FORCE

Biostratigraphy, conventional and beyond

chronosurveys

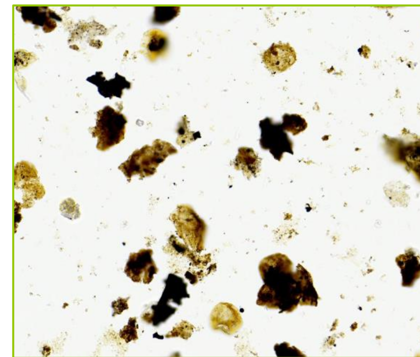
Biostrat AI - a new artificial intelligence-assisted tool to analyse palynology slides

Gil Machado, Frederik Strothmann, Laurenz Strothmann

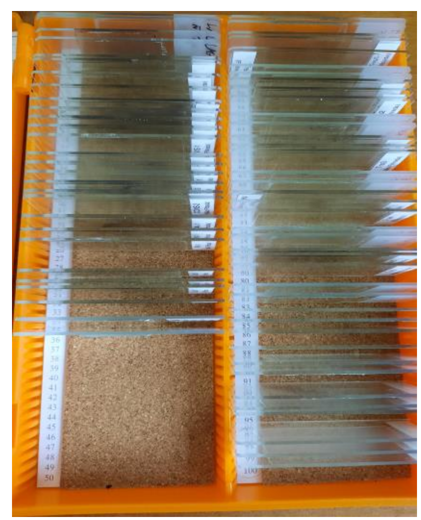
Standard palynology work

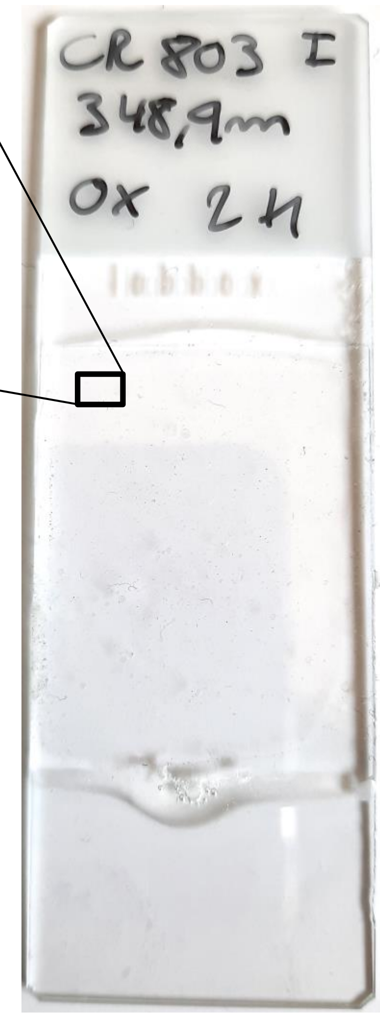
- Process rock samples to obtain a slide with microfossils.
- Analyse microfossil content of rocks under the microscope (glass slides).
- Identify types of particles and species.
- Determine sedimentation age and setting (lake, beach, open sea, paleoecological interpretation, etc).
- Unchanged method for over 100 years.

Standard palynology work



- Analysis under the microscope is laborious – hours per sample, in projects that typically have dozens or hundreds of samples.
- Analysis difficult to reproduce - documentation is done by photographing a selection individual specimens.
- Interpretation is potentially biased
- Requires specialist with Msc or PhD
- Few specialists exist worldwide

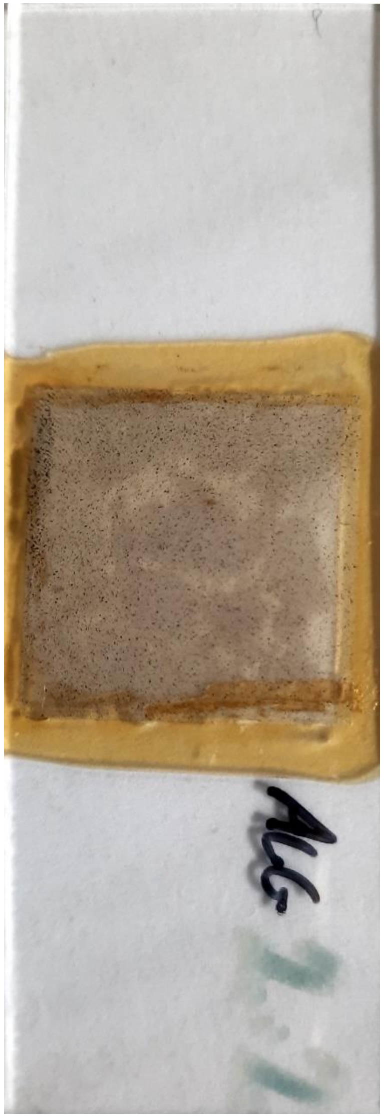




Standard palynology work

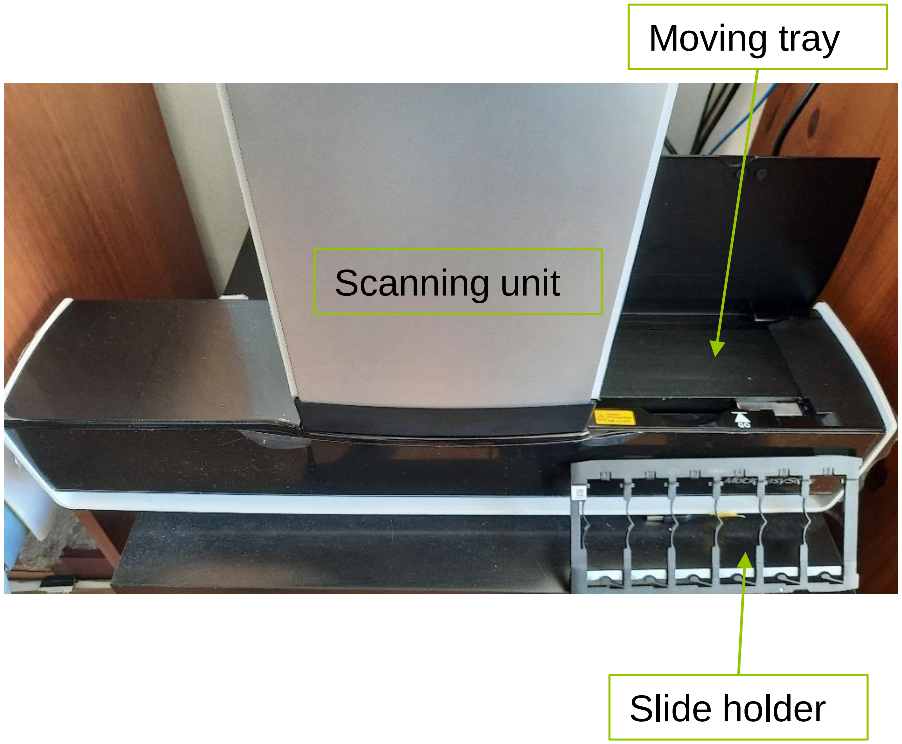
- For vintage slides, physical degradation is an issue:
  - Yellowing of mounting medium (resin) – observation quality decreases
  - Mounting medium drying – does not allow observation
- Slides not recorded by modern techniques (into a software) hold stranded data (stratigraphy, source rock, paleoenvironment).
- Holotypes may be lost.
- Samples are frequently re-processed, re-analysed, with consequent costs and resources spent.

