

# Modelling the effects of domestication in Wheat through novel computer-vision techniques

Nathan Hughes

March 16, 2018

# Outline

- 1 Description of Topic
- 2 Materials and Image Capture Setup
- 3 Completed Work
- 4 Work in Progress
- 5 Additional Aims
- 6 References

# What is the project?

## Description

The project is aiming to use computational methods to answer biologically significant questions on wheat grain morphology and domestication using  $\mu$ CT images.

## How?

To do this, I will be using:

- Computer vision on 3D image sets
- Statistical analysis and data science
- Scientific theory to create reproducible results

## Research Question:

Is it possible to use  $\mu$ CT imaging to answer questions about Wheat domestication?

*I hope so!*

### Null-Hypothesis

- $H_0$  = Domestication has no effect on the morphometric properties of wheat

*\*Hulling is a trait associated with domestication*

# Research Question:

Is it possible to use  $\mu$ CT imaging to answer questions about Wheat domestication?

*I hope so!*

## Null-Hypothesis

- $H_0$  = Domestication has no effect on the morphometric properties of wheat
- $H_0$  = Ploidy has no effect on the morphometric properties of wheat

*\*Hulling is a trait associated with domestication*

## Research Question:

Is it possible to use  $\mu$ CT imaging to answer questions about Wheat domestication?

*I hope so!*

### Null-Hypothesis

- $H_0$  = Domestication has no effect on the morphometric properties of wheat
- $H_0$  = Ploidy has no effect on the morphometric properties of wheat
- $H_0$  = There is no difference in hulled and non-hulled genotypes

*\*Hulling is a trait associated with domestication*

# Extracted Features

## Features List

The features I am collecting are:

- Length

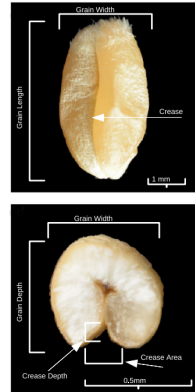


Figure 1: Major features extracted from analysis

# Extracted Features

## Features List

The features I am collecting are:

- Length
- Width

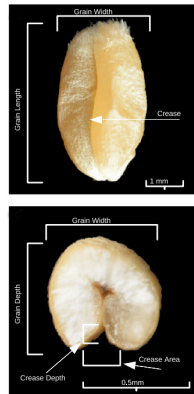


Figure 1: Major features extracted from analysis



# Extracted Features

## Features List

The features I am collecting are:

- Length
- Width
- Depth

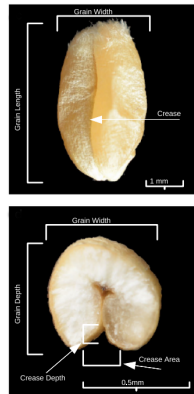


Figure 1: Major features extracted from analysis

# Extracted Features

## Features List

The features I am collecting are:

- Length
- Width
- Depth
- Volume

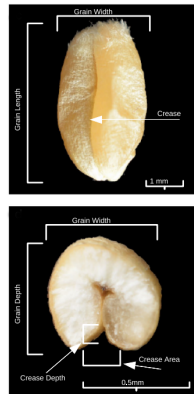


Figure 1: Major features extracted from analysis

# Extracted Features

## Features List

The features I am collecting are:

- Length
- Width
- Depth
- Volume
- Surface Area

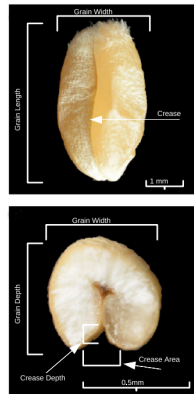


Figure 1: Major features extracted from analysis

# Extracted Features

## Features List

The features I am collecting are:

- Length
- Width
- Depth
- Volume
- Surface Area
- Crease Depth

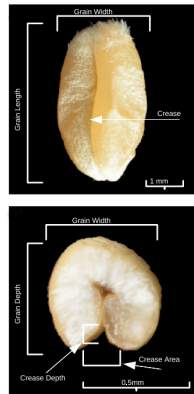


Figure 1: Major features extracted from analysis

# Extracted Features

## Features List

The features I am collecting are:

