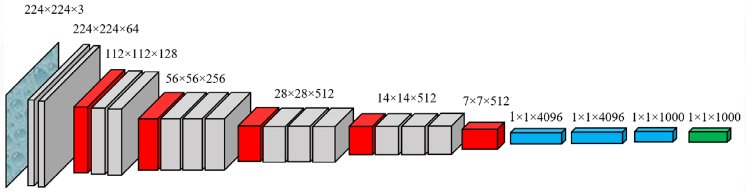
Leveraging data in chemical- and minerals-processing industries

Department of Chemical Engineering
Tobi Louw

Photo by Stefan Els

Undeniable impact of data science



AlphaGo

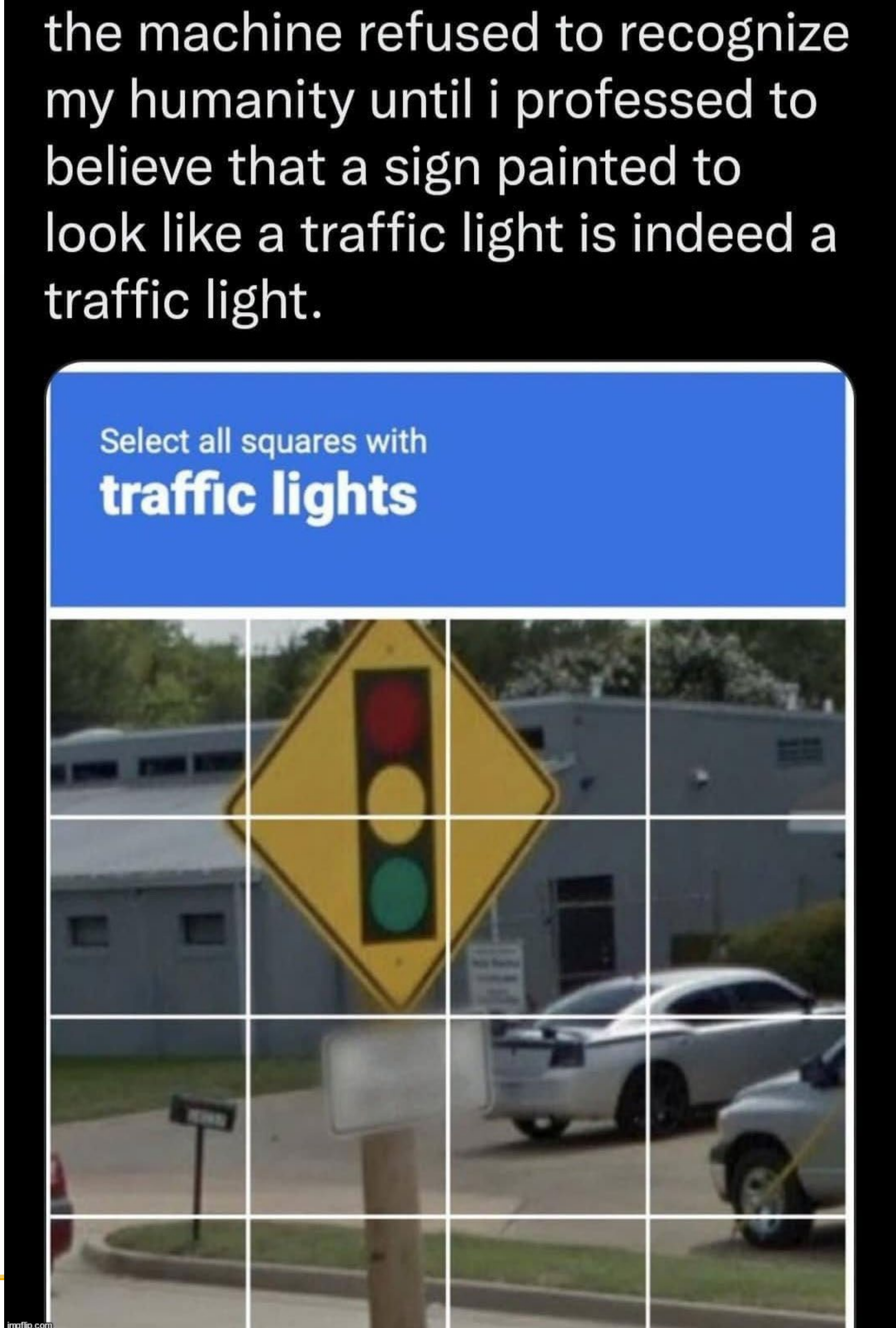
OpenAI

The success of data science
+ machine learning in…
…pattern recognition for images,
…reinforcement learning for games,
…or generative large language models,

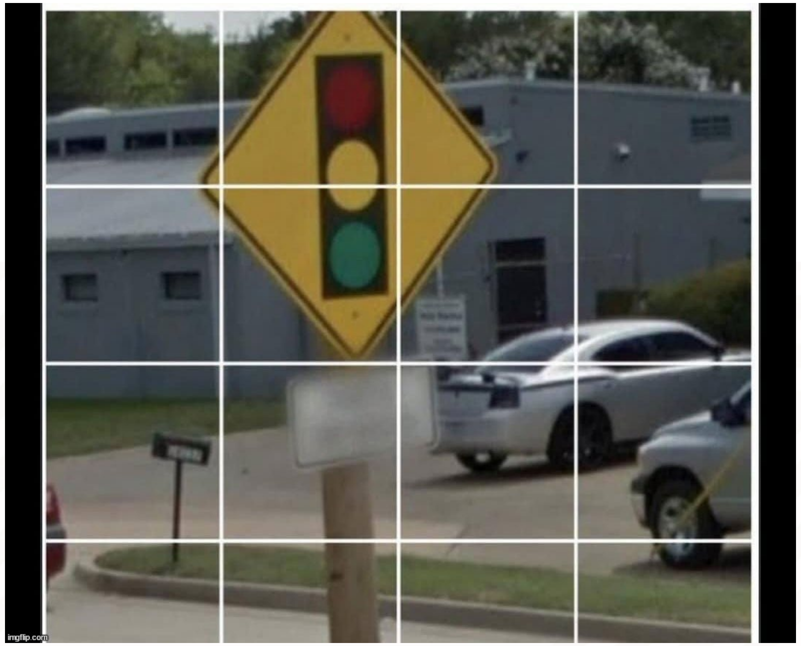
does not immediately imply utility for the process industry

Challenges in process industry

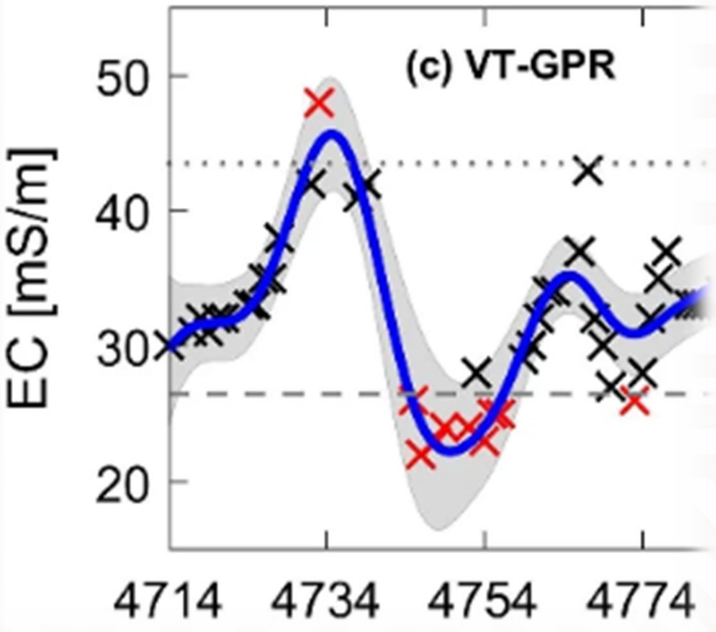
Major successes supported
by very large datasets (AlexNet, ChatGPT),
or accurate simulations (AlphaGo)



Challenges in process industry



Major successes supported
by very large datasets (AlexNet, ChatGPT),
or accurate simulations (AlphaGo)



Process data:

