

USING MACHINE LEARNING AND GROWTH MODELS TO STREAMLINE LARGE SCALE ROOT PHENOTYPING

Guillaume Lobet, Manu Noll, Jonathan
Atkinson, Darren Wells, Xavier Draye,
Patrick Thaon, Iko Koevoets

g.lobet@fz-juelich.de

www.rosi.science

 [@guillaumelobet](https://twitter.com/guillaumelobet)



HOW CAN WE ANALYSE
LARGE NUMBER
OF
COMPLEXE
ROOT SYSTEMS

IMAGE
ANALYSIS



ROOT
MODELS



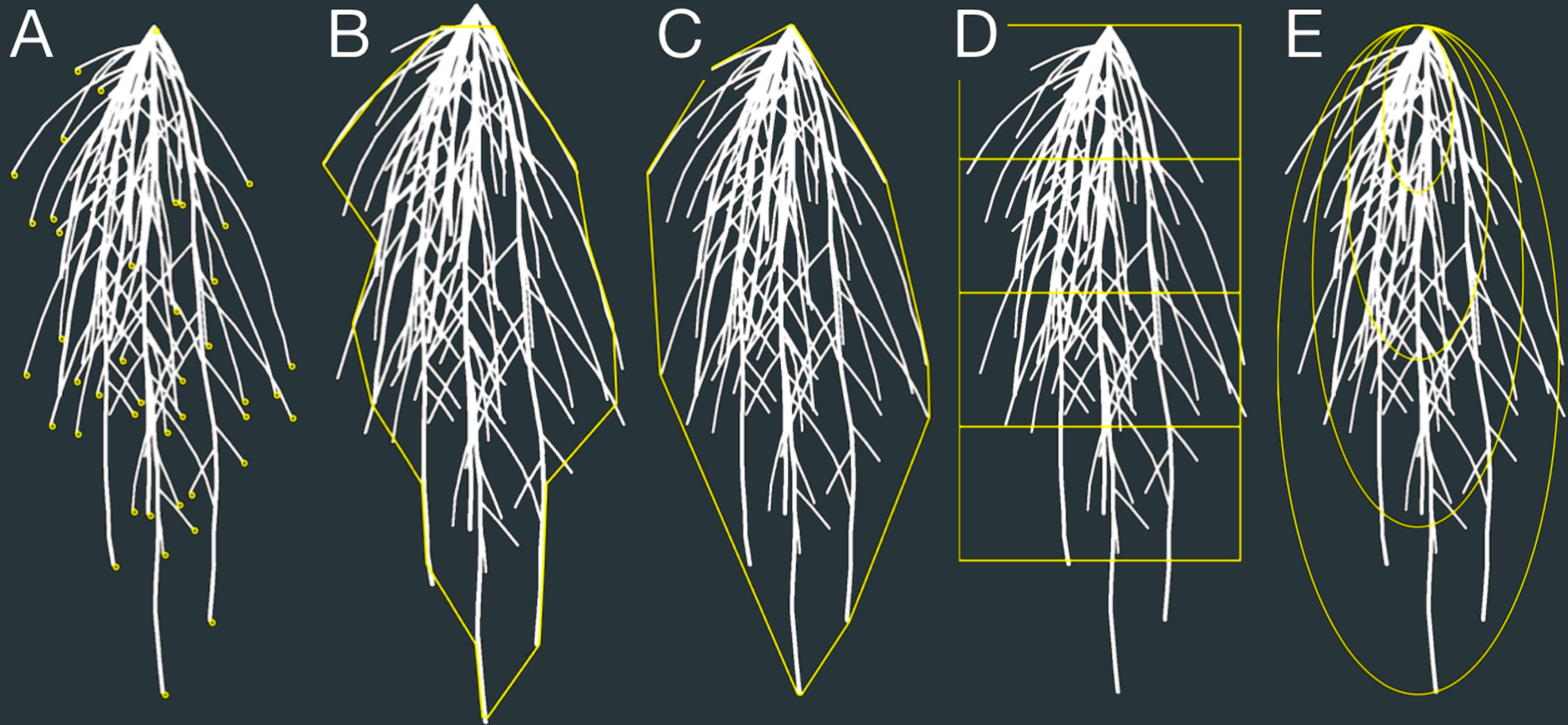
MACHINE
LEARNING



EXTRACT
USEFUL
INFORMATION
FROM ROOT
IMAGES



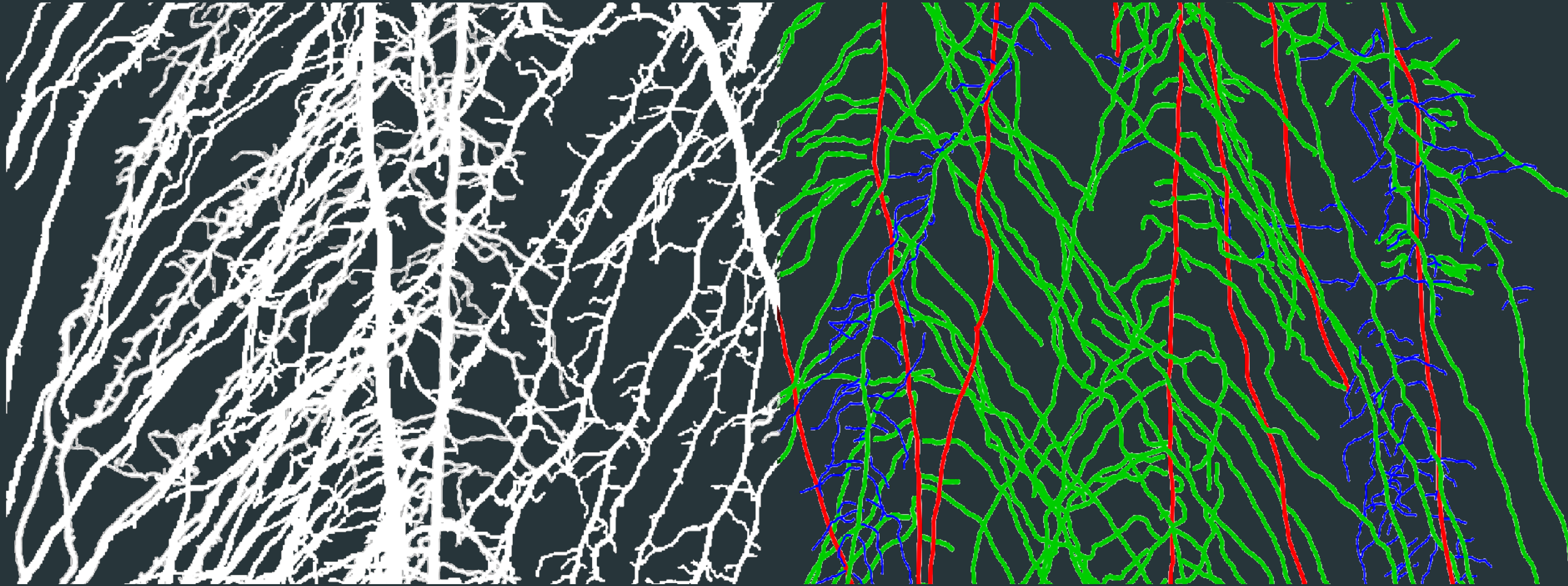
CHEAP VARIABLES = AUTOMATED



DIRT | WINRHIZO | **RIA-J** | GIA-ROOT | ROOTGRAPH



EXPENSIVE VARIABLES = **MANUAL**



SMARTROOT | DART | **ROOTNAV** | EZ-RHIZO



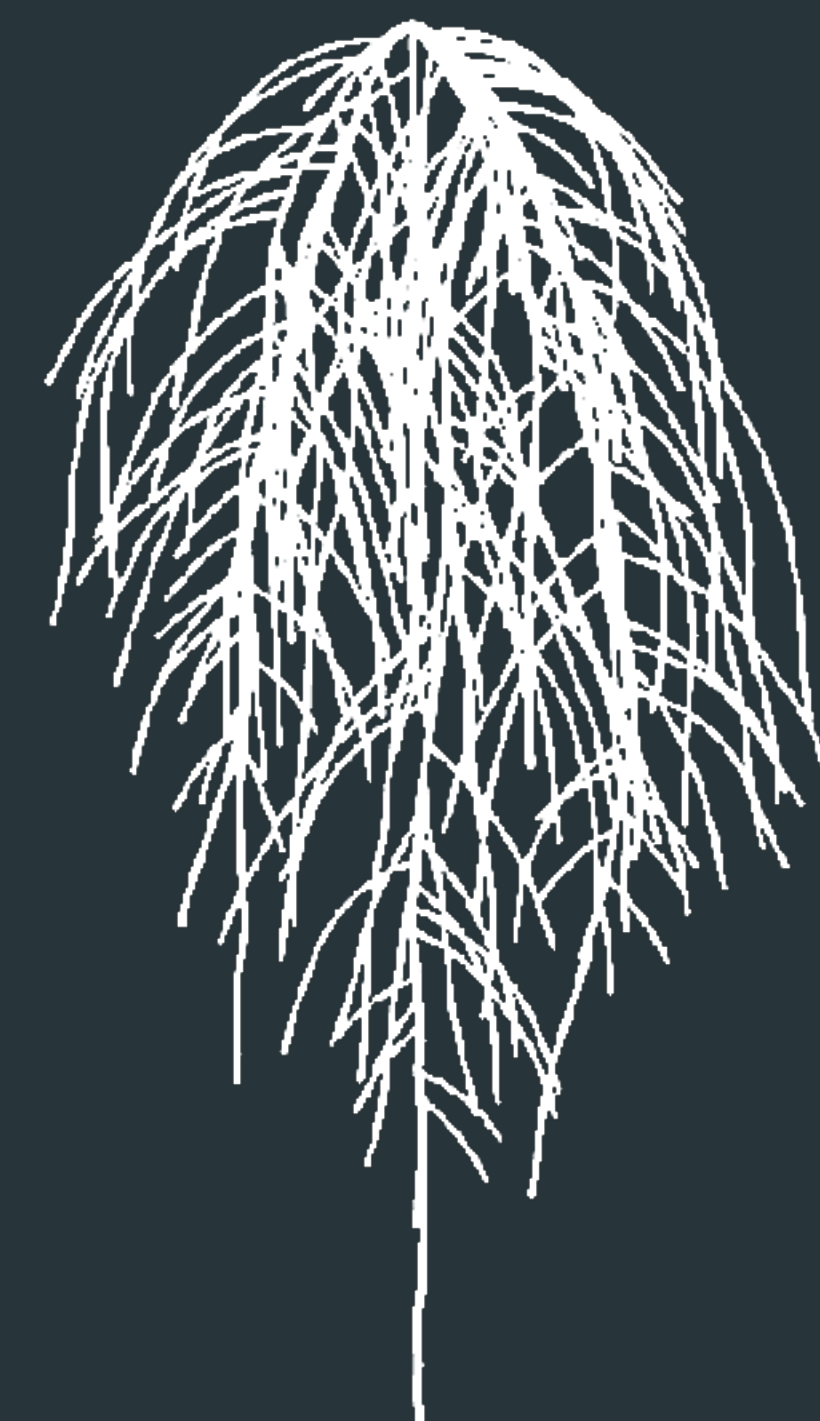
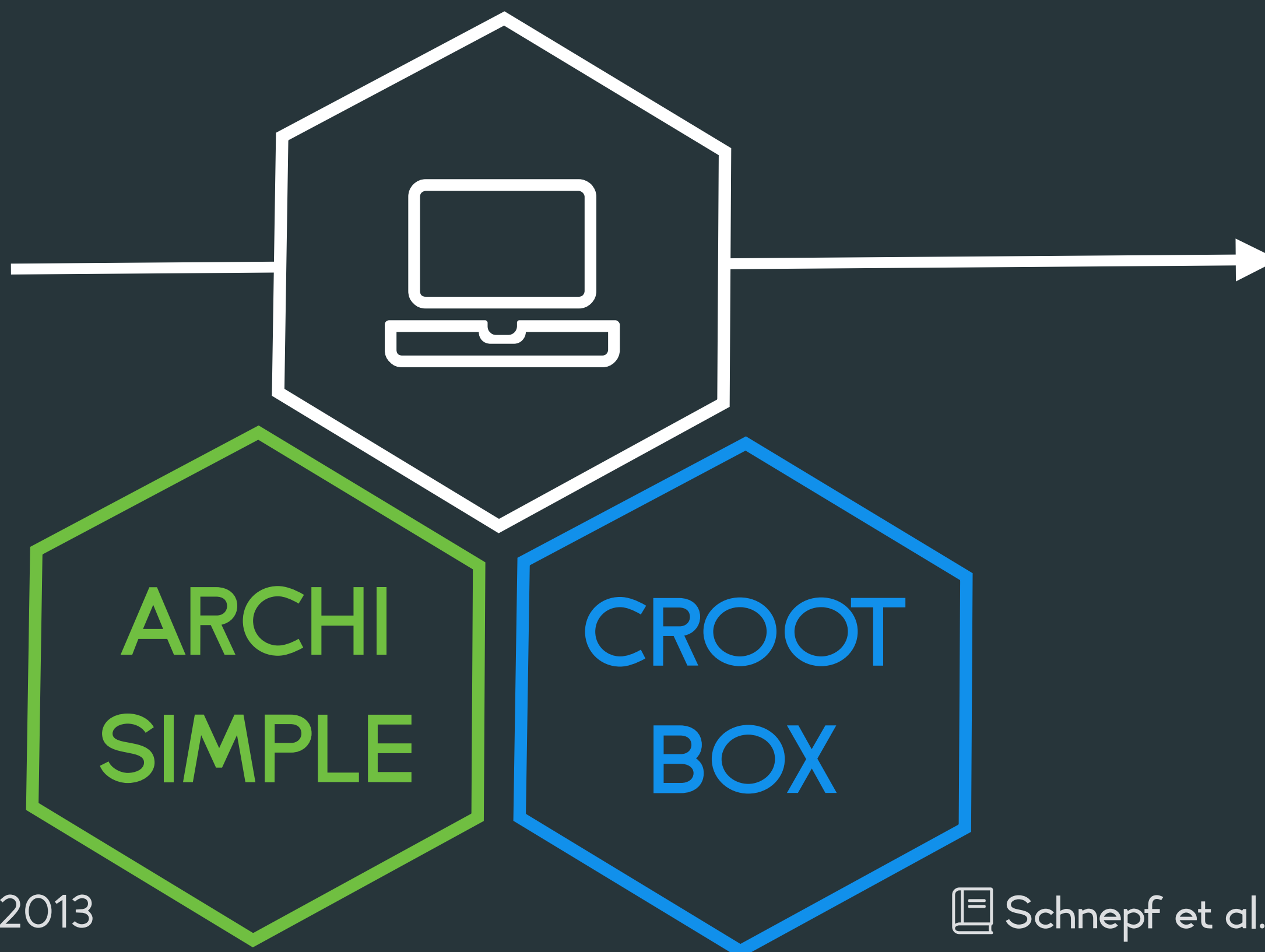
ROOT **STRUCTURAL** MODEL

