CEDR

Conférence Européenne
des Directeurs des Routes

Conference of European
Directors of Roads

Data driven support for improved decision systems to help protect soils at road construction sites

Joint WP2-WP3 results of the ROADSOIL project

Attila Nemes
Lorena Chagas Torres
Loraine ten Damme
Teodora Todorcic Vekic
Thomas Keller

Contents

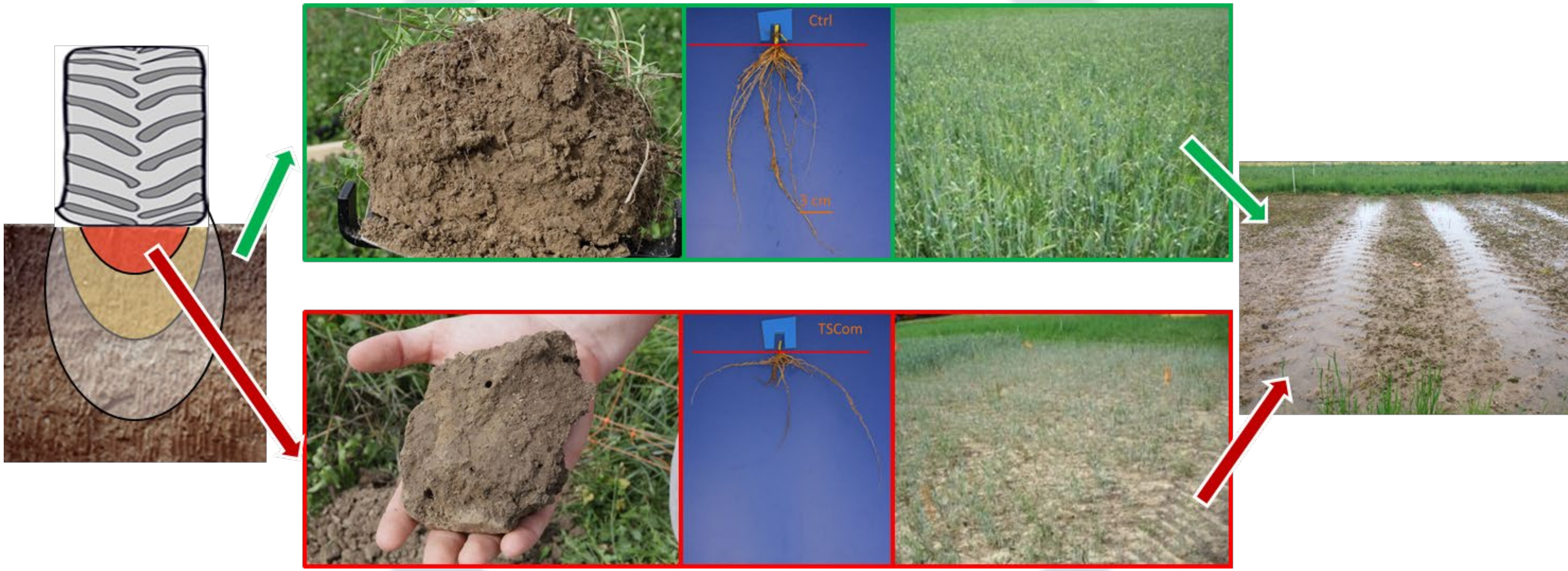
- Introduction
- Methodology
  - The chosen machine learning tool
  - Data collection and processing steps
- Results
  - Model choice
  - The prediction model
  - Challenges, the way forward
- Application options (prior to and during work)
- Final remarks

Compaction negatively affects…

soil structure

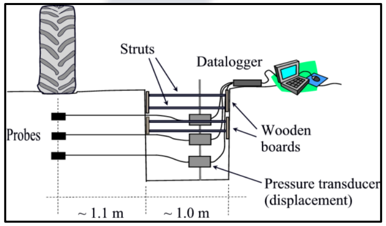
Compaction negatively affects…

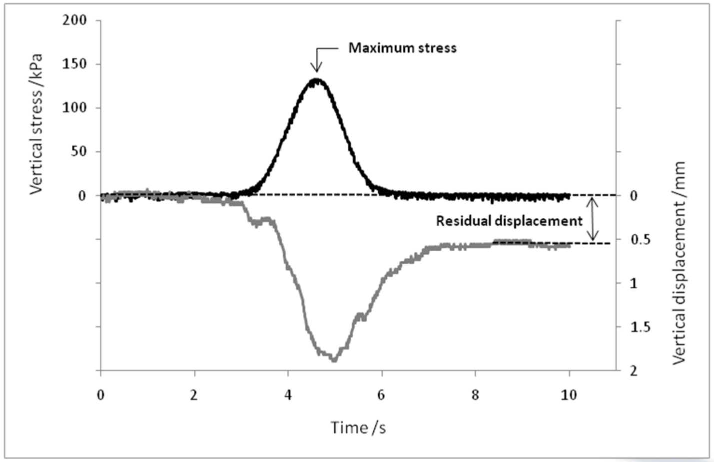
a range of soil functions



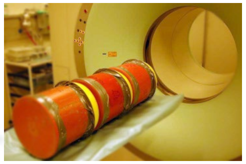
Vastly different time scales for compaction and recovery

