



# Decision support: using machine learning through MATLAB to analyze environmental data

David W. Nadler<sup>1</sup>

Published online: 14 August 2019

© AESS 2019

## Abstract

Machine learning is not a tool that is available for use by computer scientists, but one that can and should be used by all researchers in this technological era. Gone are the days of humans solely relying on older techniques for decision support. The age of information we live in is filled with countless pieces of data and we need to use the correct tools to help make sense of it all. Using MATLAB and its machine learning tools is an excellent resource for environmental scientists to conduct deep-dives into their data. We use this software title to demonstrate some of its capabilities to enhance our research projects. Regression learning examines the capability of developing the best linear regression model based upon the selected independent and dependent variables. Clustering analysis displays how data can be grouped by similar characteristics and how distant they are from one another. Classification analysis can predict future outcomes depending upon historical input data, a crucial tool in developing models for impending environmental events. It is suggested that environmental scientists who have not incorporated machine learning into their research to begin to add it to their data analyses.

**Keywords** Decision support · Cluster analysis · Neural networks · Regression · Machine learning

## Introduction

MATLAB is used to showcase a number of machine learning techniques that environmental researchers may consider incorporating into their research agenda. Regardless of a researcher's computing abilities, machine learning provides the ability to take in large sets of data and return models relatively quickly and easily. World Health Organization data was used in this example.

Using the variables female life expectancy, continent, access to clean water in rural areas, and particulate matter concentration, machine learning algorithms displayed the ability to conduct deeper analysis for decision support. This paper is intended to introduce the techniques to get the most information out of collected data rather than to propose quantitative analyses of the data. Table 1 presents key terms used hereafter.

The concept of machine learning has been around for several decades and its earliest definition is a component of

computer science in which computers are given the ability to learn without being programmed to do so (Samuel 1959). Machine learning has evolved drastically since first being described in literature, with abilities such as deep learning available to almost anyone with a personal computer. Deep learning can uncover intricate structures in data and dictates how a machine should change algorithms to compute a model (LeCun et al. 2015).

Machine learning can assist any environmental researcher with a number of practical tools, including adaptability, recovery from missing data, confounding control, and causal discovery through regression learning (Pearl 2019). Some example questions a research can apply machine learning to include

- Where are the most effective places to install rain gardens to maximize stormwater capture and natural treatment? (Younos et al. 2019)

A researcher can approach this question quite systematically. Using the rational method, where the peak discharge of stormwater is equal to the product of rainfall intensity, surface area and a runoff coefficient, areas in a community that prone to flooding can be quickly identified. Capturing the data of flooded sites, storm parameters, water depth, and time for complete drainage, machine

---

✉ David W. Nadler  
dnadler@nyit.edu

<sup>1</sup> New York Institute of Technology, Old Westbury, NY 11710, USA

**Table 1** Glossary of key terms

Term	Definition
Machine learning	A mathematical model for data that is used to make predictions dependent upon future data (Bishop 2016)
Support vector machine (SVM)	A machine learning algorithm that uses historical data to create regression models and classifications (“How Efficient Is Twitter: Predicting 2012 U.S. Presidential Elections Using Support Vector Machine via Twitter and Comparing against Iowa Electronic Markets” 2017)
Classification/confusion matrix	“A matrix that summarizes the performance of a classifier with respect to some test data” (Ting 2010)
Training	Application of a learning algorithm to a data set (Ting)

learning can provide infrastructure agencies develop effective strategies for installing rain gardens to mitigate this issue.

- What is the carbon footprint of a region where data availability is limited? (Habans et al. 2019)

Understanding a community’s waste management system allows for its carbon footprint to be determined using a tool such as the Environmental Protection Agency’s Waste Reduction Model (WARM). The foundation is based upon knowing the tonnage of every waste stream is generated (i.e., cardboard, yard waste, rubber) and its fate (i.e., landfill, incineration, composting). Some data is not readily available or recorded, making traditional modeling adequate but with a lower confidence of prediction. With machine learning, historical data such as population, waste stream rates, transportation and final destination can be used to “increase the probabilistic relationships recovered from incomplete data (Pearl)”.

- What are the most effective ways of addressing climate change impacts which meet social objectives? (Dulal 2019)

Climate change brings the increased risk of major events such as flooding, drought, and crop loss. One social object to be addressed is the malnourishment of children in areas where drought is more likely to occur. Under- and malnutrition are precursors to a child’s ability to grow, learn, and rise out of poverty (“The Social Dimensions of Climate Change” n.d.). Classification matrices can be used to help predict areas in greatest need for this sustainable develop goal to be addressed.

## Review of relevant literature

Incorporating environmental data into machine learning methods will enhance researchers’ abilities to generate stronger developed models. A significant application of this is the forecasting of environmental events, ranging from fires and floods to land and water quality. Geologists are using multivariate data analysis through machine learning techniques to

predict more accurately future occurrences of landslides from historical data and patterns (Korup and Stolle 2014). Landslide displacement prediction modeling has been performed through machine learning and causal factors (Zhou et al. 2018). Machine learning methods have identified two central landslide predictors, unfailed cliff slope angle, and fault proximity, which may support management approaches, that include all predictor information (Dickson and Perry 2016).

Environmental toxicity has been shown to be predictable using neural networks in cluster analysis. Research has shown that machine learning can find new knowledge that is encoded within a chemical’s molecular structure (Mayr et al. 2016). Applications such as this will allow for a decreased risk in chemical safety testing (“Prediction of human population responses to toxic compounds by a collaborative competition” *Nature Biotechnology* 2015).

