

Biodiversity knowledge synthesis:
an introduction to
meta-analyses and systematic reviews
- Quantitative data extraction -

4/10/23 - Montpellier

Dakis-Yaoba Ouédraogo (PatriNat)

dakis-yaoba.ouedraogo@mnhn.fr



Extraction of quantitative data

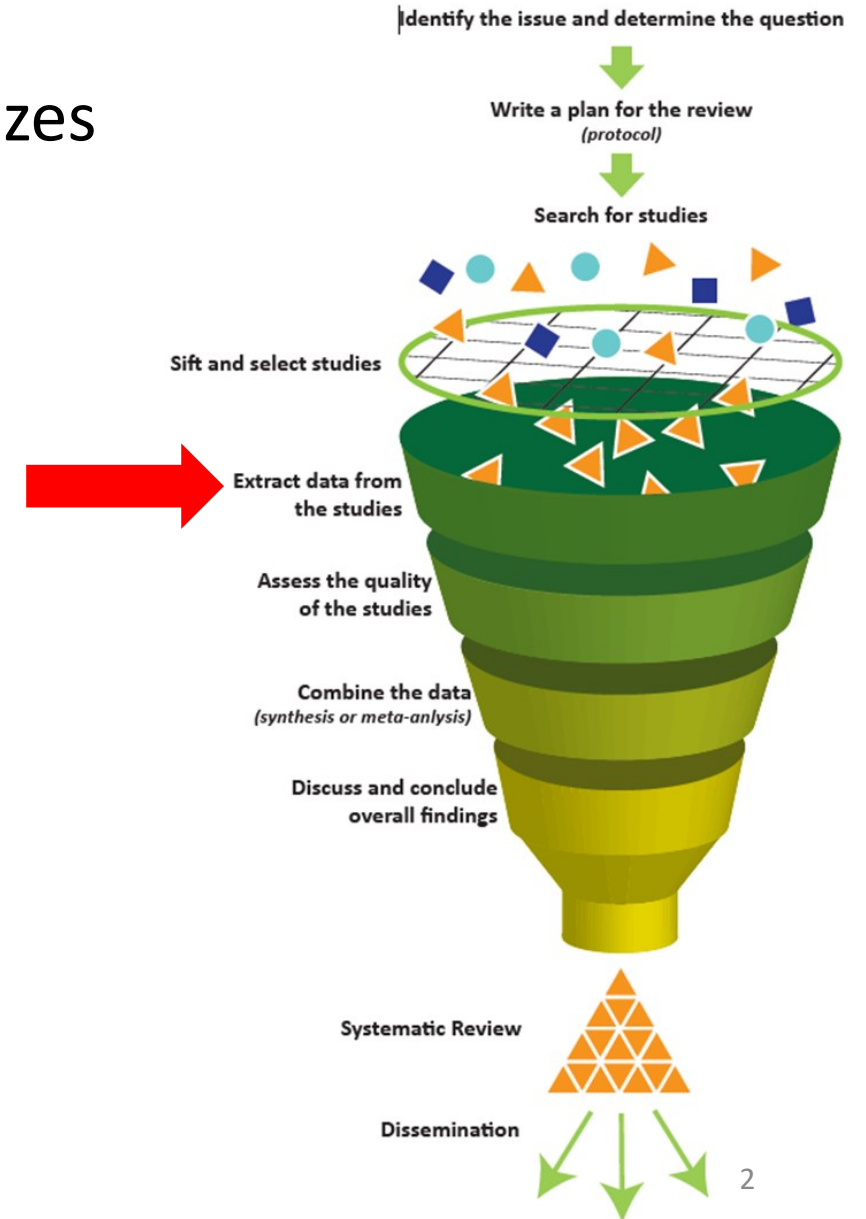
Extraction of the data needed to calculate effect sizes (e.g. mean, sample size, sd/se/95% CI)

+ extraction of variables that could explain the heterogeneity of effect sizes (*effect modifiers*)

Extraction from

- text
- table
- figure
- supp. mat.

+ calculations may be needed (keep a record)



Extraction of quantitative data

! Warning !

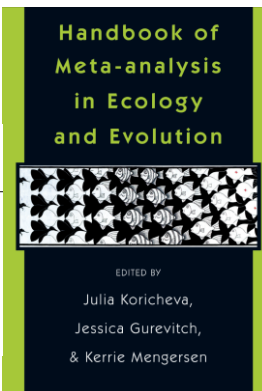
Data extraction is time-consuming: clearly define the extraction grid and the effect modifiers to be extracted

Importance of **testing** the extraction sheet on a sample of articles to check that it matches the content of the studies

Document the work/decisions (transparency, repeatability)

Decide what to do in case of **missing information** ("missing data", contact the authors, imputation)

13
Recovering Missing or Partial Data from Studies:
A Survey of Conversions and Imputations
for Meta-analysis
Marc J. Lajeunesse



Consistency check

To be sure that data extraction is objective / robust:

- Data extraction of each study carried out **independently** by 2 people
- if several people share the work, **check the consistency of data extraction** between people on a sample before starting the actual extraction (discuss any disagreements)
- if only 1 person, have someone to check a sample of the extracted data at the start of the work (discuss any disagreements)

Example of extraction sheet

A study = a taxon × an exposure × an outcome

Case study level: ex. several concentrations of a chemical

IDdata	ID_map	author	...	taxon	Population_descri	Life_stage	Type_system	Tempera	pH
ScreenTA_9680	880	Cantin, N.E. ...		<i>Acropora tenuis</i>	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	880	Cantin, N.E. ...		<i>Acropora tenuis</i>	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	884	Cantin, N.E. ...		<i>Acropora valida</i>	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	884	Cantin, N.E. ...		<i>Acropora valida</i>	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	889	Cantin, N.E. ...		<i>Pocillopora damicornis</i>	Colonies	Adult	500 L outdoor tank	27.5	NA
ScreenTA_9680	889	Cantin, N.E. ...		<i>Pocillopora damicornis</i>	Colonies	Adult	500 L outdoor tank	27.5	NA

Treatment_description	Control_description	Solvent	Concentration_nom	Concentration_eff	Duration	Measured_variable	Time_after
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	53 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	53 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	90 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	90 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	67 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	67 days	Symbiodinium density / total protein	NA

Example of extraction sheet

Treatment_description	Control_description	Solvent	Concentration_nom	Concentration_eff	Duration	Measured_variable	Time_after
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	53 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	53 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	90 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	90 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	1 µg/L	0.91 µg/L	67 days	Symbiodinium density / total protein	NA
Diuron	Unfiltered oceanic seawater	No	10 µg/L	8.8 µg/L	67 days	Symbiodinium density / total protein	NA

Metaanalyse_data	unit	ID_experiment	ID_case	ID_common_control	N_c	Mean_c	Type_variation_c	Variation_c	N_t	Mean_t	Type_variation_t	Variation_t
OK (Fig3, SE, n=6)	x 10 ⁶ / mg protein	1	3	1	6	4.2	sd	1.714642819	6	3.5142857	sd	0.979795897
OK (Fig3, SE, n=6)	x 10 ⁶ / mg protein	1	4	1	6	4.2	sd	1.714642819	6	3.6	sd	0.524890659
OK (Fig3, SE, n=6)	x 10 ⁶ / mg protein	2	7	2	6	0.928571	sd	0.454905237	6	1.4142857	sd	0.979795897
OK (Fig3, SE, n=6)	x 10 ⁶ / mg protein	2	8	2	6	0.928571	sd	0.454905237	6	1.3142857	sd	0.699854212
OK (Fig3, SE, n=6)	x 10 ⁶ / mg protein	3	11	3	6	1.714285	sd	0.699854212	6	2.2285714	sd	0.699854212
OK (Fig3, SE, n=6)	x 10 ⁶ / mg protein	3	12	3	6	1.714285	sd	0.699854212	6	0.9142857	sd	0.244948974

Method_extraction	Source	Comment_extraction	Name_data_extraction
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO
Figure	Figure 3	NA	DYO

Tools for extracting data from figures

<https://automeris.io/WebPlotDigitizer/>

WebPlotDigitizer
Web based tool to extract data from plots, images, and maps

Home | Blog | Tutorials | Citation | Privacy

File Help | 100% Fit

Web Application
English
Launch Now!

Desktop Version
Windows | macOS | Linux

View Source
GitHub

It is often necessary to reverse engineer images of data visualizations to extract the underlying numerical data. WebPlotDigitizer is a semi-automated tool that makes this process extremely easy.

- Works with a wide variety of charts (XY, bar, polar, ternary, maps etc.)
- Automatic extraction algorithms make it easy to extract a large number of data points
- Free to use, opensource and cross-platform (web and desktop)
- Used in hundreds of published works by thousands of users
- Also useful for measuring distances or angles between various features
- More to come soon...

Version 4.5 Released (August 15, 2021)
[\[Release Notes \]](#)

WebPlotDigitizer - Copyright 2010-2019 Ankit Rohatgi

File Help | 100% Fit

Image
Axes
Datasets
Measurements

Symbiodinium spp. density (x10⁶) mg⁻¹ protein

control

Choose Plot Type

- 2D (X-Y) Plot
- 2D Bar Plot
- Polar Diagram
- Ternary Diagram
- Map With Scale Bar
- Image

Align Axes | Cancel

Align Bar Chart Axes

Click on two known points (P1, P2) on the continuous axes along the bars

Proceed

WebPlotDigitizer - Copyright 2010-2019 Ankit Rohatgi

File Help | 100% Fit

Image
Axes
Datasets
Measurements

Symbiodinium spp. density (x10⁶) mg⁻¹ protein

control

Bar Chart Calibration

Enter the values at the two points selected on the continuous axes along the bars

Point 1	Point 2	Log Scale
0	5	<input type="checkbox"/>

Rotated axes (not exactly vertical or horizontal)

OK

Axes Calibration
Click points to select and use cursor keys to adjust positions. Use Shift+Arrow for faster movement. Click complete when finished.