Compaction: seconds (10^0 s)

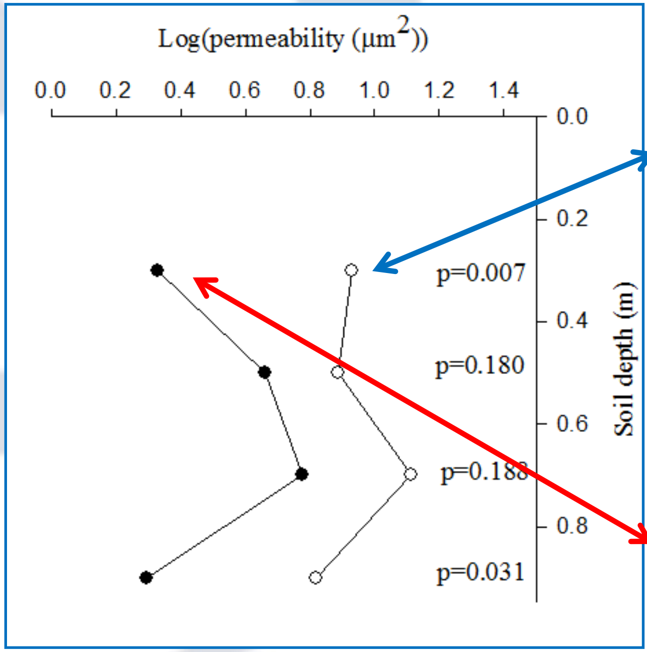


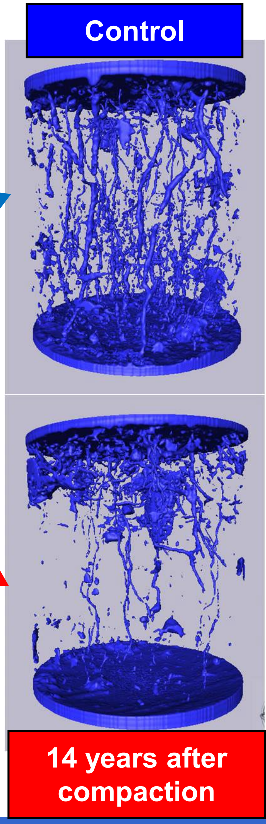


Keller *et al*. 2012. *Soil Sci*. 177.



Compaction recovery: decades (10^9 s) (?)

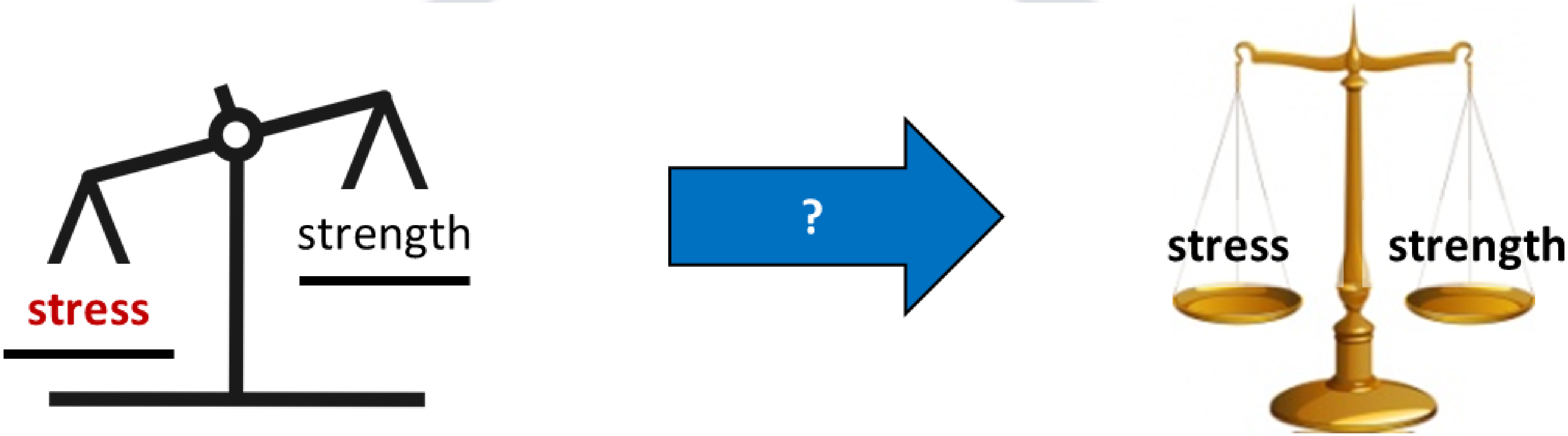




Berisso *et al*. 2012. *Soil Tillage Res.* 122
Lamandé *et al*., 2013. *Soil Sci*., 178.