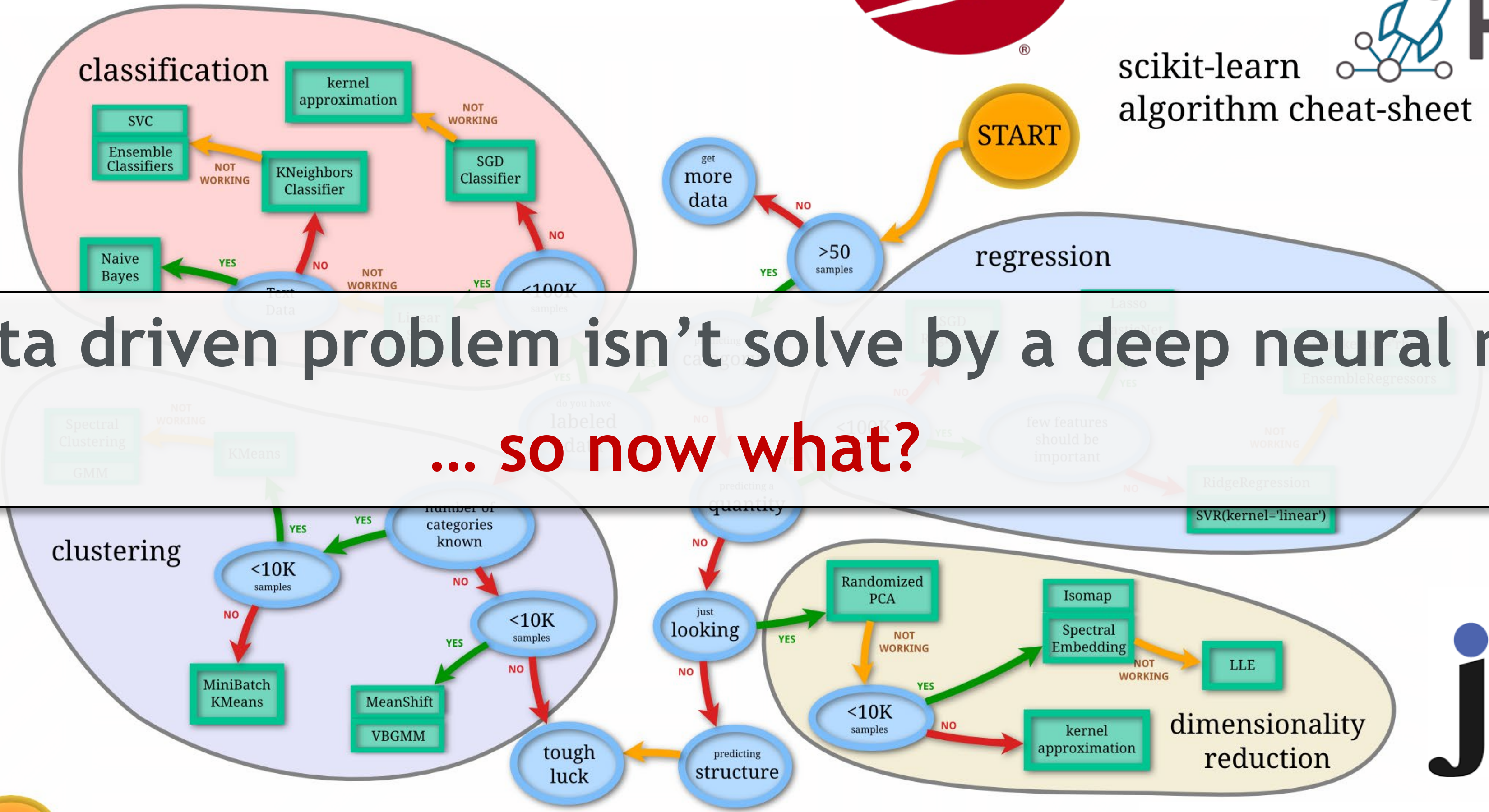
- Dynamic (comparatively) low volume
- Noisy, missing, censored, faulty
- Low observability (measurements vs states)
- Amenable to simulation (sometimes)

Challenges in process industry

Every data driven problem isn’t solve by a deep neural net…

Challenges in process industry

scikit-learn
algorithm cheat-sheet



Every data driven problem isn't solve by a deep neural net…

… so now what?

Back

Variety and (perceived) opacity of ML algorithms lead to faulty insights (sometimes)

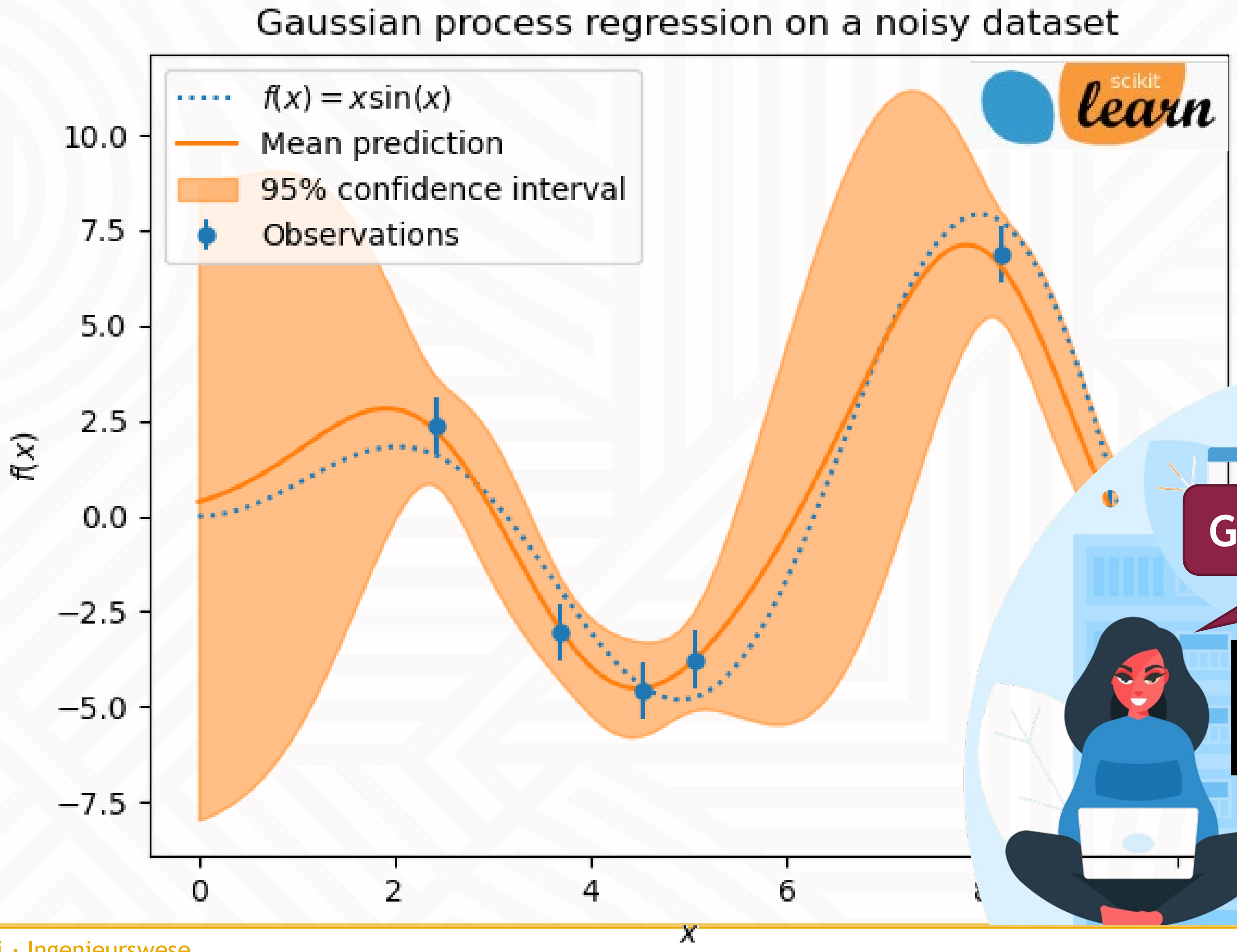
Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

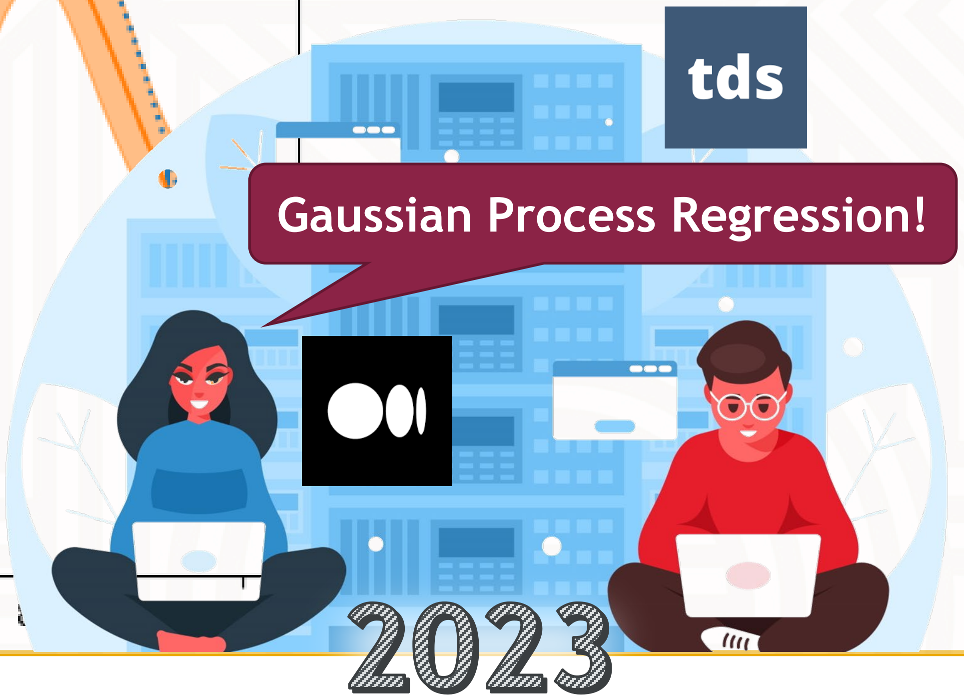
- Recent paper considered bio-energy yields of crops (not SA)
- Variety of ML algorithms assessed
- Feature selection, cross-validation for hyperparameter optimization, hold-out test set
- Best performing model:
  - k-nearest neighbour with k = 2
  - Features = province, crop type, weather and humidity

