A case study comparing machine learning and vegetation indices for assessing corn nitrogen status in an agricultural field in Minnesota.

Aicam Laacouri, Tyler Nigon, David Mulla and Ce Yang
Department of Soil, Water, and Climate, University of Minnesota, St. Paul MN

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018 Montreal, Quebec, Canada

Abstract*: Compact hyperspectral sensors compatible with UAV platforms are becoming more readily available. These sensors provide reflectance in narrow spectral bands while covering a wide range of the electromagnetic spectrum. However, because of the narrow spectral bands and wide spectral range, hyperspectral data analysis can benefit greatly from data mining and machine learning techniques to leverage its power. In this study, rainfed corn was grown during the 2017 growing season using four nitrogen treatments (between 0 and 200 kg N/ha using 67 kg/ha increment steps) where 200 kg/ha represented the economically optimum nitrogen rate (EONR). This design generated four ordinal classes of N deficiency (dN) that were confirmed using chlorophyll readings (SPAD), leaf nitrogen content, and soil nitrate all collected at V5 corn growth stage, on the same day the hyperspectral data was collected. Hyperspectral images were collected using a line scanner sensor on board a hexacopter UAV platform. Following radiometric correction, image segmentation, spectral extraction and preprocessing, eight machine learning algorithms were compared for their accuracy in determining the four classes of nitrogen deficiency levels based on the DEONR (dN =dEONR). These algorithms are the logistic regression (LR), support vector machine (SVM), random forest (RF), gradient boosting (GB), naïve Bayes (NB), decision tree (DT) and Multi-Layer Perceptron (MLP). A confusion matrix based on a 30% test set (unseen by the models) was used to determine the performance of these classifiers and to illustrate the type and magnitude of errors for each classification method. Narrow-bands and simulated broad-band vegetation indices (VIs) were also derived from the hyperspectral data and compared to machine learning algorithms for the prediction of dN. Our findings confirm the challenge of using VIs to assess early season (V5) corn nitrogen status. VIs correctly estimated N stress on only 50% of plant samples, and caused nearly 30% of plants to be under-fertilized. On the other hand, hyperspectral machine learning significantly improved the assessment of corn nitrogen stress at V5, and achieved more than 90% classification accuracy. Particularly, SVM, LR, MLP and GB showed promising results. These findings illustrate the huge potential for UAV-compatible hyperspectral sensors to improve in-season corn nitrogen management for timely variable rate sidedress application.*

The authors are solely responsible for the content of this paper, which is not a refereed publication. This paper should be cited as: Laacouri A., Nigon T., Mulla D. and Yang C. 2018. *A case study comparing machine learning and vegetation indices for* *assessing corn nitrogen status in an agricultural field in Minnesota*. In Proceedings of the 14th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

Introduction

Nitrogen management for Midwestern row crops continues to pose a challenge for society because of societal, environmental, agronomic and economic ramifications of mismanagement (Keeler et al., 2016). Crops under-fertilized with nitrogen (N) produce lower yields and are undesirable to growers, while fields that are over-fertilized can cause offsite pollution of surface and ground water. Current best management practices for nitrogen management recommend an in-season sidedress application using crop sensing to improve N efficiency, a task usually undertaken around V5-V6 corn growth stage because of farm logistics. In-season canopy sensing has been traditionally conducted using broadband vegetation indices (VIs) such as the ubiquitous Normalized Difference Vegetative Index (NDVI). However, these broadband VIs are only weakly correlated to corn N status at early growth stages. The weak relationship between VIs and corn N status at early corn growth stages has been attributed to the crop low early demand for nitrogen (Abendroth *et al.*, 2011), and poses a hurdle for accurate sidedress rate computation. Recent advances in sensor development and UAVs offers a unique opportunity for the timely capture of crop spectral signatures in the entire visible and infrared light spectrum, thus potentially detecting early nitrogen deficiencies missed by using broadband VIs. However, because of the high spectral and spatial dimensions, hyperspectral data require data mining techniques to extract key features that provide information about N stress. The objective of this study is to compare the accuracy of corn N status detection based on broadband VIs to detection based on machine learning (ML) classification algorithms using UAV-based hyperspectral sensing.