Recovery is very slow

-> it is better to prevent compaction



Option 1: reduce soil stress

Option 2: increase soil strength

More general: make sure that stess ≤ soil strength

The ‘*workability*’ or ‘*workable soil*’ concept

Methodology - I

Main aims (from the proposal):

- To collect available quantitative and quantifiable information on drivers and effects of soil degradation in the context of soils impacted by road construction

  Manually, from published, available data sources

- To derive quantitative descriptive and predictive relationships between human induced (road work and operation) and natural (soil, hydro-climatic, etc.) driving factors and soil degradation and functioning

  Using a chosen machine learning method

Methodology - II

CART: Classification and Regression Tree

- One of the “supervised” machine learning methods
- Better ones exist, and there are complex sub-types of it, e.g. boosted trees, random forests
- **....however, we wanted one that is easy to be *read by humans*, but also possible to be *machine read* (programmed).**

Table 3
Comparison of Different Mathematical Predictive Models, ++ = Good, + = Fair, and— = Poor (Adapted From Hastie et al., 2001)

Feature	Class PTF	MLR, GLM	GAM	Regression tree	Random forests	Neural net	SVM	Nearest neighbor
Parsimony	++	++	—	++	—	—	+	—
Interpretability of the model	++	++	+	++	—	—	—	—
Variable selection	—	++	—	++	++	—	—	—
Nonlinearity	—	—	++	++	++	++	++	++
Handling of mixed data type (qualitative and quantitative)	+	+	+	++	++	—	++	+
Computational efficiency (large data)	++	++	+	++	—	—	+	++
Predictive power	—	+	+	+	++	++	++	++

Van Looy et al. 2017. Reviews of Geophysics, 55, 1199–1256. https://doi. org/10.1002/2017RG000581

Methodology - III

https://en.wikipedia.org/wiki/Decision_tree_learning

Data collection and processing - I

- **Literature search** for soil mechanical data (focus on pre-compression stress) + auxiliary soil, environmental and methodological information
- **Quality assessment** and pre-selection
- **Harmonization** of soil texture (particle-size information) that adhere to different international standards
- **Descriptive statistics**, correlation analysis, visual assessment, exploratory tree-models
- **Hierarchical approach** to include more inputs and assess the data requirements vs. their benefits in predictions

Data collection and processing - II

- Data of **129 studies** between 1992-2021, 4776 entries
- **12 US Soil taxonomy orders**, a wide range of land uses
- **Europe: data from 10 countries** (SUI, GER, SWE, DEN, FRA, NOR, EST, ROM, BEL, UK)
- Top/subsoils: 63/37%
- Dominantly available auxiliary soil data: texture, bulk density, organic carbon content, soil moisture status*
  *(*moisture content _or_ soil moisture tension)*
- MS Excel -> MS Access, to be published
- **Working data set: 907 entries, all from Europe**

CEDR
Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads

Data collection and processing - III

Data subsets with different degrees of input availability from European data sources

Model no.	Required input variables	N
1	USDA texture class (all data)	907
2	USDA texture class (limited to data only at 60 hPa moisture tension)	540
3	Sand-silt-clay content (SSC)	907
4	SSC + wetness*	841
5	SSC + soil moisture tension**	841
6	SSC + soil moisture tension + bulk density***	633
7	SSC + soil moisture tension + bulk density + soil organic carbon content****	475
8	*SSC + gravimetric water content*******	*238*
9	*SSC + gravimetric water content + bulk density*	*142*
10	*SSC + gravimetric water content + bulk density + soil organic carbon content*	*89*

*wetness = 1 if ψ<100 hPa, =2 if 100<=ψ<1000 hPa, =3 if ψ>=1000 hPa

** Soil moisture tension in hPa

***Bulk density (BD) in g/cm^3

**** Soil organic carbon (SOC) content in g/g % (if soil organic matter content is given, divide by 1.724 to get SOC)

***** Gravimetric water content in g/g as a fraction (multiply by BD for volumetric water content)

CEDR
Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads

Results – model choice

Drivers of choosing a model: data availability, simplicity, benefit

Model no.	Required input variables	N	RMSE (*kPa*)
1	USDA texture class (all data)	907	78.26
2	USDA texture class (limited to data only at 60 hPa moisture tension)	540	38.75
3	Sand-silt-clay content (SSC)	907	71.35
4	SSC + wetness*	841	45.03
5	SSC + soil moisture tension**	841	45.61
6	SSC + soil moisture tension + bulk density***	633	41.85
7	SSC + soil moisture tension + bulk density + soil organic carbon content****	475	40.82
8	*SSC + gravimetric water content*******	*238*	102.44
9	*SSC + gravimetric water content + bulk density*	*142*	53.53
10	*SSC + gravimetric water content + bulk density + soil organic carbon content*	*89*	45.24

