02445
Statistical evaluation of artificial intelligence systems

Nikolaj S. Povlsen
Rasmus J. Pedersen

Technical University of Denmark (DTU)

January 24, 2020

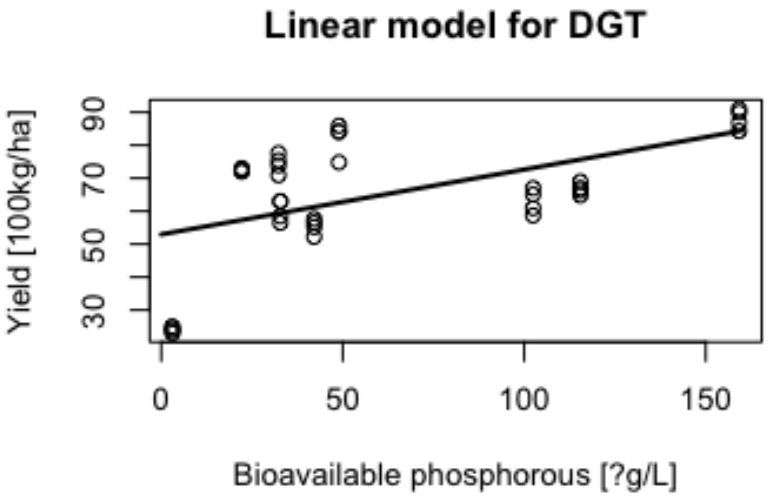
Project 2

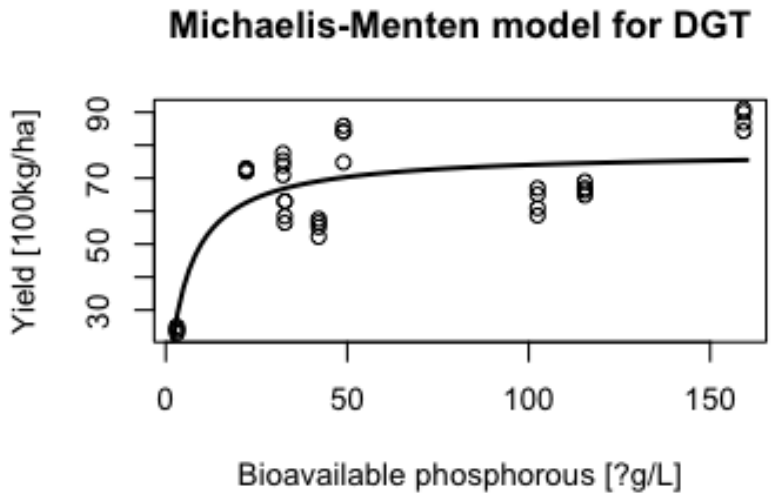
Choosing between Olsen P and DGT measurements

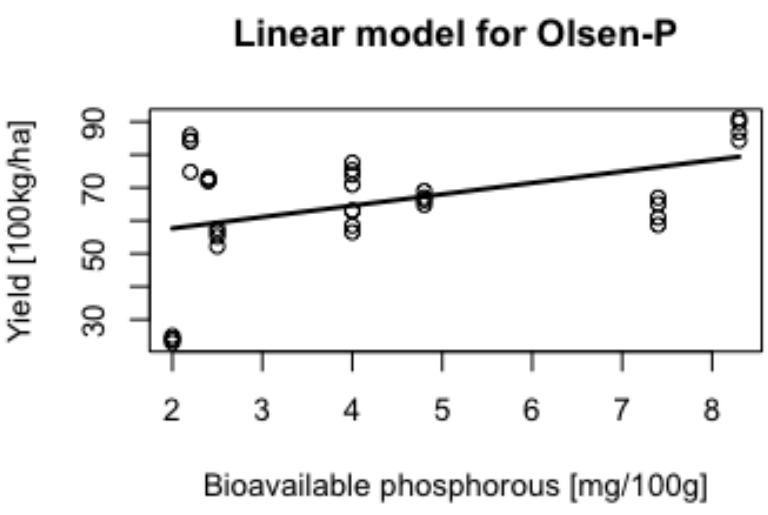
- Evaluating models
- Testing for significance
- Analyzing residuals