Different names for the same thing…

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

https://scikit-learn.org/stable/auto_examples/gaussian_process/plot_gpr_noisy_targets.html

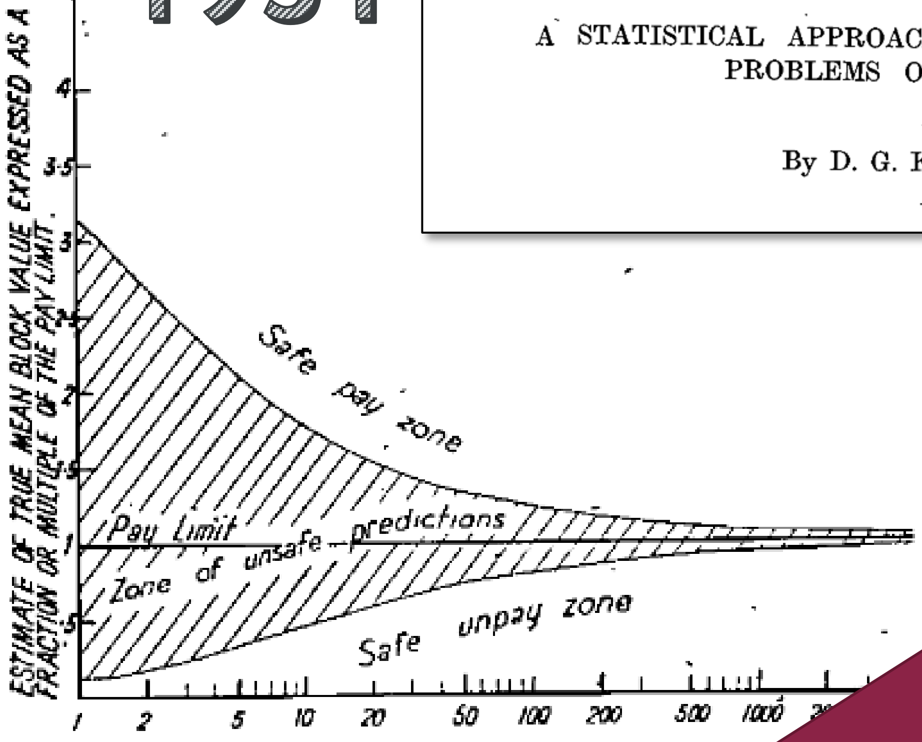


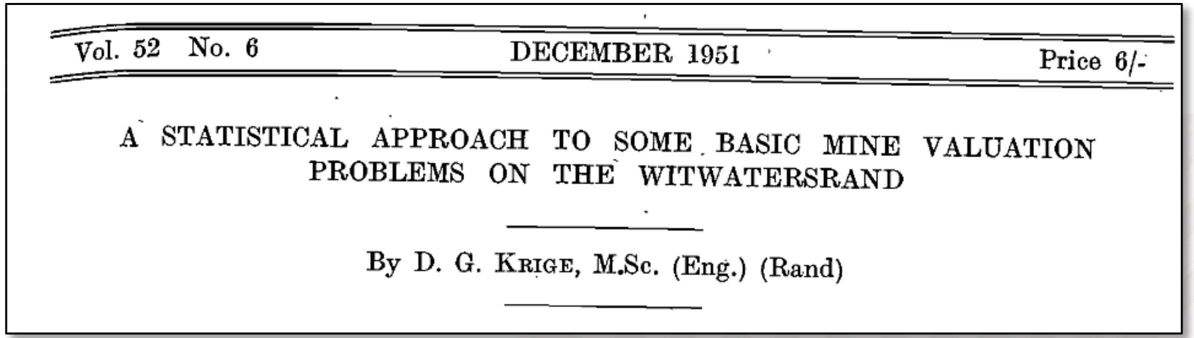


Different names for the same thing…

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

1951

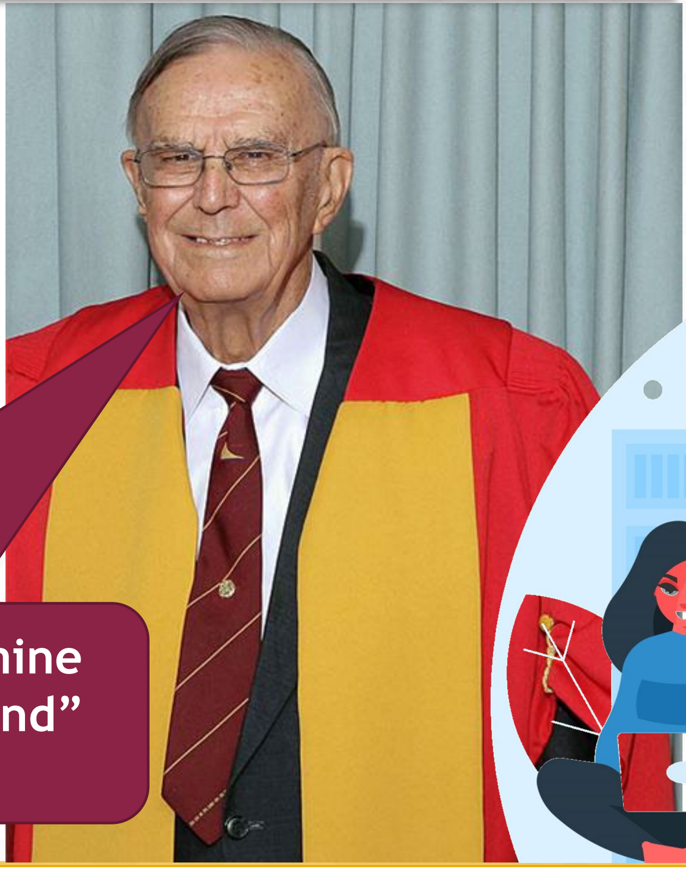




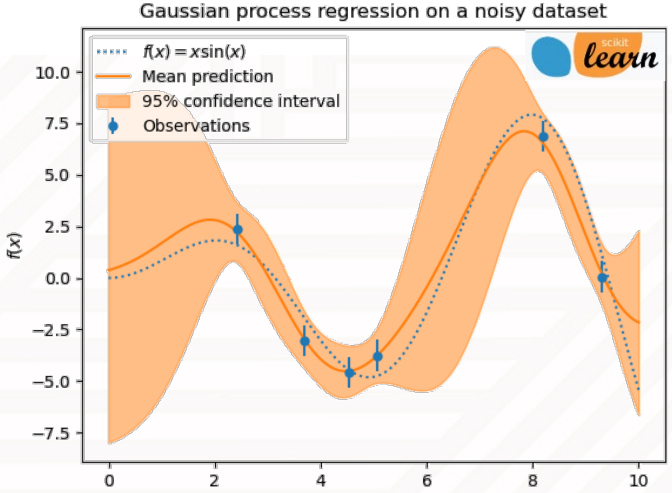
Vol. 52 No. 6 DECEMBER 1951 Price 6/-

A STATISTICAL APPROACH TO SOME BASIC MINE VALUATION PROBLEMS ON THE WITWATERSRAND

By D. G. Krige, M.Sc. (Eng.) (Rand)