*wetness = 1 if ψ<100 hPa, =2 if 100<=ψ<1000 hPa, =3 if ψ>=1000 hPa

** Soil moisture tension in hPa

***Bulk density (BD) in g/cm^3

**** Soil organic carbon (SOC) content in g/g % (if soil organic matter content is given, divide by 1.724 to get SOC)

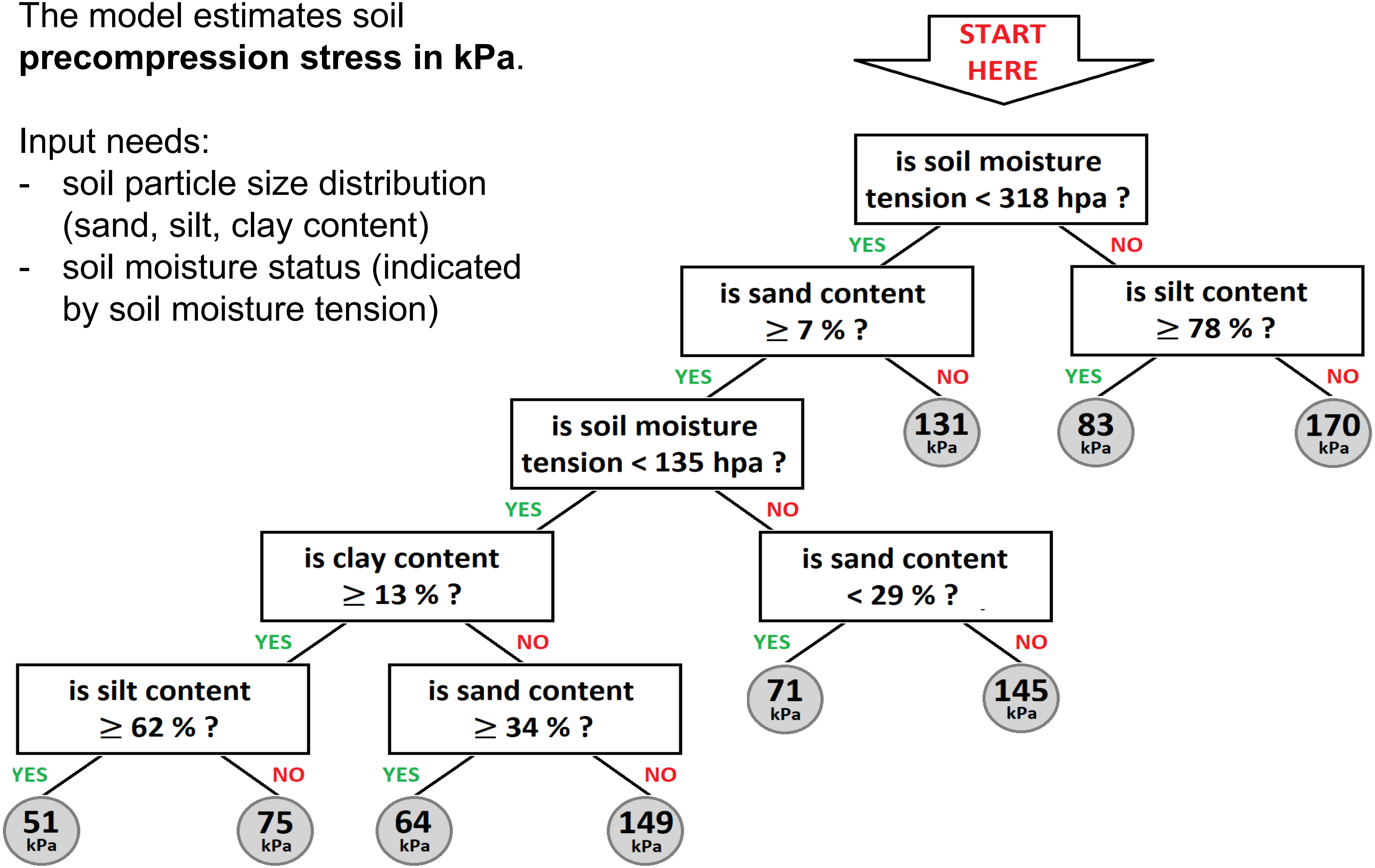
***** Gravimetric water content in g/g as a fraction (multiply by BD for volumetric water content)

Results – The prediction model

CEDR
Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads

The model estimates soil **precompression stress in kPa**.

Input needs:
- soil particle size distribution (sand, silt, clay content)
- soil moisture status (indicated by soil moisture tension)



Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads

Challenges, the way forward

- There is a grave need for more measurements to support such models. (more data, more and coexisting properties)
- Europe lags behind on such measurements!
- Land use or soil type specific estimations were impossible, due to a shortage of such data.
- There is a need to standardize methodology, classification systems.
- Moisture status measurements by users should also be harmonized. (Hot-spots need particular attention!)
- _*Volumetric*_ soil moisture is rarely published with soil mechanical measurements – which is a loss!
- PS: The need for more complex but more accurate models will also depend on the willingness to use them…