CIMP slide collection
Florinites dissacoides holotype



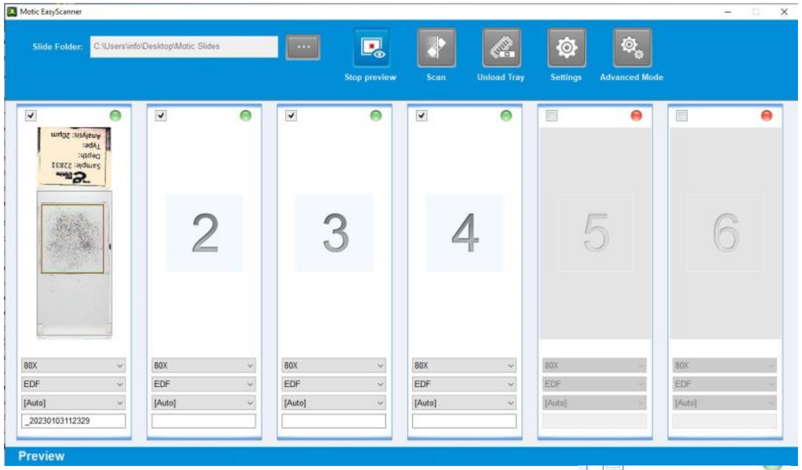
New tools - digitalization

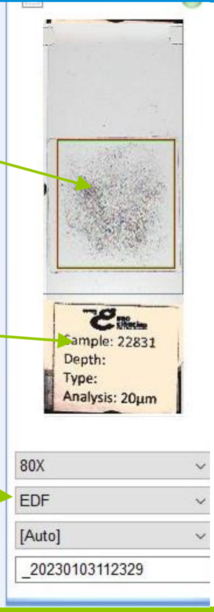
- High-resolution slide scanners scan whole slides.
- Allows permanent record of the slide, remote access, and analysis.
- Typical slide takes ca. 1,5 h to scan – high impact on efficiency and consistency.
- Produces large files – one slide corresponds to a 3,5 Gb file.
- Multiple brands and models exist in the market – scanning time/slide is similar.



New tools - digitalization

- Different models, from 1 to hundreds of slides' batches
- Scan details can be customized.
- Metadata is included (magnification, date, name, scan type, etc).
- Scan types include:
  - Extended depth of field (multiple focus levels are flattened to a single layer) – ideal for routine work. ~3,5 Gb file
  - Multi-Z (multiple focus levels are kept in different layers). Allows focus in and out as in a microscope. Useful for detailed taxonomic work. Generates larger file > 10 Gb.





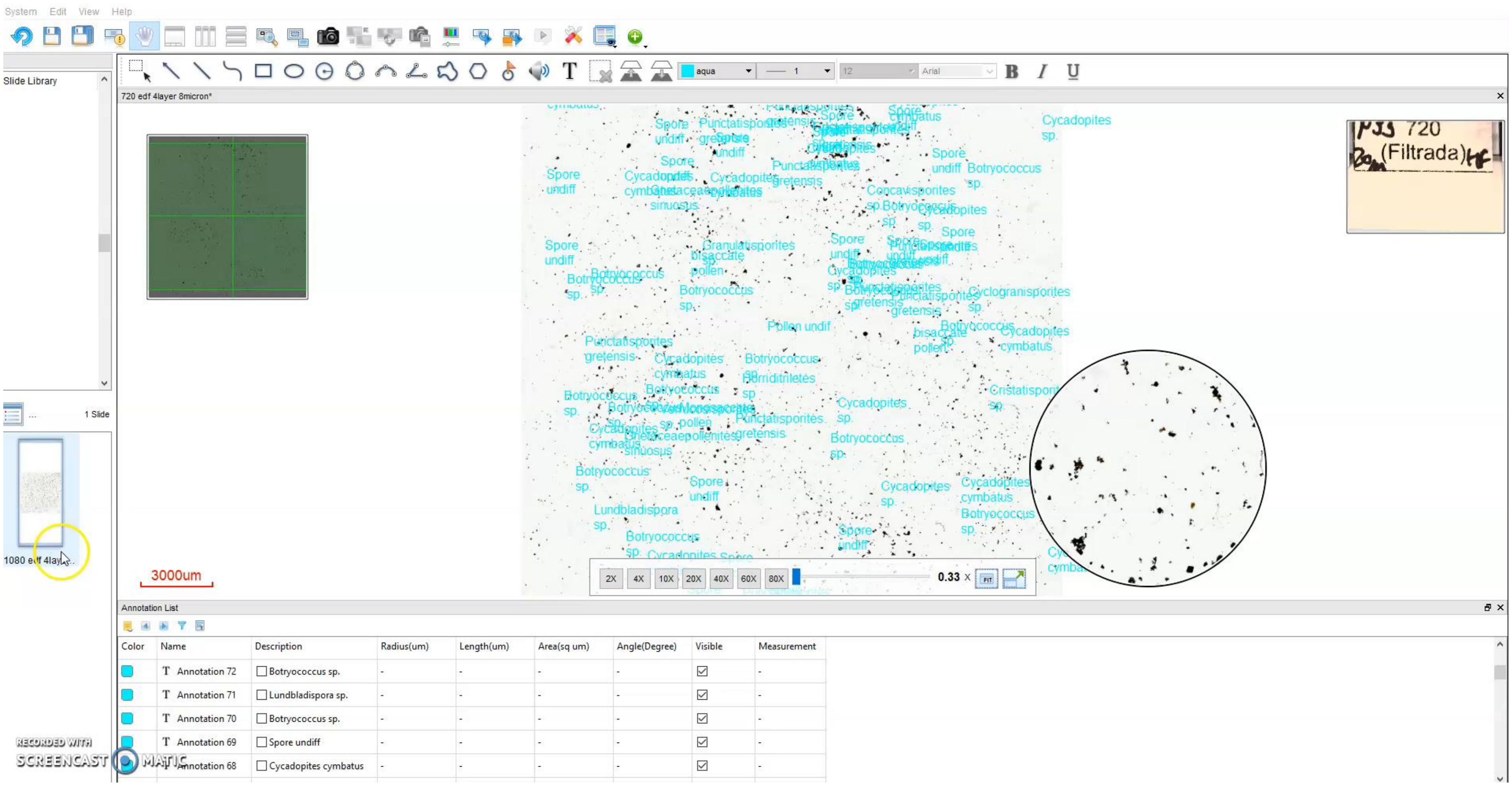
Scan area can be customized – typically edges are excluded and any degraded area