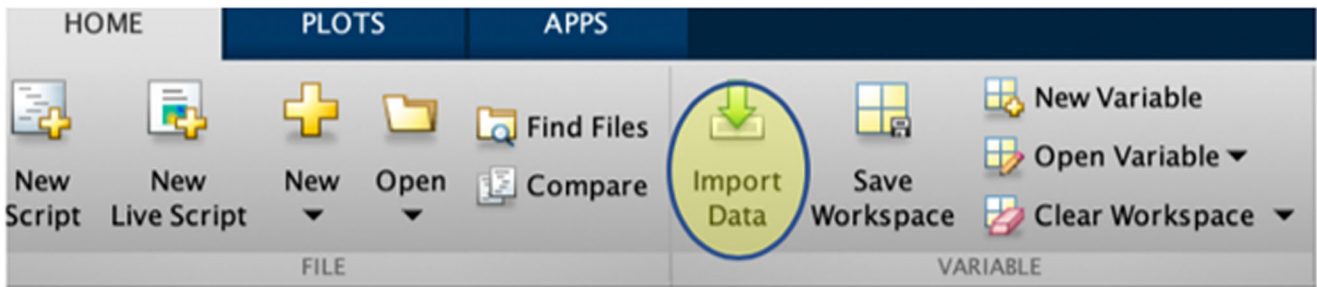
Rain gardens are being implemented in cities as part of a green infrastructure push to help manage the effects of stormwater runoff (Liao et al. 2017; Malaviya et al. 2019). In New York City’s Jamaica Bay and Tributaries zone, over 169 million gallons of stormwater have been managed over 135 equivalent green acres of this type of green infrastructure (“NYC Green Infrastructure 2018 Annual Report” 2019). Selecting locations is tedious work due to several variables including types of trees, size of sidewalks, size of drainage area, location of public transportation, and underground utilities. Machine learning has the potential of streamlining the process.

The methods for calculating carbon footprint emission from solid waste have a great deal of variance, resulting in a lower confidence of prediction (Robinson et al. 2018). In India, there is very “limited data available for waste generation patterns” in the largest urban areas (Joshi and Ahmed 2016). While machine learning cannot be directly used to curb a society’s waste generation, it can be used to predict waste models with limited data (Pearl).

Climate change will have an impact on social objectives, particularly equity in developing nations, in the near future. In Tanzania, social equity in rural area adaptation must need to be addressed (Smucker et al. 2015). Social objectives in

**Table 2** Machine learning techniques and associated null hypothesis

Technique	Null hypothesis
Regression Learner	Linear support vector machine will have the best-fit regression line from the data Female life expectancy is not influenced by environmental factors
Classification Learner	Machine learning will be able to classify a continent at a 60 % success rate by using three predictor variables



**Fig. 1** Toolbar location for data import

regions that cover several nations, such as the Brahmaputra River Basin in South Asia, can affect a population of tens of millions inhabitants (Yang et al. 2016). Machine learning can provide policymakers with better guidance on where to develop hydroelectric facilities, water supply basins, and farmland.

World Health Organization environmental health data has been the subject of machine learning in other studies. Ground-based particulate matter readings were trained in a machine learning algorithm to estimate average concentrations of the air pollutant (Lary et al. 2015). This technique was used to demonstrate its relationship with mental illness-related emergency room visits in Baltimore, Maryland. The outcome for particulate matter exposure presented in this paper is female life expectancy, so the researcher may use any relevant predicted variable. The National Institute of Environmental Health Sciences (NIEHS) has identified machine learning as a method for the analysis of multi-pollutant environmental epidemiology studies (Taylor Kyla et al. 2016).

## Methodology

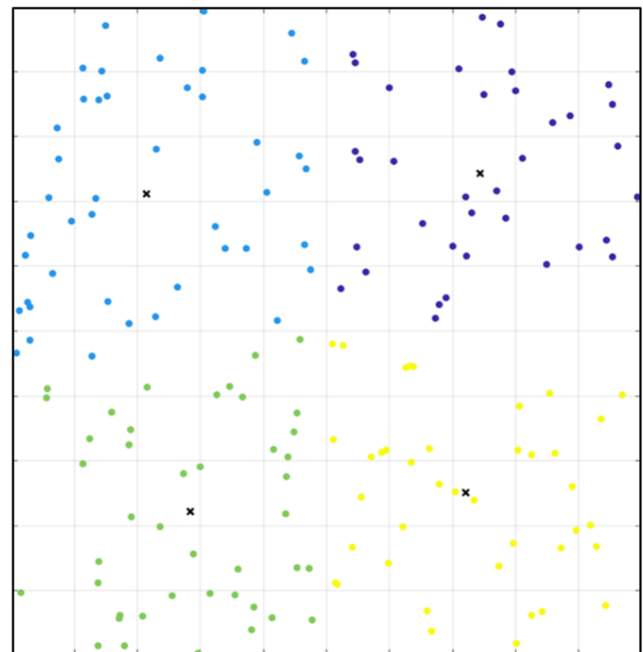
### Data setup

The data of 162 nations was extracted from the World Health Organization (WHO) “Global Health Observatory” data repository. The study population was composed from the data made publicly available by the WHO. Each country was documented as a record along with a number of other variables, including (1) average life expectancy, (2) average life expectancies for males and females, (3) percentage of population with access to improved potable water, (4) percentage of population with access to improved sanitation, (5) average exposure concentration of PM10, (6) number of deaths per one-

hundred thousand citizens, and (7) continent. Data was saved as a Microsoft Excel 2016 file.

### Example hypotheses

Table 2 provides examples of machine learning techniques and a null hypothesis to demonstrate the practicality of incorporating MATLAB into common research studies. The intent is for the reader to be able to apply some to his/her own research.



**Fig. 2** Clustered classification data

**Table 3** Model outcomes from regression learning

Model	RMSE	$R^2$	Mean absolute error
1.1 Linear SVM	4.0041	0.79	2.9338
1.2 Quadratic SVM	4.5865	0.72	3.3752
1.4 Fine Gaussian SVM	4.6869	0.71	3.3708
1.5 Medium Gaussian SVM	3.5632	0.83	2.7045
1.6 Coarse Gaussian SVM	4.2148	0.76	3.0343
2 Stepwise Linear	3.9346	0.79	2.9208