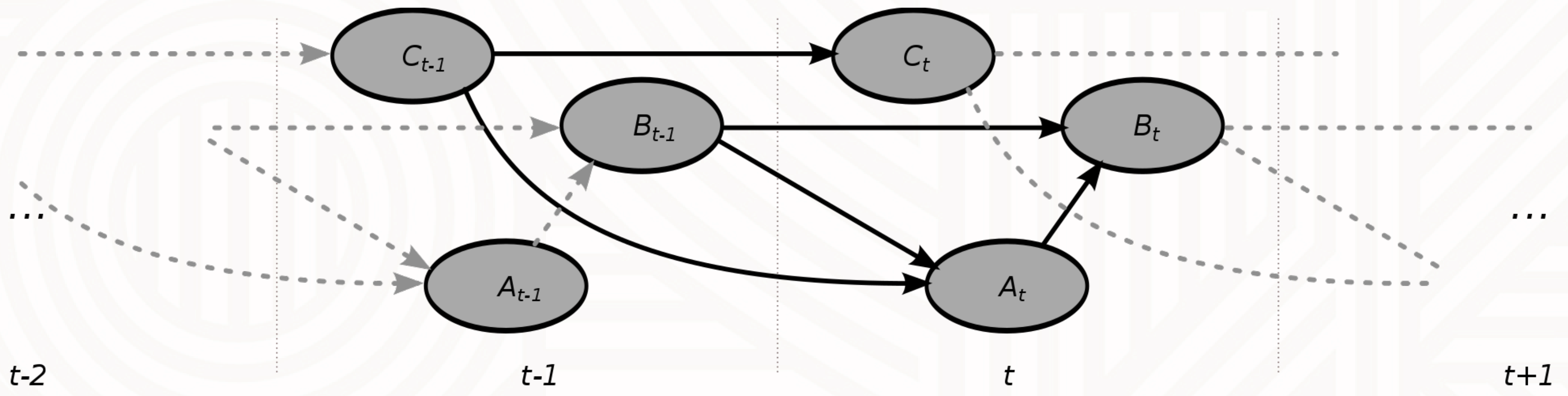


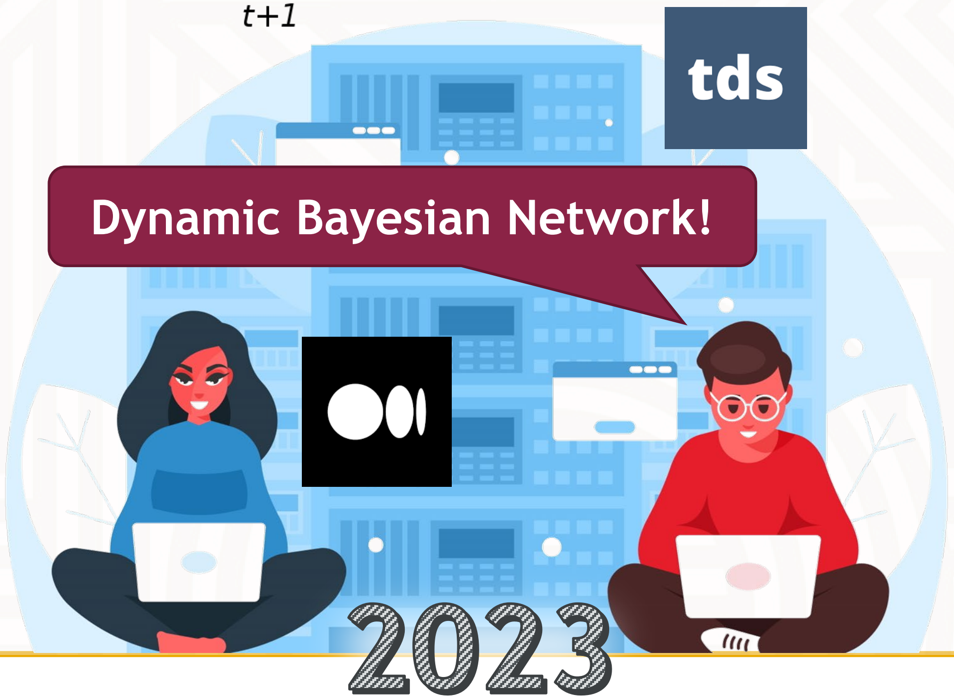
“A statistical approach to some basic mine valuation problems on the Witwatersrand”
Danie Krige (Kriging)



