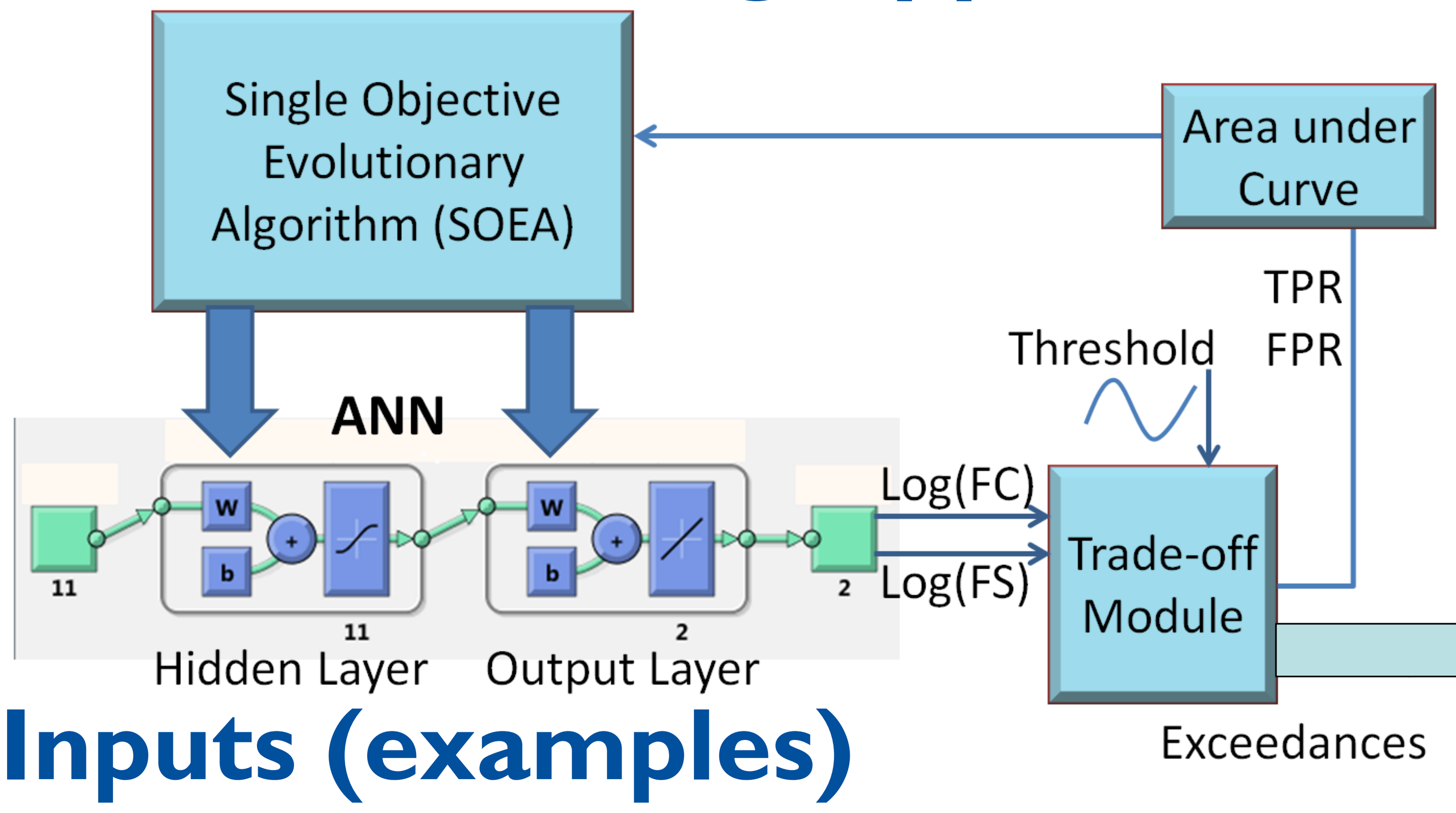
UNIVERSITY OF EXETER Centre for Water Systems