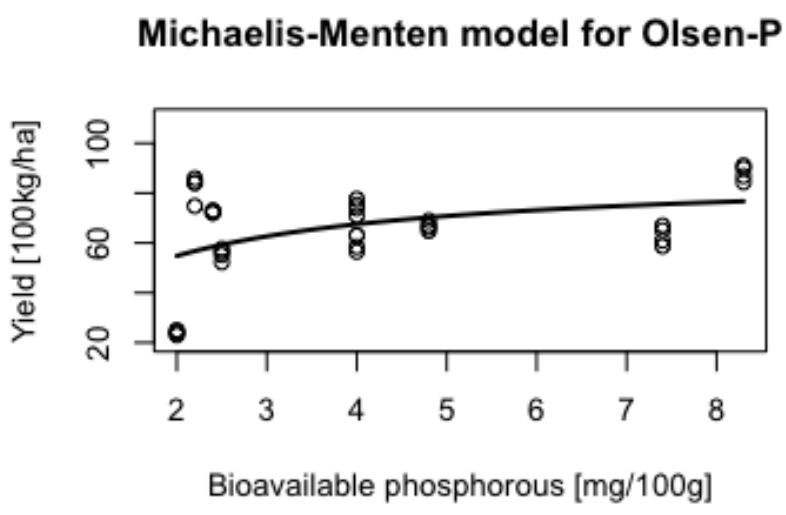
Does the amount of bio-available phosphorous influence the harvest yield?

Choosing between Olsen P and DGT measurements









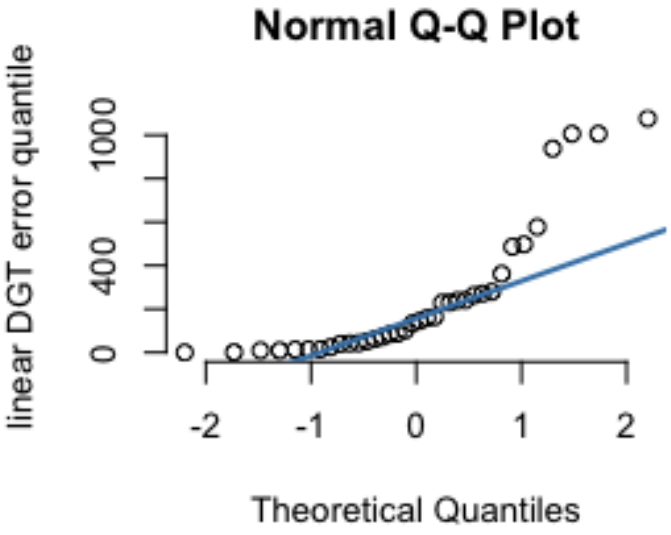
Choosing between Olsen P and DGT measurements

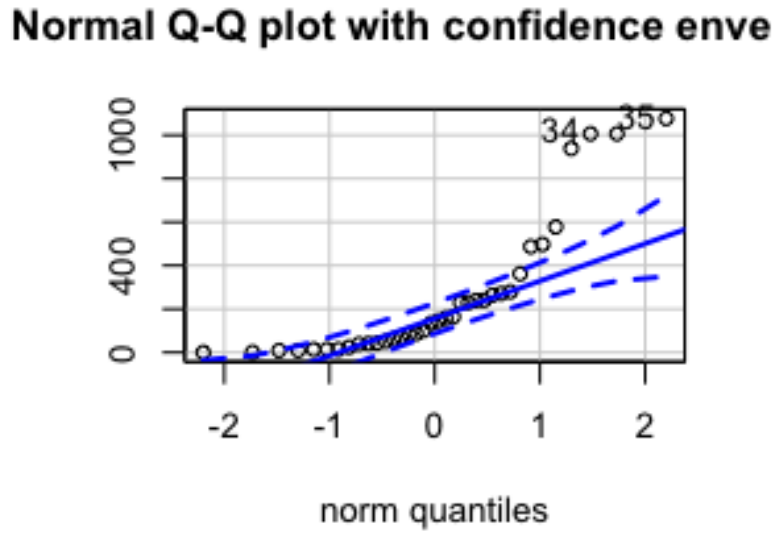
Model	Linear DGT	Linear Olsen-P	Non-linear DGT	Non-linear Olsen-P
Std squared error	15.37	16.55	10.58	16.33
p-values	DGT = 0.000685	Olsen-P = 0.0103	Alpha = 2e-16 Beta = 0.0014	Alpha = 1e-9 Beta = 0.0432