Complete!

A. tenuis *A. valida* *P. damicornis*

Tools for extracting data from figures

WebPlotDigitizer - Copyright 2010-2019 Ankit Rohatgi

File Help

+ - 100% Fit

Image
Axes
Bar

Datasets

Default Dataset

Measurements

Dataset

Axes: Bar

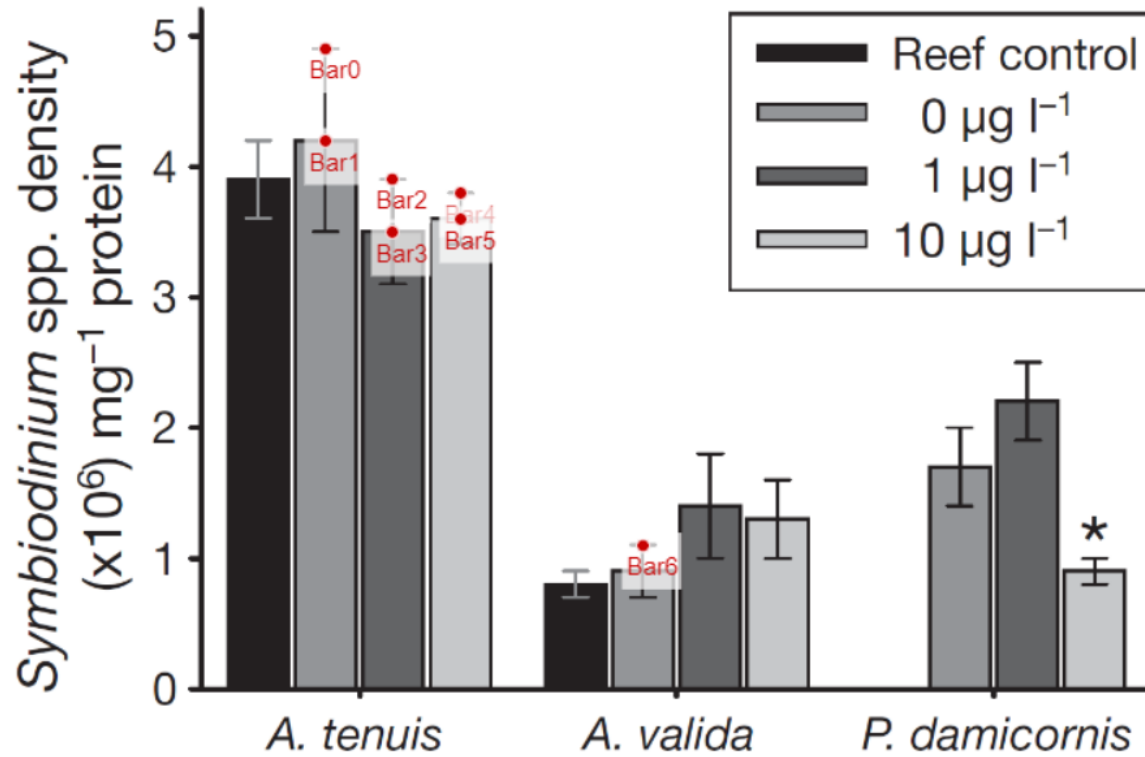
Rename Dataset

Delete Dataset

View Data

Clear Data

Data Points: 7



[1.1070e+0]

Manual Extraction

Add Point (A) Adjust Point (S)

Delete Point (D) Edit Labels (E)

Automatic Extraction

Mask Box Pen Erase View

Color Foreground Color

Distance 120 Filter Colors

Algorithm Bar Extraction

ΔX 30 Px

ΔVal 10 Px

Run

Visualizations to extract the underlying numerical data. WebPlotDigitizer is a semi-automated tool for extracting data from various types of plots (line, bar, scatter, ternary, maps etc.) and desktop applications. It is used by thousands of users and has many features.