PARAMETER SET

MODEL

ROOT ARCHITECTURE

[P1, P2, P3, ..., P_n]



 Pagès et al. 2013

DOI: [10.1016/j.ecolmodel.2013.11.014](https://doi.org/10.1016/j.ecolmodel.2013.11.014)

 Schnepf et al. 2017

DOI: [10.1101/139980](https://doi.org/10.1101/139980)



MACHINE LEARNING: **RANDOM FORESTS**

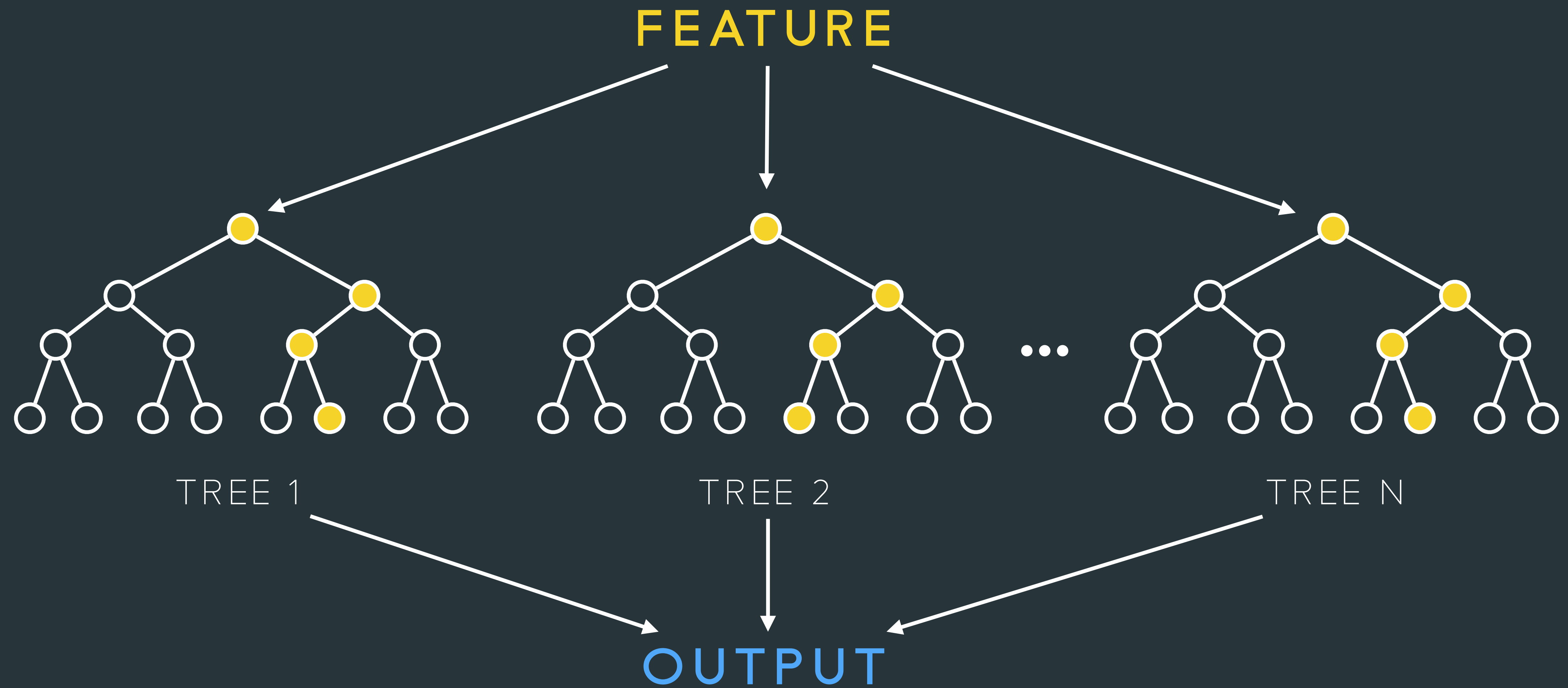


IMAGE
ANALYSIS



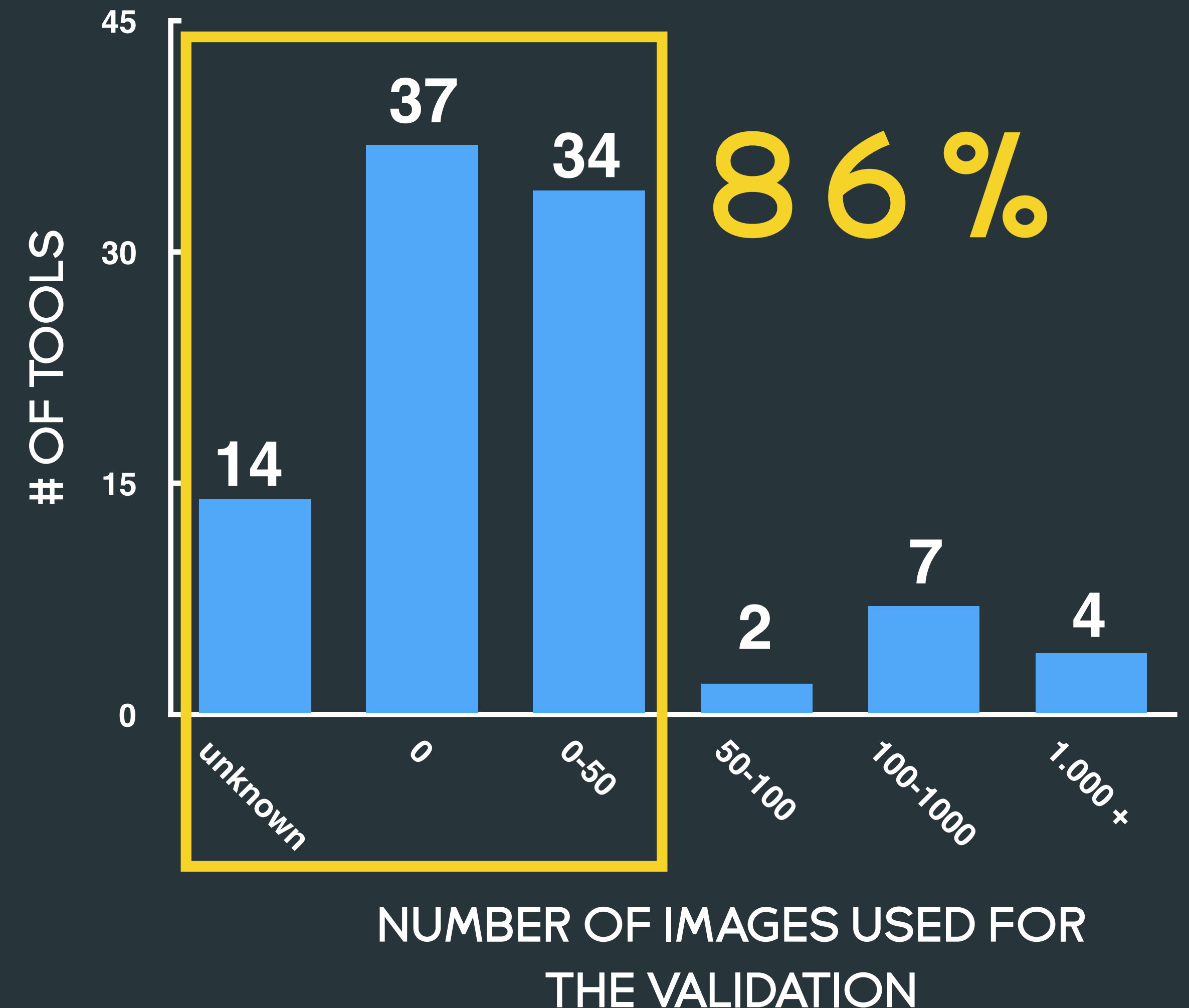
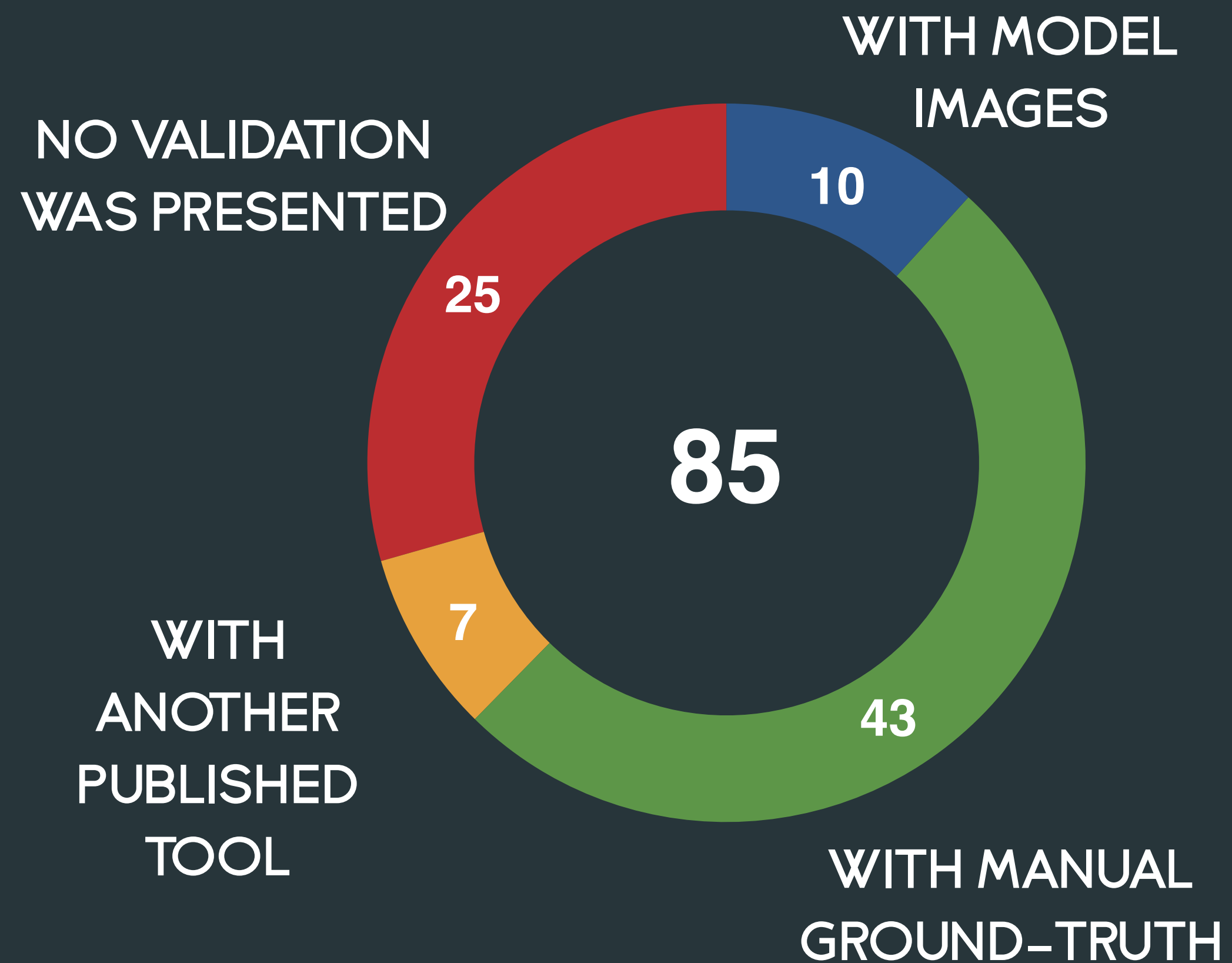
ROOT
MODELS