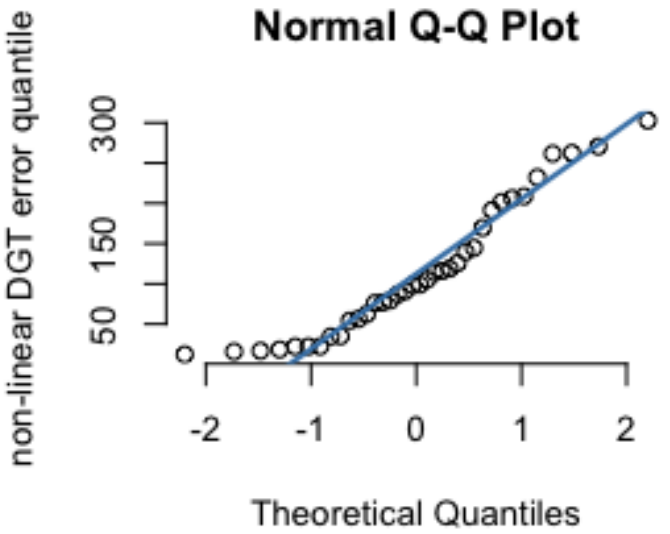
Choosing between Olsen P and DGT measurements

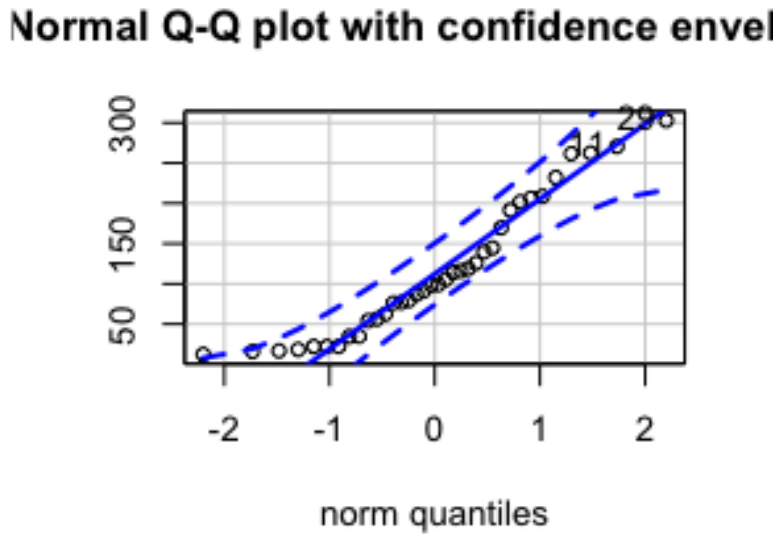
Paired t-test between:	**t-statistic**	**df**	**p-values**
Non-Linear DGT - Non-linear Olsen-P	-2.694	35	**0.011**
Non-Linear DGT - Linear Olsen-P	-2.481	35	**0.018**
Non-Linear DGT - Linear DGT	-2.381	35	**0.023**
Linear DGT - Linear Olsen-P	-1.874	35	0.069
Linear DGT - Non-linear Olsen-P	-1.590	35	0.12
Linear Olsen-P - Non-linear Olsen-P	0.3065	35	0.76