1.4.5 Released (August 15, 2021)
[\[Release Notes\]](#)

Tools for extracting data from figures

<https://plotdigitizer.com/>

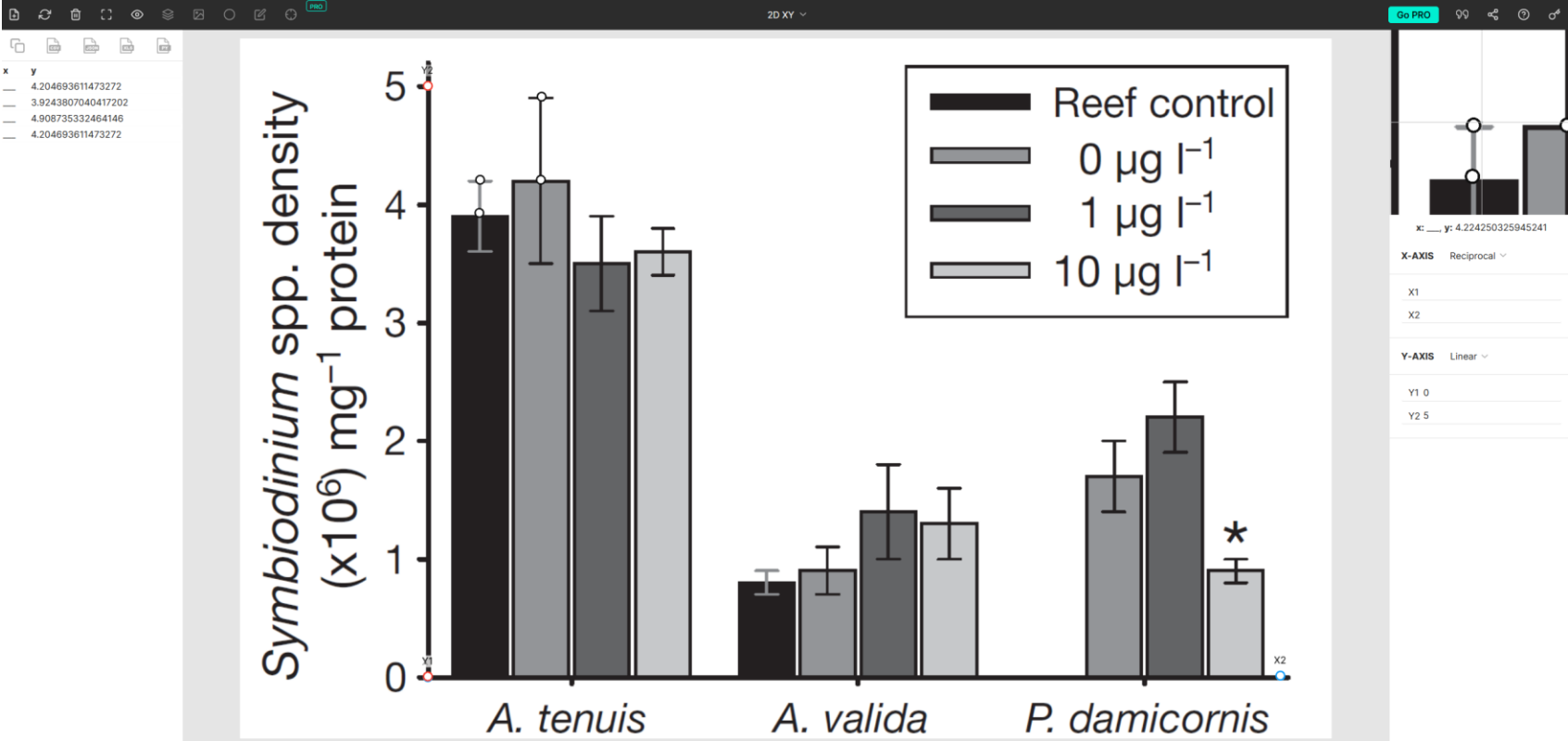

PlotDigitizer

Free Online App Features Download Buy Now

All-in-One Tool to Extract Data from Graphs, Plots & Images

Plotdigitizer is an online data extraction tool that allows users to extract data from images in numerical format. In short, it reverse-engineers your visual graphs into numbers. The software comes with plenty of useful and time-saving features.

Launch App Buy Now



Tools for extracting data from figures: metaDigitise



Received: 13 July 2018 | Accepted: 12 October 2018

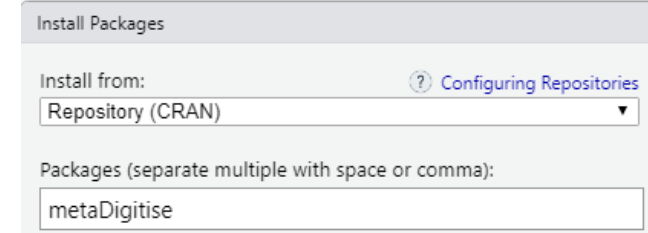
DOI: 10.1111/2041-210X.13118

APPLICATION

Methods in Ecology and Evolution 

Reproducible, flexible and high-throughput data extraction from primary literature: The METADIGITISE R package