Material and Methods

This study was conducted in the 2017 growing season on a portion of a 40-acre field owned by the University of Minnesota that was used for a larger water quality study. The field is characterized by a clay-loam soil texture (major soil types at the site include Webster silty clay loam (fine-loamy, mixed, superactive, mesic Typic Endoaquolls) and Nicollet clay loam (fine-loamy, mixed, superactive, mesic Aquic Hapludolls)) and relatively flat topography. Soil samples were taken before planting and at sidedress time to confirm soil nitrogen level and crop stress. Corn was planted on May 12 and V5 growth stage was attained on June 15. Four rates of nitrogen (0, 67, 134 and 200 kg/ha) were applied in a two-block experimental design. Each plot was 30 m x 14 m in size and crop row spacing was 0.76 m, so there were 18 rows per plot. The highest rate of 200 kg/ha represents the University of Minnesota's economically optimum nitrogen rate (EONR) for corn following corn. Nitrogen rates were applied one day before planting using a toolbar spreader. These plots were used as part of a larger study to develop nitrogen sidedress rates for the entire field. When corn reached growth stage V5, SPAD measurement and tissue samples were collected from each plot and used to classify the plots into four nitrogen deficit (dN) classes that would require different rates of sidedress (Table 1). With this approach, plots that received zero preplant nitrogen are the most deficient and are labeled Class 1, followed by Class 2 which contains plants from plots that received 67 kg/ha preplant, and Class 3 with plants that received 134 kg/ha. The fourth class is represented by plots receiving 200 kg/ha N. These treatments resulted in four different nitrogen levels in corn that were confirmed with SPAD measurements, soil N tests and leaf tissue N analysis all measured at V5 corn growth stage (Table1).

Table 1: Description of the four classes of corn nitrogen stress based on SPAD, Soil N and tissue N

Ordinal classes	N treatment	dN (200- N trt)	SPAD (V5)	Tissue (V5)	Soil N (V5)
Class 1 (very deficient)	0 kg/ha	200	34	2.6 % N	2.94 ppm N
Class 2	67	134	39.1	3	4.89
Class 3	134	67	46.1	3.5	49.1
Class 4 (EONR)	200	0	49.3	3.9	15.3

Image acquisition: Hyperspectral aerial images (2.1 nm spectral resolution ranging from 395 nm to 885 nm) were captured with a gimbal-stabilized Pika II line-scanning hyperspectral camera (Resonon, Inc.; Bozeman, MT) mounted on an unmanned hexacopter (DJI Matrice 600 Pro, Nanshan District, Shenzhen, China). DJI Ground Station Pro (iPad app) was used to create and execute flight plans for controlling altitude, heading, and ground speed. The camera framerate was set to 99 frames per second, and the hexacopter passed over the target area at a speed of 2.5 mps and an altitude of 30 m, resulting in an image pixel size of approximately 2.5 cm. Gray reference panels with known reflective properties were placed in the study area prior to image capture; panels were 60 x 60 cm and the surface was 50% BaSO4/50% grey paint by weight.

Image processing: Radiometric correction was performed via SpectrononPro software (Resonon, Inc.; Bozeman, MT) using a calibration file provided by Resonon for the specific camera and lens that were used. Grey reference panels that were placed in the study area prior to image capture were used to convert spectral radiance to surface reflectance across all images. Soil background was removed using a supervised classification leaving only pure vegetation pixels. Only spectral data from the pure vegetation pixels in each plot were extracted and converted to tabular data for analysis.

Vegetation indices: The spectral data were collected at V5 corn grown stage. The conventional VIs (broad-band) were computed from the hyperspectral data by averaging over the corresponding VI spectral reflectance (30 nm bandwidth). These indices included the NDVI, Green Relative Vegetative Index (GRVI), Normalized Difference Red Edge (NDRE) and Green Normalized Difference Vegetative Index (GNDVI). Additionally, a narrow-band version of these VIs was also computed using the central hyperspectral narrow-bands of the corresponding VI. These narrow-bands VIs are termed $NDVI_H$, $GRVI_H$, $NDRE_H$ and $GNDVI_H$. Each of these indices was regressed on dN (dN= 200- Applied N) to establish the relationship between each VI and N status and thus was used to predict the N stress class. We used 70% of measured data to develop the regression equations and we proceeded toward estimating the stress level (class) with the remaining 30% of the data to assess the fit of each VI for corn N status quantification. When the prescribed N stress level (dN) is overestimated, we labeled it "over-fertilized" and when it is under-estimated we labeled it "under-fertilized".