Early Warning System for Bathing Water Quality

Environment Agency

Andrew Duncan, Ed Keedwell, Slobodan Djordjević, Dragan Savić

Modelling Approach

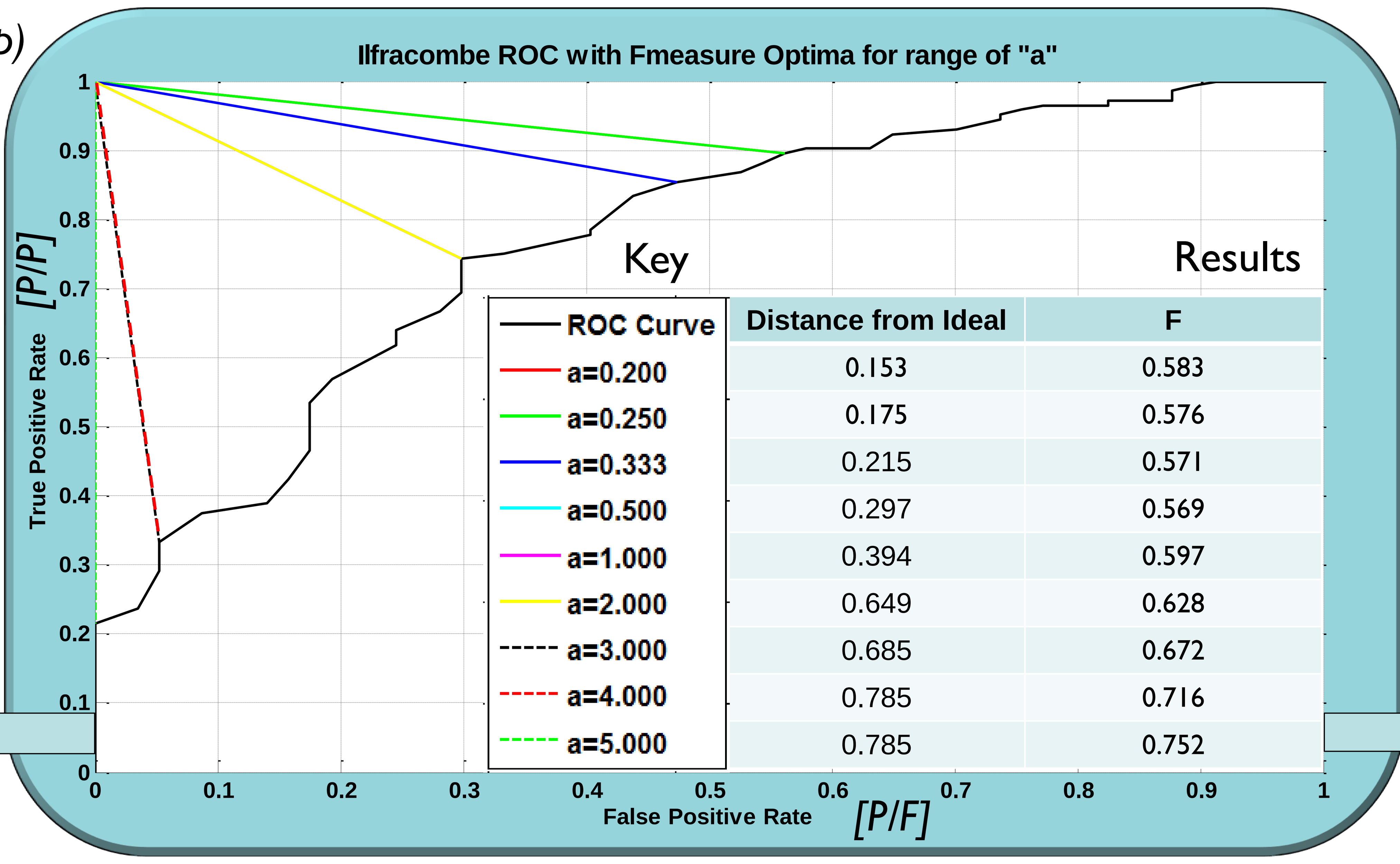


Inputs (examples)

1. Time of sample w.r.t. HW
2. Tidal Height at HW
3. Est Tidal Height at Sampling Time
4. Tidal Range at Standard Port
5. Tide Level Class (Spring/Mean/Neap)
6. 24hr Antecedent Rainfall Total
7. 48hr Antecedent Rainfall Total
8. 72hr Antecedent Rainfall Total
9. 96hr Antecedent Rainfall Total
10. 120hr Antecedent Rainfall Total
11. Salinity

Case study beaches

Trade-off curve: False negatives [F/P] vs False positives [P/F]



Key	Distance from Ideal	F
a=0.200	0.153	0.583
a=0.250	0.175	0.576
a=0.333	0.215	0.571
a=0.500	0.297	0.569
a=1.000	0.394	0.597
a=2.000	0.649	0.628
a=3.000	0.685	0.672
a=4.000	0.785	0.716
a=5.000	0.785	0.752

Terminology

[Predicted/Actual] is used throughout:

- *[P/F] = predicted pass / actual fail = false positive*
- *[F/P] = predicted fail / actual pass = false negative*
- *[P/P] = predicted pass / actual pass = true positive*
- *[F/F] = predicted fail / actual fail = true negative*

Risks from bathing water quality model errors

An ideal classifier model would have both zero false positive rate (FPR)[P/F] and zero false negative rate (FNR)[F/P] (top left corner of plot). In practice a trade-off exists between these. False positives [P/F] mean that the model predicts it's safe to swim, when a sample would exceed Bacti threshold (2006/7/EC) i.e. a potential public health issue. Conversely false negatives [F/P] mean that the model predicts it is unsafe to swim, when a sample would pass, i.e. economic issues (tourism). Assigning a relative importance factor between these two ('a' in formula to right) – sets an optimum operating point on the trade-off curve for the model (centre). Stidson et al. (SEPA report) and others have used a=4 (dashed-red-line). This corresponds with a threshold for the classifier trade-off module (top).

F = measure of skill of model

a = Relative importance of negatives [failures] to positives [passes]

TN = count of true negatives [F/F]

FP = count of false positives [P/F]

$$F = \frac{(1+a)TN}{(1+a)TN + aFP + FN}$$

(Stidson et al., 2012, SEPA report)

Benefits of ANN approach

1. Use of Artificial Neural Network (ANN) allows non-linear relationships between inputs and water quality to be modelled
2. Model can be trained for new data / beaches, automatically
3. Most suitable operating point for any trade-off 'a' (FP:FN) found
4. Model can be optimised using area under the trade-off curve to maximise accuracy
5. Flexible options for model inputs used
6. Use of forecast of catchment rainfall would allow Bacti forecasting
7. Principal Component Analysis (PCA) can be applied to select the input signals giving the most skilful model

Further Information

For further information please contact Prof. Dragan Savić (d.savic@exeter.ac.uk)

UNIVERSITY OF EXETER College of Engineering, Mathematics and Physical Sciences