## MATLAB

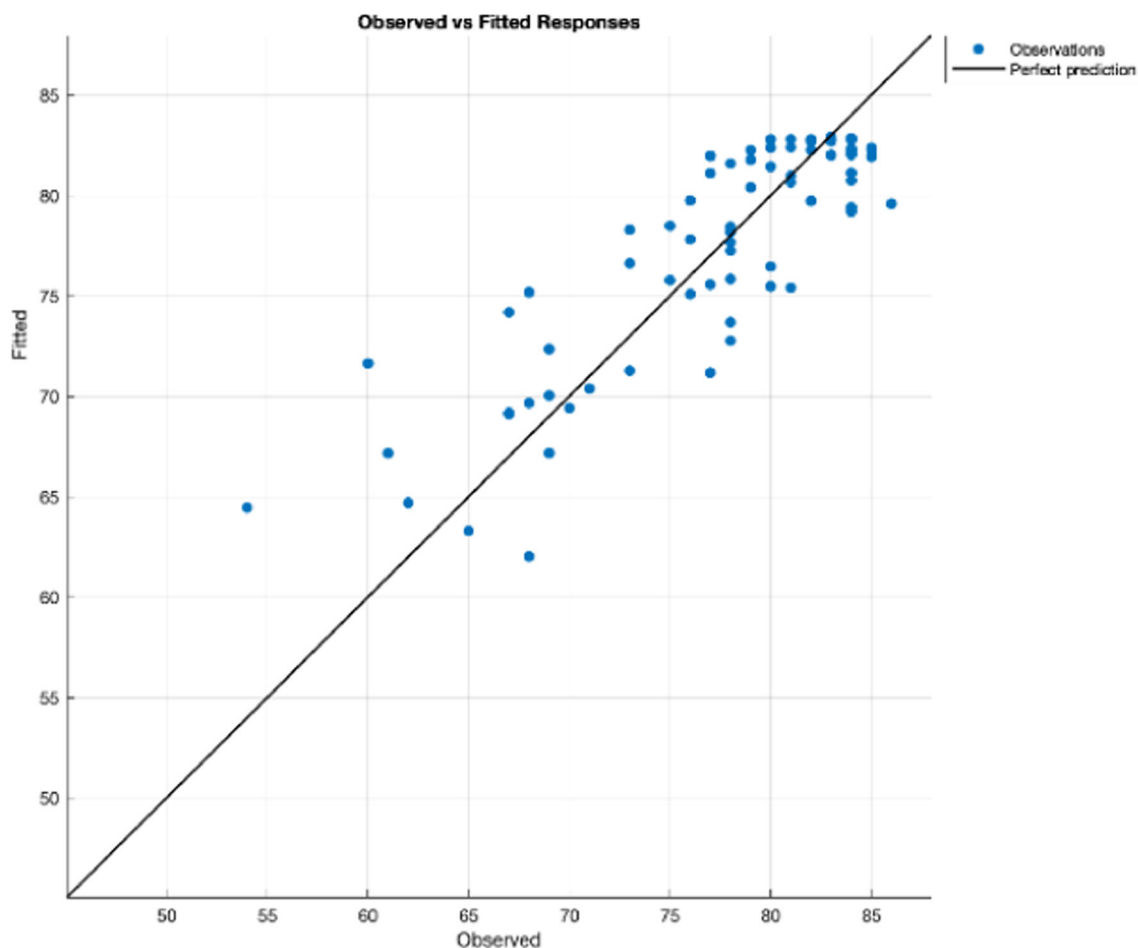
MATLAB, which stands for “Matrix Laboratory,” is a software title designed for scientists and engineers to conduct high-level research through function generation, model building, plotting data, application development, and more (Gilat 2017; Hahn and Valentine 2016). Users are able to add on a number of applications to conduct higher level mathematical analysis, including Machine Learning and Data Analysis.

Within MATLAB, the Import Data function is used to bring a data set into the session’s memory (Fig. 1).

## Regression Learner

MATLAB’s Regression Learner app is accessible from the APPS ribbon and is used to train regression models to predict data. It can also be directly called from the command Regression Learner. Using this app, you can survey your data, train models, and evaluate outcomes. Users have the ability to “perform automated training to search for the best regression model, including linear regression models, regression trees, Gaussian process regression models, support vector machines, and ensembles of regression trees” (MATLAB (version 2019a) 2019).

Once a data set is in MATLAB’s memory, with independent and dependent variables already decided upon, users can begin to conduct machine learning. Independent variables are used to train a model that generates predicted outcomes in the general form of