Joel L. Pick  | Shinichi Nakagawa | Daniel W. A. Noble 



(+) possible to save, trace and modify data extraction

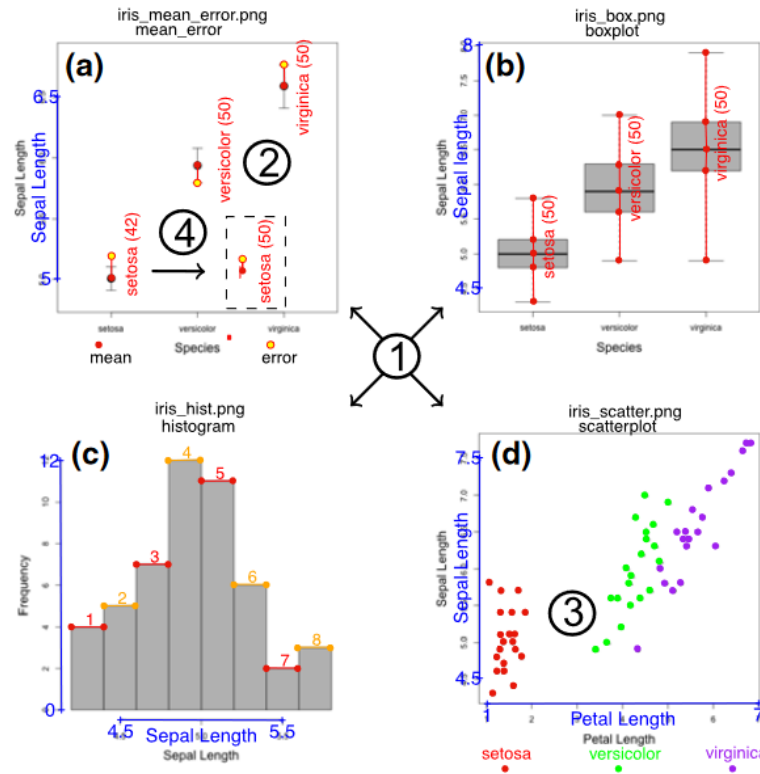
(-) no zoom

<https://cran.r-project.org/web/packages/metaDigitise/vignettes/metaDigitise.html>

Tools for extracting data from figures: metaDigitise

metaDigitise

Plot interface



Data output

filename	variable	group_id	mean	sd	n	r	plot_type
iris_box.png	Sepal length	setosa	5.01	0.317	50	NA	boxplot
iris_box.png	Sepal length	versicolor	5.93	0.497	50	NA	boxplot
iris_box.png	Sepal length	virginica	6.49	0.603	50	NA	boxplot
iris_hist.png	Sepal Length	setosa	4.95	0.364	50	NA	histogram
iris_mean_error.png	Sepal Length	setosa	5.01	0.680	50	NA	mean_error
iris_mean_error.png	Sepal Length	versicolor	5.94	1.025	50	NA	mean_error
iris_mean_error.png	Sepal Length	virginica	6.59	1.251	50	NA	mean_error
iris_scatter.png	Petal Length	setosa	1.44	0.215	20	0.109	scatterplot
iris_scatter.png	Sepal Length	setosa	5.03	0.427	20	0.109	scatterplot
iris_scatter.png	Petal Length	versicolor	4.29	0.415	20	0.786	scatterplot
iris_scatter.png	Sepal Length	versicolor	5.97	0.603	20	0.786	scatterplot
iris_scatter.png	Petal Length	virginica	5.66	0.668	20	0.932	scatterplot

FUNCTIONALITY

① Different plot types

Capable of handling A) mean error plots, B) boxplots, C) histograms and D) scatterplots

② Entry of Metadata

Enter sample sizes variable and group names while digitising that are displayed on plot

③ Grouped Data

Enter as many groups as needed to capture descriptive statistics for sub-samples of data

④ Digitise, edit or replot digitisations

Simple user interface to guide user. Can digitise new images, edit digitisations or easily replot previous digitisations and metadata by cycling through images or choosing specific images

⑤ Summarising data

Get descriptive statistics automatically calculated for all plot types or use raw x,y data, if desired

⑥ Multiple image processing

Process as many images at once as needed and of varying types efficiently and quickly. New plots automatically plotted for digitisation

Question: estimate the effects of chemicals on the photosynthetic performance (*maximum quantum yield, Fv/Fm*) of tropical reef-building corals

Sample of 3 articles

A **study** = a taxon × an exposure × an outcome

A **case study** = one tested concentration-duration

Effect size = standardized mean difference

If monitoring over time, extraction of the longest duration of exposure

Extract data using the metaDigitise package

Question: estimate the effects of chemicals on the photosynthetic performance (*maximum quantum yield, Fv/Fm*) of tropical reef-building corals