Label is scanned in lower resolution. Allows bar code

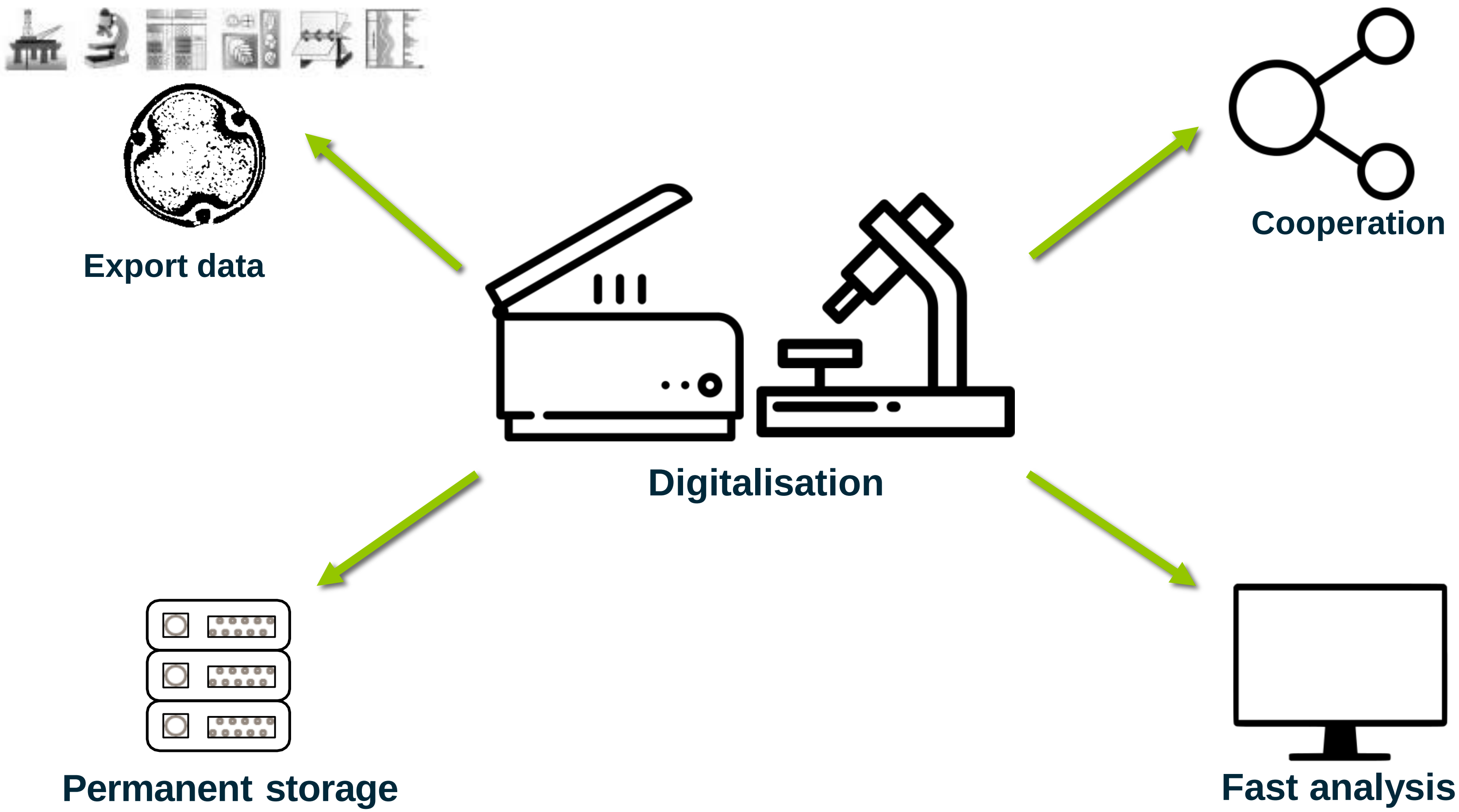
Magnification, scan type and file name

New tools - digitalization

Short video



New tools - digitalization



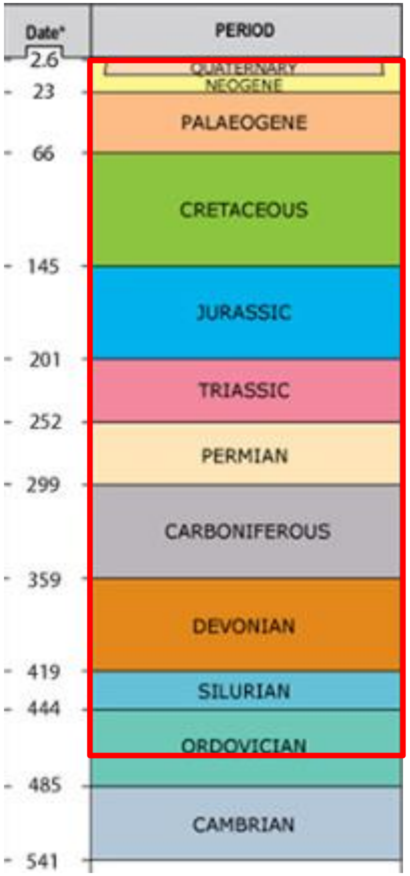
Workflow streamlined, now used as standard.