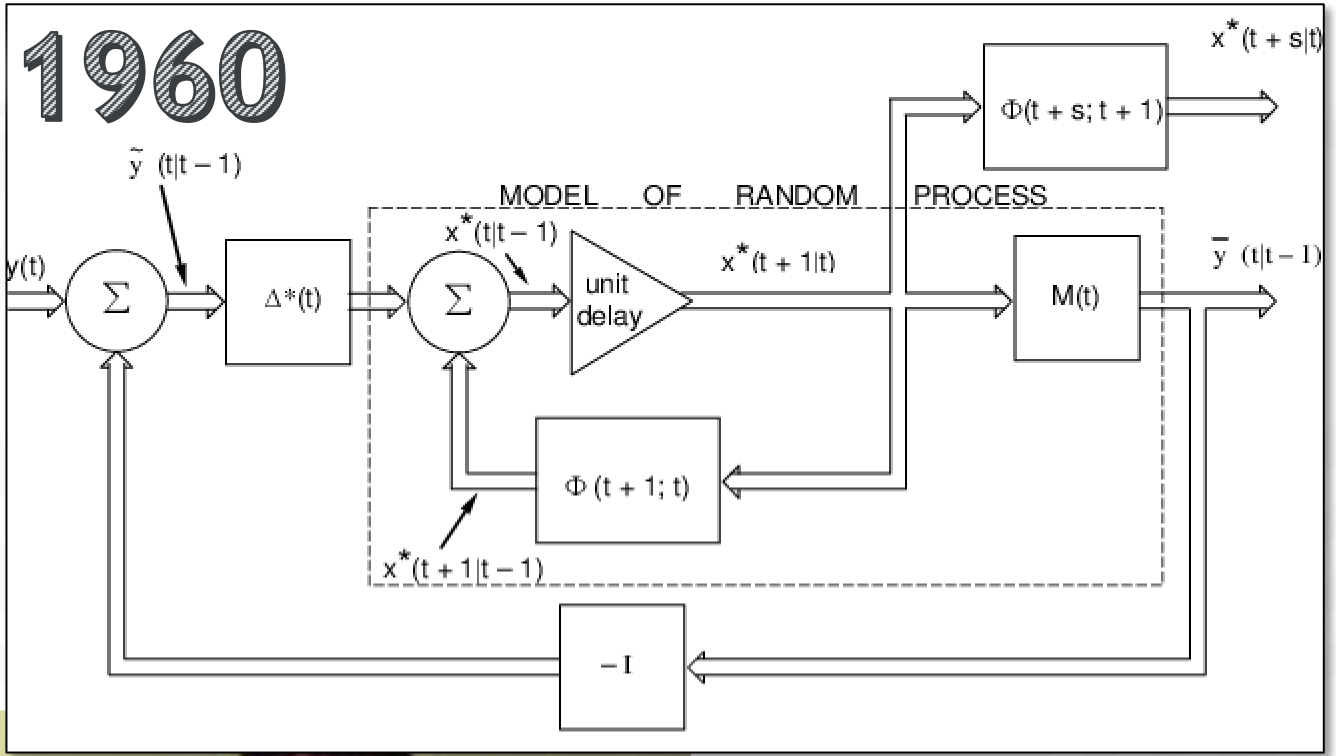
2023

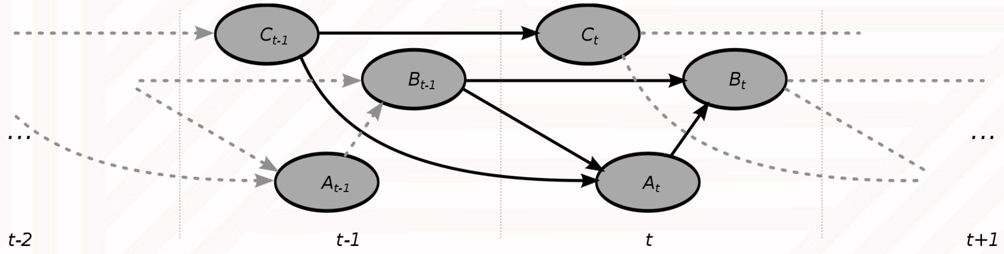
Different names for the same thing…

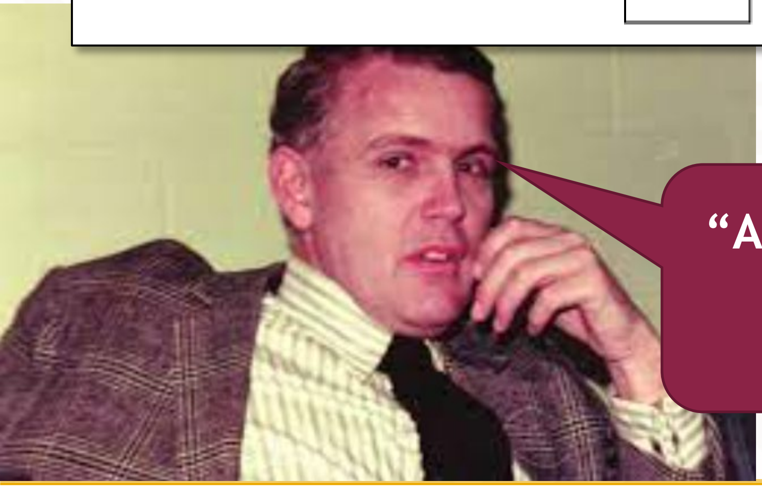




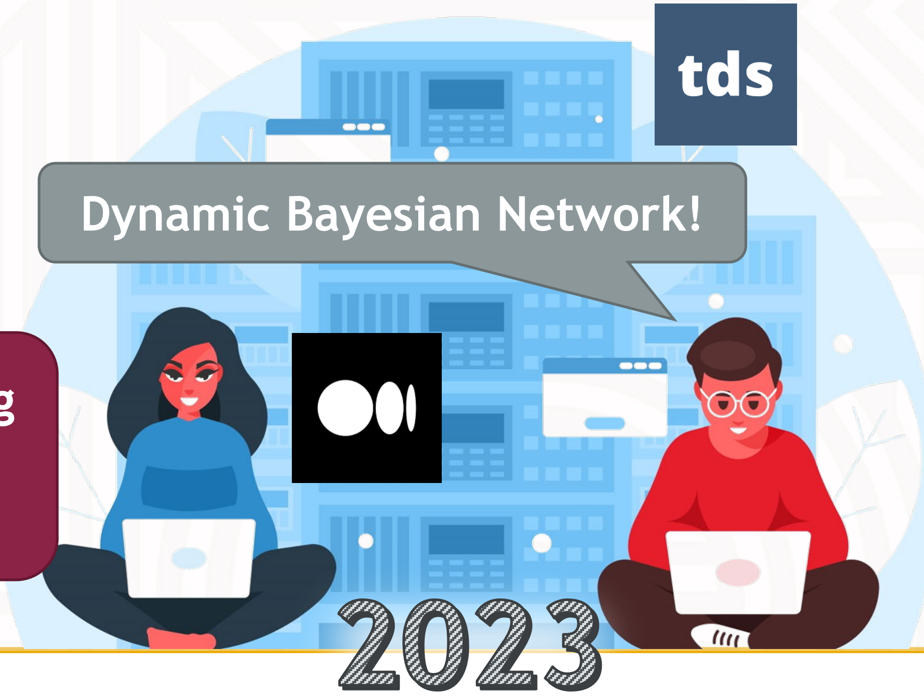
Different names for the same thing…







"A new approach to linear filtering and prediction problems"
Rudolf Kalman



Dynamic Bayesian Network!

Bringing innovations in ML to bear on process modelling, monitoring, and control

As statisticians, our exposition will naturally reflect our backgrounds and areas of expertise. However in the past eight years we have been attending conferences in neural networks, data mining and machine learning, and our thinking has been heavily influenced by these exciting fields. This influence is evident in our current research, and in this book.

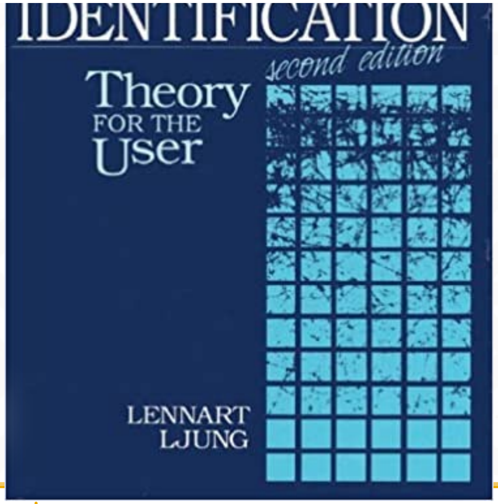
Bringing innovations in ML to bear on process modelling, monitoring, and control

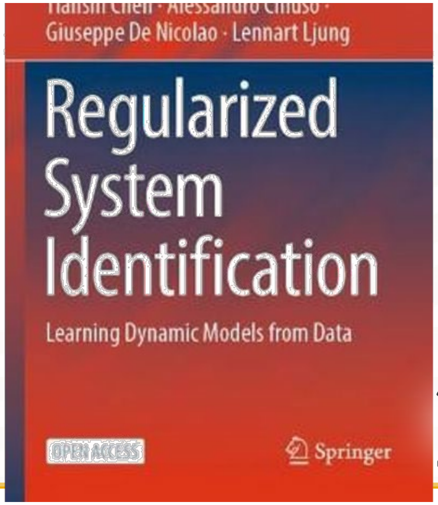
Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

- Starting from a firm foundation in established fields (system identification, state estimation, statistical process control)
- Devel ing, unde lems

As statisticians, our exposition will naturally reflect our backgrounds and areas of expertise. However in the past eight years we have been attending conferences in neural networks, data mining and machine learning, and our thinking has been heavily influenced by these exciting fields. This influence is evident in our current research, and in this book.