Machine learning algorithms: Eight ML classification algorithms were compared for their performance in classifying corn plants into the four classes of N deficiency. These algorithms are decision tree (DT), random forest (RF), gradient boosting trees (GB), K nearest neighbor (KNN), naïve Bayes (NB), support vector machine (SVM), and multi-layer perceptron (MLP). Mueller and Guido (2016) and Brieman *et al.* (1984) provide more details about these algorithms and parameter selection for each. We fitted these algorithms on the high dimension hyperspectral data (240 bands) both with and without feature extraction. Feature extraction was evaluated using RF to rank the 15 most important bands. RF can guide in the identification of important bands by identifying those used as nodes (or split). This method can therefore be used to identify bands for designing multispectral sensors. In a second feature extraction technique we examined the spectral signature of all the treatments and selected 8 bands that best separated the treatments. We also evaluated principal component analysis (PCA) as a dimension reduction technique. For validation, data were split into 70% training set and 30% test set. We opted for this technique of validation in lieu of the popular cross-validation because we wanted to evaluate the confusion matrix for the mis-labeled pixels to assess the over- and under-fertilization potential. Using the cross-validation method will produce a confusion matrix for each run instead of one confusion matrix for each classification technique. Overfitting, a serious issue in ML, was evaluated by comparing the accuracy on the test set to the training set. A perfect classification of the training set is highly indicative of overfitting. Scikit-learn (Pedregosa *et al.* (2011)) and python were used for ML implementation following the guidance of Mueller and Guido (2016) for parameter optimization.

Results and Discussion

Fig. 1 shows a λ-λ plot illustrating the strength of the relationship between pairs of the 240 bands collected by the Resonon hyperspectral line scanner. The plot illustrates the strong relationship between bands in the NIR range which could potentially be used to reduce bands from the data. However, this multicollinearity does not impact the prediction power of most ML algorithms (De Veau and Ungar, 1994). Also noticeable in Fig. 1 is the weak relationship between the red (660 - 698 nm) and NIR region of the spectrum, and between the violet (395 - 400 nm) and the rest of the spectrum. Weak relationships between bands are desirable because the information in each band is complementary and leads to better diagnosis of N stress.

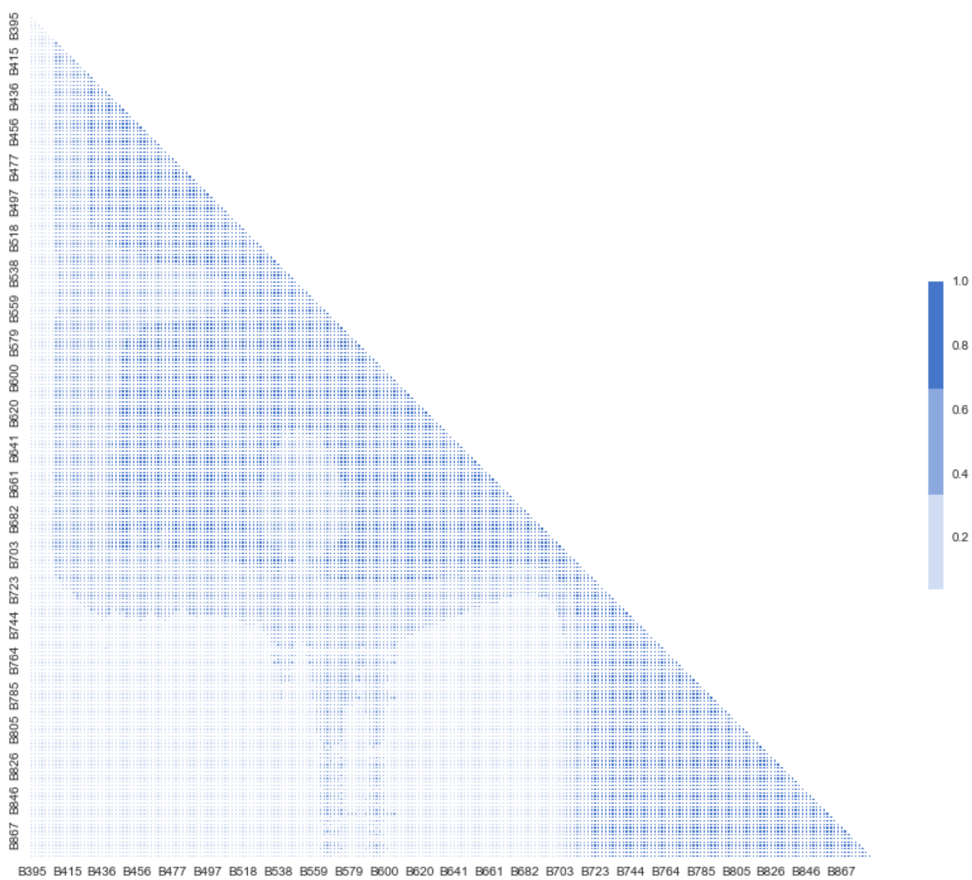

Fig. 1: λ-λ correlation plot between the different bands with darker blue indicating stronger correlation.