ARTIFICIAL INTELLIGENCE

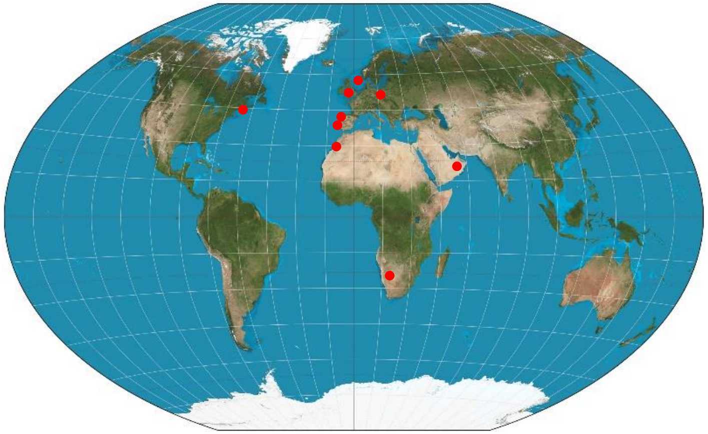
Artificial Intelligence – how and why?

- Artificial intelligence has developed greatly in the past few years, with applications in many areas.
- For geological applications, commonly the expertise is **computer vision** – algorithms and models to interpret visual data.
- To allow automated identification and interpretation, a supervised algorithm is needed, which needs to be trained.
- This requires a large training dataset – big data.
- The result is a software package that takes scanned images and identifies the different types of palynological particles, with minimal supervision.
- It will not replace the biostratigraphers, but it will boost their productivity.

Artificial Intelligence – building the model



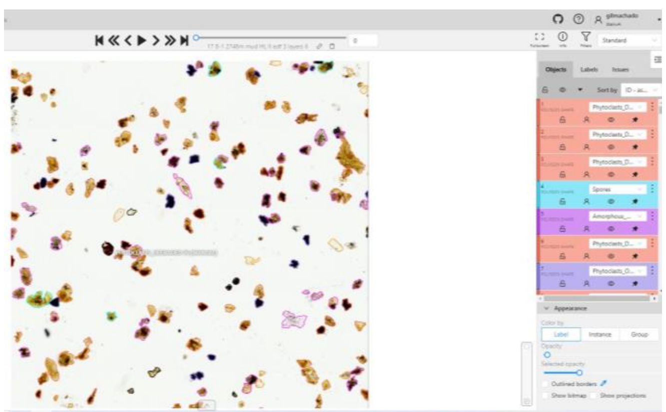
- A palynological slide training dataset was created:
  - Ages: Ordovician to Miocene
  - Environment: Turbidites, carbonate platform, evaporites, continental clastics, etc
  - Location: Oman, Portugal, UK, Canada, Namibia, Morocco,…
- Sensitive information is not included (well, depth, location, etc).
- Classes are defined (24 particle types – AOM, phytoclast, spores,…) – not species or genera

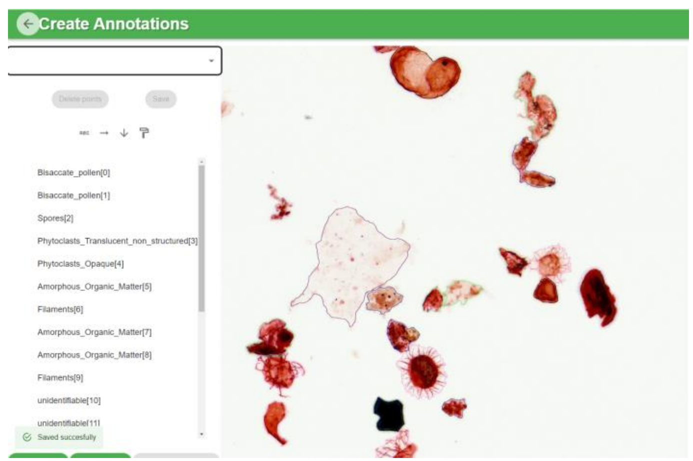


Artificial Intelligence – building the model

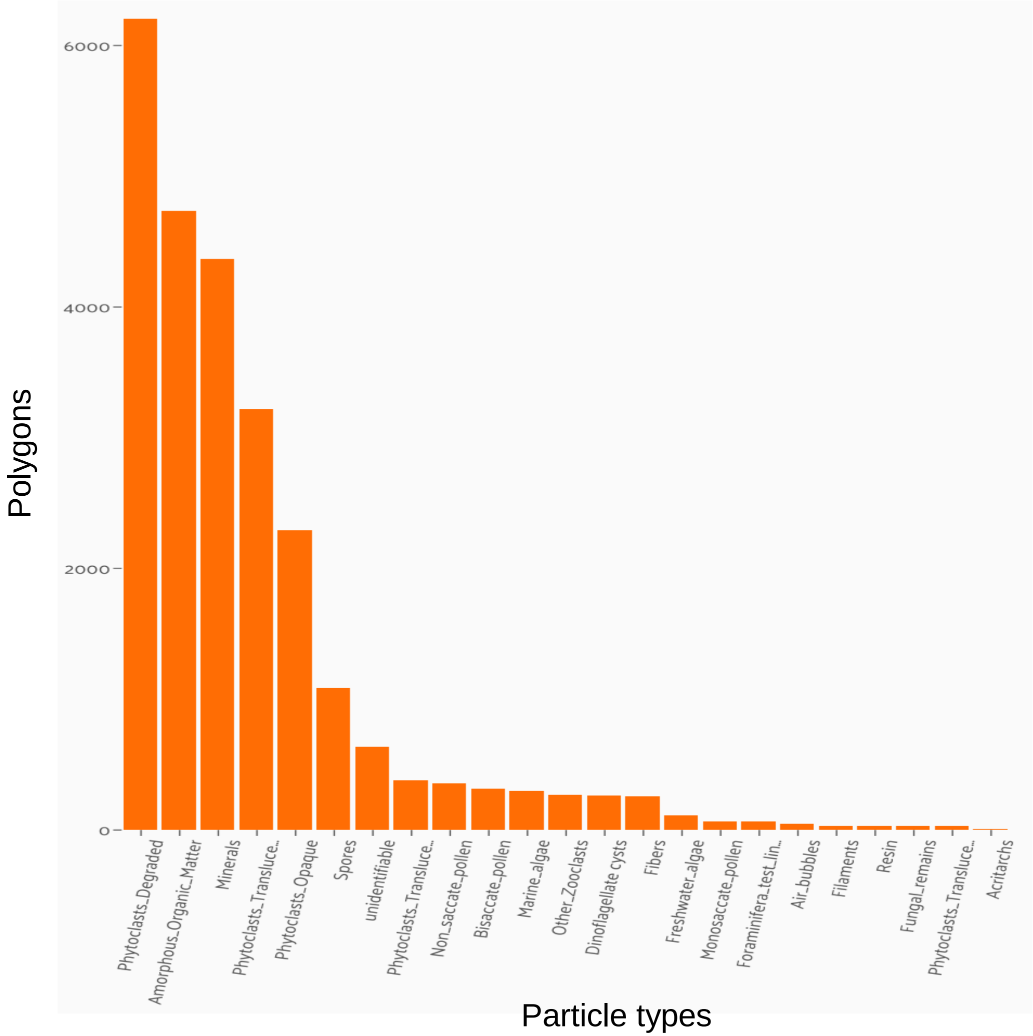
- Graphical interface takes scanned images.
- A segmentation model is trained to detect each object separately.
- A human-in-the-loop process is used to pre-annotate all data (palynologist).
- A total of 2921 images (4096x4096px) were already annotated with 31717 polygons of microfossils.
- Feeds the AI model with training data.
- The model is embedded into feature-rich statistical post-processing to extract meaningful data.