- Length
- Width
- Depth
- Volume
- Surface Area
- Crease Depth
- Crease Volume

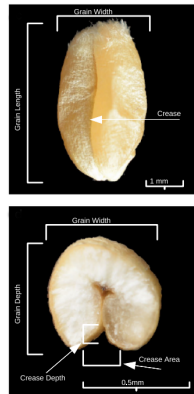


Figure 1: Major features extracted from analysis

# Extracted Features

## Features List

The features I am collecting are:

- Length
- Width
- Depth
- Volume
- Surface Area
- Crease Depth
- Crease Volume
- X,Y,Z coordinates of grains

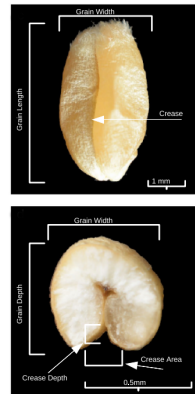


Figure 1: Major features extracted from analysis

# Aims

## Primary Aims

I am wanting to produce:

- A software library (in Python) which can be used to help analysis of  $\mu$ CT scanned seeds

# Aims

## Primary Aims

I am wanting to produce:

- A software library (in Python) which can be used to help analysis of  $\mu$ CT scanned seeds
- A GUI application for researchers to use to auto analyse seeds



# Aims

## Primary Aims

I am wanting to produce:

- A software library (in Python) which can be used to help analysis of  $\mu$ CT scanned seeds
- A GUI application for researchers to use to auto analyse seeds
- Answers to my hypothesis

# Outline

- 1 Description of Topic
- 2 Materials and Image Capture Setup
- 3 Completed Work
- 4 Work in Progress
- 5 Additional Aims
- 6 References

# Materials (Plant)

## Wheat information

I have a wide range of Wheat genotypes, these are:

- Ranged between diploid, tetraploid and hexiploid
- 12 total genotypes
- Divided between domestication status



Figure 2: A Primitive Wheat Spike

# Methods (Image Generation)

## Capture Methods

In order to generate this data a  $\mu$ CT100 scanner (ScancoMedical, Switzerland) was used Each spike was:

- ~1000 slices (51 slices per stack),
- 125 projections/180° were taken and a
- binning of 6 was used.
- Output images are a 0.2 megapixel (512 × 512) resolution (68.8  $\mu$ /pixel)

# Outline

- 1 Description of Topic
- 2 Materials and Image Capture Setup
- 3 Completed Work
- 4 Work in Progress
- 5 Additional Aims
- 6 References

# Data Gathering

## MATLAB Software

Data has been extracted from the  $\mu$ CT images using a method I previously developed at IBERS [1]. A lot of additions and novel augmentations have been added in order to work with the range of Wheat genotypes.

- This is done via separating non-grain from grain ROIs in the image data
- Measurements are taken on a per grain level
- Measurements are in mm,  $\text{mm}^2$  and  $\text{mm}^3$  based on type
- Data is exported as CSV files

# Improved Watershedding Algorithm

## Algorithm

$$|x_1 - x_2| + (\sqrt{2} - 1), |x_1 - x_2| > |y_1 - y_2| (\sqrt{2} - 1) |x_1 - x_2|, \text{ otherwise} \quad (1)$$