1



MACHINE
LEARNING

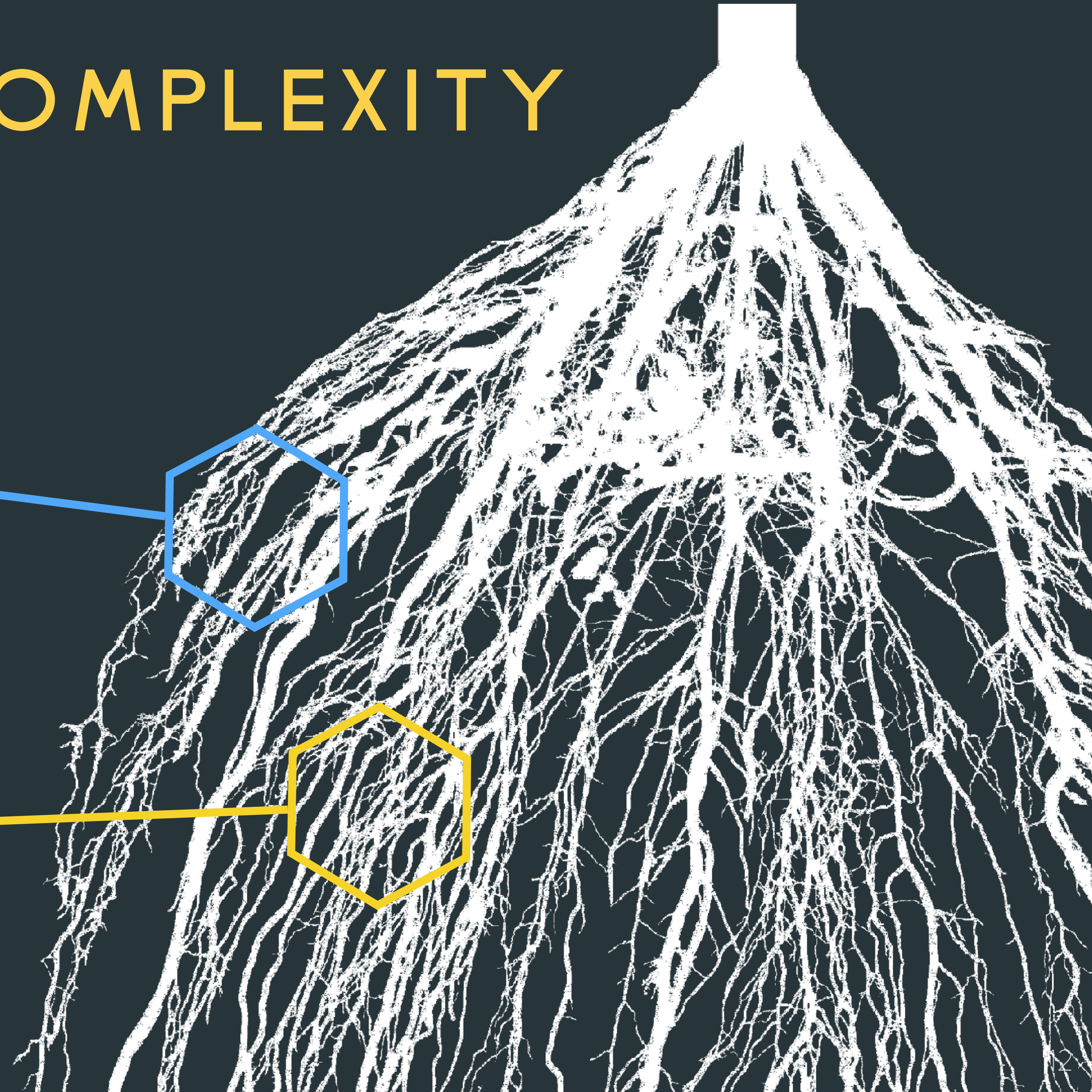
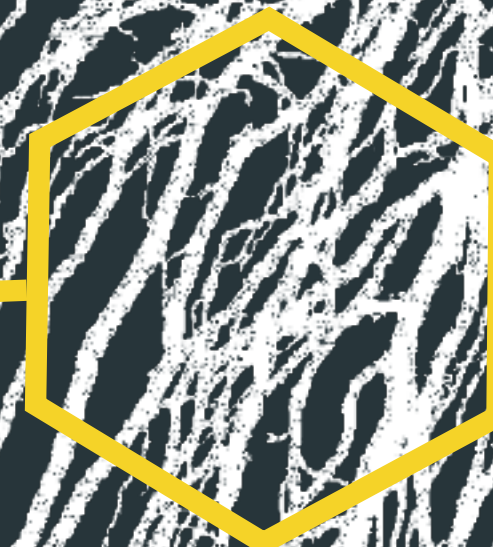
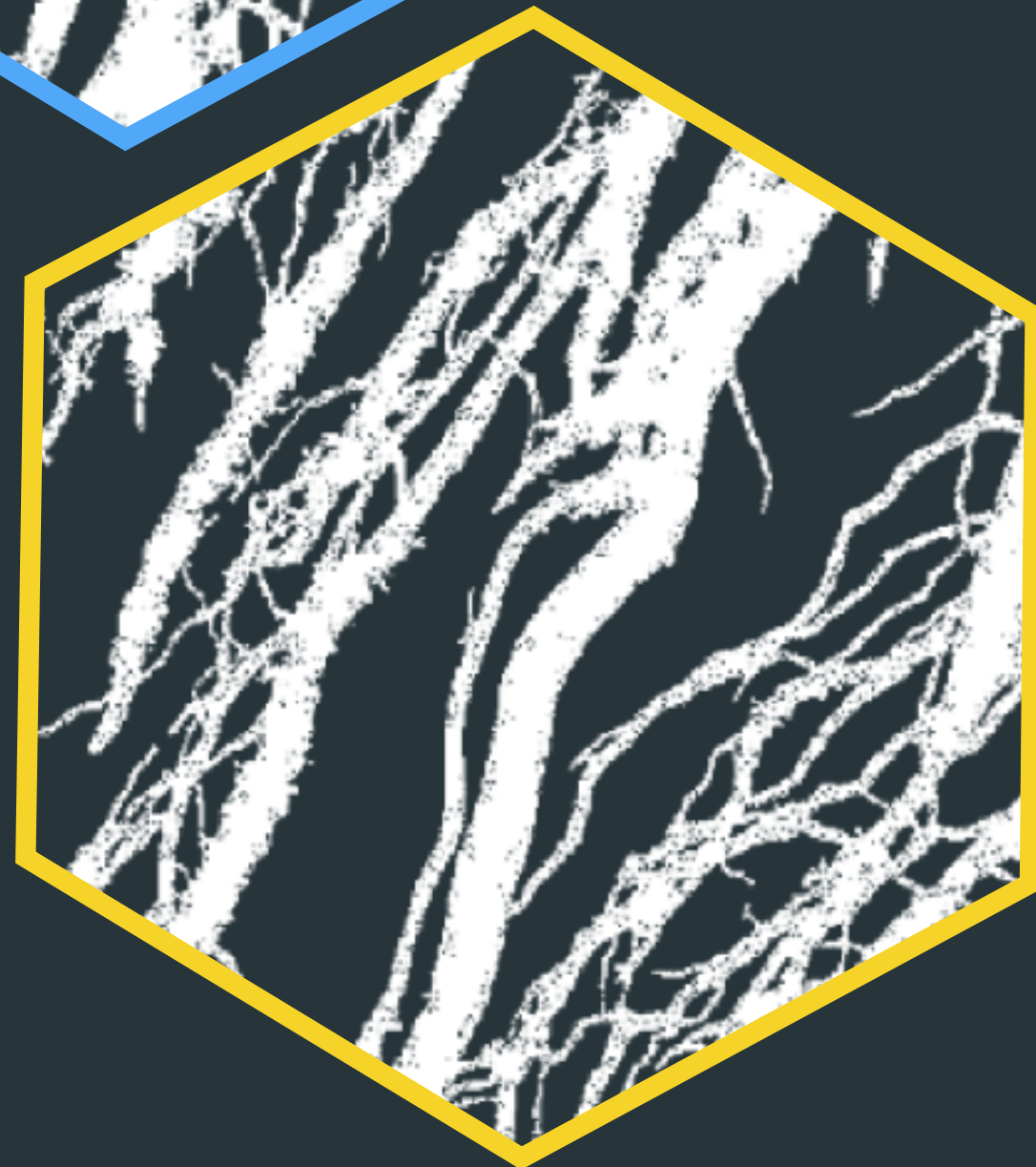
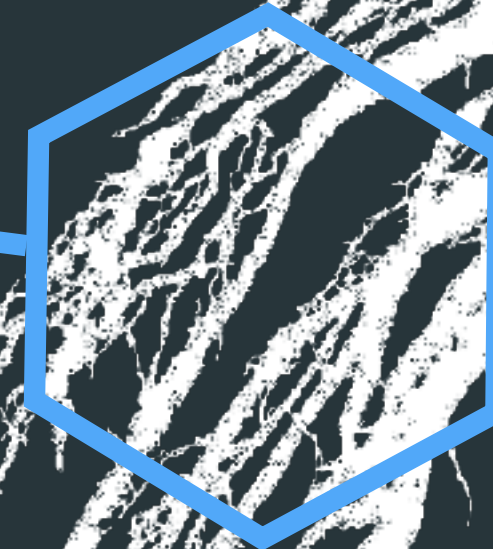
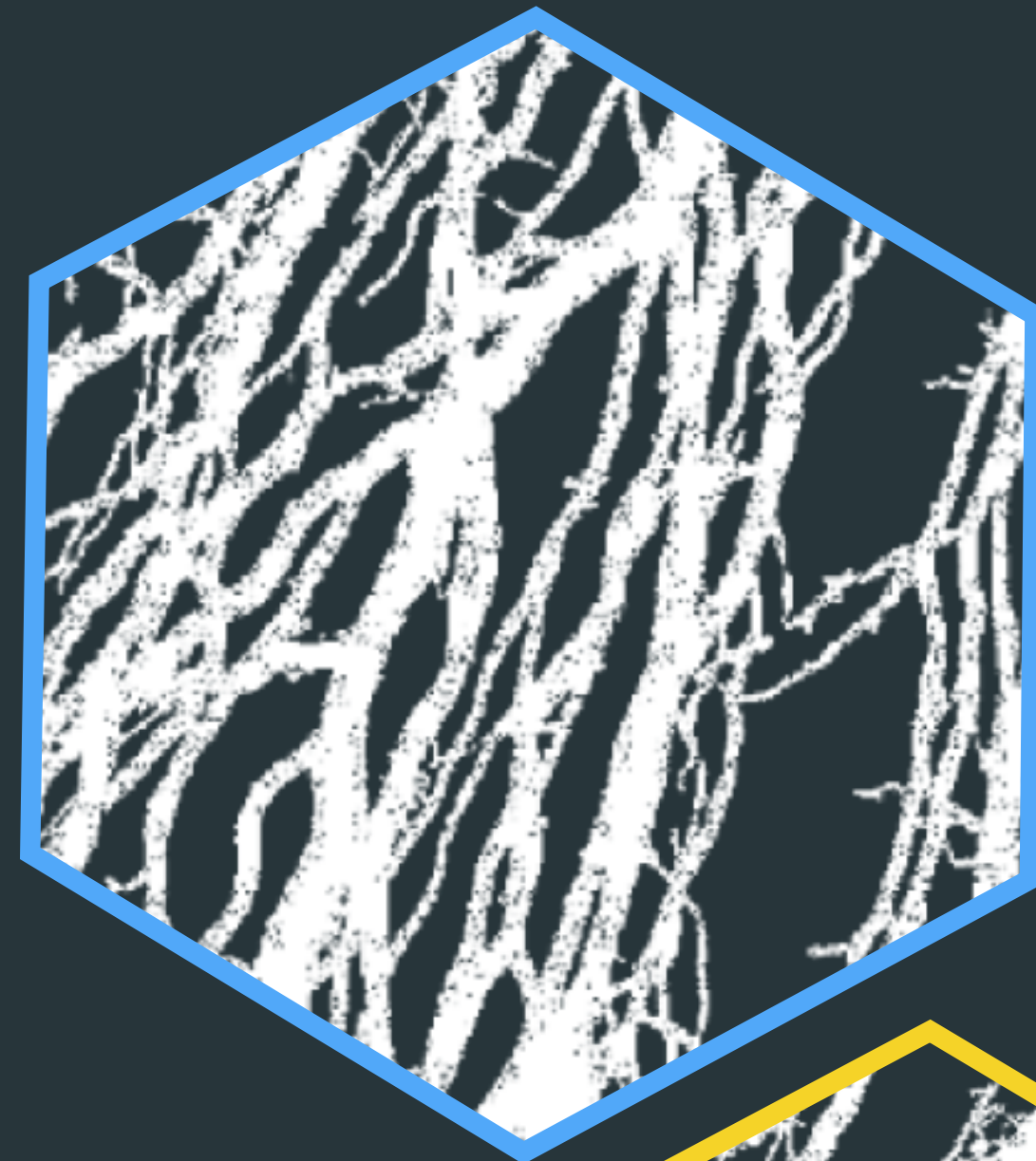
VALIDATION IN **EXISTING** TOOLS