Spores
Pollen
Bisaccate pollen
Monosaccate pollen
Non-saccate pollen
Acritarchs
Dinoflagellate cysts
Chitinozoans
Amorphous Organic Matter (AOM)
Fungal remains
Filaments
Foraminifera test linings
Other Zooclasts
Freshwater algae
Marine algae
Phytoclasts
Opaque equant (length:width ratio <2)
Opaque lath (length:width ratio >2)
Degraded
Translucent cuticle
Translucent structured
Translucent non-structured
Resin
Minerals
Fibers
air bubbles
unidentifiable



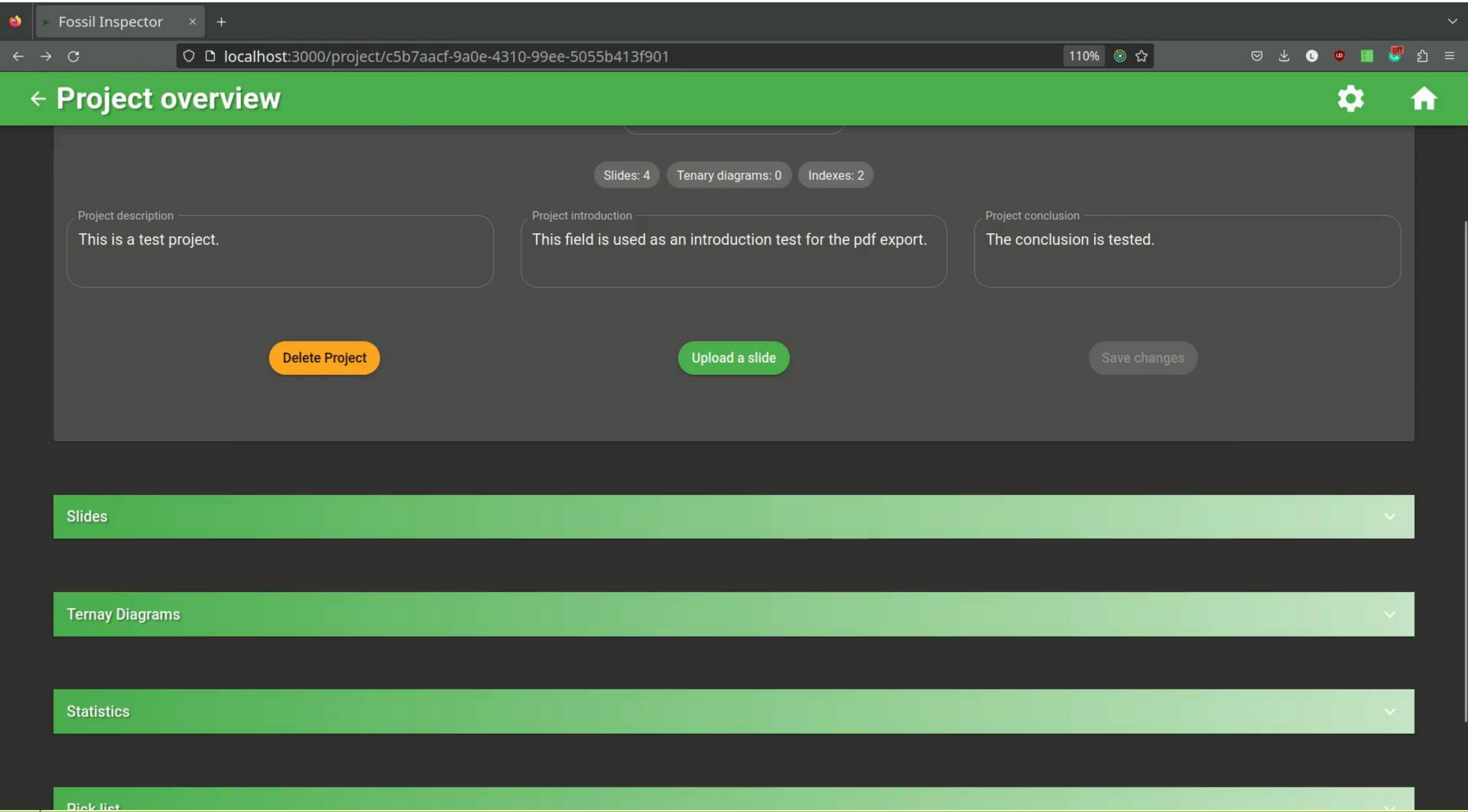


Artificial Intelligence – building the model



- Although the model is able to identify particle types after a few dozens of examples have been show, high accuracy levels are only obtained with hundreds to thousands of examples.
- While some particle types are common, others are not so much.
- Further development of the training dataset is needed.

Artificial Intelligence



Artificial Intelligence

- Outputs of the AI processing include:
  - Visualization of the slide, with automatically generated annotations (allows QC)
  - Statistics of the several particle types encountered.
  - Paleoenvironmental interpretation, based on multiple plots (Tyson ternary diagram, marine indexes, etc), automatically generated.
  - Kerogen typing – inert, gas or oil-prone particle proportions done automatically, used in conjunction with organic geochemistry
  - Automatic display of stratigraphically relevant particles: spores, pollen, chitinozoans to ease biostratigraphers' work.