Application options

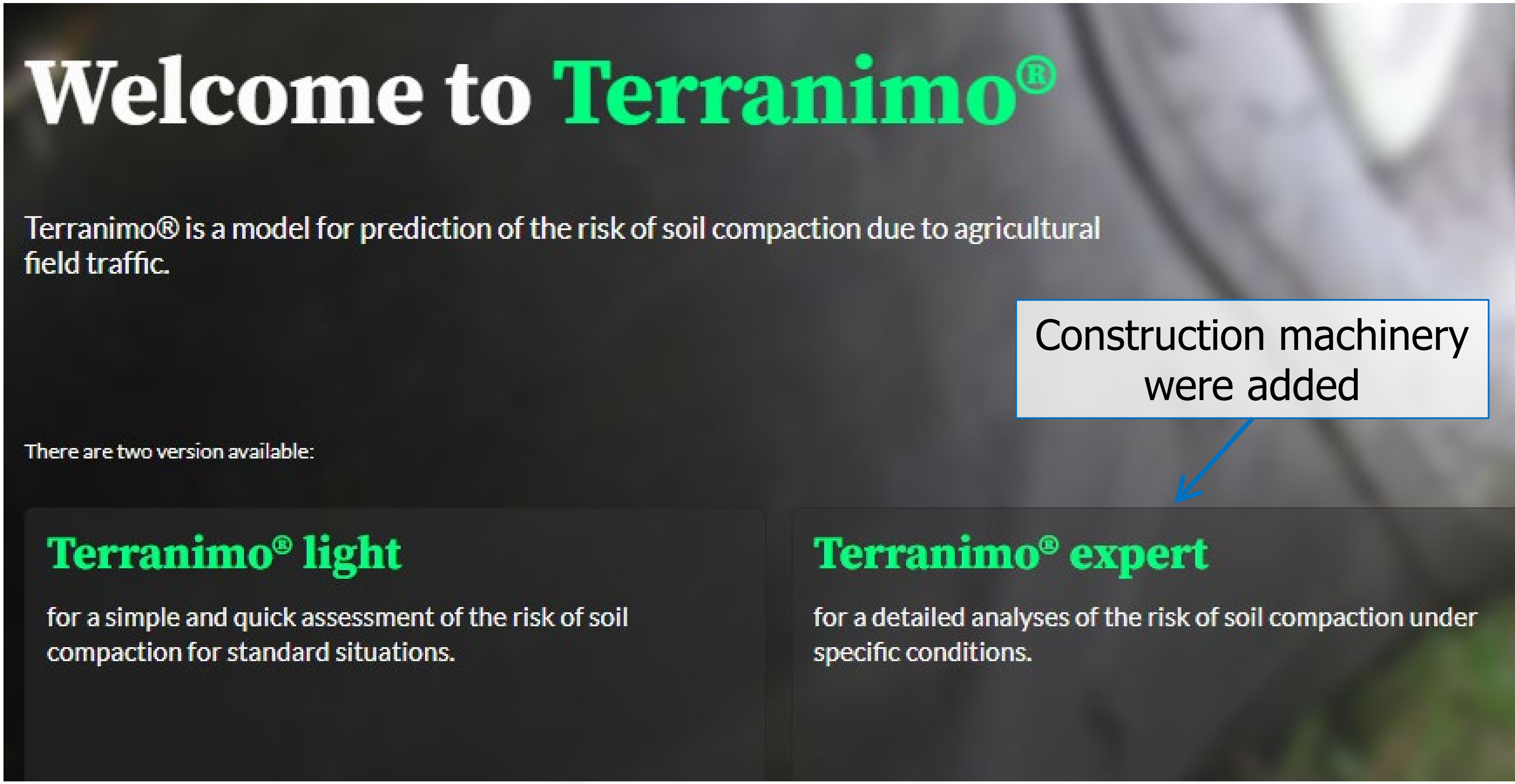
The decision tool can be applied:

- In the planning phase: help prepare for scenarios, machine selection, etc.