INCREASING COMPLEXITY



INCREASING COMPLEXITY



HOW TO
BENCHMARK
ROOT IMAGE
ANALYSIS TOOLS?

```
data/arabido/rsml/arabido-sim-1-1
data/arabido/rsml/arabido-sim-1-2
data/arabido/rsml/arabido-sim-1-3
data/arabido/rsml/arabido-sim-1-4
data/arabido/rsml/arabido-sim-1-5
data/arabido/rsml/arabido-sim-1-6
data/arabido/rsml/arabido-sim-1-7
data/arabido/rsml/arabido-sim-1-8
data/arabido/rsml/arabido-sim-1-9
data/arabido/rsml/arabido-sim-1-10
1% of root systems generated
```

```
----- Mean simulation time = 2.68
----- Total simulation time = 2.68
```

```
-----
data/arabido/rsml/arabido-sim-2-1
data/arabido/rsml/arabido-sim-2-2
data/arabido/rsml/arabido-sim-2-3
data/arabido/rsml/arabido-sim-2-4
data/arabido/rsml/arabido-sim-2-5
data/arabido/rsml/arabido-sim-2-6
data/arabido/rsml/arabido-sim-2-7
data/arabido/rsml/arabido-sim-2-8
data/arabido/rsml/arabido-sim-2-9
data/arabido/rsml/arabido-sim-2-10
```

```
2% of root systems generated
----- Mean simulation time = 2.91
----- Total simulation time = 5.83
```

```
-----
data/arabido/rsml/arabido-sim-3-1
data/arabido/rsml/arabido-sim-3-2
data/arabido/rsml/arabido-sim-3-3
data/arabido/rsml/arabido-sim-3-4
data/arabido/rsml/arabido-sim-3-5
data/arabido/rsml/arabido-sim-3-6
```

WE USED
ARCHISIMPLE
TO
AUTOMATICALLY
CREATE
GROUND-TRUTHED
IMAGES

📖 Lobet, Koevoets, Noll et al. 2017
PMID: [28421089](#)

📖 Pagès et al. 2013
DOI: [10.1016/j.ecolmodel.2013.11.014](#)

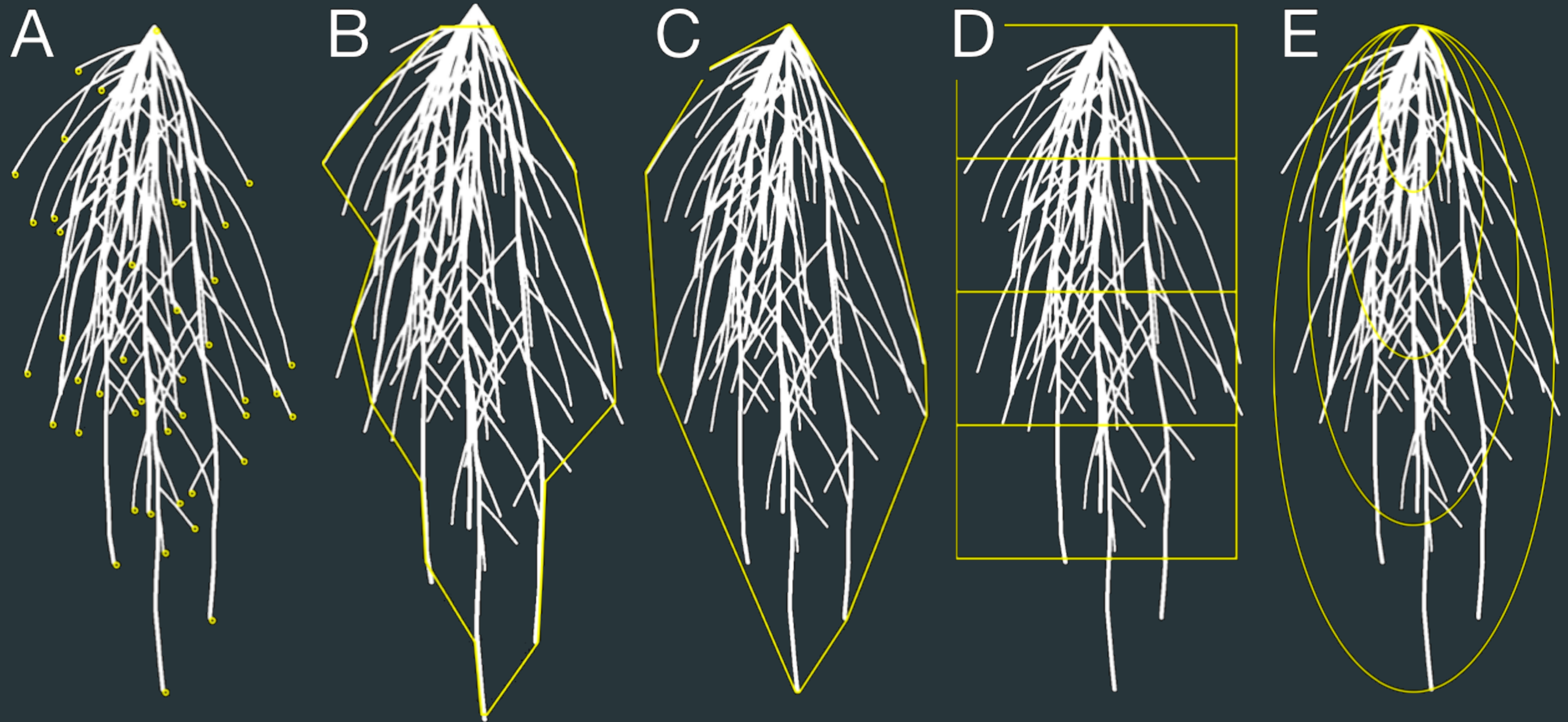
SIMULATED ROOT IMAGE LIBRARY



bit.ly/root-lib

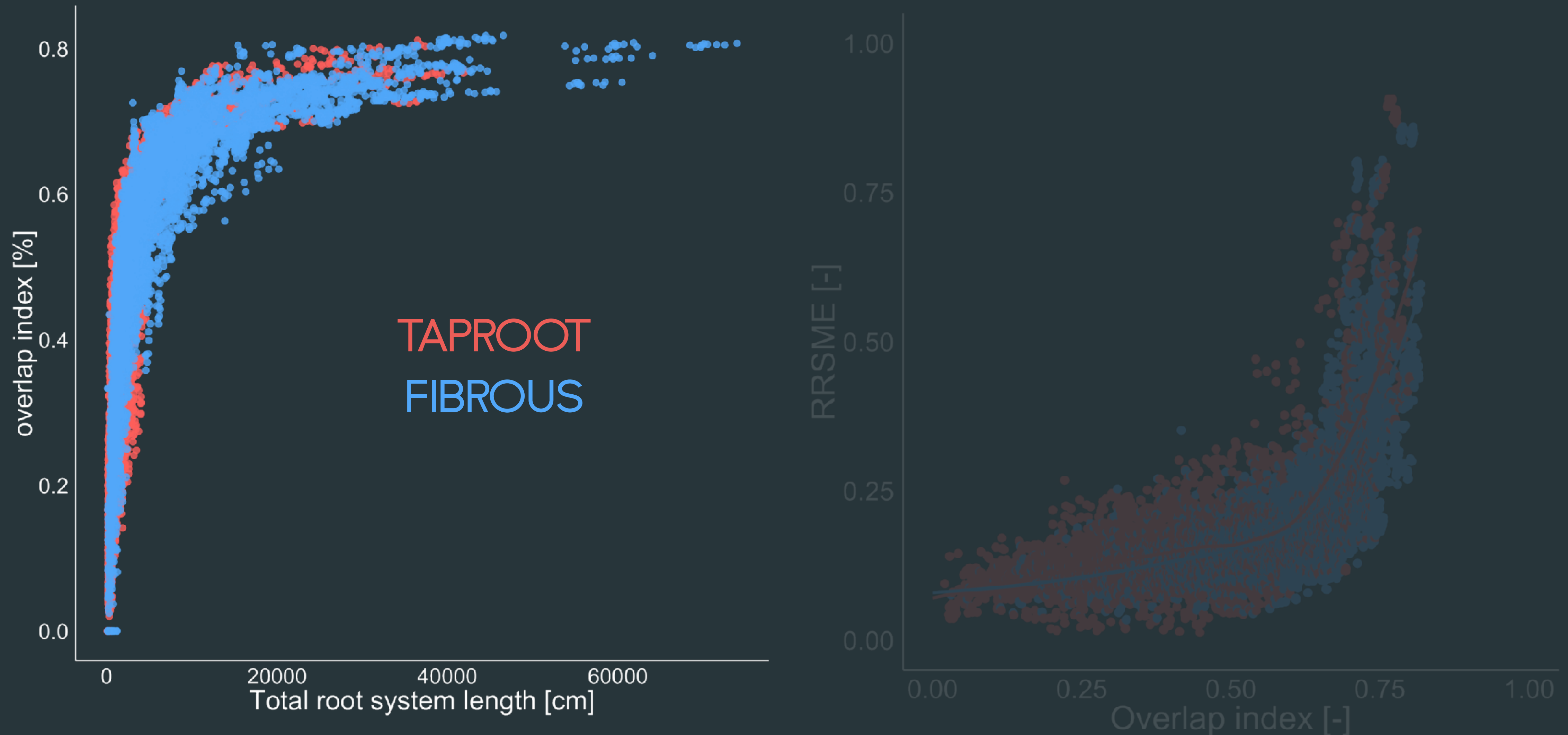


WE EXTRACT **CHEAP VARIABLES**



COMPARED WITH **GROUND-TRUTH**

ERROR INCREASES WITH SIZE



☒ Lobet, Koevoets, Noll et al. 2017

PMID: [28421089](#)