**Fig. 3** Observed versus fitted responses of female life expectancy

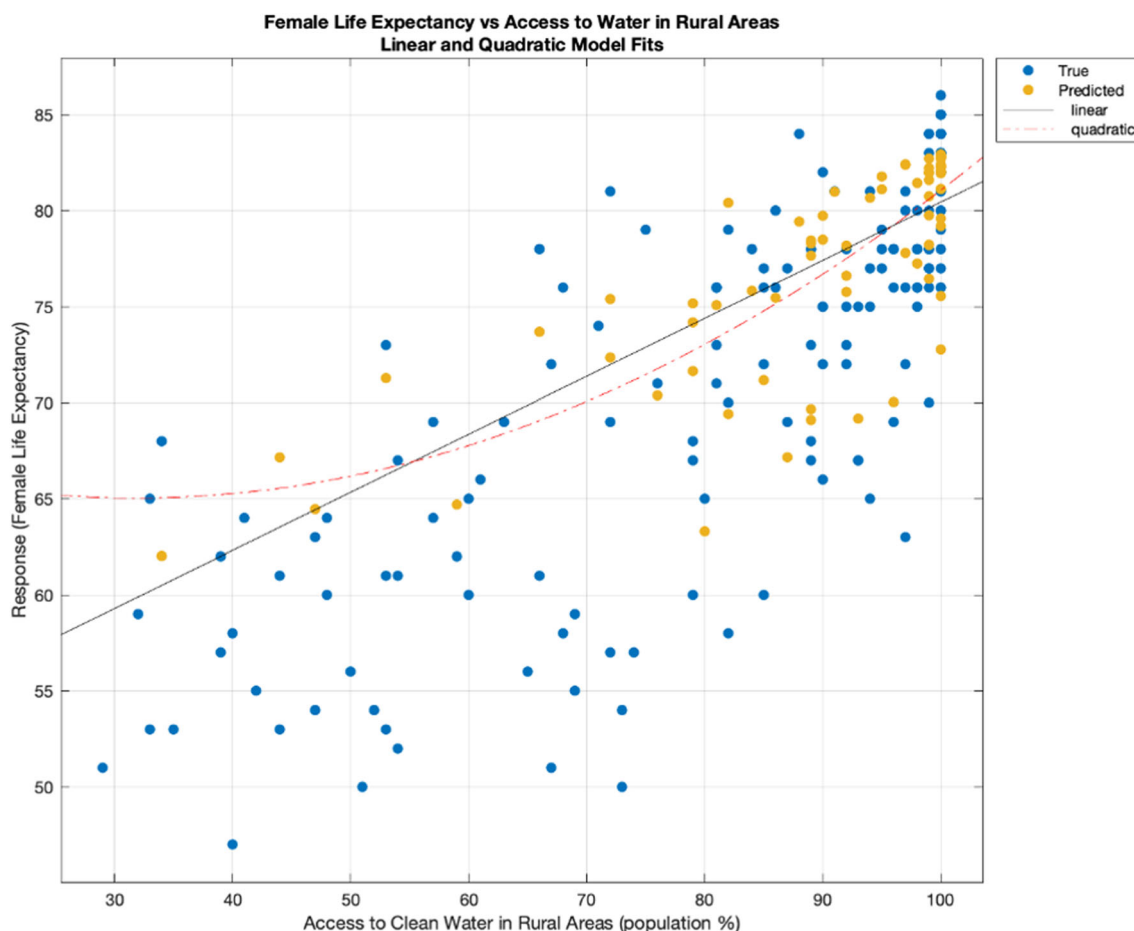


Fig. 4 Linear and quadratic models for female life expectancy as predicted by access to water in rural areas

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n,$$

where

- $\hat{y}$  is the predicted outcome,
- $\beta_0$  is the y-intercept of the best-fit regression line,
- $\beta$  coefficients represent the change in outcome relative to its  $x$  value, and
- $x$  variables represent distinct predictor values (Johnson and Bhattacharyya 2018).

Data can be imported later into the Regression Learner for predictions and, very importantly, continued model learning. As more data is added to the model, its predictions will become more accurate and reliable.

### Classification Learner

The Classification Learner app is accessible through the APPS ribbon and is used to visualize how well variables can group together based on commonalities. Machine learning allows for the clustering of data which, in essence, divides all samples in a database to be grouped together (Alpaydin 2014). Software algorithms find densely populated areas of data points to find

commonalities (Daniel 2013). Figure 2 displays a hypothetical data set gathered around a cluster center, represented by the four “x” symbols.

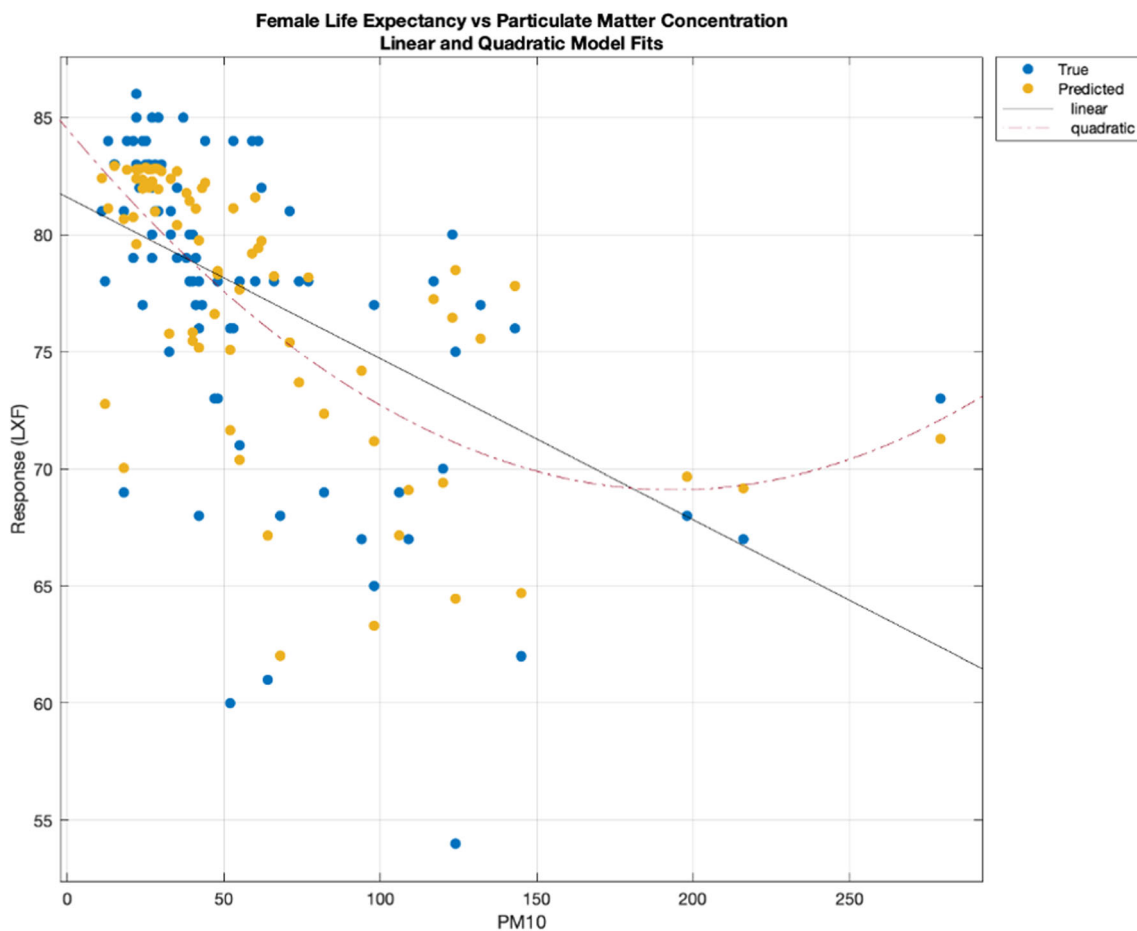
Some examples questions that classification can help answer include

- How do water quality parameters influence algae blooms?
- Are there clusters of environmentally-related illnesses in a region where particulate matter fluctuates?
- To what scale do small generators of toxic substances impact the environmental justice of a community? (Collins et al. 2016)
- What zip codes present the greatest impact on stormwater runoff?