Residuals from linear and non-linear DGT measurements









Project 1

Classifying 3D motions

- Comparing models

Does the experiment influence the motion

- Test statistics
- PCA
- T^2

Resuts

Model accuracy

Model	CI of Generalization Accuracy
ANN	0.708 ± 0.0078
KNN	0.644 ± 0.0066

Paired t-test

Test	Test Statistic	p-value
Paired t-test	10.934	$8e-12$

Testing for normally distributed variables

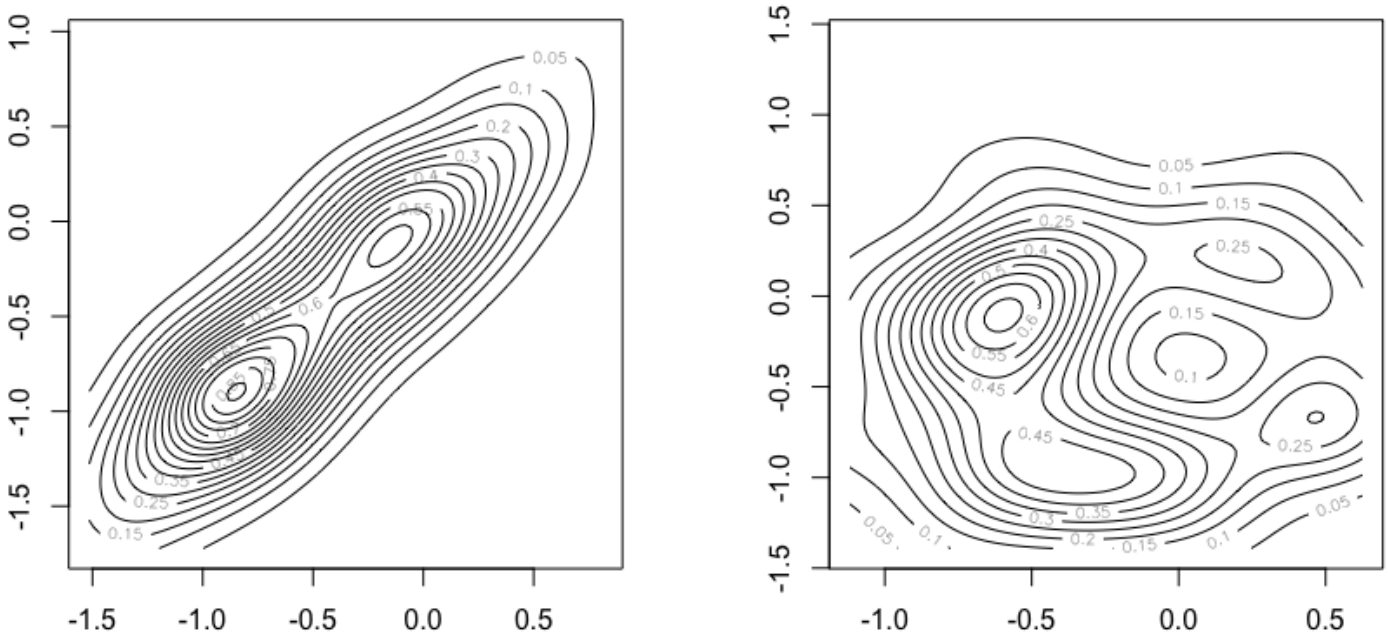


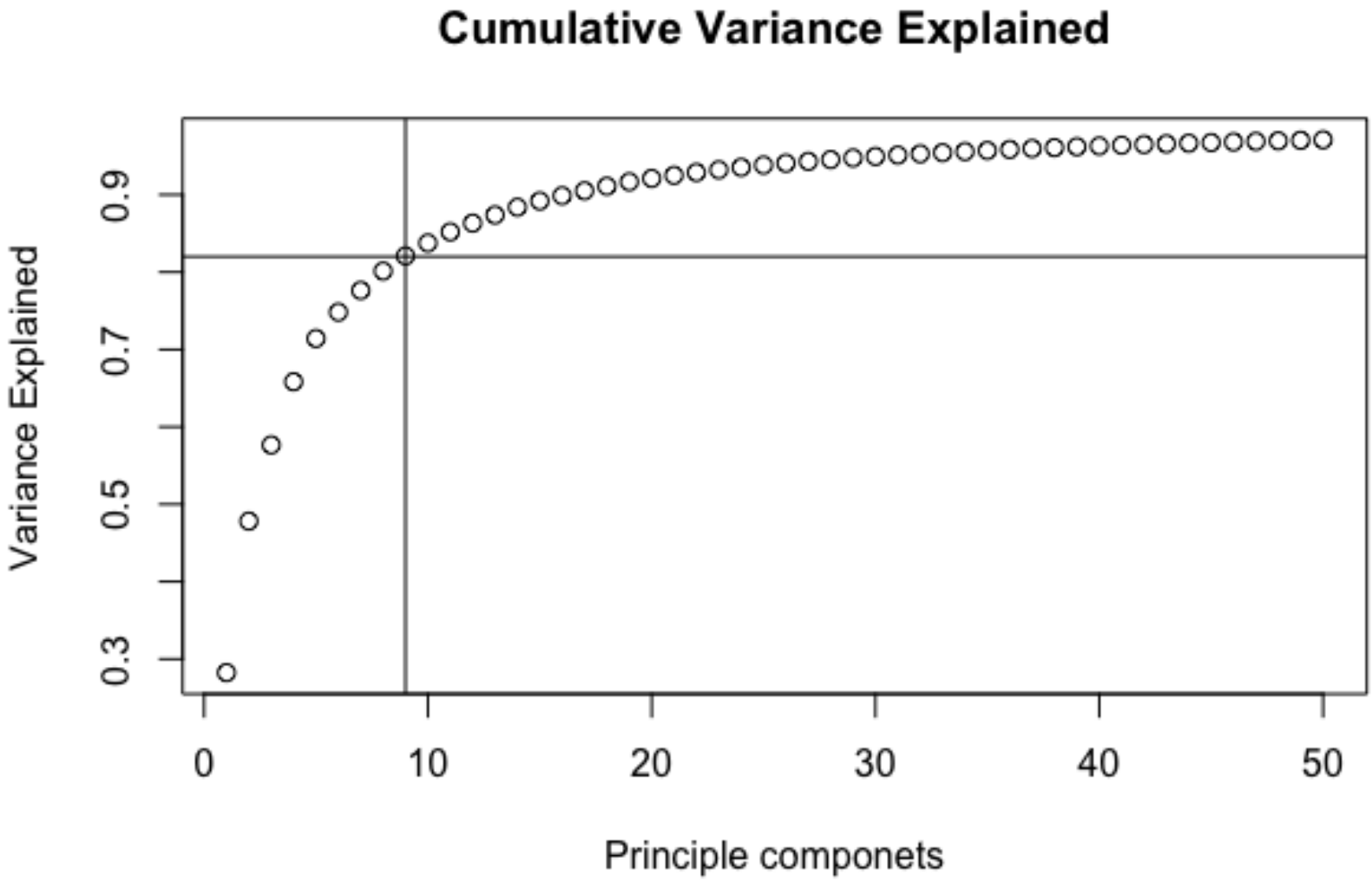
Figure: Two univariate variables forming a multivariate distribution. Left pane shows approximately a multivariate normally distribution, the right does not.

Central Limit Theorem

Equals a 90% reduction in sample size.

PCA

But 9 principal components explain 82% of the variance



5 out of 120 pairs of tests were not significant

Experiment pair	**p-value**
1, 4	0.137
4, 7	0.177
7, 10	0.278
5, 8	0.0855
6, 9	0.162

The experiments

	S	M	T
15.0 cm	1	2	3
22.5 cm	**4**	5	6
30.0 cm	**7**	8	9
37.5 cm	10	11	12
45.0 cm	13	14	15