- In the work phase: help daily decisions

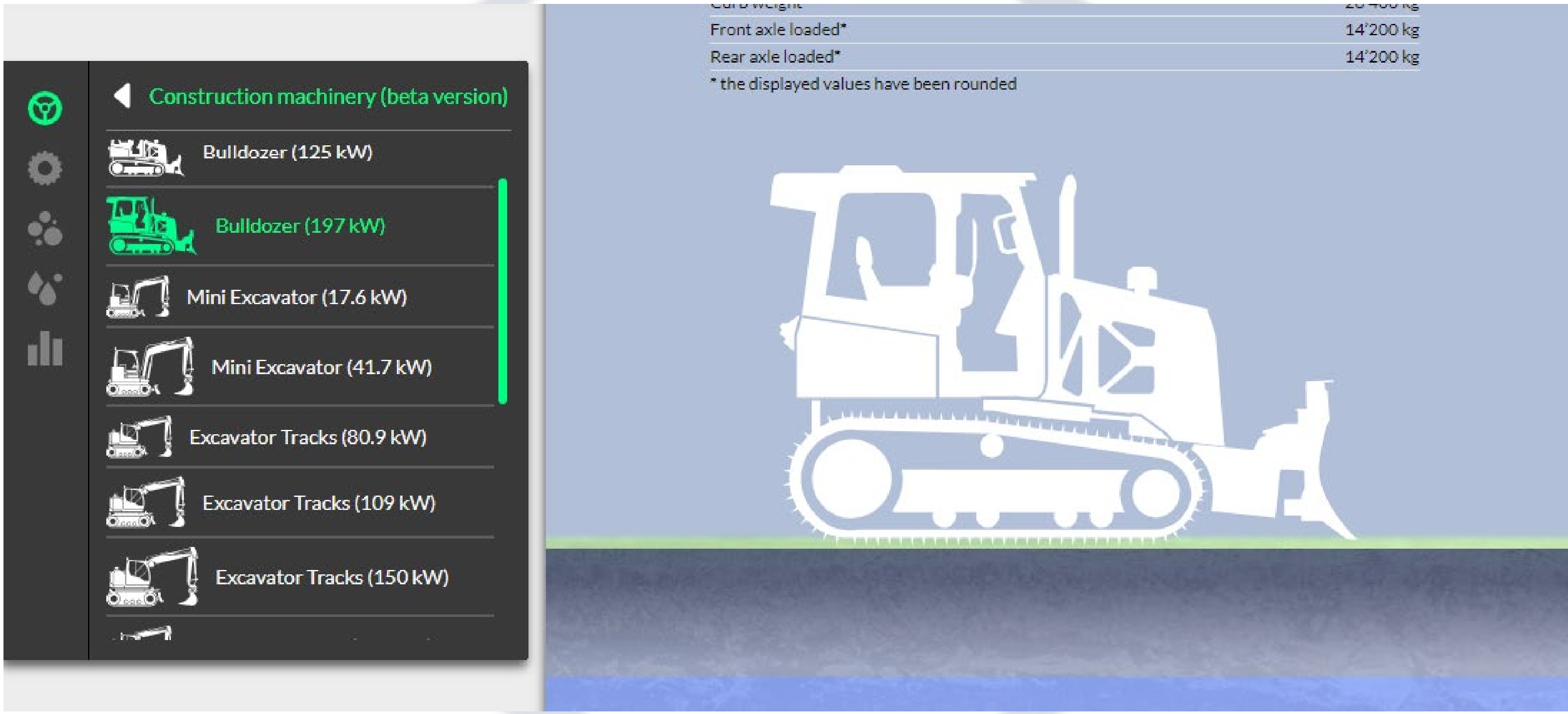
CEDR
Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads

www.terranimo.se



Construction machinery were added

www.terranimo.se - an example



www.terranimo.se - an example

choose soil moisture

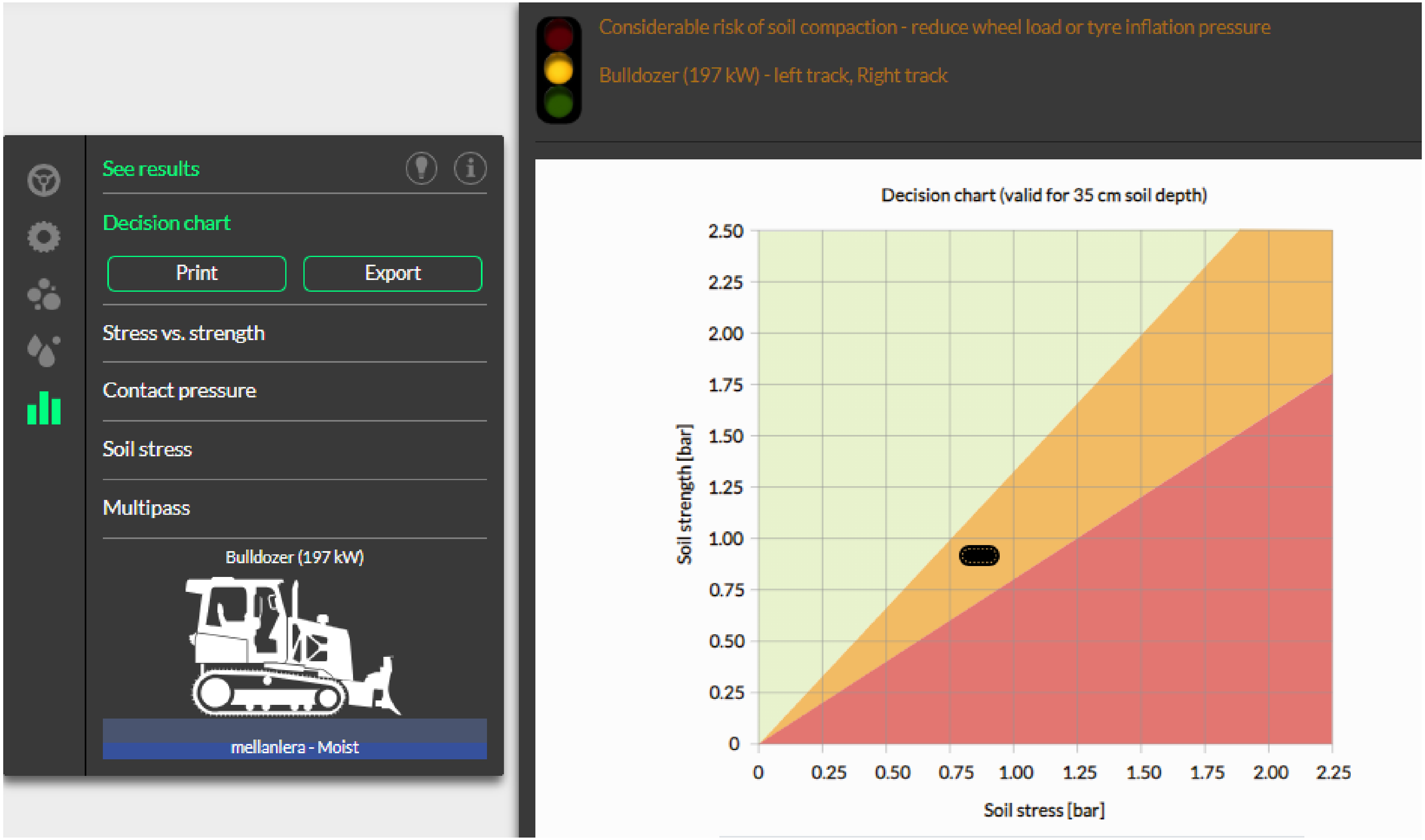
- Dry
- Moist
- Wet

Moist

No.	Bottom	Matric potential
1	10 cm	10 cbar
2	20 cm	10 cbar
3	30 cm	10 cbar
4	40 cm	10 cbar
5	50 cm	10 cbar
6	60 cm	9 cbar
7	70 cm	8 cbar
8	80 cm	7 cbar
9	90 cm	6 cbar
10	100 cm	5 cbar

1 cbar = 10 hPa

www.terranimo.se - an example



Application – during work

- Appendix 1 of Deliverable D2.2-3.2 provides a detailed description of the model's use, as well as a printable version of the decision tree model and its quick guide.
- Input requirements:
  - (1) soil particle-size distribution (one time determination),
  - (2) soil moisture tension information (<u>upon each evaluation</u>),
  - (3) soil stress exerted by a given machine (as in its specs).
- If the soil gets wet, frequent evaluations (*and proper response*) may be necessary.

CEDR
Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads

Assessment methodologies and mitigation measures for the impacts of road projects on soils – ROADSOIL

Using machine learning to improve the prediction of soil mechanical properties

Deliverable D2.2-3.2
29 March 2023

APPENDIX 1

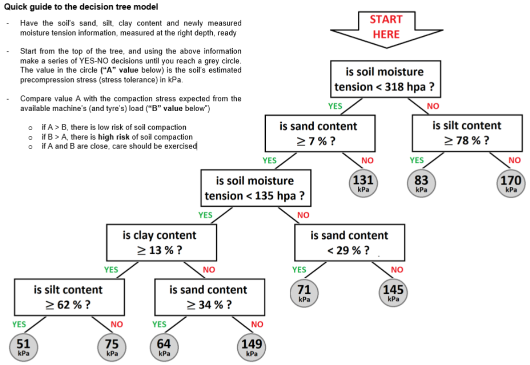
Practical guide to the in-field decision aid tool to assess soil strength

Executive summary

Soil compaction negatively affects soil functions including food production, water cycling, climate regulation, and biodiversity. To prevent soil compaction by construction machinery, mechanical soil stresses exerted by machinery need to be evaluated against the soil's tolerance to compression stress (soil strength) prior to soil work. Soil strength is a dynamic soil property that may change day-to-day because it does not only depend on soil texture but also on the actual soil moisture content. Here a decision tree model is presented for daily in-field use by entrepreneurs. Knowledge of soil texture (sand, silt, clay content) and soil moisture status (represented by soil moisture tension) is required for using the model. The model will estimate soil strength (represented by precompression stress in kPa) that is to be compared with soil stress exerted by the machinery to be used. This decision aid tool will help deciding which of the machines – if any – can be used on the given soil in its given moisture condition.

Quick guide to the decision tree model

- Have the soil's sand, silt, clay content and newly measured moisture tension information, measured at the right depth, ready
- Start from the top of the tree, and using the above information make a series of YES-NO decisions until you reach a grey circle. The value in the circle (**"A" value** below) is the soil's estimated precompression stress (stress tolerance) in kPa.
- Compare value A with the compaction stress expected from the available machine's (and tyre's) load (**"B" value** below")
  - if A > B, there is low risk of soil compaction
  - if B > A, there is **high risk** of soil compaction
  - if A and B are close, care should be exercised



Final remarks

- Soil protection is our joint interest, we all need to invest in it
- To succeed in protecting soils, it is critical to build this knowledge-base further
- It is easier to measure soil moisture than soil tension but we don't have sufficient data background yet to suit such predictive models
- Our models need to be applied together with other measures before and during road construction projects
- Engaging a soil scientist with relevant background may pay off

