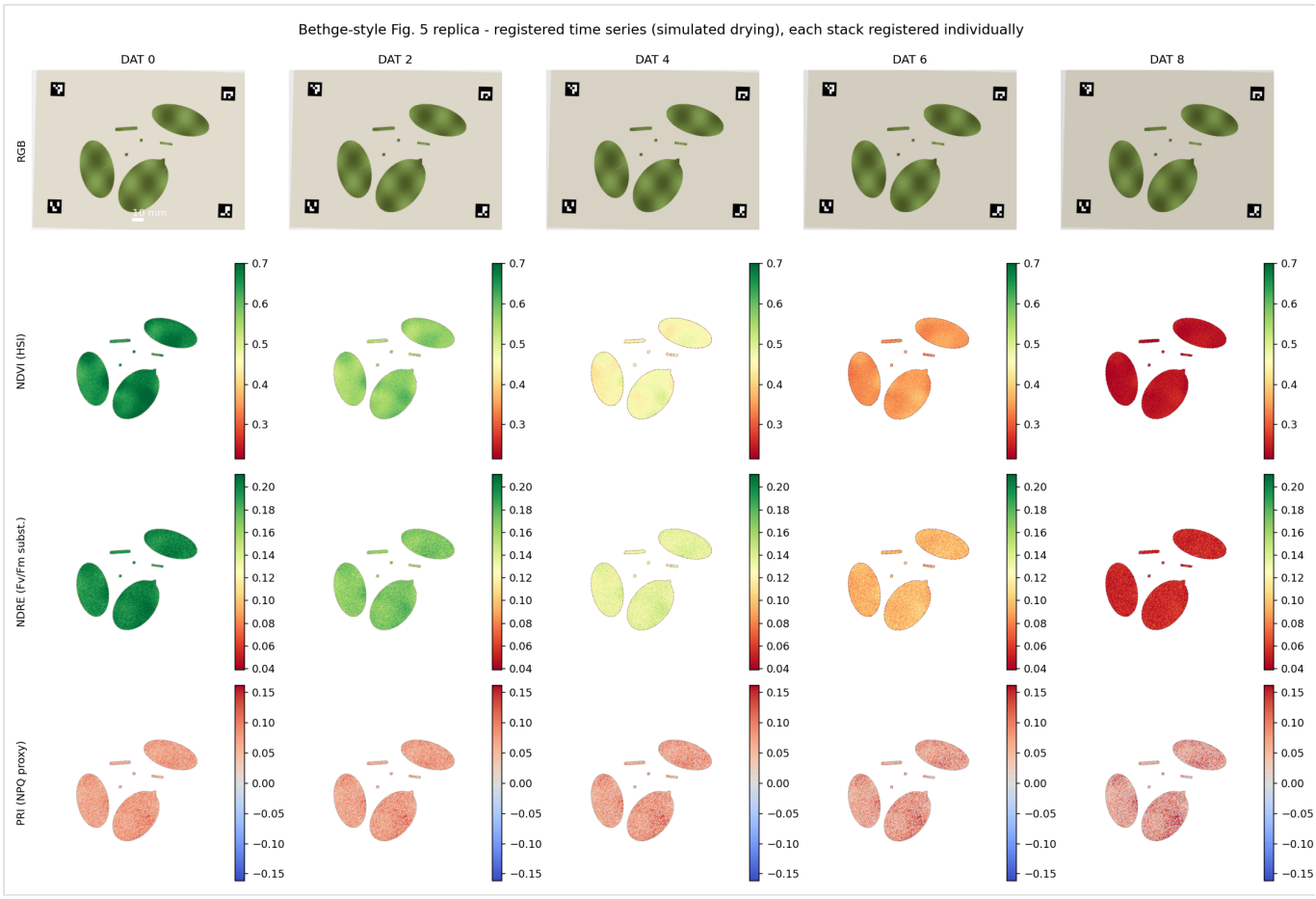

Figure 22. Registered time series demonstration: a simulated drying experiment, with every time point brought into the same frame before comparison.