1999



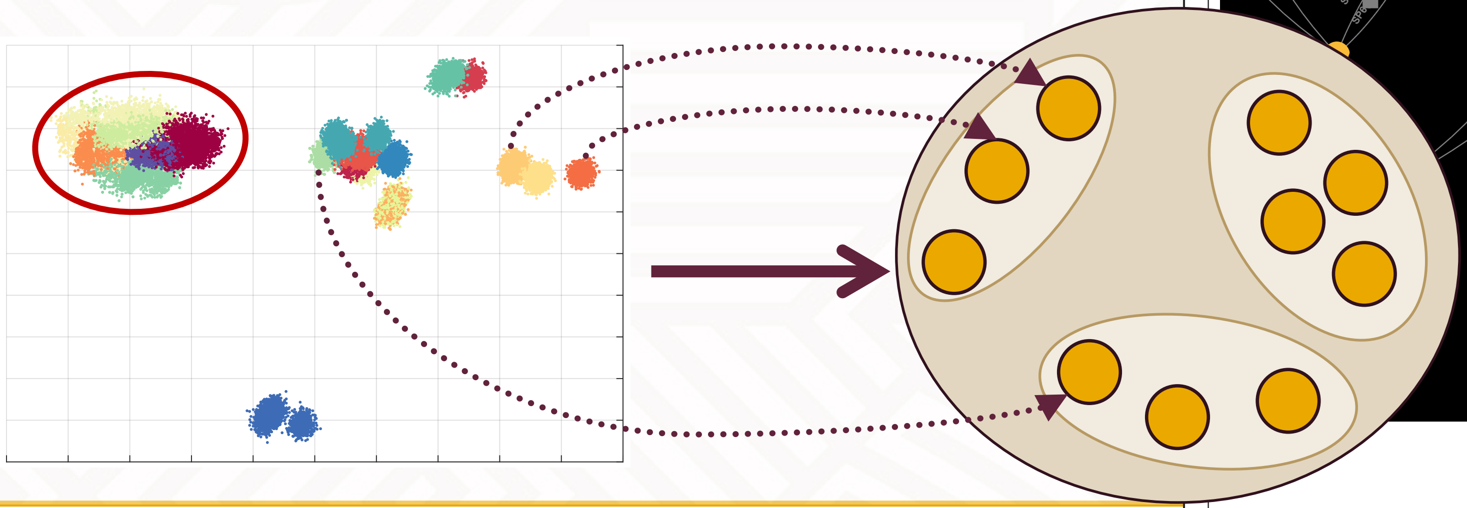


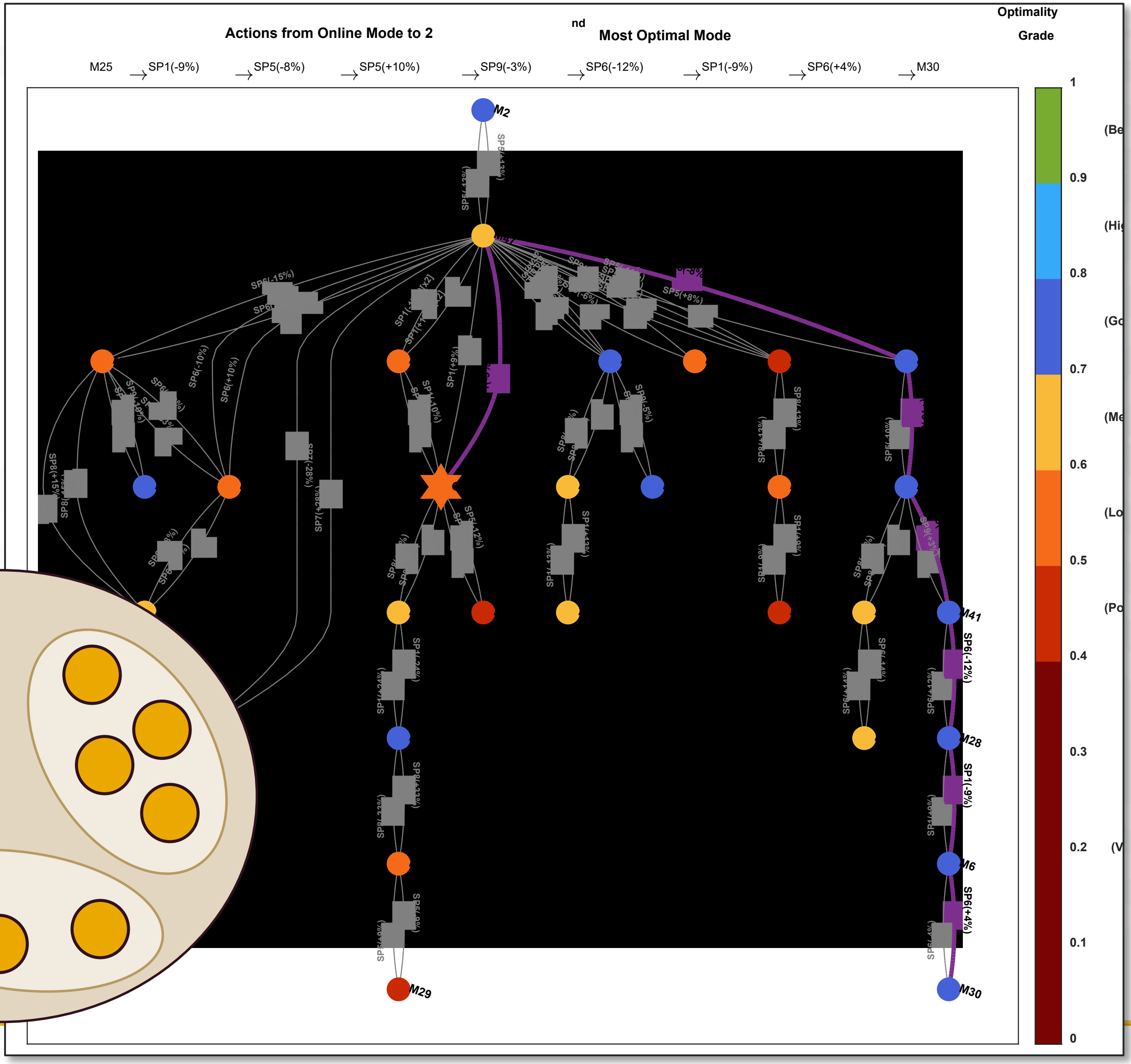
2022

Bringing innovations in ML to bear on process modelling, monitoring, and control

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

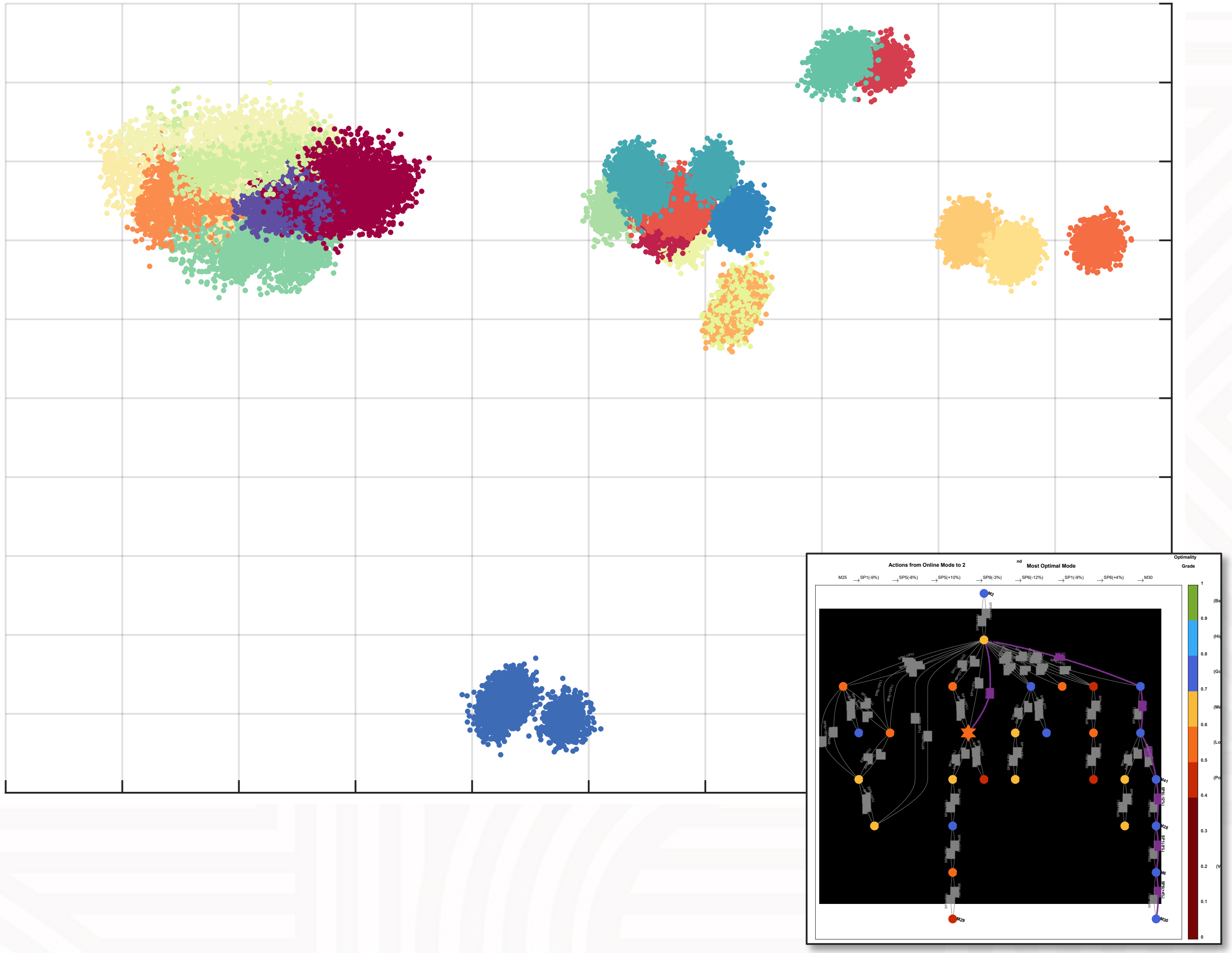
- Purely data-driven approach to identify process modes linked to (retrospective) KPIs and modal shifts
- Online identification of current process mode as well as optimal reachable mode





Bringing innovations in ML to bear on process modelling, monitoring, and control

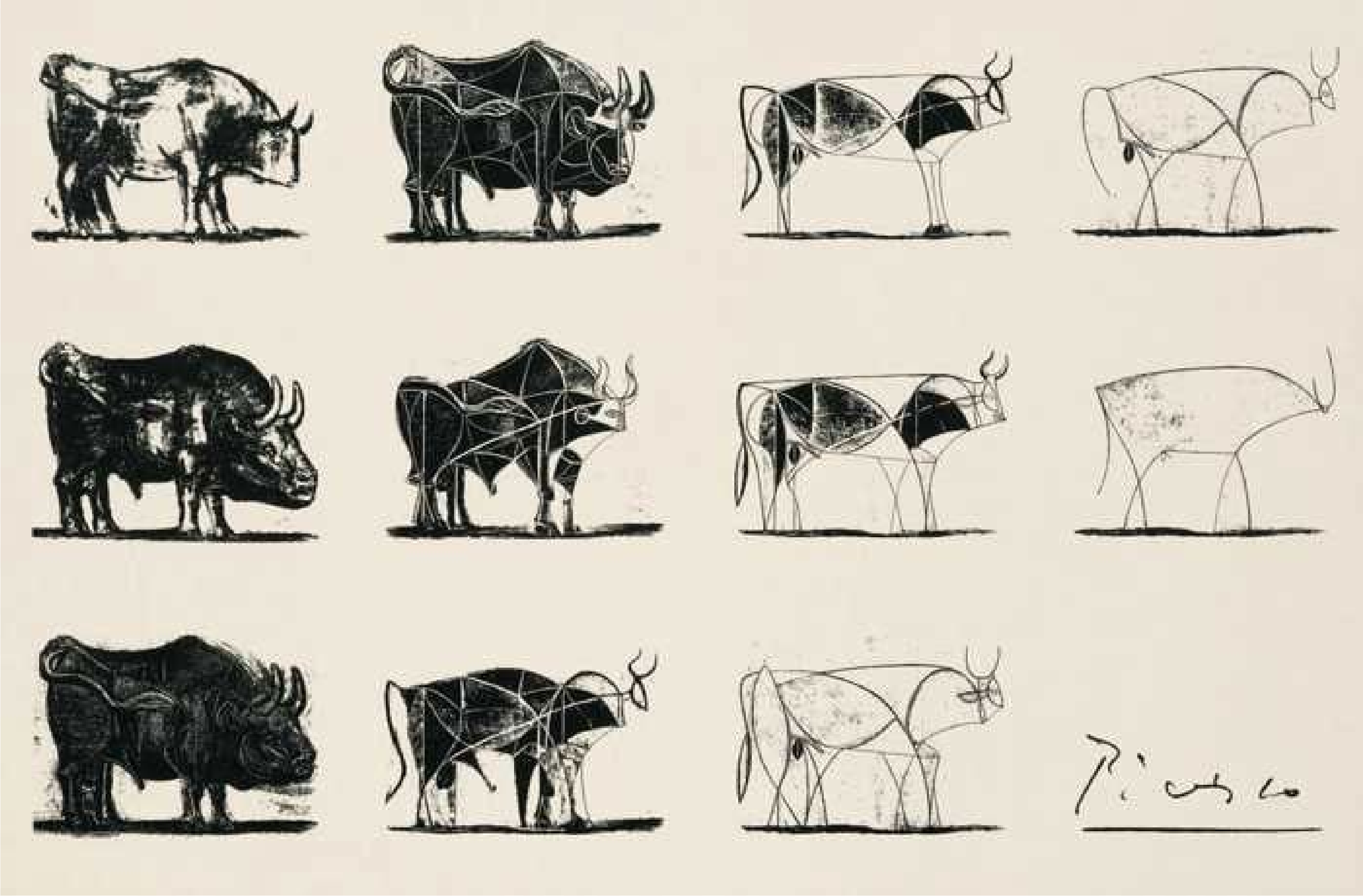
- Purely data-driven approach to identify process modes linked to (retrospective) KPIs and modal shifts
- Online identification of current process mode as well as optimal reachable mode
- Data overlap limits modal identification