IMAGE
ANALYSIS



ROOT
MODELS

2



MACHINE
LEARNING

$\begin{bmatrix} \text{cheap var. 1 ... N} \\ \dots \\ \text{cheap var. 1 ... N} \end{bmatrix}$

$\begin{bmatrix} \text{groundtruth 1 ... N} \\ \dots \\ \text{groundtruth 1 ... N} \end{bmatrix}$

RANDOM
FORESTS

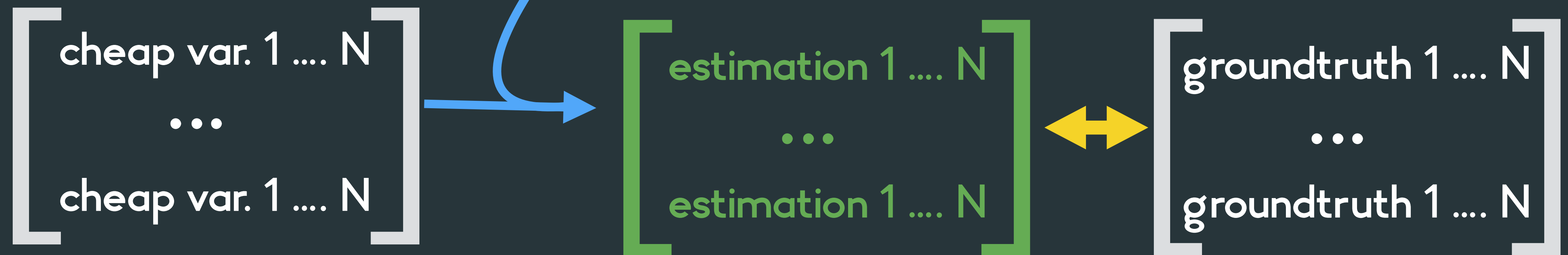
$\begin{bmatrix} \text{gt 1} = f(d1, \dots, dn) \\ \dots \\ \text{gt n} = f(d1, \dots, dn) \end{bmatrix}$

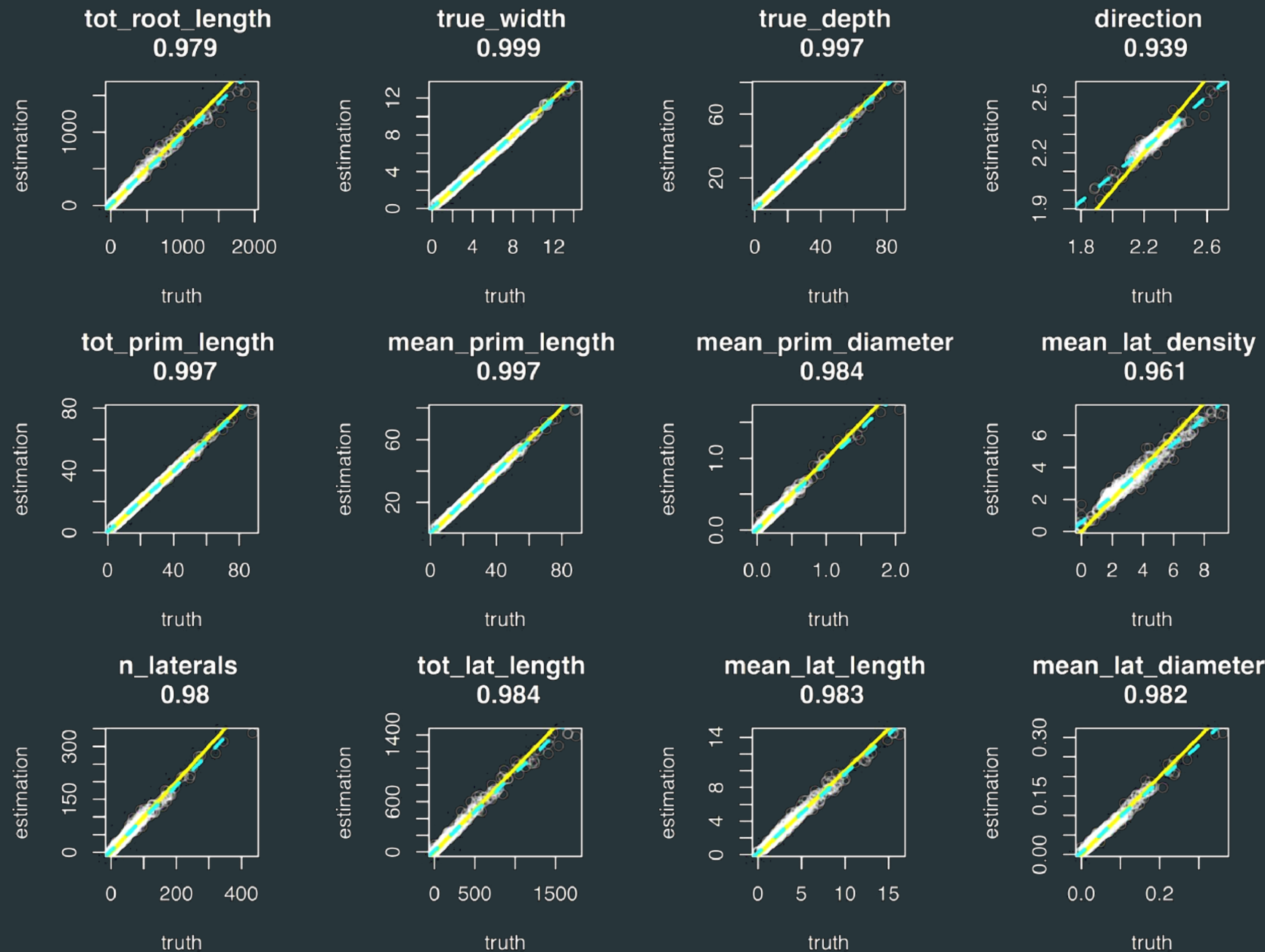
USE LIBRARY TO **TRAIN** RANDOM FOREST

TRAINING



TESTING





AVERAGE
R-SQUARED
0.92

AVERAGE
ANALYSIS
TIME
3 SEC/IMG

📖 Lobet, Koevoets, Noll et al. 2017
PMID: **28421089**

IMAGE
ANALYSIS



ROOT
MODELS

3



MACHINE
LEARNING

LARGE EXPERIMENTAL DATASET

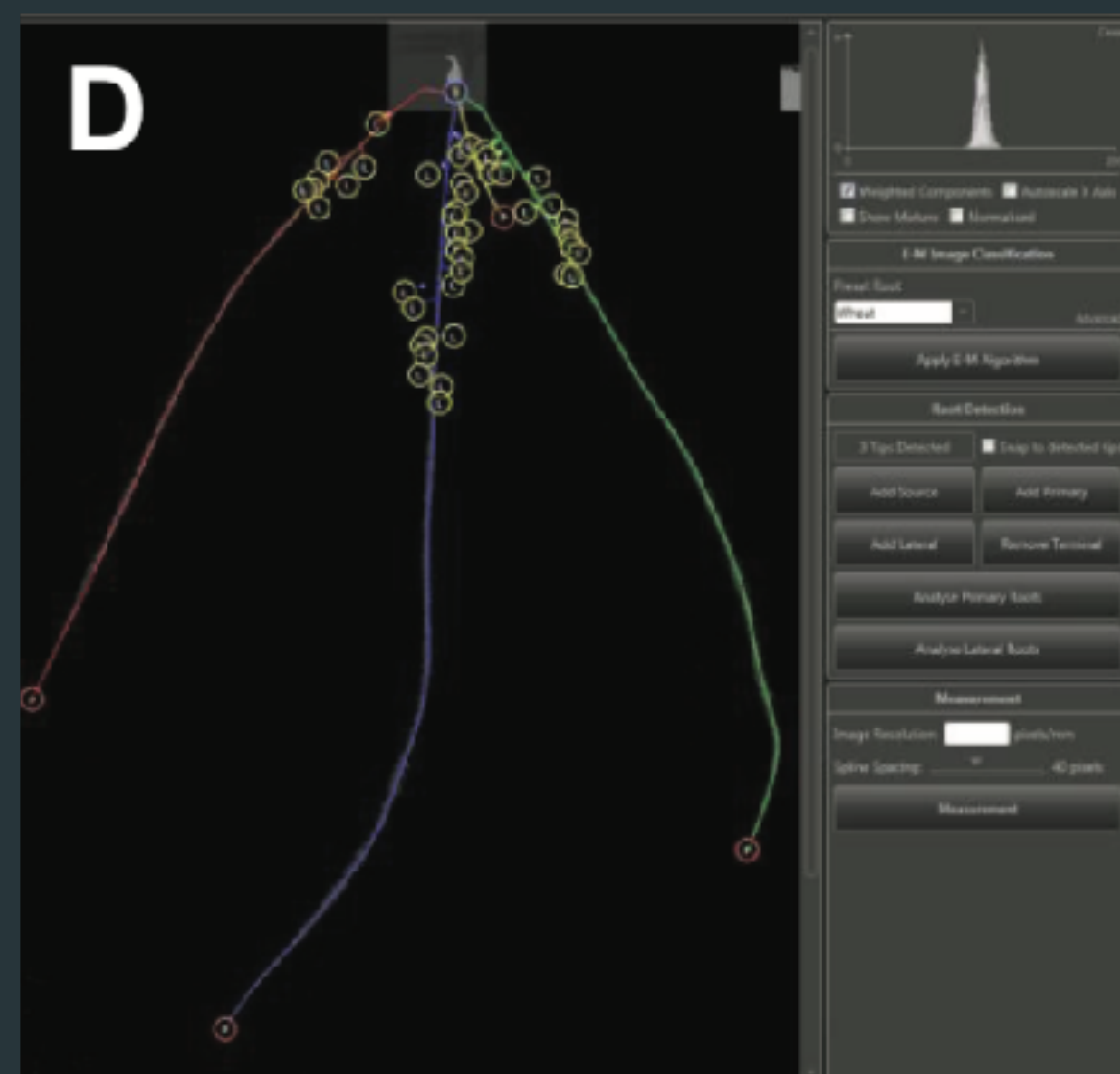
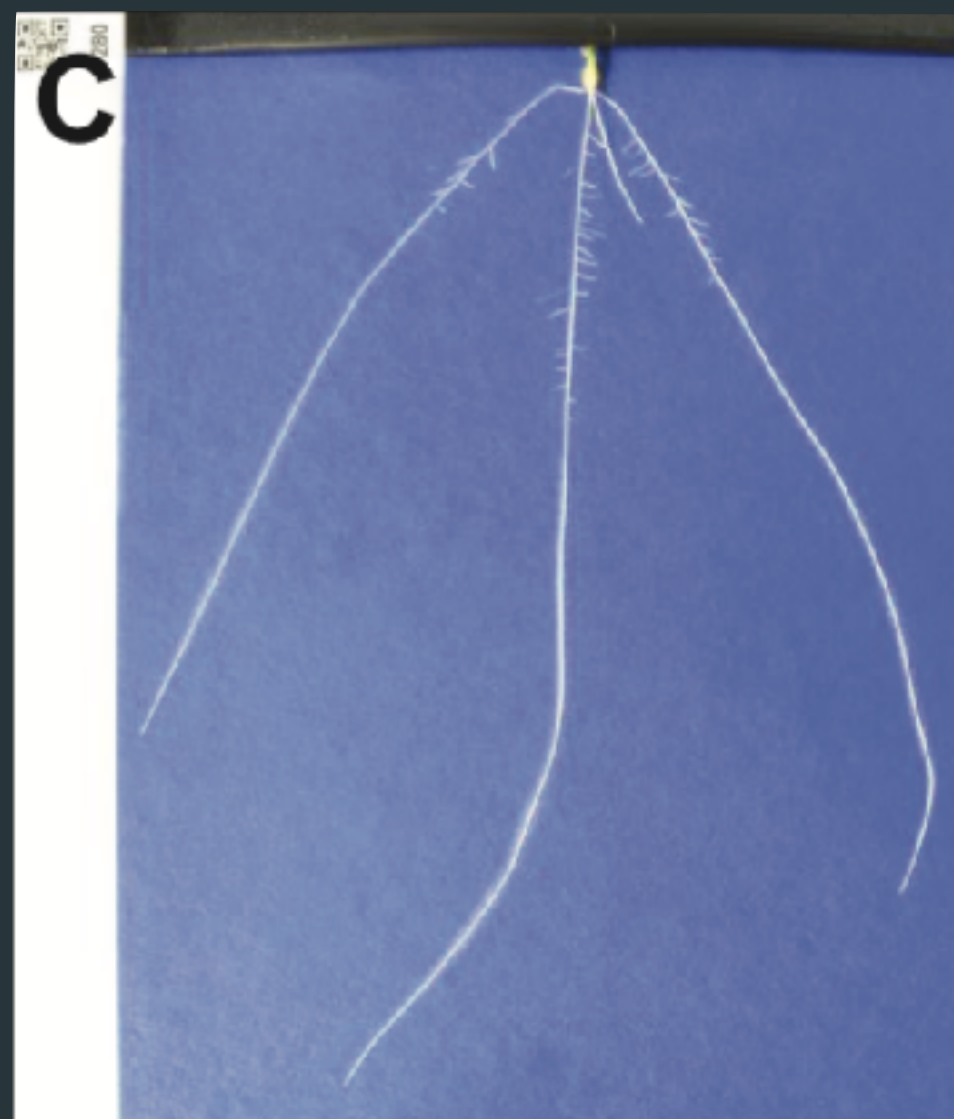