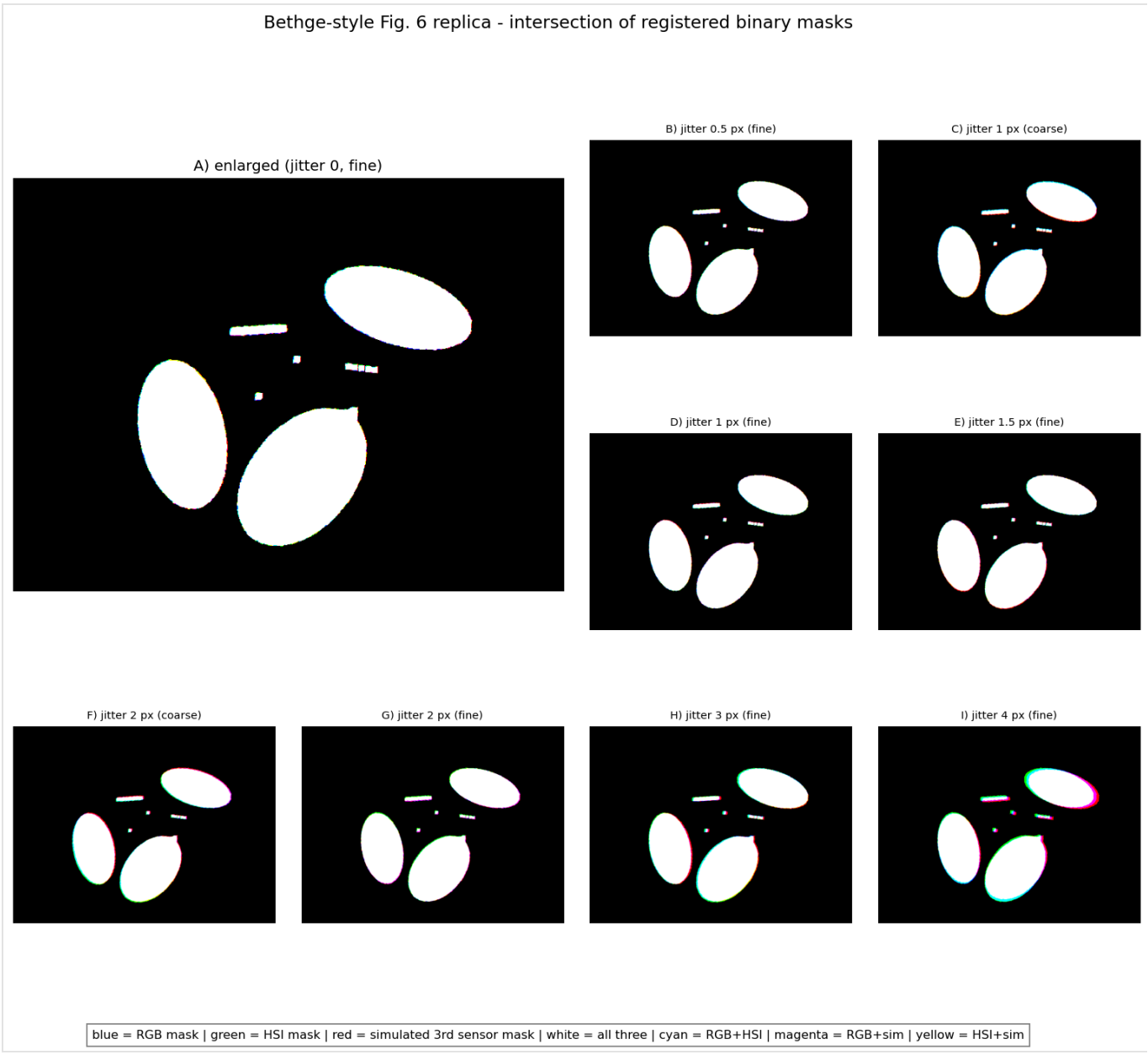

Figure 23. Intersection of registered binary masks, a compact visual way to audit how well the specimen outlines agree after registration.