### Results

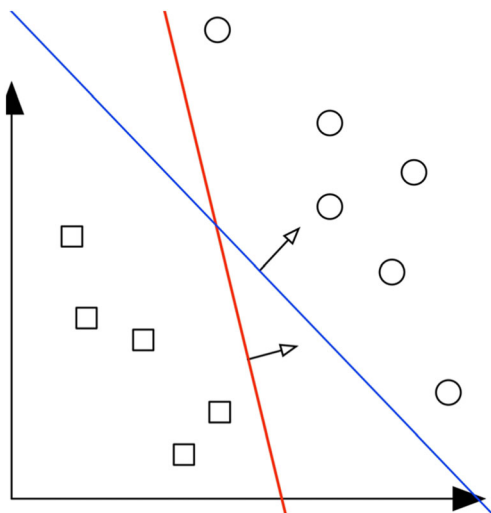
#### Regression Learner outcomes

Eight independent variables were used in this machine learning session for the prediction of the outcome



**Fig. 5** Linear and quadratic models for female life expectancy as predicted by particulate matter concentration

variable, “Female Life Expectancy”. These variables were “continent,” “PM10,” “Rural Water,” “Urban Water,” “Total Water,” “Rural Sanitation,” “Urban Sanitation,” and “Total Sanitation.” The hypothesis of “female life expectancy is not influenced by environmental factors” was tested in this technique.



**Fig. 6** Linear SVM separating data through a best-fitting line

The chosen model picked from machine learning was the Medium Gaussian Support Vector Machine and is shown along with a number of other model outcomes in Table 3. The coefficient of multiple determination ( $R^2$ ) was highest for linear regression Model 1.5 (0.83). This implies that 83% of female life expectancy variability can be explained by this linear model. The plot of this model is represented in Fig. 3.

As this model is studied in more details, the machine learning computes that only three input variables are required to predict female life expectancy (LXF): continent, access to water in rural areas, and particulate matter concentration.

The linear model may be explained as:

$$\hat{y} = \beta_0 + \beta_1 \text{Continent} + \beta_2 \text{Access to Water in Rural Areas} + \beta_3 \text{PM10}$$

Generated unadjusted (crude) linear and quadratic models are presented in Figs. 4 and 5. For Fig. 4, the linear model may be explained as:

$$\hat{y} = \beta_0 + \beta_1 \text{Access to Water in Rural Areas,}$$

while Fig. 7 may be explained by

$$\hat{y} = \beta_0 + \beta_1 PM_{10}$$

Clearly, the more of the population that has access to clean water, the average life expectancy increases (Fig. 4 positive slope) while as the concentration of particulate matter increases, the average life expectancy decreases (Fig. 5 negative slope).

Linear SVM was the postulated model for this technique but the machine learning has provided a stronger one in a Medium Gaussian SVM model. Linear SVM is capable of separating data with a straight-line vector while the Gaussian SVM method uses an additional dimension. The differences between the two SVMs are presented in Figs. 6 and 7.

### Classification Learner outcomes

Additional testing was conducted through machine learning to evaluate how well a continent can be classified through the data provided within the female life expectancy, access to water in rural areas, and particulate matter concentration variables. Figure 8 presents the confusion matrix table for the classification through machine learning. Six continents/regions (AF = Africa; AS = Asia; EU = Europe, NA = North American; OC=Oceania; and SA = South America) are represented as true classes in rows and predicted classes in columns.

This output shows that there is an 80% positive rate in the data that the female is from Africa by knowing the values of the three predictor variables, water, sanitation and particulate matter concentration. Twenty percent constitutes false predictions. Classifying other continents based on the input factors is less reliable for this model. Seven percent of the data classified as Africa was falsely identified as Asia.

The hypothesis for this technique was “machine learning will be able to classify a continent at a sixty percent

success rate by using three predictor variables.” Data from African (80%) and European (72%) nations supported this hypothesis but the machine did not learn for the other continents. Oceania can be explained by having fewer nations with data for the machine to learn robustly. Data similarities between North America, South America, and Asia likely resulted in collinearity during regression learning.

Classification was conducted using the percentages of access (1) to clean drinking water and (2) to improved sanitation to find commonalities between continents. Figure 9 shows how the data results in clustering among African nations as opposed to the greater variability of European and North American countries.

### Conclusions

The main goals were to demonstrate readily available machine learning modeling techniques using historical environmental health data (Feldman et al. 2016). This was not an attempt to propose a fully functioning model for predicting female life expectancy dependent upon environmental infrastructures and pollution exposures. Providing researchers with an introduction to how they may use machine learning in their own research is the priority of this research brief.

MATLAB is a powerful software title that is often provided to research faculty free of charge through their institution’s academic license. Several add-ons enable research projects to be enhanced while having data evaluated quickly and accurately. A few basics of machine learning have been presented which can be easily incorporated into a researcher’s toolbox with some practice. World Health Organization environmental health data was used to demonstrate some techniques.

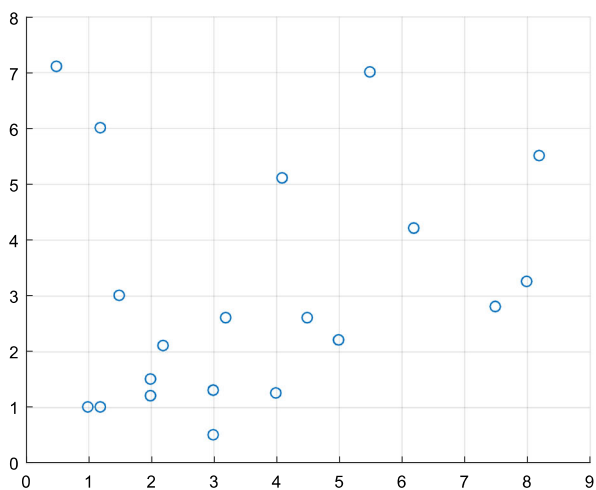
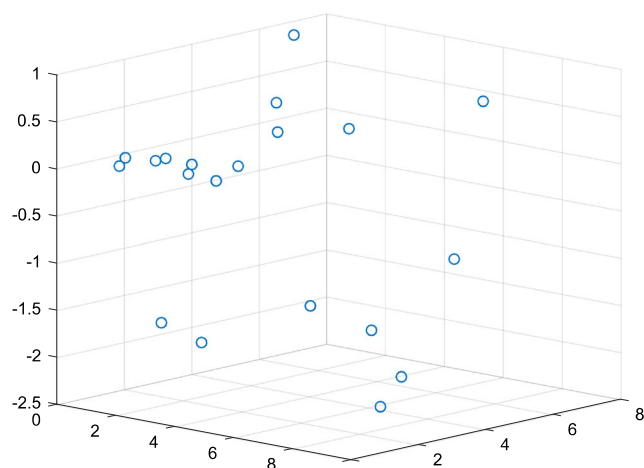