WHEAT

SAVANNAH

X

RIALTO



2500

IMAGES

ROOTNAV

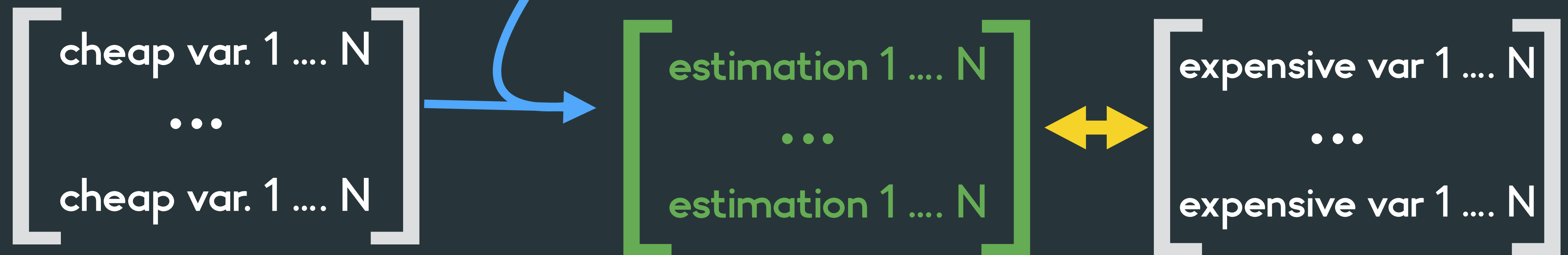
RSML

USE LIBRARY TO **TRAIN** RANDOM FOREST

TRAINING

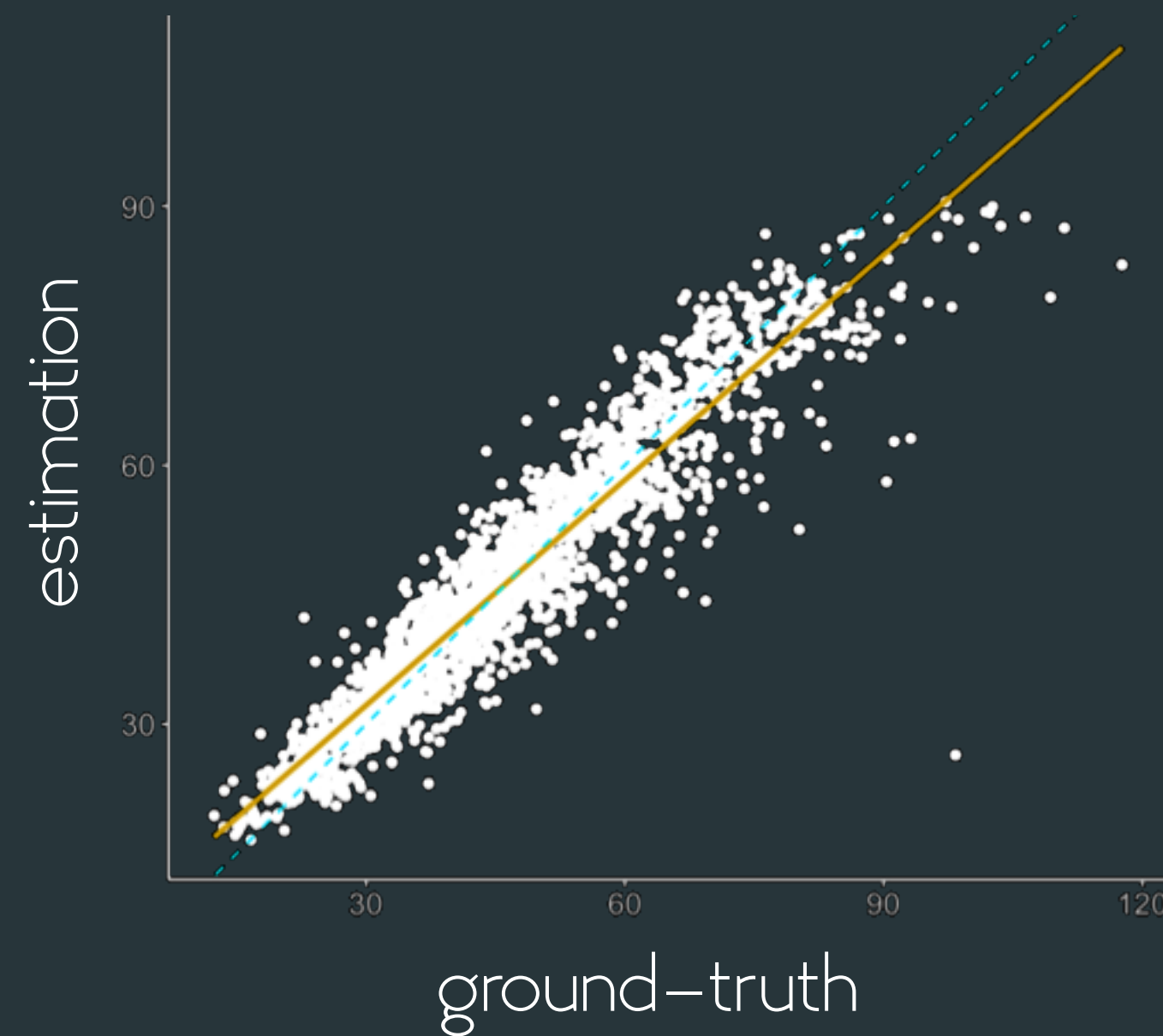


TESTING

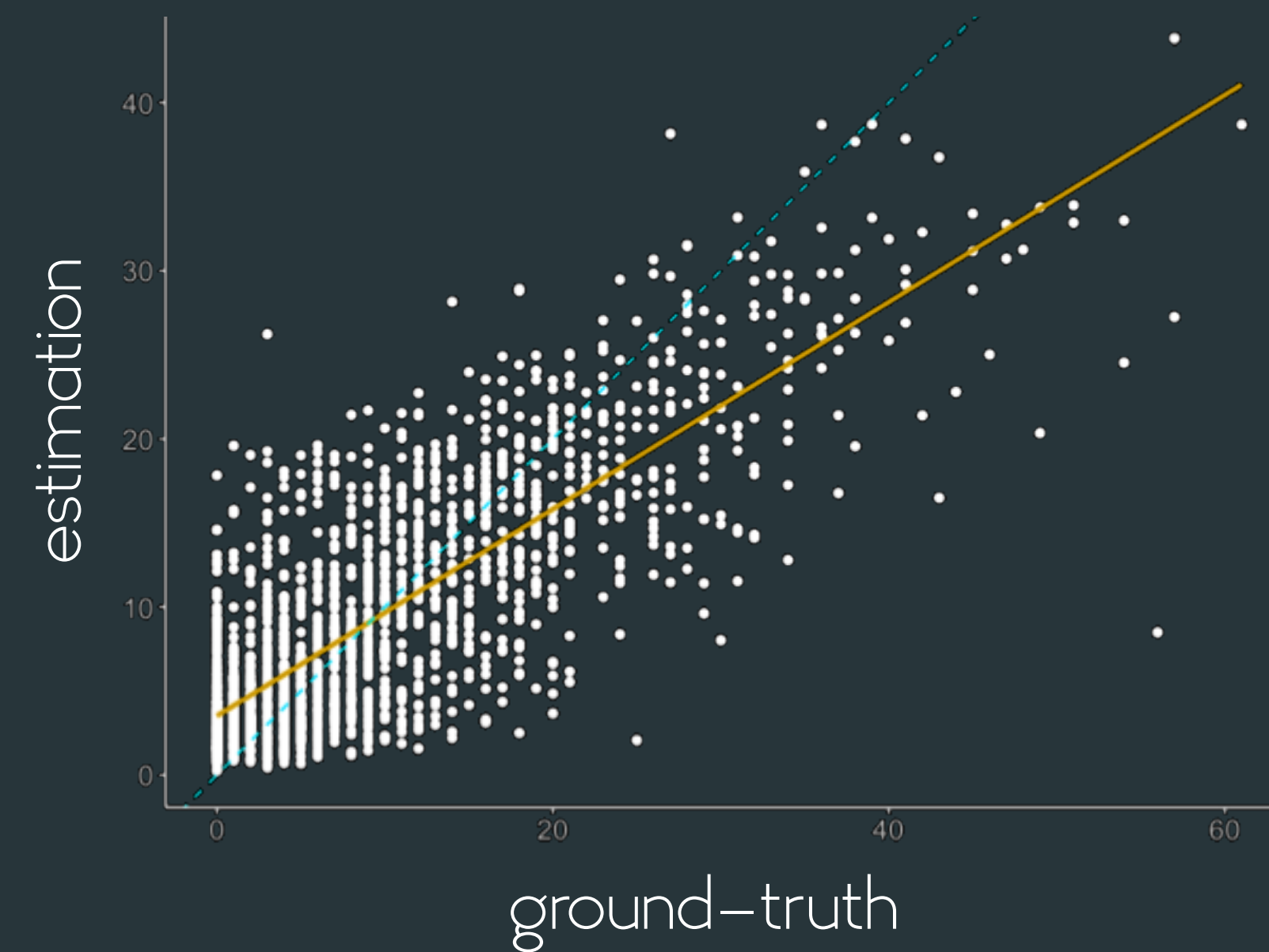


ANALYSIS
TIME
–75%

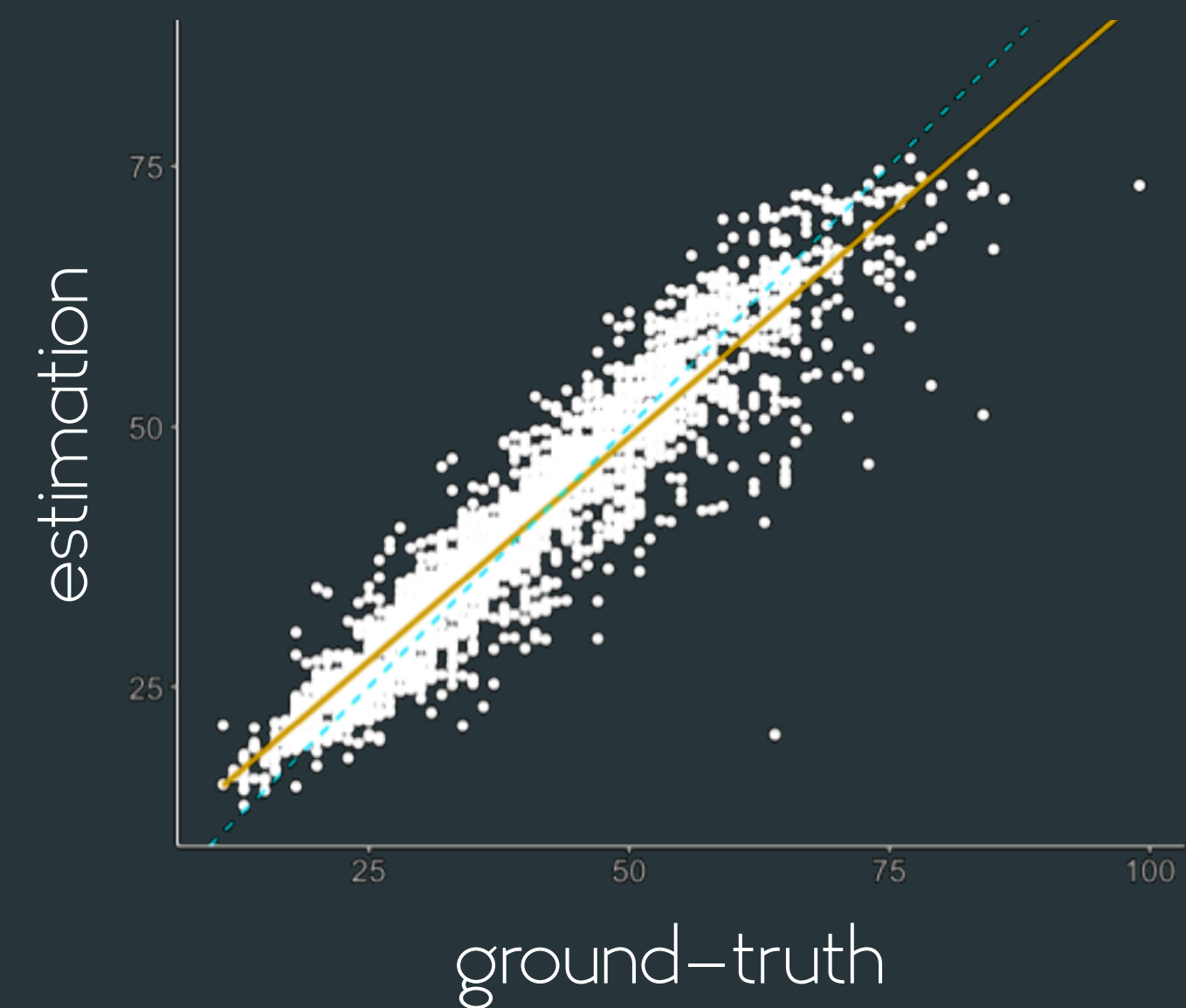
TOTAL ROOT LENGTH



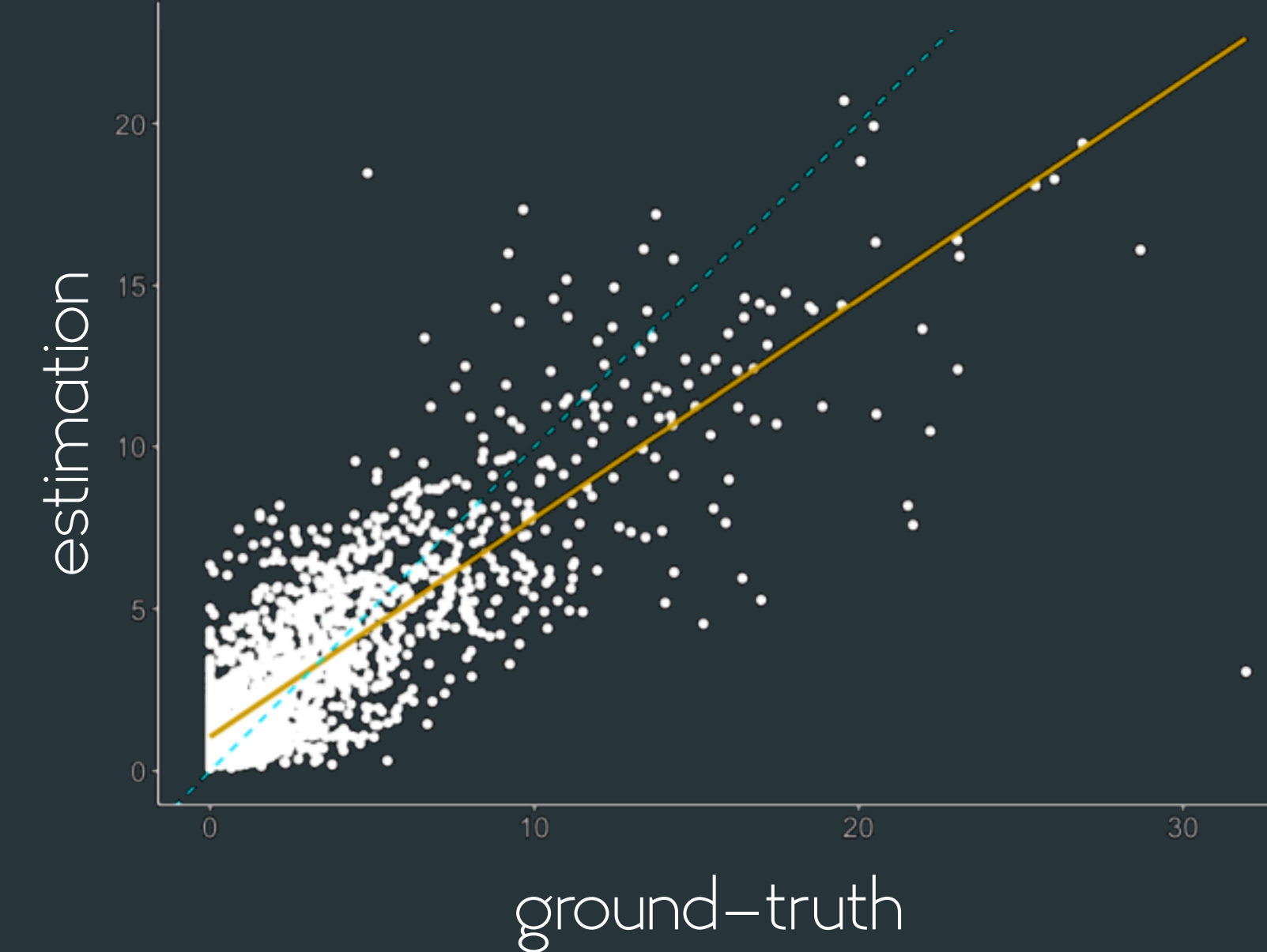
NUMBER OF LATERALS