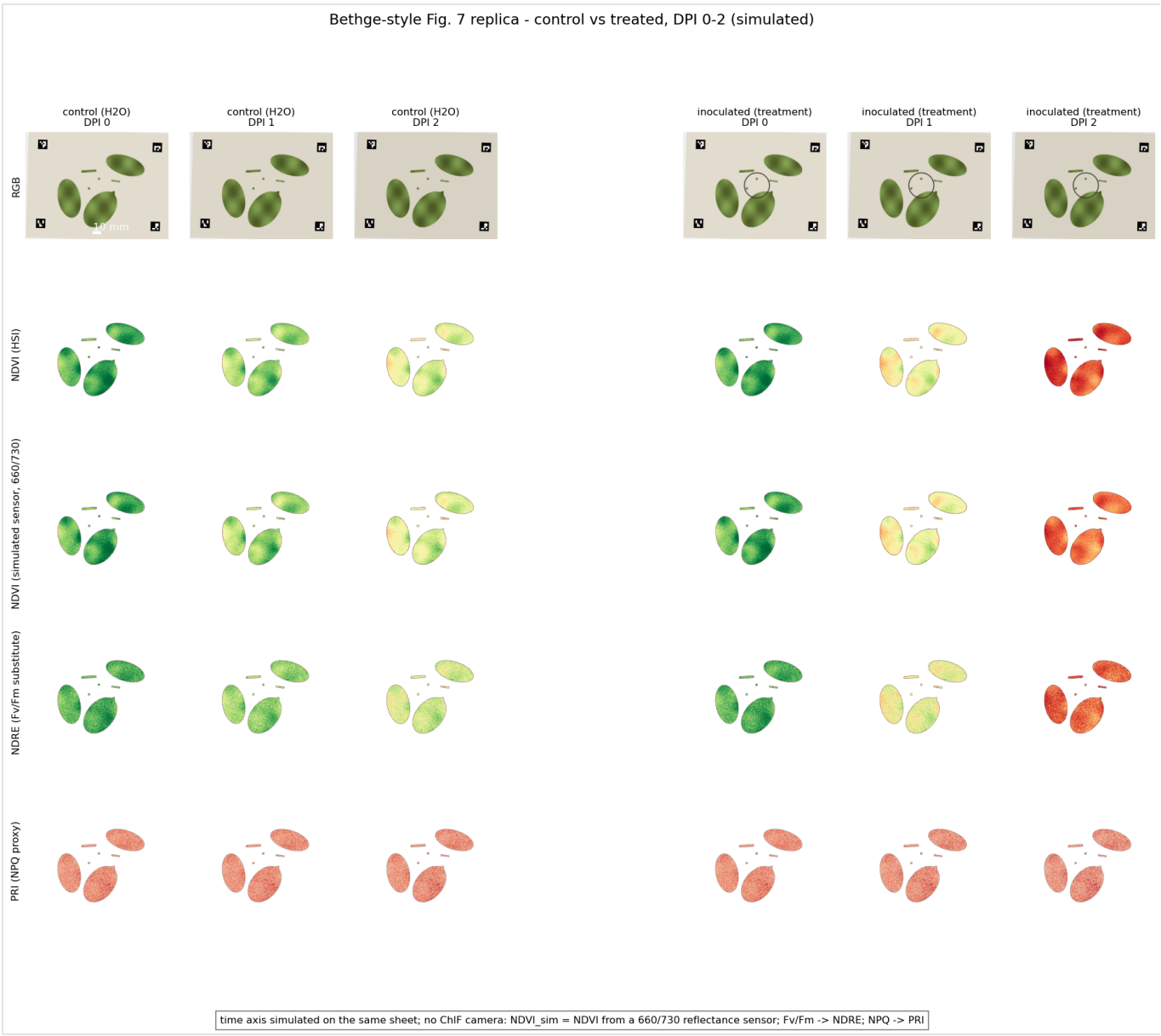

Figure 24. Control versus treated sheets compared at different imaging resolutions (simulated demonstration), the kind of comparison the registered stack enables.

Every step of the workflow carries its own number, so quality is measured rather than assumed: calibration reports its reprojection error, registration reports fiducial, landmark, overlap and information metrics, segmentation reports its overlap against ground truth, the XRF join reports its position error, and the models report cross validated performance. The approach is instrument agnostic, works with pressed 2D specimens, and extends to other pressed or flat samples where image, spectrum and chemistry need to be brought together.

References

1. Bethge, H.L. et al. (2024). Automated image registration of RGB, hyperspectral and chlorophyll fluorescence imaging data. Plant Methods 20:175. https://doi.org/10.1186/s13007-024-01296-y
2. Sharma, N. et al. (2023). An Open-Source Package for Thermal and Multispectral Image Analysis for Plants in Glasshouse. Plants 12(2):317. https://doi.org/10.3390/plants12020317
3. Rana, S. et al. (2024). Comprehensive Evaluation of Multispectral Image Registration Strategies in Heterogenous Agriculture Environment. J. Imaging 10(3):61. https://doi.org/10.3390/jimaging10030061
4. Tavares, T.R. et al. (2020). Combined Use of Vis-NIR and XRF Sensors for Tropical Soil Fertility Analysis: Assessing Different Data Fusion Approaches. Sensors 21(1):148. https://doi.org/10.3390/s21010148

5. Kreissig, K.J. et al. (2021). Characterisation and chemometric evaluation of 17 elements in ten seaweed species from Greenland. PLOS ONE 16(2):e0243672. https://doi.org/10.1371/journal.pone.0243672
6. Lehmann, M.K. et al. (2022). GLORIA, a global dataset of remote sensing reflectance and water quality from inland and coastal waters. PANGAEA. https://doi.org/10.1594/PANGAEA.948492
7. Chang, S.J. et al. (2024). HyperPRI: a dataset of hyperspectral images for underground plant root study. Computers and Electronics in Agriculture. https://doi.org/10.1016/j.compag.2024.109307
8. Seaweed photographs from Wikimedia Commons, free licences. Portable EDXRF macroalgae dataset: own measurements, ten samples by 13 elements.