## Visual Example

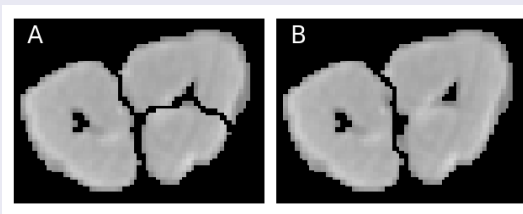


Figure 3: A showing the chessboard method, B improved quasi-euclidean method

# Grain Analyser Library example output - 1

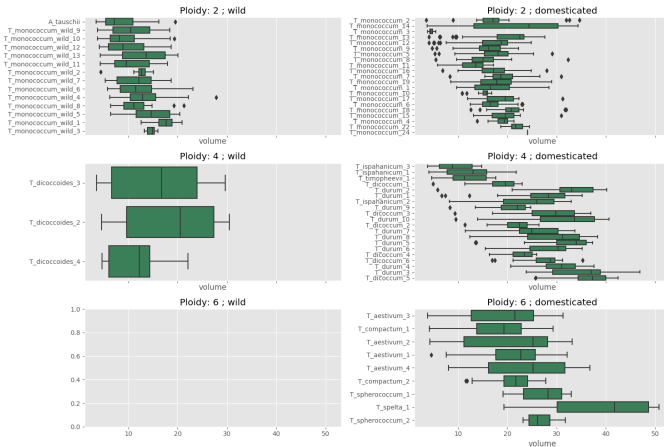


Figure 4: Statistical Python Library Boxplot output



# Grain Analyser Library example output - 2

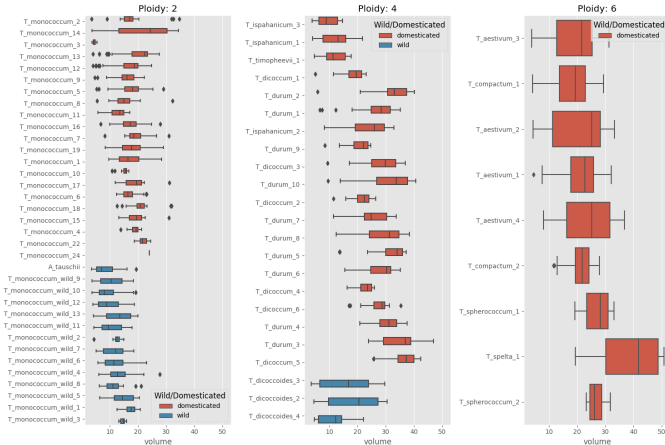


Figure 5: Statistical Python Library Boxplot output style 2

# Grain Analyser Library example output - 3

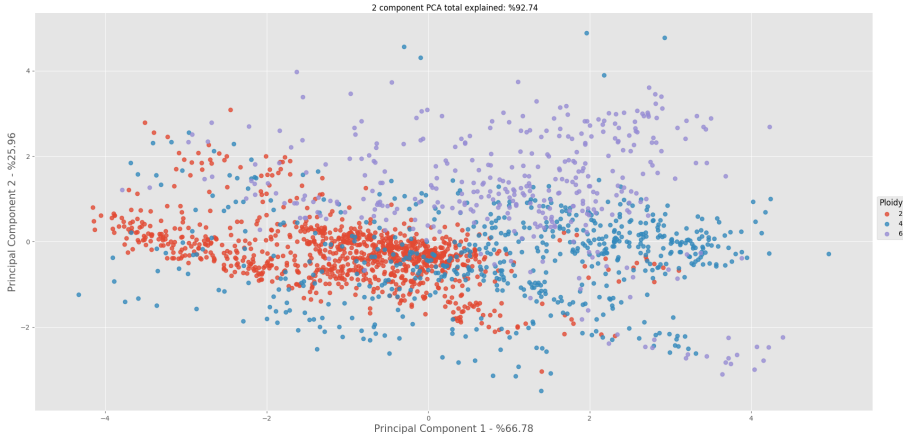


Figure 6: Statistical Python Library PCA output

# Grain Analyser Library example output - 4



Figure 7: Statistical Python Library PCA output

# Outline

- 1 Description of Topic
- 2 Materials and Image Capture Setup
- 3 Completed Work
- 4 Work in Progress
- 5 Additional Aims
- 6 References

# Data Analysis *WIP*

## Statistical Methods to be implemented

These will be implemented in Python using the library *SciPy* [2]

- ANOVA
- MANOVA
- GLMs
- $\chi^2$  Tests
- PCA
- T/F-Tests

## Graphing Methods to be implemented

All Graphing is being implemented using Matplotlib [3]