Bringing innovations in ML to bear on process modelling, monitoring, and control

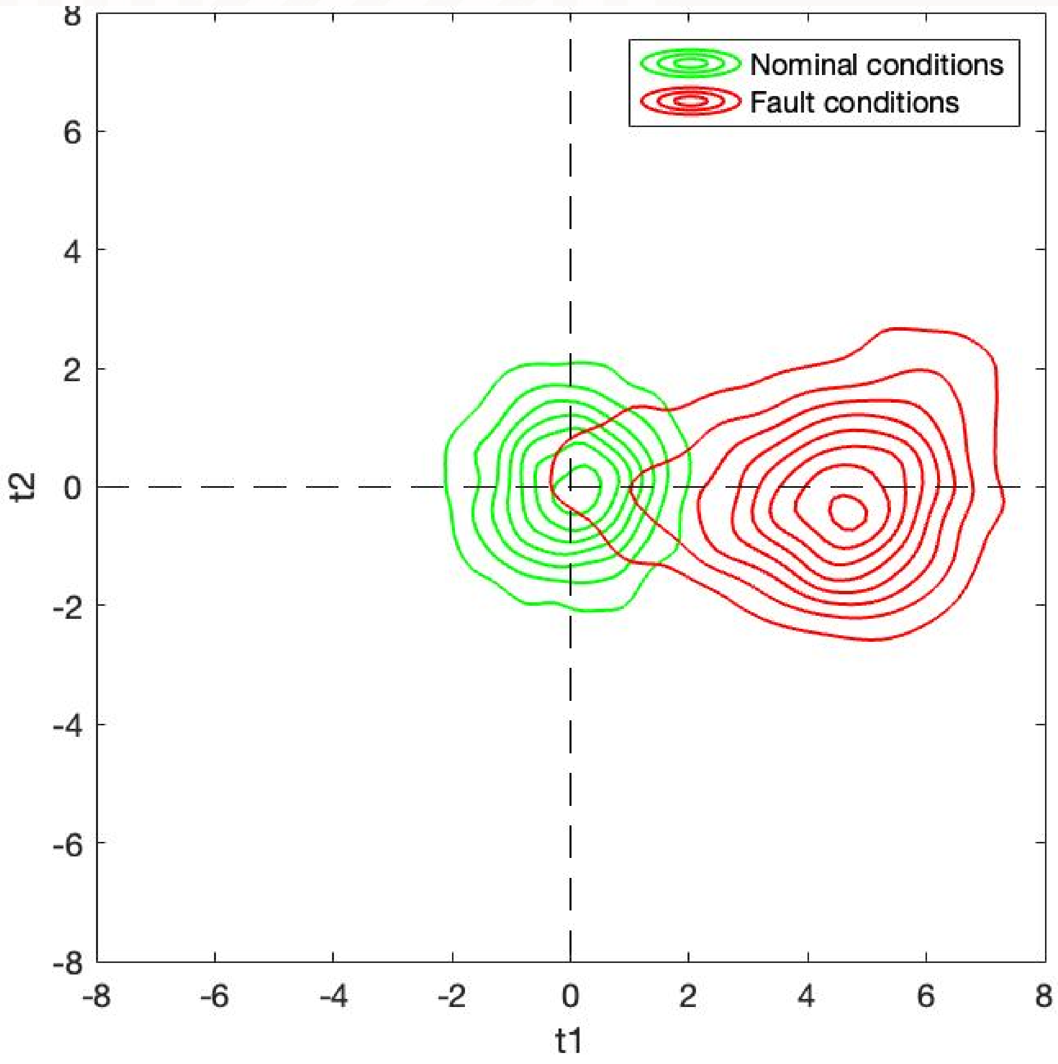
Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

- Feature extraction to provide separability:
  - Kernel methods transform data to infinite dimensional space
  - Deep learning extracts features through multiple non-linear transformations



Bringing innovations in ML to bear on process modelling, monitoring, and control

- Feature extraction to provide separability:
  - Kernel methods transform data to infinite dimensional space
  - Deep learning extracts features through multiple non-linear transformations
- State estimation can be used for feature extraction to enhance classification
- Example: hazardous event detection
- Future work: combine with advances in Dynamic Bayesian Networks to estimate discrete states (e.g., sensor health)



Bringing innovations in ML to bear on process modelling, monitoring, and control

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

- Variational inference in Gaussian Process Regression enables computationally efficient co-kriging (multiple outputs)

Variational Learning of Inducing Variables in Sparse Gaussian Processes

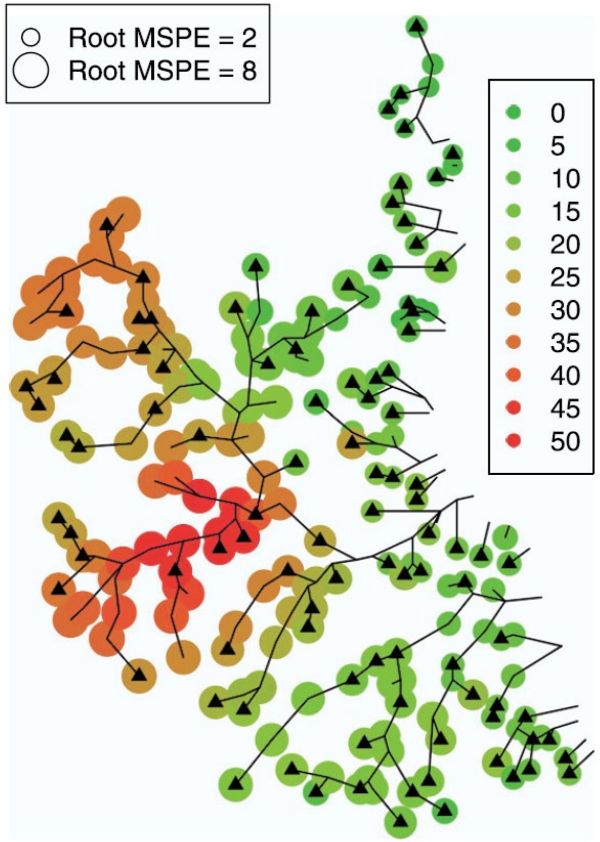
Michalis K. Titsias
School of Computer Science,
University of Manchester, UK
mtitsias@cs.man.ac.uk

Bringing innovations in ML to bear on process modelling, monitoring, and control

- Variational inference in Gaussian Process Regression enables computationally efficient co-kriging (multiple outputs)
- Coupled with kriging developed for flow connected systems (e.g., pollutants in river network)