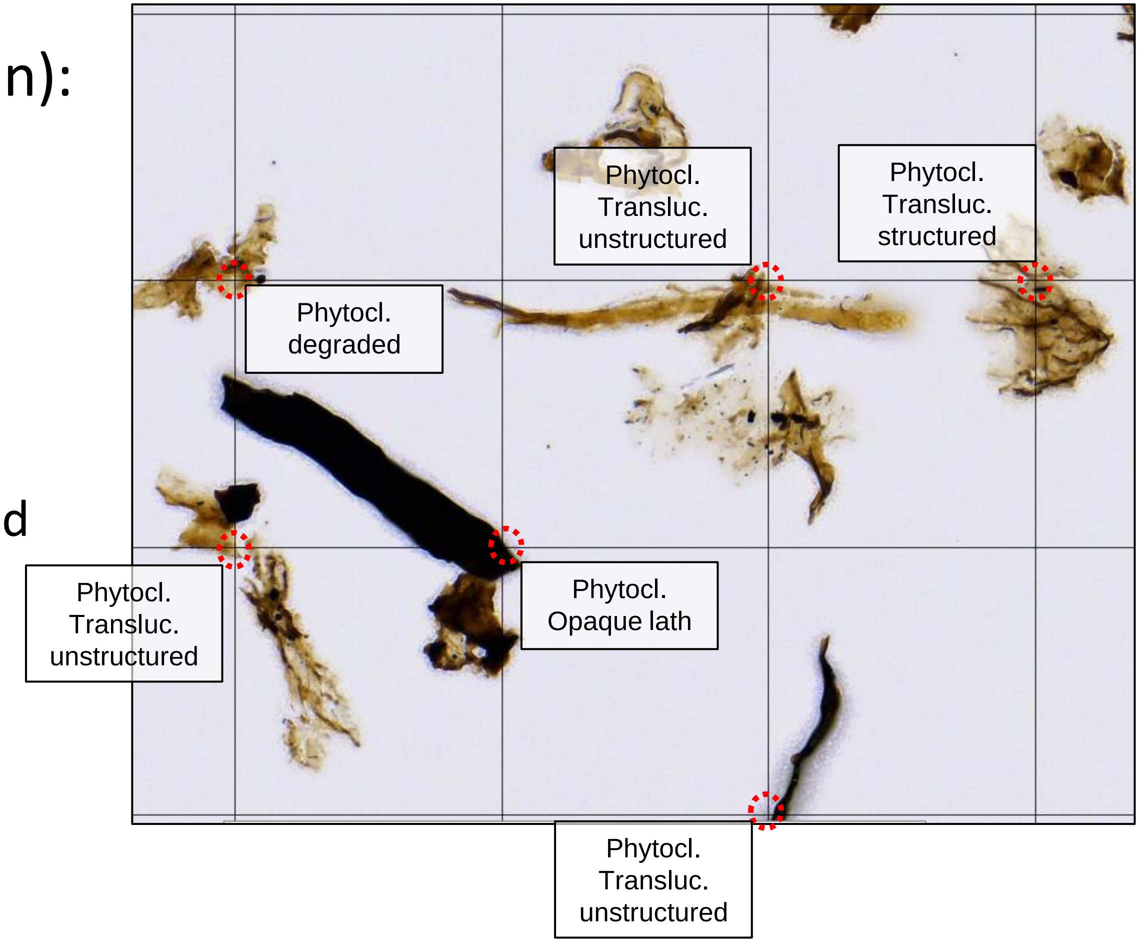
CASE STUDY

Artificial Intelligence – Case study

- 9 outcrop samples
- Late Jurassic
- Slope to basin low-density turbidites
- Samples unknown to AI model
- Good and bad quality slides: clean and mineral-rich slides

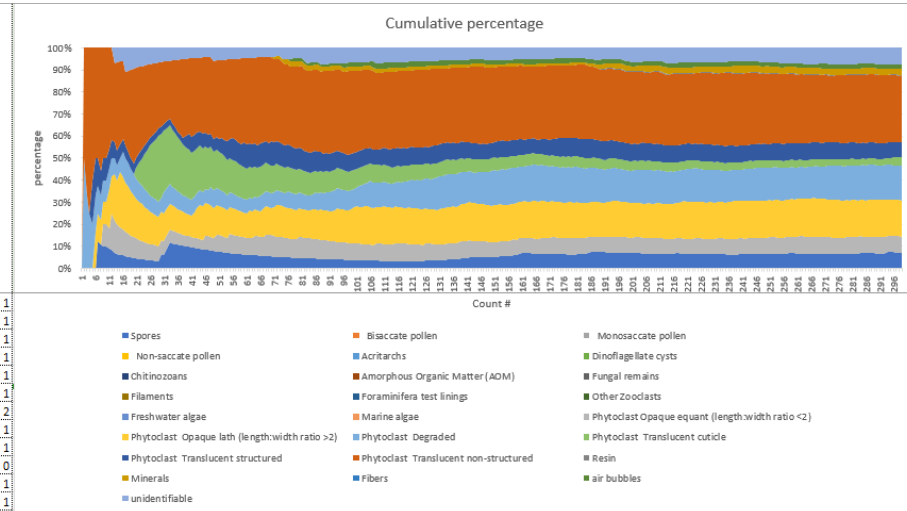
Artificial Intelligence – Case study

- Counting method (Human):
  - Scanned slides
  - Computer screen
  - 100 µm grid
  - 24 particle types considered (same as model considers)
  - Counts at crossings
  - 300+ counts in each slide