Fig. 7 Gaussian SVM separating data in an extra dimension



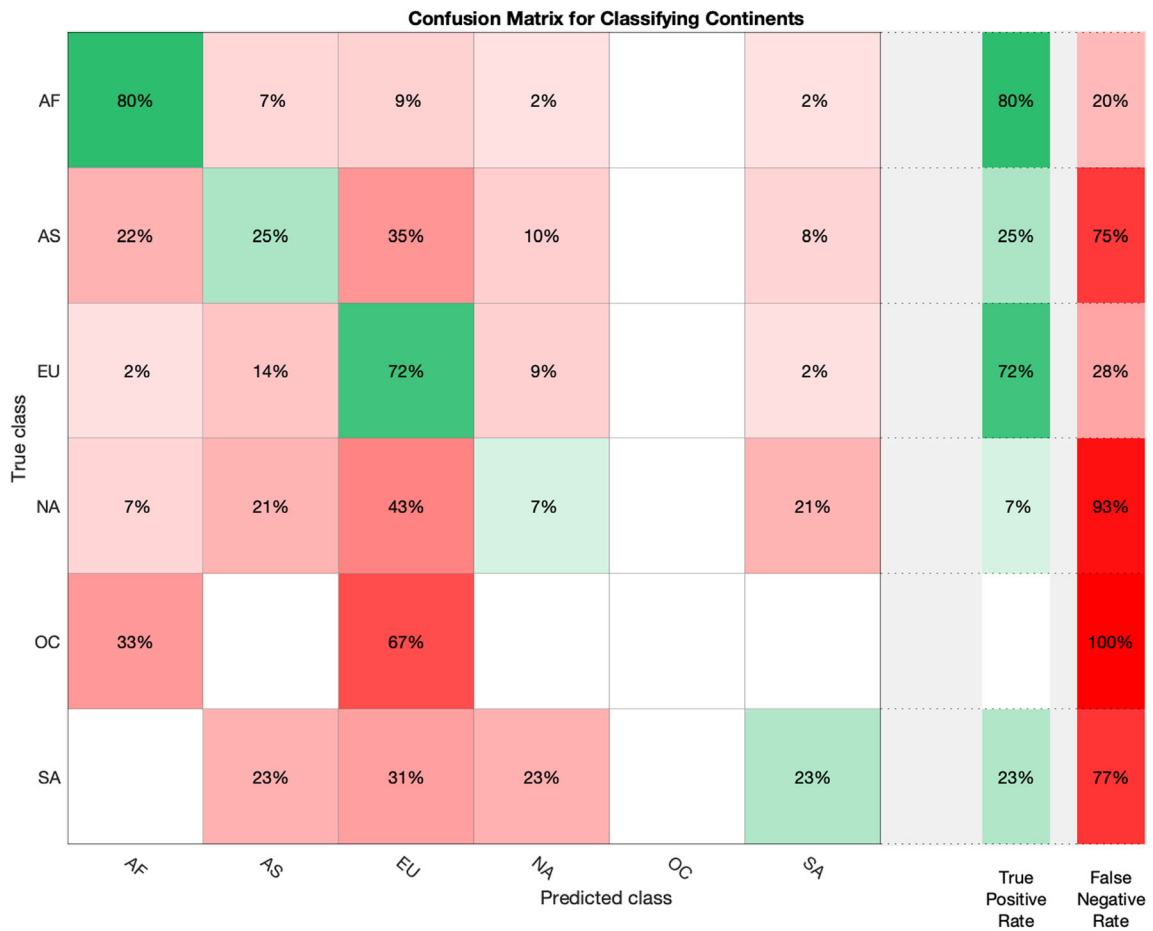
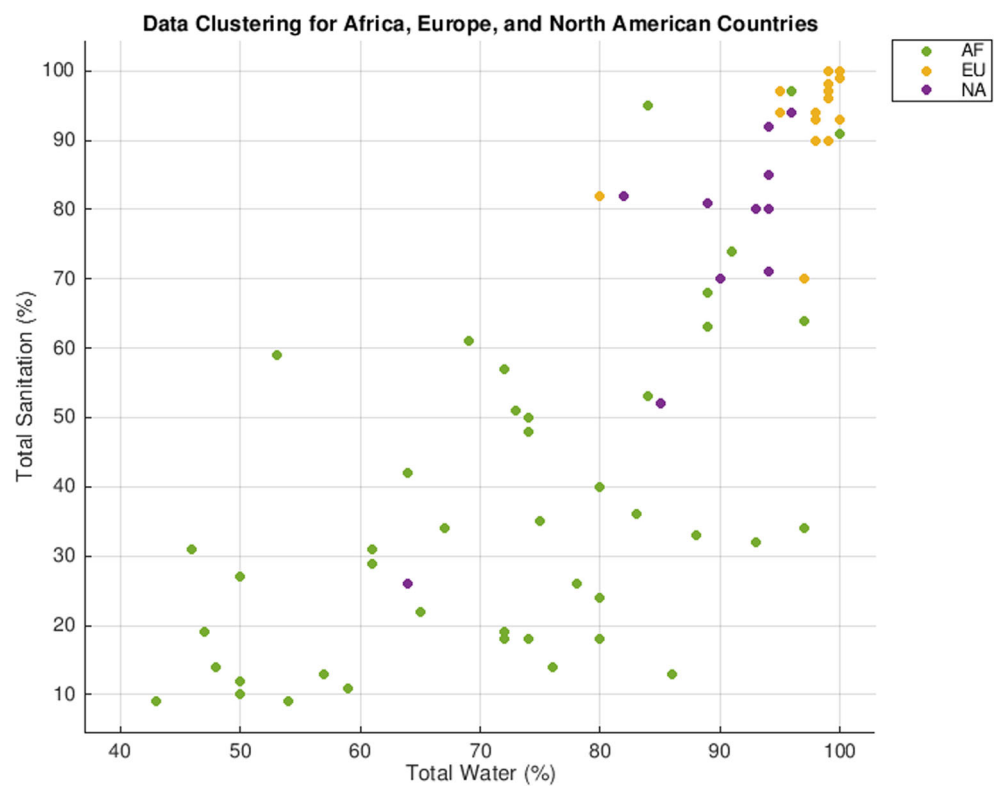