Spatial Prediction on a River Network

Noel CRESSIE, Jesse FREY, Bronwyn HARCH, and Mick SMITH



Bringing innovations in ML to bear on process modelling, monitoring, and control

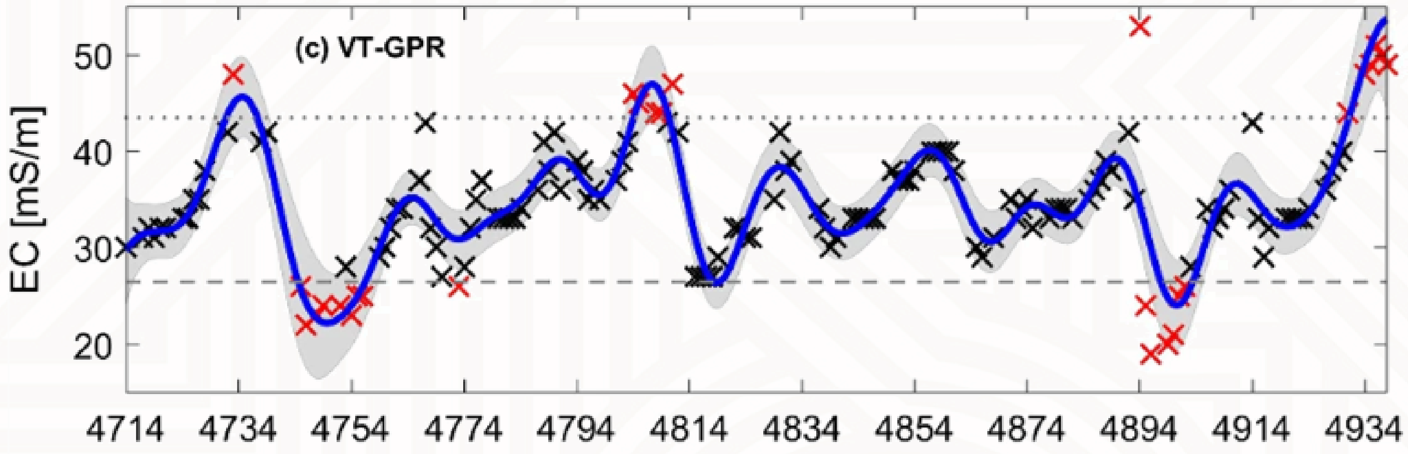
- Variational inference in Gaussian Process Regression enables computationally efficient co-kriging (multiple outputs)
- Coupled with kriging developed for flow connected systems (e.g., pollutants in river network)
- Inclusion of censored data: measurements outside of measurement range (not zero, but not known)

Original Paper | Open Access | Published: 31 March 2023

Variational Tobit Gaussian Process Regression

Marno Basson, Tobias M. Louw & Theresa R. Smith

Statistics and Computing **33**, Article number: 64 (2023) | Cite this article



Bringing innovations in ML to bear on process modelling, monitoring, and control

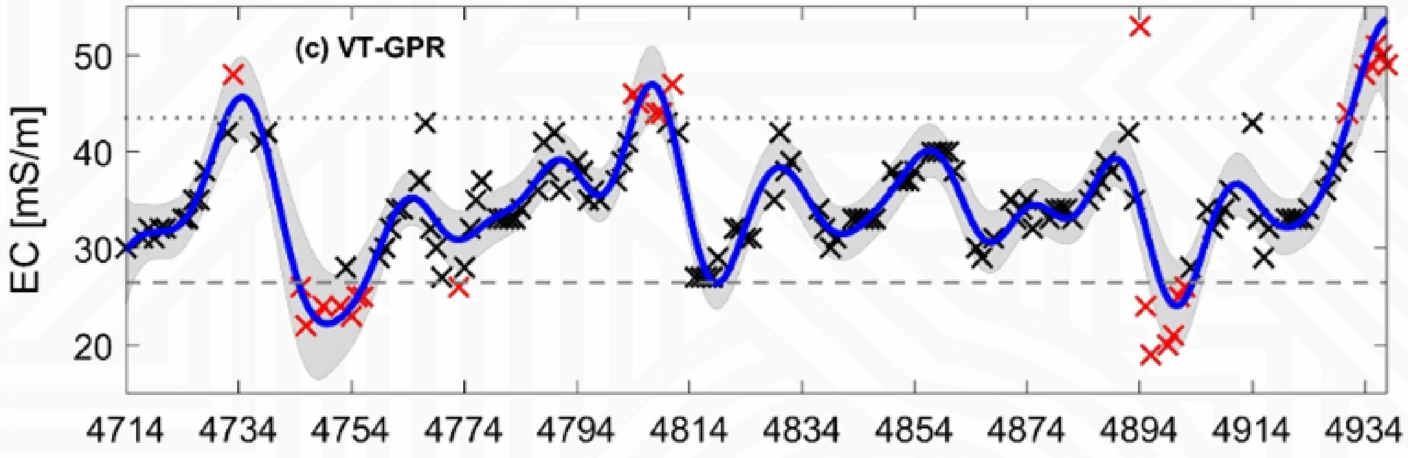
- Variational inference in Gaussian Process Regression enables computationally efficient co-kriging (multiple outputs)
- Coupled with kriging developed for flow connected systems (e.g., pollutants in river network)
- Inclusion of censored data: measurements outside of measurement range (not zero, but not known)

Original Paper | Open Access | Published: 31 March 2023

Variational Tobit Gaussian Process Regression

Marno Basson, Tobias M. Louw & Theresa R. Smith

Statistics and Computing **33**, Article number: 64 (2023) | Cite this article



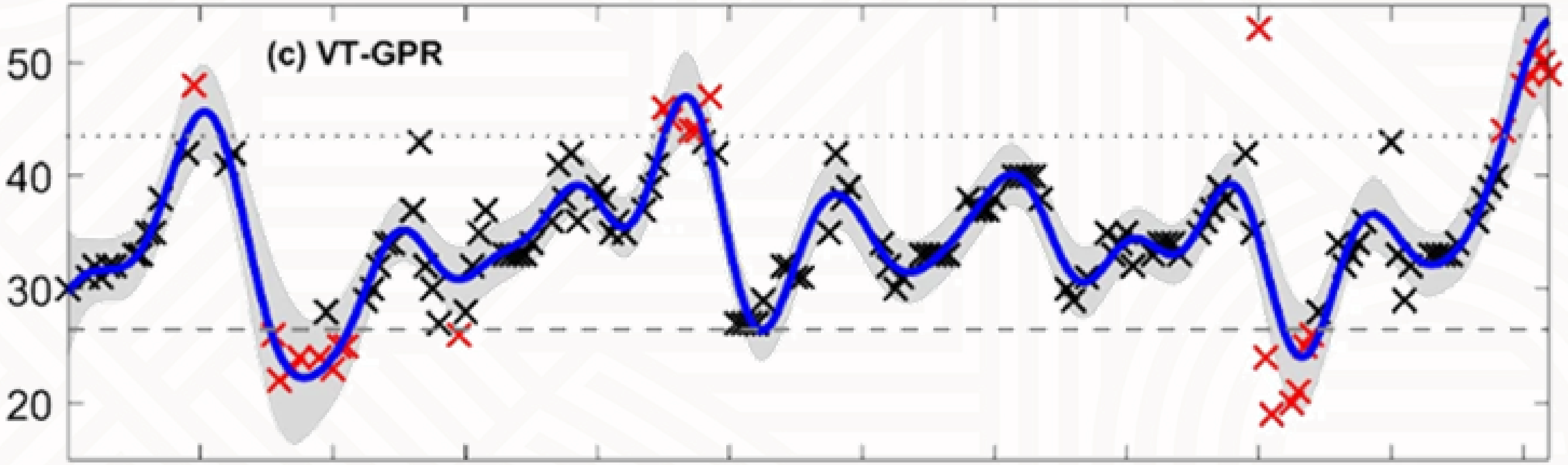
Future work: ensure inferred measurements obey conservation laws (i.e., data reconciliation)

VALUE PROPOSITION

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

We aim to develop the tools to replace process historians containing noisy, corrupt, censored, faulty, measured data

with labelled distributions over fundamental state variables satisfying mass/energy balances and physical laws



Data Analytics and Machine Learning at Chemical Engineering

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

- Continually identifying new opportunities to apply innovations in ML to process industry through PhD and research M.Eng
- 2024: structured M.Eng (Chemical) with focus area Data Analytics
- Collaboratively developed with industry
- Aimed at working engineers studying part-time
- Equip engineers with fundamentals of data science enabling application in the context of integrated industrial processes

Part-time students	Term 1	Term 2	Term 3	Term 4	*Full-time students*
Year 1	Data Science	Applied Machine Learning	Plantwide Dynamics andControl		Year 1
Year 2	Numerical Methods	Optimisation		Data Analysis for Dynamic Processes	
Year 3	Advanced Topics in Eng. Management Integrated Process Data Analysis	Research project			Year 2

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

Thank you
Enkosi
Dankie

Photo by Stefan Els