Artificial Intelligence – Case study

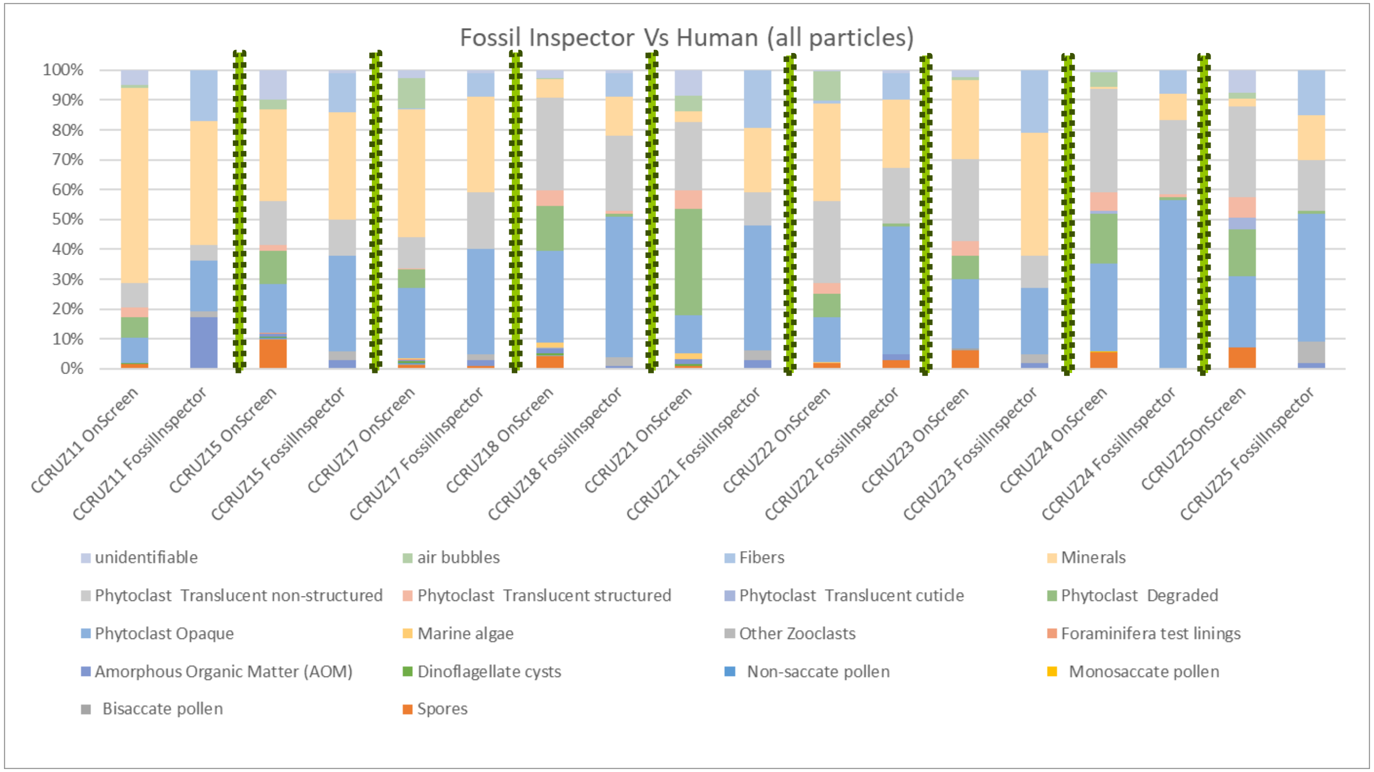
Field of view	Spores	Bisaccate pollen	Monosaccate pollen	Non-saccate pollen	Acritarchs	Dinoflagellate cysts	Chitinozoans	Amorphous Organic Matter (AOM)	Fungal remains	Filaments	Foraminifera test linings	Other Zooclasts	Freshwater algae	Marine algae	Phytoclast Opaque equant (length:width ratio <2)	Phytoclast Opaque lath (length:width ratio >2)	Phytoclast Degraded	Phytoclast Translucent cuticle	Phytoclast Translucent structured	Phytoclast Translucent non-structured	Resin	Minerals	Fibers	air bubbles	unidentifiable	Totals
1																	1									1
3																				1						1
4																				1						1
5																				1						1
6																			1							1
7																1										1
8	1																			1						2
9																				1						1
10															1											1
11																										0
12																1										1
13															1											1
14																1									1	2



- Counts in Excel, with real-time cumulative percentage
- In most samples a “stable” percentage was achieved at ~150 counts
- Phytoclast-dominated assemblage, with minor spores, marine algae, dinoflagellate cysts and pollen. Not all particle types were detected.