Fig. 8 Confusion matrix for the classification data by continent

Fig. 9 Classification clustering between Africa, Europe, and North America





Regression learning displayed the ease of importing a dataset into the MATLAB memory and then defining which variables are chosen as predictors and which will be the predicted. Support Vector Machine learning allowed for the running of a number of linear models of the selected data and quickly evaluated the best one.

Classification of data was demonstrated through the confusion matrix. Through the analysis of a number of variables, the algorithm can learn what outcomes may be in future analyses. This case showed that there was an 80% chance of predicting a female in Africa depending upon the predictor variables used in this example. Clustering analysis was used to demonstrate how the software can group data points together by similar characteristics into clusters and then show relative distances between all identified clusters.

Example research questions were provided in which machine learning can be applied for deeper understanding of issues. Clustered classification can be used to determine locations for the installation of rain gardens to help mitigate flooding. Carefully chosen variables, such as latitude/longitude, may provide the ability to overlay graphed clusters on a map to identify where rain gardens would be most efficient. Regression learning would be a practical technique to model the carbon footprint of a community where there is incomplete municipal waste data. The regression line will become more significant as more data is put into a database for the model to learn. The confusion matrix can be used to predict if an underserved community is experiencing a lack of equity based upon the variables incorporated into the machine learning environment.

Environmental science is experiencing an increase in the number of studies using machine learning (Best et al. 2018; Fuchs et al. 2019; Feldman et al. 2016; Gómez-Losada et al. 2018; Griffin et al. 2019; Sung Kyun et al. 2017; Hoos et al. 2019). Machine learning will account for uncertainties in environmental science studies, allowing for an increase in credibility of future results (Strith et al. 2019). Incorporating some of these tools and techniques can enhance our decision-making and support for our planet's needs (Hill et al. 2019; Tasdighi et al. 2018).

## References

- Alpaydin E (2014) Introduction to machine learning. MIT Press, Cambridge MA
- Best ÜSN, Van der Wegen M, Dijkstra J, Willemsen PWJM, Borsje BW, Roelvink DJA (2018) Do salt marshes survive sea level rise? Modelling wave action, morphodynamics and vegetation dynamics. *Environ Model Softw* 109(November):152–166. <https://doi.org/10.1016/j.envsoft.2018.08.004>
- Bishop CM (2016) Pattern recognition and machine learning. Springer, New York
- Collins MB, Munoz I, JaJa J (2016) Linking ‘toxic outliers’ to environmental justice communities. *Environ Res Lett* 11(1):015004. <https://doi.org/10.1088/1748-9326/11/1/015004>
- Daniel G (2013) Principles of artificial neural networks, 3rd edn. World Scientific, Singapore
- Dickson ME, Perry GLW (2016) Identifying the controls on coastal cliff landslides using machine-learning approaches. *Environ Model Softw* 76(February):117–127. <https://doi.org/10.1016/j.envsoft.2015.10.029>
- Dulal HB (2019) Cities in Asia: how are they adapting to climate change? *J Environ Stud Sci* 9(1):13–24. <https://doi.org/10.1007/s13412-018-0534-1>
- Feldman D, Contreras S, Karlin B, Basolo V, Matthew R, Sanders B, Houston D et al (2016) Communicating flood risk: looking back and forward at traditional and social media outlets. *International Journal of Disaster Risk Reduction* 15(March):43–51. <https://doi.org/10.1016/j.ijdrr.2015.12.004>
- Fuchs S, Heiser M, Schlögl M, Zischg A, Papathoma-Köhle M, Keiler M (2019) Short communication: a model to predict flood loss in mountain areas. *Environ Model Softw* 117:176–180. <https://doi.org/10.1016/j.envsoft.2019.03.026>
- Gilat A (2017) MATLAB: an introduction with applications. John Wiley & Sons, Incorporated, Hoboken
- Gómez-Losada Á, Pires JCM, Pino-Mejías R (2018) Modelling background air pollution exposure in urban environments: implications for epidemiological research. *Environ Model Softw*, Special issue on environmental data science. Applications to air quality and water cycle 106(August):13–21. <https://doi.org/10.1016/j.envsoft.2018.02.011>
- Griffin LP, Griffin CR, Finn JT, Prescott RL, Faherty M, Still BM, Danylchuk AJ (2019) Warming seas increase cold-stunning events for Kemp's Ridley Sea turtles in the Northwest Atlantic. *PLoS One* 1
- Habans R, Clement MT, Pattison A (2019) Carbon emissions and climate policy support by local governments in California: a qualitative comparative analysis at the county level. *J Environ Stud Sci*. <https://doi.org/10.1007/s13412-019-00544-1>
- Hahn B, Valentine D (2016) Essential MATLAB for engineers and scientists. Academic Press, Cambridge MA
- Hill G, Kolmes S, Humphreys M, McLain R, Jones ET (2019) Using decision support tools in multistakeholder environmental planning: restorative justice and subbasin planning in the Columbia River basin. *J Environ Stud Sci* 9:170–186. <https://doi.org/10.1007/s13412-019-00548-x>
- Hoos AB, Wang SH, Schwarz GE (2019) Adapting a regional water-quality model for local application: a case study for Tennessee, USA. *Environ Model Softw* 115(May):187–199. <https://doi.org/10.1016/j.envsoft.2019.01.001>
- How efficient is twitter: predicting 2012 U.S. presidential elections using support vector machine via twitter and comparing against Iowa electronic markets. 2017. 2017 Intelligent Systems Conference (IntelliSys), Intelligent Systems Conference (IntelliSys), 2017, 646. <https://doi.org/10.1109/IntelliSys.2017.8324363>
- Johnson RA, Bhattacharyya GK (2018) Statistics: principles and methods. John Wiley & Sons, Hoboken
- Joshi R, Ahmed S (2016) Status and challenges of municipal solid waste Management in India: a review. Edited by Carla Aparecida Ng. *Cogent Environ Sci* 2(1):1139434. <https://doi.org/10.1080/23311843.2016.1139434>
- Korup O, Stolle A (2014) Landslide prediction from machine learning. *Geol Today* 30(1):26–33. <https://doi.org/10.1111/gto.12034>
- Lary DJ, Lary T, Sattler B (2015) Using machine learning to estimate global PM2.5 for environmental health studies. *Environ Health Insights* 9s1(January):EHI.S15664. <https://doi.org/10.4137/EHI.S15664>

- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553): 436–444. <https://doi.org/10.1038/nature14539>
- Liao, K-H, Deng S, Tan PY (2017) Blue-green infrastructure: new frontier for sustainable urban stormwater management. In: Puay Yok Tan and Chi Yung Jim (eds) *Greening cities: forms and functions*. *Advances in 21st Century Human Settlements*. Springer Singapore, Singapore, pp 203–26. [https://doi.org/10.1007/978-981-10-4113-6\\_10](https://doi.org/10.1007/978-981-10-4113-6_10)
- Malaviya P, Sharma R, Sharma PK (2019) Rain gardens as Stormwater management tool. In: Shachi S, Venkatramanan V, Prasad R (eds) *Sustainable green technologies for environmental management*. Springer Singapore, Singapore, pp 141–166. [https://doi.org/10.1007/978-981-13-2772-8\\_7](https://doi.org/10.1007/978-981-13-2772-8_7)
- MATLAB (version 2019a). 2019. Mathworks. <https://www.mathworks.com/help/stats/regressionlearner-app.html>
- Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. *Frontiers in Environmental Science* 3. <https://doi.org/10.3389/fenvs.2015.00080>
- NYC green infrastructure 2018 annual report. 2019. <https://www1.nyc.gov/assets/dep/downloads/pdf/water/stormwater/green-infrastructure/gi-annual-report-2018.pdf>
- Pearl J (2019) The seven tools of causal inference, with reflections on machine learning. *Commun ACM* 62(3):54–60. <https://doi.org/10.1145/3241036>
- Prediction of human population responses to toxic compounds by a collaborative competition | *Nature Biotechnology*. 2015. 2015. <https://www.nature.com/articles/nbt.3299>
- Robinson OJ, Tewkesbury A, Kemp S, Williams ID (2018) Towards a universal carbon footprint standard: a case study of carbon management at universities. *J Clean Prod* 172(January):4435–4455. <https://doi.org/10.1016/j.jclepro.2017.02.147>
- Samuel AL (1959) Some studies in machine learning using the game of checkers. *IBM J Res Dev*:71–105
- Smucker TA, Wisner B, Mascarenhas A, Munishi P, Wangui EE, Sinha G, Weiner D, Bwenge C, Lovell E (2015) Differentiated livelihoods, local institutions, and the adaptation imperative: assessing climate change adaptation policy in Tanzania. *Geoforum* 59(February):39–50. <https://doi.org/10.1016/j.geoforum.2014.11.018>
- Stritih A, Bebi P, Grêt-Regamey A (2019) Quantifying uncertainties in earth observation-based ecosystem service assessments. *Environ Model Softw* 111(January):300–310. <https://doi.org/10.1016/j.envsoft.2018.09.005>
- Sung Kyun Park, Zhao Z, Mukherjee B, Park SK, Zhao Z (2017) Construction of environmental risk score beyond standard linear models using machine learning methods: application to metal mixtures, oxidative stress and cardiovascular disease in NHANES. *Environ Health Glob Access Sci Source* 16(September):1–17. <https://doi.org/10.1186/s12940-017-0310-9>
- Tasdighi A, Arabi M, Harmel D, Line D (2018) A Bayesian Total uncertainty analysis framework for assessment of management practices using watershed models. *Environ Model Softw* 108(October):240–252. <https://doi.org/10.1016/j.envsoft.2018.08.006>
- Taylor Kyla W, Joubert Bonnie R, Braun Joe M, Caroline D, Chris G, Russ H, Heindel Jerry J, Rider Cynthia V, Webster Thomas F, Carlin Danielle J (2016) Statistical approaches for assessing health effects of environmental chemical mixtures in epidemiology: lessons from an innovative workshop. *Environ Health Perspect* 124(12):A227–A229. <https://doi.org/10.1289/EHP547>
- The Social Dimensions of Climate Change (n.d.) In: World Health Organization. Accessed June 4, 2019. <https://www.who.int/globalchange/mediacentre/events/2011/social-dimensions-of-climate-change.pdf>
- Ting KM (2010) Confusion matrix. In: Sammut C, Webb GI (eds) *Encyclopedia of machine learning*. Springer US, Boston, MA, pp 209–209. [https://doi.org/10.1007/978-0-387-30164-8\\_157](https://doi.org/10.1007/978-0-387-30164-8_157)
- Yang Y, Ethan C, Wi S, Ray PA, Brown CM, Khalil AF (2016) The future Nexus of the Brahmaputra River basin: climate, water, energy and food trajectories. *Glob Environ Chang* 37(March):16–30. <https://doi.org/10.1016/j.gloenvcha.2016.01.002>
- Younos T, Lee J, Parece T (2019) Twenty-first century urban water management: the imperative for holistic and cross-disciplinary approach. *J Environ Stud Sci* 9(1):90–95. <https://doi.org/10.1007/s13412-018-0524-3>
- Zhou C, Yin K, Cao Y, Intriери E, Ahmed B, Catani F (2018) Displacement prediction of step-like landslide by applying a novel kernel extreme learning machine method. *Landslides* 15(11):2211–2225. <https://doi.org/10.1007/s10346-018-1022-0>