The spectral signature for each of the four classes of nitrogen stress is presented in Fig. 2. According to the figure, the visible light spectrum (420-700 nm) is only capable of separating the very deficient plants from the remaining three classes. On the other hand, the NIR region can separate all four classes of stress. Interesting also is the violet light (395-400 nm) where signatures of the four classes can be distinguished. Results of the regression of dN against the different VIs was best described with a quadratic model and the R squared varied between 0.3 ($NDRE_H$) and 0.42 (simulated broadband NDVI). The prediction power of these different VIs is summarized in Table 2. Overall, the different VIs were only capable of predicting 50% of the samples correctly. About 28% of the samples were under-fertilized and 22% were over-fertilized. No difference was observed between the narrow-band VIs and the broad-band VIs, and the NDRE performed the worst among the different VIs (Table 2).

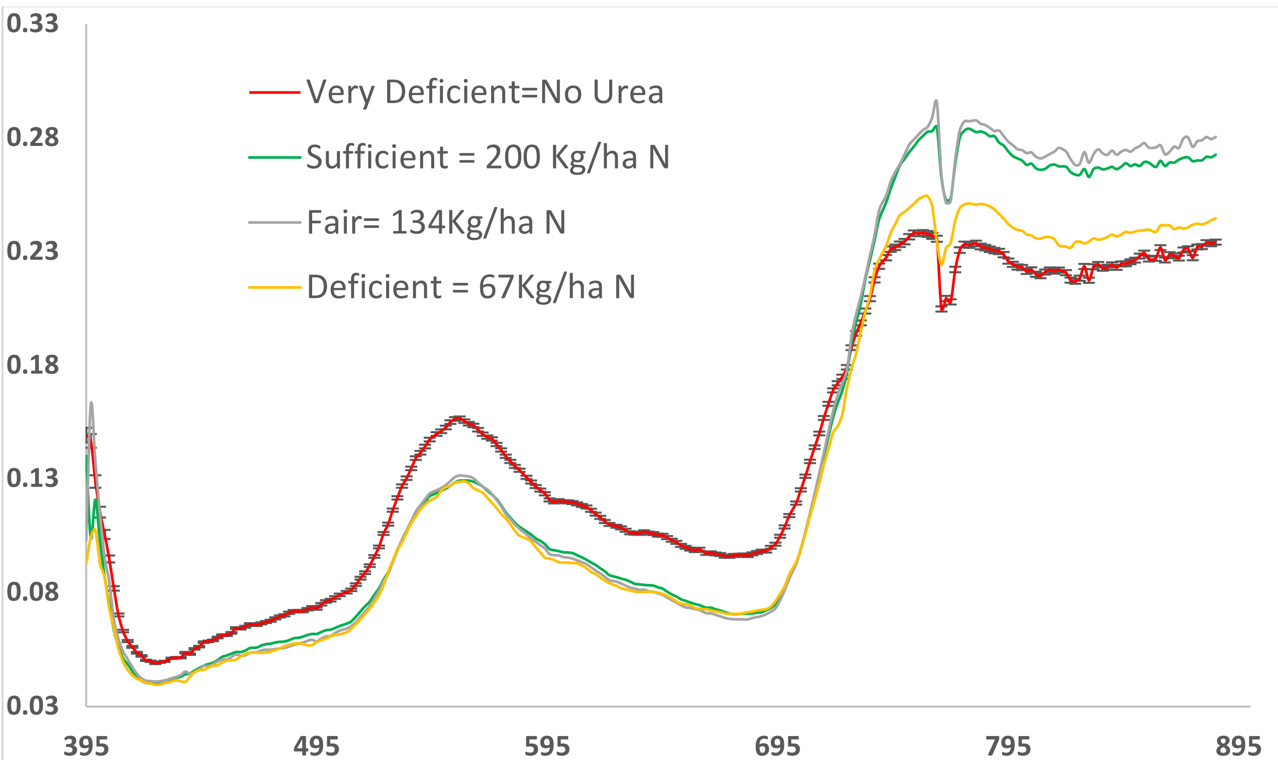

Fig.2: Spectral signature of the four classes of nitrogen stress in corn at V5 using hyperspectral sensing. Class 1 (red) is plotted with the 95% confidence interval for variance illustration.