TOTAL PRIMARY LENGTH

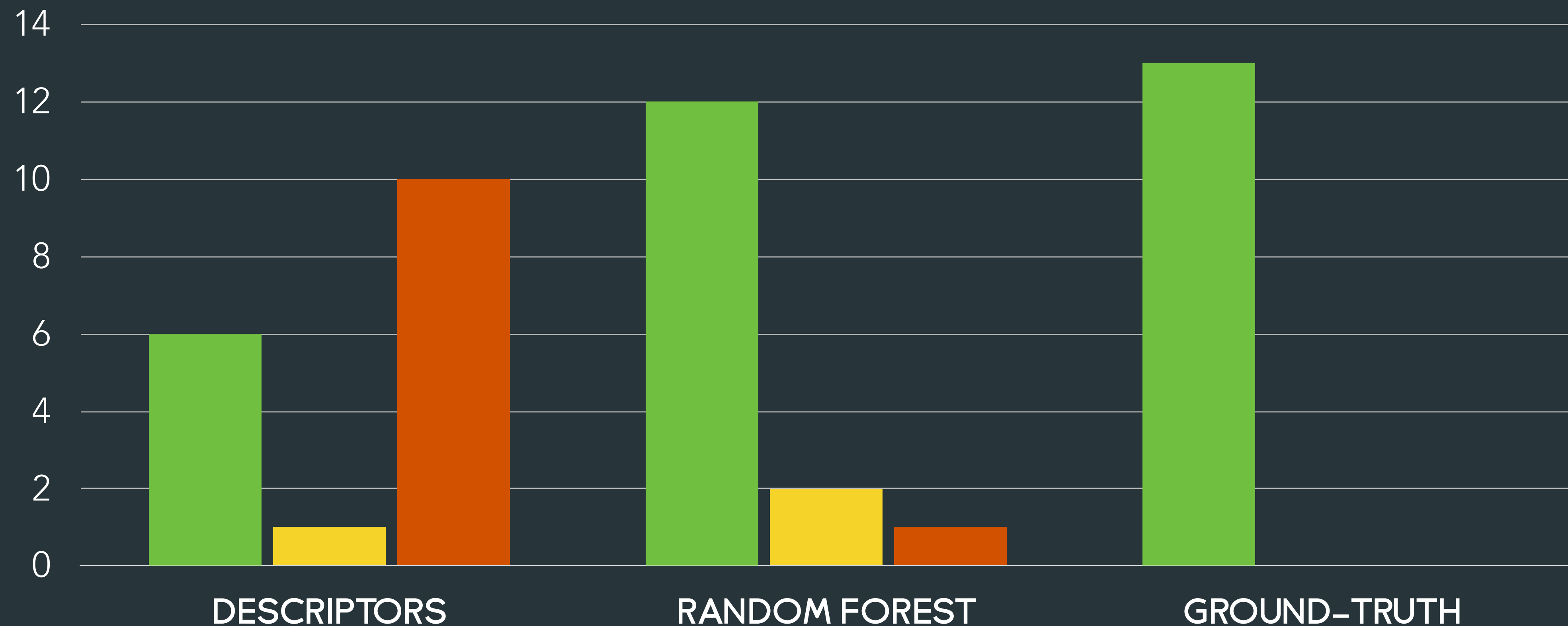


TOTAL LATERAL LENGTH



QTL ANALYSIS

CORRECTLY IDENTIFIED QTL
MISSED QTL
FALSE POSITIVE QTL



Test datatable contains 131 rows

Training datatable contains 2495 rows

Variables to use in analysis

tot_root_length

n_primary

tot_prim_length

mean_prim_length

n_laterals

tot_lat_length

Number of models to try:

5

50

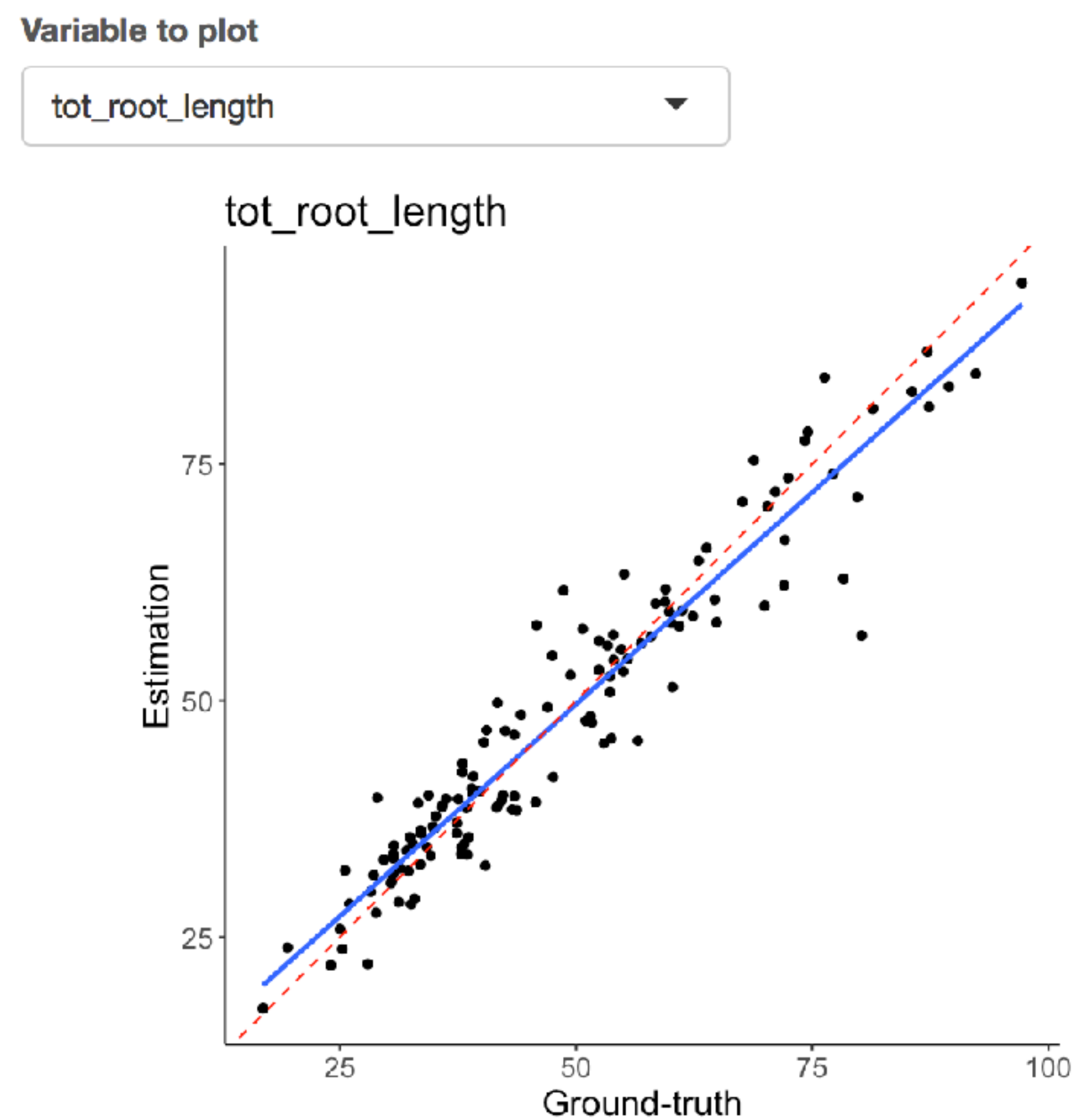
Number of trees in each model:

5

200

Train PRIMAL

If you are happy with the accuracy of the Random Forest, go to the next step. If not, you might want to include more images into your training dataset and start over.

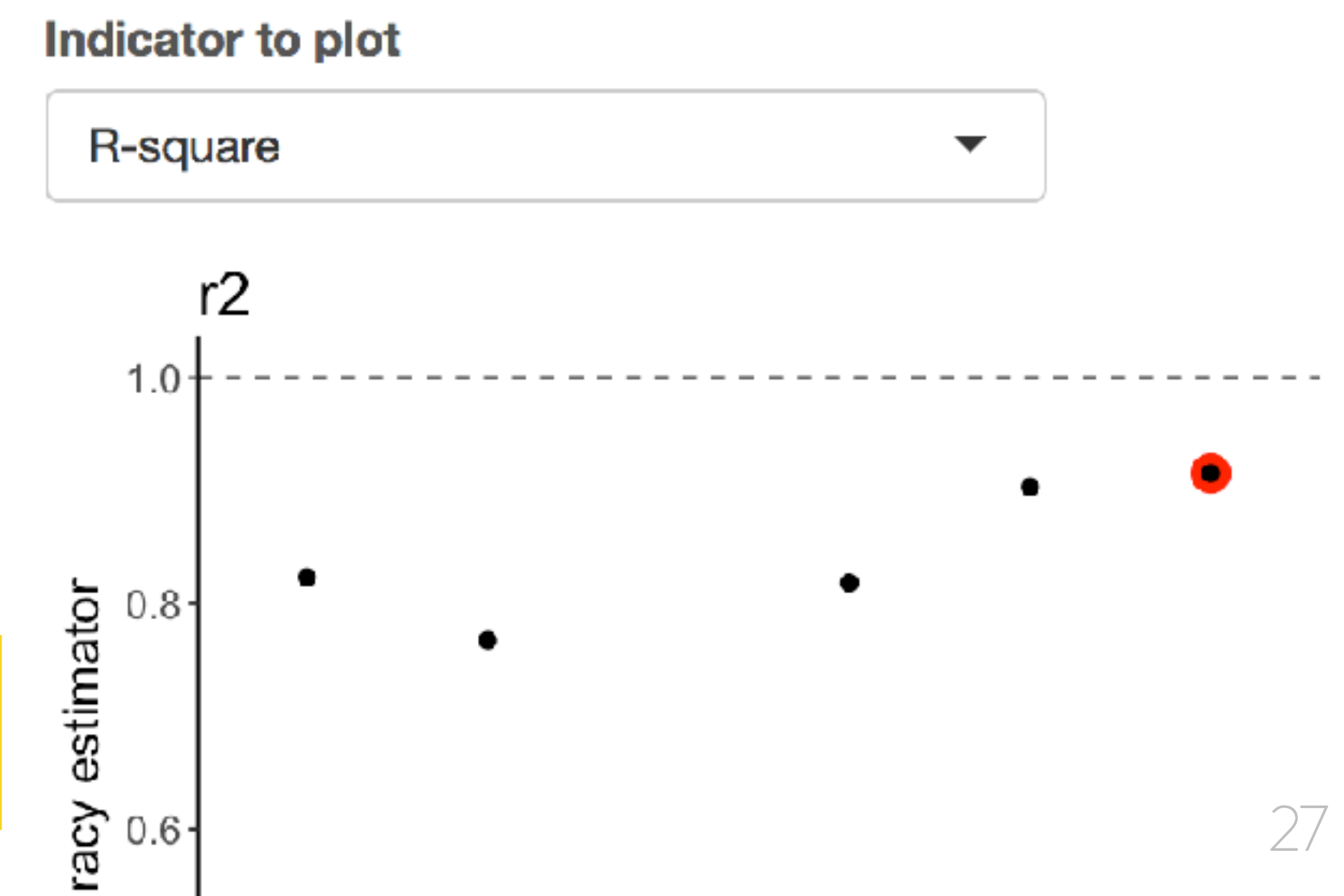


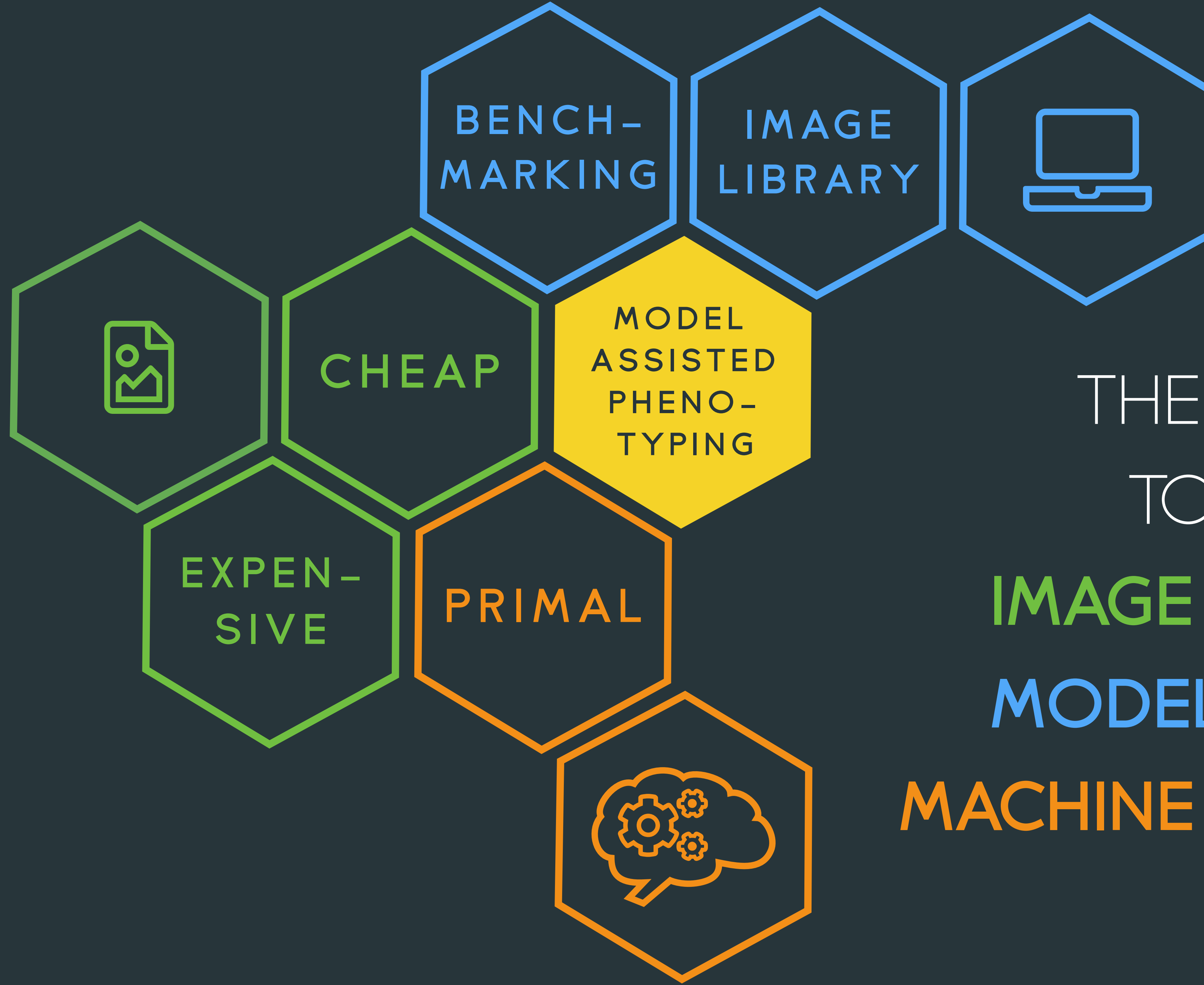
Show 5 entries Search:

	r2	rmse	pearson	sp
tot_root_length	0.915	0.083	0.957	
n_primary	0.268	0.087	0.518	
tot_prim_length	0.903	0.077	0.95	
mean_prim_length	0.823	0.112	0.907	
n_laterals	0.767	0.471	0.876	

Showing 1 to 5 of 6 entries

Previous12Next





THE FUTURE IS
TO COMBINE
IMAGE ANALYSIS,
MODELLING AND
MACHINE LEARNING



THANK YOU
ANY
QUESTIONS ?

Presentation available on **figshare**

bit.ly/lobet-tokyo-2017