Sample of 3 articles

A **study** = a taxon × an exposure × an outcome

A **case study** = one tested concentration-duration

Effect size = standardized mean difference

If monitoring over time, extraction of the longest duration of exposure

1 – Extract data

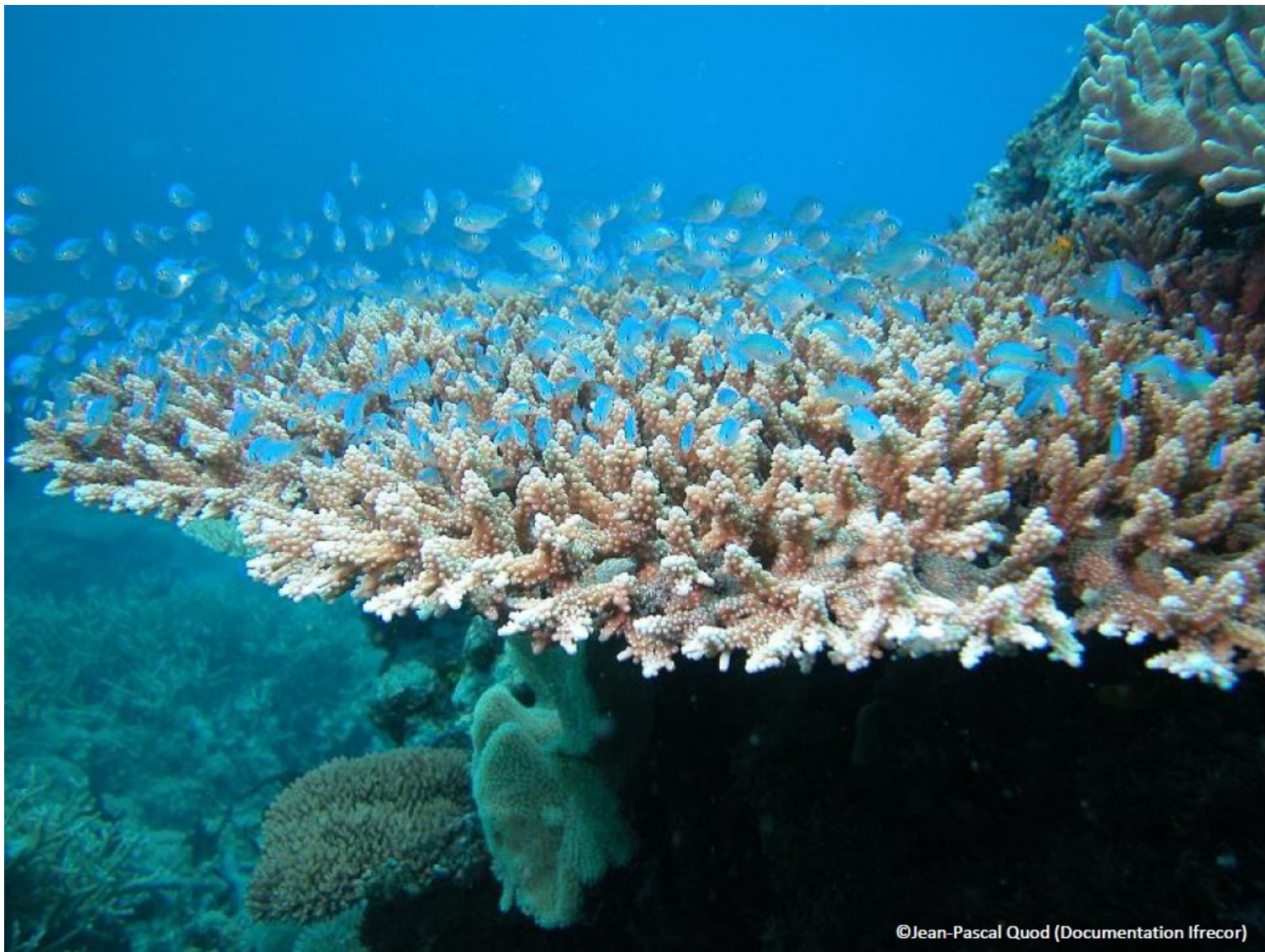
Install et load metaDigitise package; set working directory

Prepare the figure files (screen capture → .png), put them in a "figs" folder in the working directory

```
> dat <- metaDigitise(dir = "./figs")
```