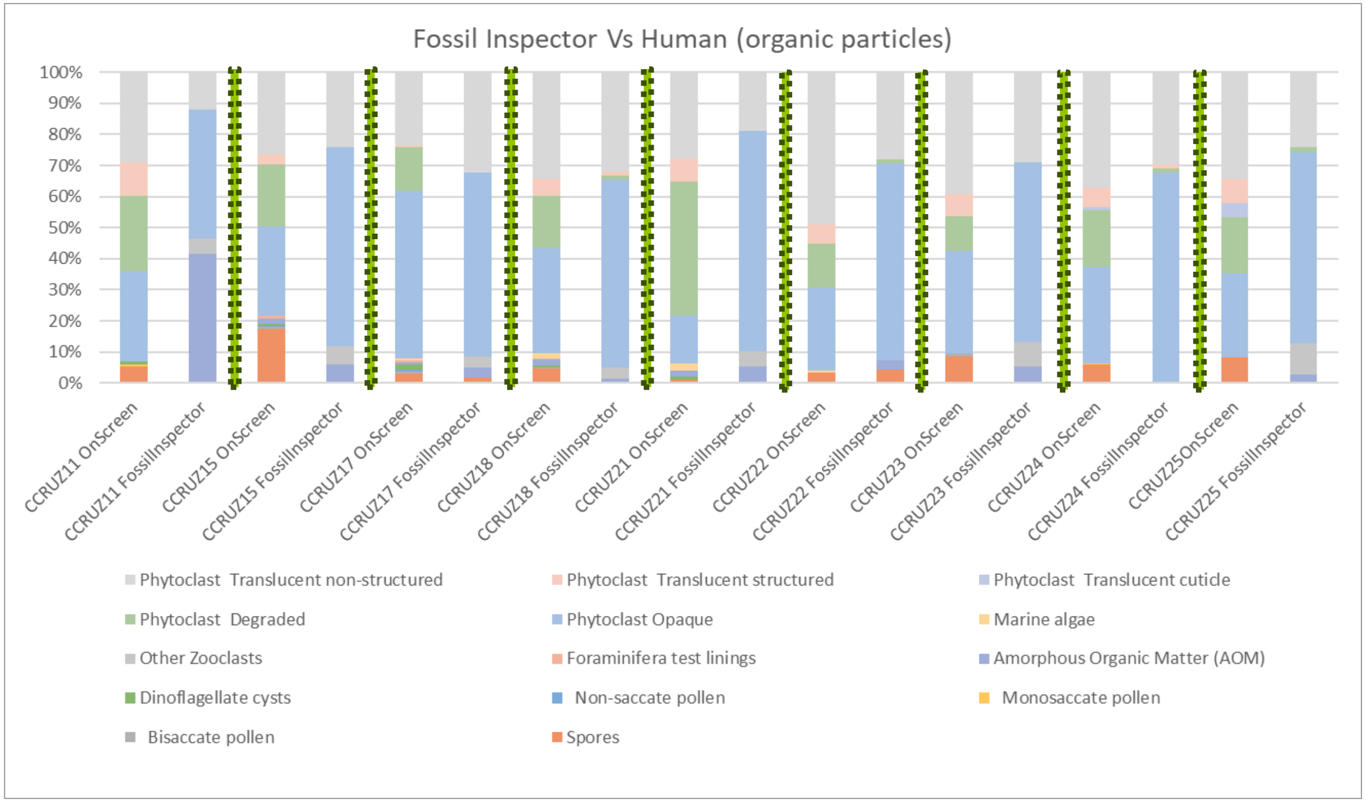
Artificial Intelligence – Case study

- Mineral-rich slides skewed overall statistics.
- Particles which are yet "poorly known" resulted in many false positives.



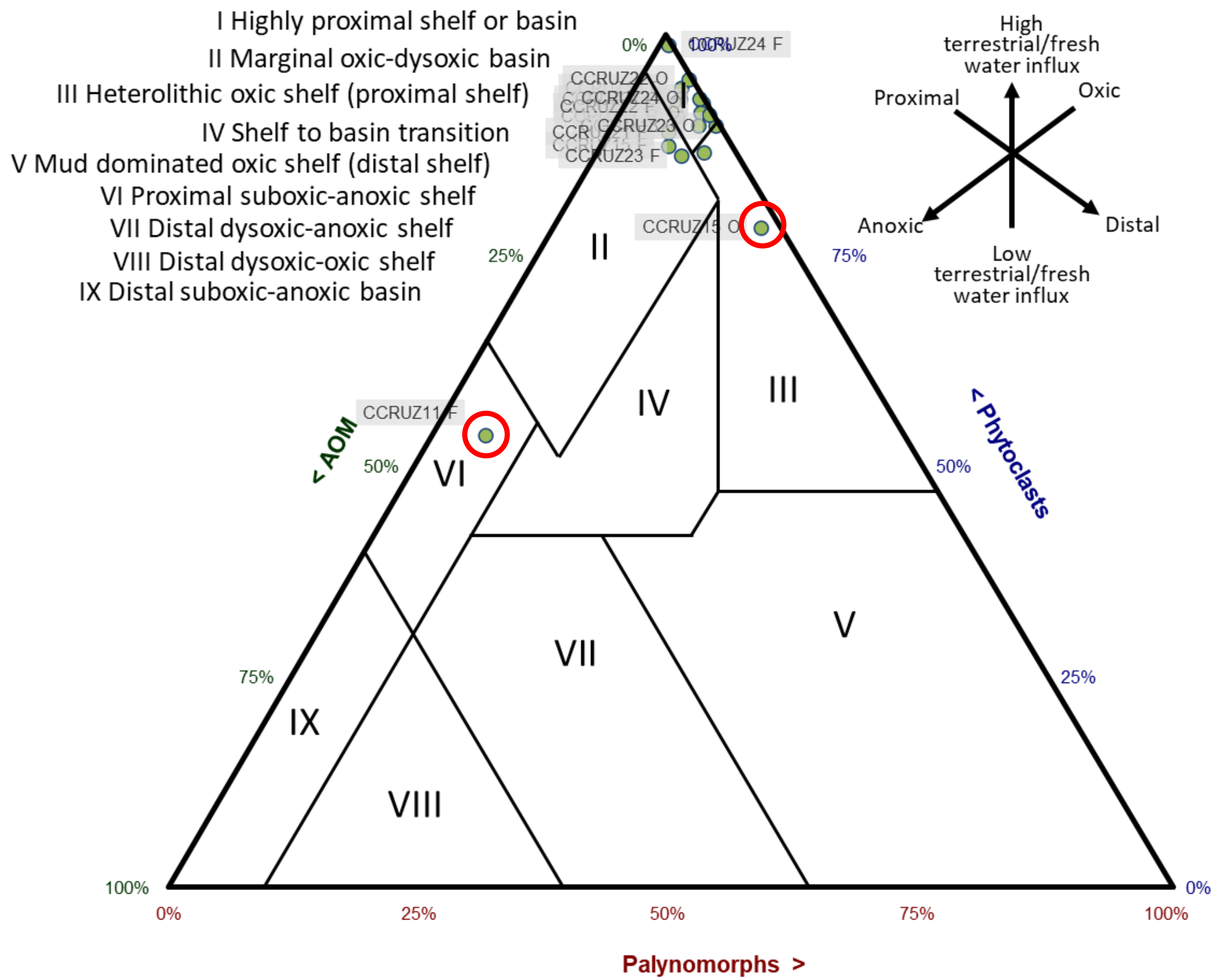
Artificial Intelligence – Case study

- Organic particles' plots show closer results (human Vs model).
- Total phytoclast counts are very similar overall
- Total palynomorph counts as well



Artificial Intelligence – Case study

- Most samples plot very closely together – validating model results when few classes are considered.
- Outliers correspond to mineral-rich samples



Artificial Intelligence

- Conclusions and next steps
  - Slide preparation quality influences the quality of the scanned image and consequently the quality of the model results.
  - Not-so-common particles (under represented in training dataset) have a detrimental impact on the model results – false positives.
  - The features the model uses to identify different classes need to be known (shape, colour, texture, etc) and possibly emphasized by the coders – e.g. opaque phytoclasts are necessarily black.
  - The current model version is able to correctly quantify simple particle types (by lumping different classes) and plotting in Ternary Diagrams, indexes, etc.
  - Mature training dataset focusing on not-so-common particles.
  - Compare with traditional method – counts under microscope considering lower number of particle types (5 to 8 instead of 24).
  - Incorporate more indexes and paleoenvironmental interpretation statistics.
  - Develop taxonomic identification help tool – plates and pick list

Acknowledgements

STAIRWAI

AIBC euroclusters

AI4EU

FPI FONDAZIONE PIEMONTE INNOVA

FORCE

Biostratigraphy, conventional and beyond

chronosurveys

Biostrat AI - a new artificial intelligence-assisted tool to analyse palynology slides

Gil Machado, Frederik Strothmann, Laurenz Strothmann
www.chronosurveys.com gil.machado@chronosurveys.com