- Box
- Swarm
- Scatter / LM
- Violin
- Correlation Maps
- Frequency/Histograms

# Reproducible Analysis GUI - 1 *WIP*

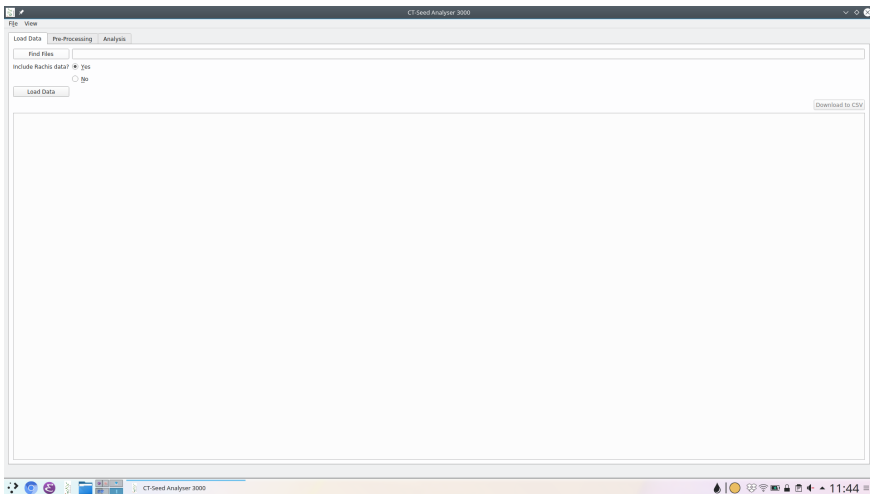


Figure 8: Initial GUI

# Reproducible Analysis GUI - 2 *WIP*

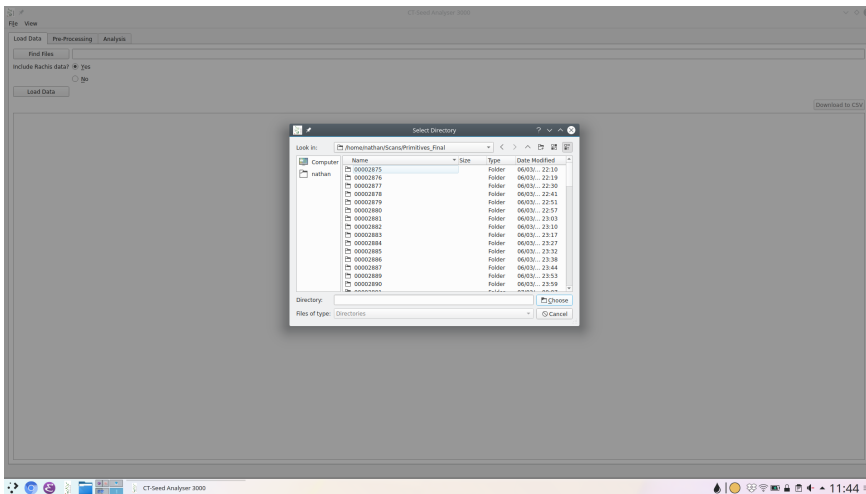


Figure 9: GUI select folders to search for data

# Reproducible Analysis GUI - 3 WIP

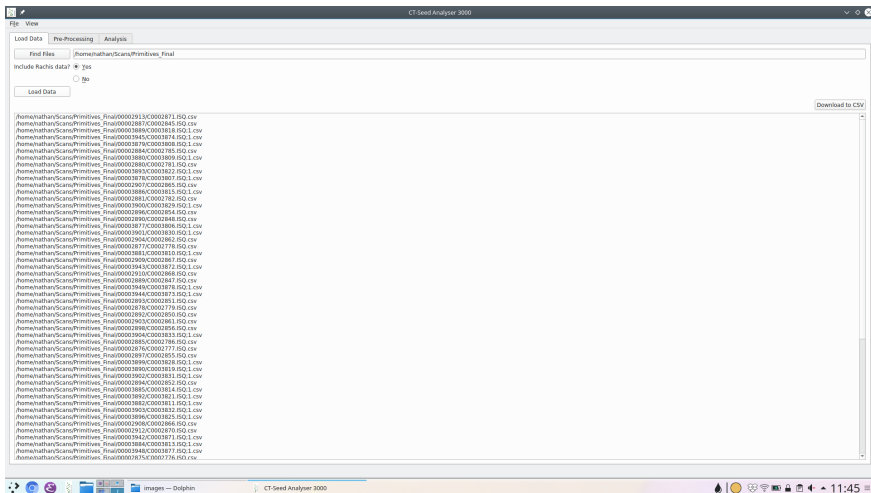


Figure 10: GUI with data loaded successfully



# Reproducible Analysis GUI - 4 *WIP*

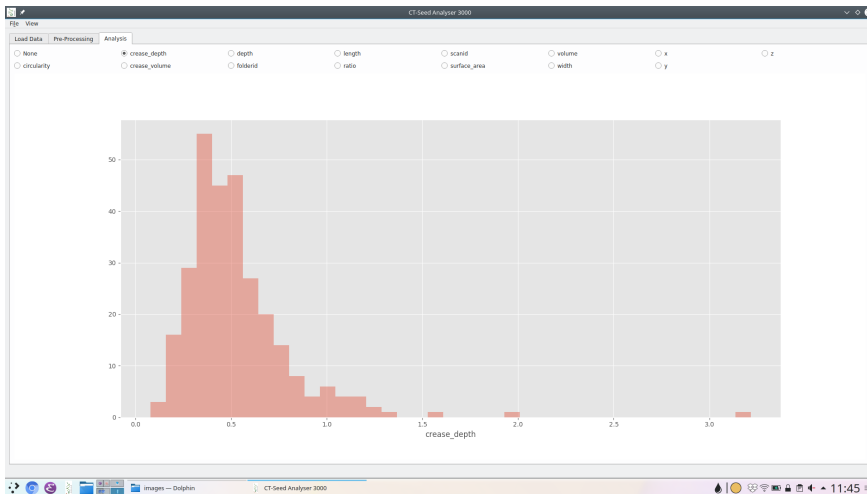


Figure 11: GUI dynamically graphing crease\_depth

# Reproducible Analysis GUI - 5 *WIP*

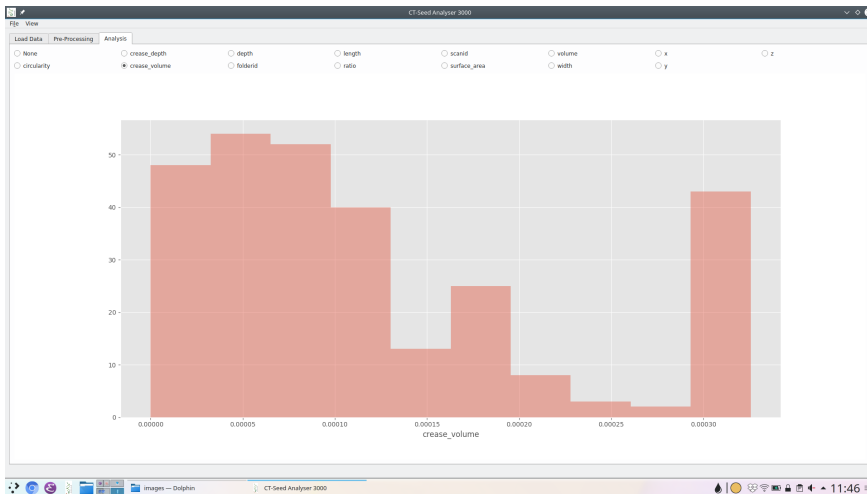


Figure 12: GUI dynamically graphing crease\_volume

# Outline

- 1 Description of Topic
- 2 Materials and Image Capture Setup
- 3 Completed Work
- 4 Work in Progress
- 5 Additional Aims
- 6 References

## Additional Aims

### Not yet started, but would like to implement by end of project

- Automate error finding and identification
- Visually Display Errors in Data
- Write testing suite for software

### Ideas for future work

- Test software library on other datasets
- Grow more plants and create more replicates for more robust study

# Outline

- 1 Description of Topic
- 2 Materials and Image Capture Setup
- 3 Completed Work
- 4 Work in Progress
- 5 Additional Aims
- 6 References

# References

- [1] Nathan Hughes, Karen Askew, Callum P Scotson, Kevin Williams, Colin Sauze, Fiona Corke, John H Doonan, and Candida Nibau.  
Non-destructive, high-content analysis of wheat grain traits using X-ray micro computed tomography.  
*Plant Methods*, 13, 2017.
- [2] Eric Jones, Travis Oliphant, and Pearu Peterson.  
{SciPy}: open source scientific tools for {Python}.  
2014.
- [3] J. D. Hunter.  
Matplotlib: A 2d graphics environment.  
*Computing In Science & Engineering*, 9(3):90–95, 2007.