Extract data into an Excel file

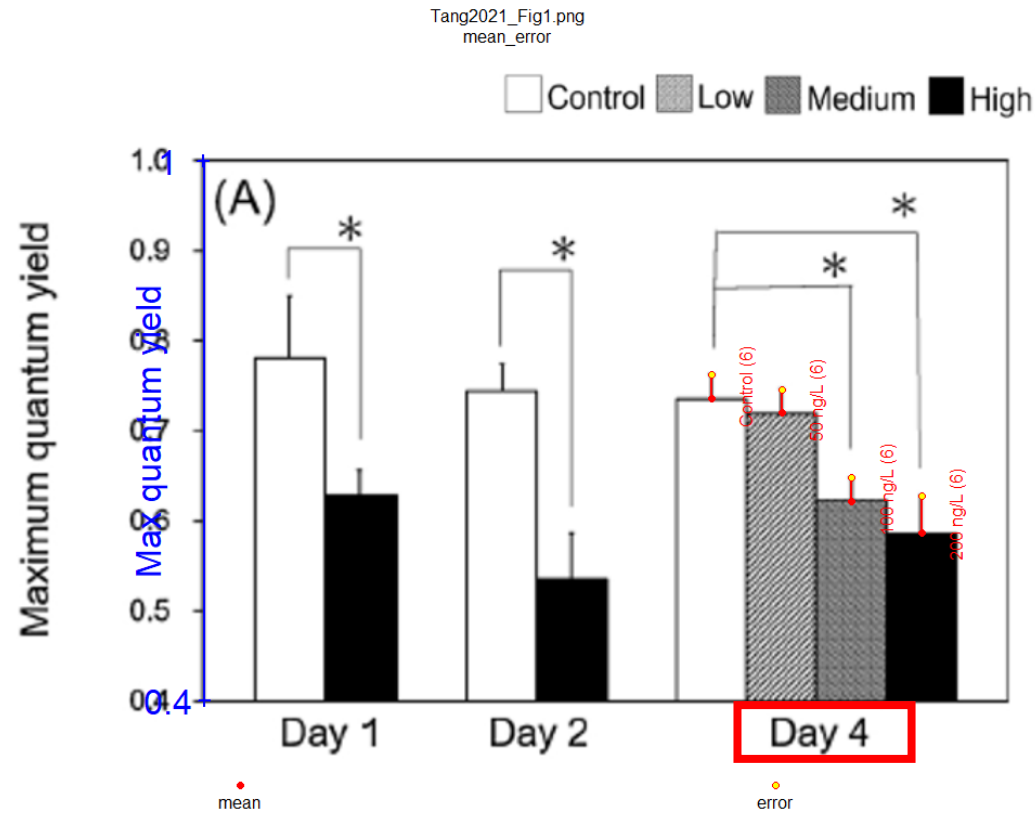
2 – Discussion



©Jean-Pascal Quod (Documentation Ifremer)

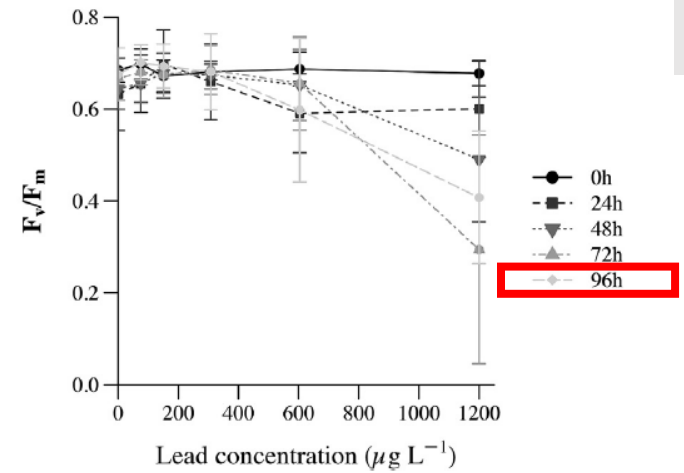
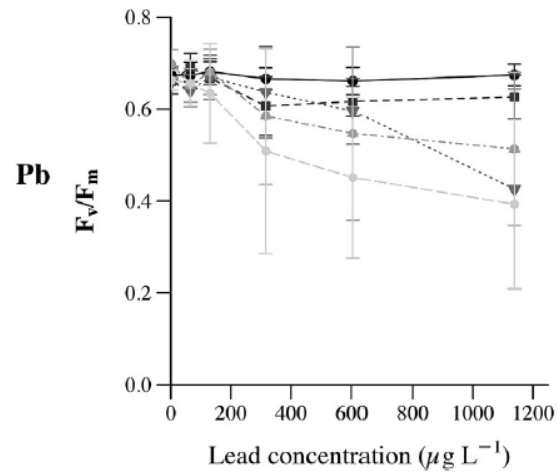
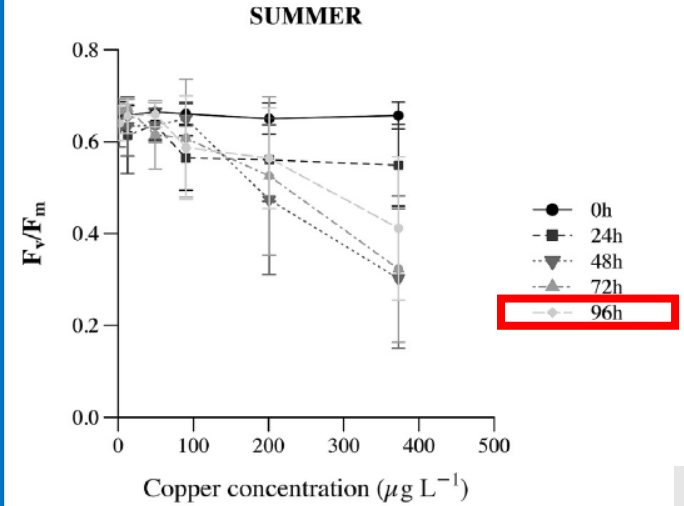
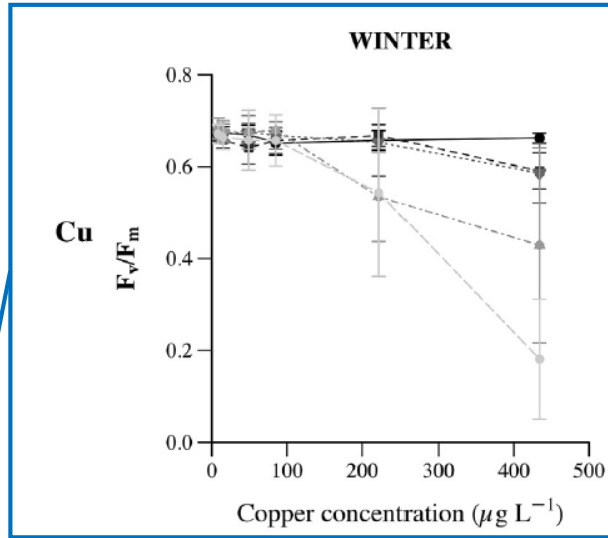
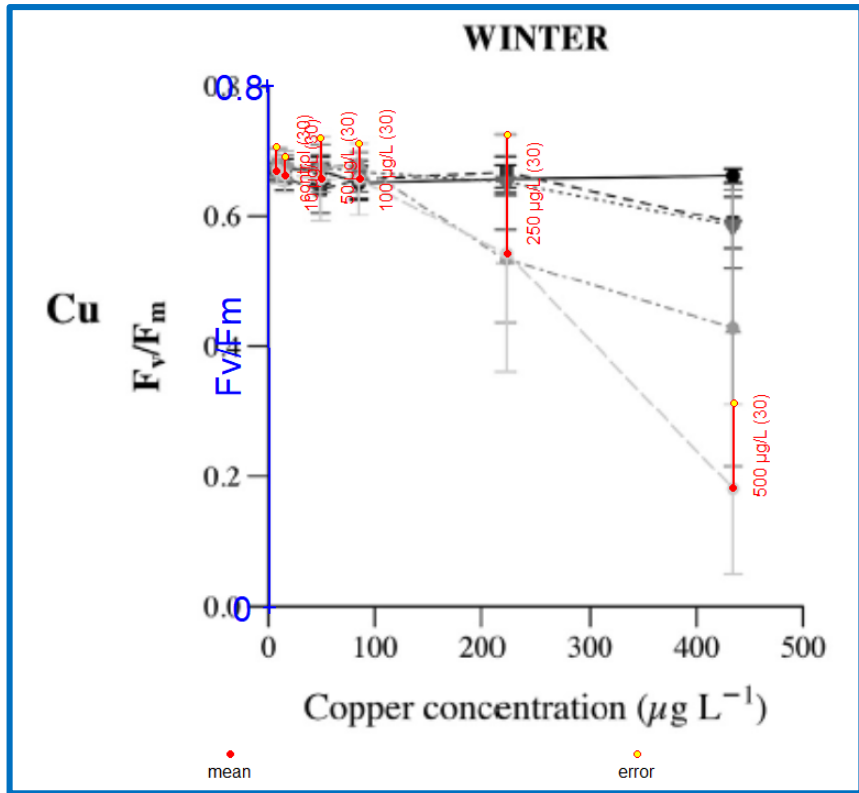
Tang et al. 2021