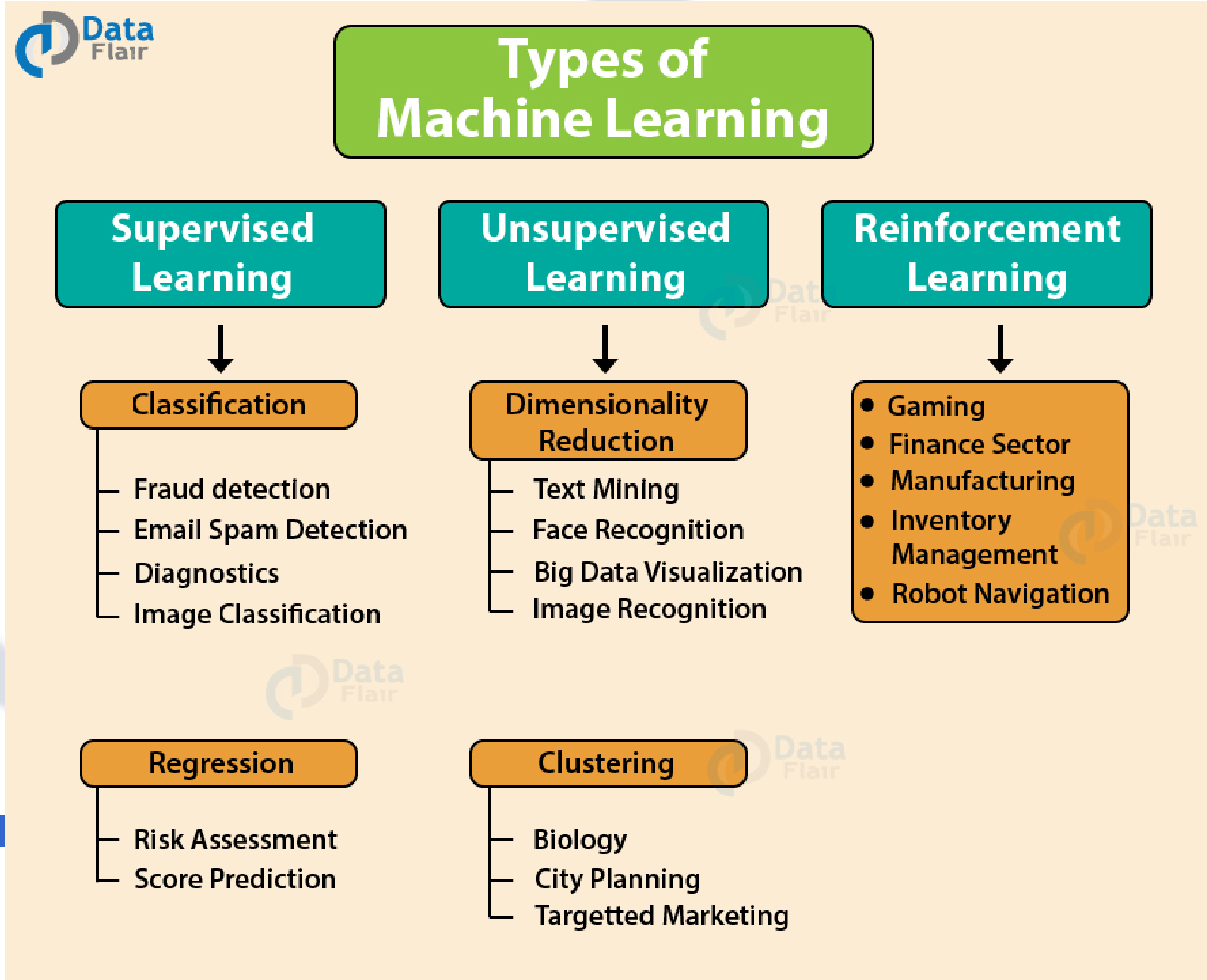
Thank you for your attention!

Methodology – support I

Machine learning: a pool of data-driven computational techniques that help unlock multi-dimensional data problems

Examples:

Regression
Decision Tree Learning
Support Vector Machines
Associated Rule Learning
Artificial Neural Networks
Inductive Logic Programming
Reinforcement Learning
Clustering
Similarity and Metric Learning
Bayesian Networks
Representation Learning



https://data-flair.training/blogs/machine-learning-tutorial/

CEDR
Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads

Methodology – support II

CART: Classification and Regression Tree

- **Input:** one output data column and at least one input data column (continuous or categorical)
- **The goal**: is to create a model that predicts the value of a target variable based on (several) input variable(s)
- **Method:** "Recursive partitioning" of the data within each node ("parent") into two additional nodes ("children") in every possible way:

  - Goal: to maximizing "variance reduction" of the node due to the split at this node.

  - Limit: when (a) the gain goes below a threshold or (b) when some set criterion is met (e.g. data support)
- **Output:** a hierarchical decision structure with logical "splitting" nodes and "terminal" nodes that give the response.
- **Evaluation**: e.g. by misclassification ("confusion") matrix (if decision tree) or RMSE or similar metric (if regression tree)