Fig. 1. Differences in photophysiological parameters of symbiotic algae in coral exposed to Irgarol levels of 50 (low), 100 (Medium) and 200 (High) ng/L. The marker “**” indicates a significant difference between the Irgarol-treated and control groups (paired *t*-test, $p < 0.05$, $N = 6$). The error bar indicates the value of the standard deviation.



N = 6
SD

Hédouin et al. 2016



N = 30
SD

Fig. 8. Dark-adapted quantum yield (F_v/F_m , mean \pm standard deviation, n = 30 nubbins per condition, -pooled data-) of *Pocillopora damicornis* corals exposed for 96 h to a range of dissolved Cu and Pb concentration in the summer and winter seasons.

Kegler et al. 2015

Table 3. Summary of *P. verrucosa* responses.

	Dark respiration [mgO ₂ h ⁻¹ cm ⁻²]	Net photosynthesis [mgO ₂ h ⁻¹ cm ⁻²]	Gross photosynthesis [mgO ₂ h ⁻¹ cm ⁻²]	Maximum quantum yield 48 h [F _v /F _m]	Maximum quantum yield 84 h [F _v /F _m]	Tissue loss after 84 h [% loss]
Control	0.019 ± 0.005	0.008 ± 0.003	0.011 ± 0.003	0.71 ± 0.02	0.71 ± 0.02	-
High temperature	0.012 ± 0.003	0.003 ± 0.001	0.009 ± 0.003	0.74 ± 0.01	0.72 ± 0.01	-
Diesel	0.015 ± 0.001	0.006 ± 0.003	0.001 ± 0.003	0.71 ± 0.02	0.71 ± 0.02	-
LAS	-	-	-	0.73 ± 0.01	-	52.5 ± 30.15
Diesel + high temperature	0.023 ± 0.003	0.008 ± 0.002	0.014 ± 0.005	0.74 ± 0.01	0.71 ± 0.01	-
LAS + high temperature	-	-	-	0.63 ± 0.13	-	92.25 ± 7.26

N = 4
SD