Prediction power of the different ML classifications based on 240 bands:

- **Decision tree (DT):** This classification algorithm has a tendency to overfit the data if restrictions are not imposed. We restricted both the depth of the tree (pruning) and the minimum number of samples for node splitting to overcome the overfitting issue. Choosing between Gini index and entropy did not impact the outcome of the classification. Overall, the DT algorithm was capable of correctly classifying 74% of the test set (79% on the training set) with 16% of the samples under-fertilized and 10% over-fertilized (Table 2).

- **Random Forest (RF):** As was the case for the DT algorithm, we used pre-pruning and imposed threshold for nodes to avoid overfitting. Additionally, we specified 100 iterations (trees). The trees were generated while allowing bootstrapping. RF outperformed DT as expected and achieved an overall classification accuracy of 78%.

- **K Nearest Neighbor (KNN):** The best performance of this algorithm was achieved using K=15 neighbors, Minkowski distance metric and uniform weights. Larger K values slows the processing, but provides better performance. KNN achieved an overall accuracy of 79%.

- **Bayes naïve (BN):** Using the Gaussian implementation of BN resulted in mediocre performance with an accuracy of 57%. This was expected as the BN assumes independence between futures (Zhang, 2004; Pedregosa *et al.* 2011) which is not the case with hyperspectral data.

- **Logistic Regression (LR):** LR achieved a 90% classification accuracy. The best result was found when we set the regularization strength to a value of 0.02, giving more weight to the correctly classified samples following the procedure outlined in Pedregosa *et al.* (2011).

- **Support Vector Machine (SVM):** data were scaled (0-1) first before SVM processing, as this algorithm is known to be sensitive to large values. SVM did extremely well classifying the unseen test set with an overall accuracy of 91%. The best performance was achieved using a linear kernel and a penalty parameter value of 1 (Chang and Lin, 2001).

- **Multi-Layer Perceptron (MLP):** Similar to SVM, we used scaled data in the MLP classification. The best result was achieved using a learning rate of 0.01. Increasing the number of iterations for the MLP did not impact performance. Overall, MLP showed the best performance with a classification accuracy of 93%.

- **Gradient Boosting (GB):** GB is similar to RF but the trees are generated using mutually exclusive features. As was the case for FR and DT, restrictions were imposed on this ML algorithm to avoid overfitting, and they included pruning, number of features per tree, learning rate and number of samples per split. GB achieved an overall classification accuracy of 89% (Table 2)

Feature extraction:

We examined various univariate feature selection techniques including $\chi 2$ test, strength of correlation between the different bands and the response (dN), and the K-best features based on variance (Pedregosa *et al.* 2011). However, the classification accuracy of these techniques was very poor (based on 15-20 features/bands). On the other hand, when we used RF to rank the most important features (based on nodes), better performance was achieved (Table 2). With the top 15 most important bands (873, 397, 768, 762, 813, 760, 688, 682, 645, 432, 612, 592, 822, 797, and 828 nm) based on RF classification, both MLP and GB achieved 81% classification accuracy. Similarly, using the spectral signature of the four classes to guide the selection of 8 bands (397, 559, 684, 746, 756, 768, 795, and 850 nm), we achieved 75-76% classification accuracy using MLP and GB. We also examined principal component analysis (PCA) as a dimension reduction method. However, we achieved lower classification accuracy with both 3 PC's and 6 PC's (Table 2)

Table 2: Three metrics for comparing the different methods of accessing corn N stress.

Methods	Algorithm	Accuracy	Under-fertilized	Over-fertilized
Broad-band VIs	**NDVI**	0.49	0.31	0.20
	GNDVI	0.50	0.30	0.20
	GRVI	0.51	0.26	0.23
	NDRE	0.46	0.30	0.24
Narrow-band VIs	**$NDVI_H$**	0.49	0.30	0.21
	$GNDVI_H$	0.51	0.28	0.21
	$GRVI_H$	0.51	0.27	0.22
	$NDRE_H$	0.45	0.31	0.24
ML on 240 bands	**DT**	0.74	0.16	0.10
	RF	0.78	0.18	0.03
	KNN	0.79	0.19	0.03
	BN	0.57	0.33	0.10
	LR	**0.90**	0.07	0.03
	SVM	**0.91**	0.05	0.04
	MLP	**0.93**	0.04	0.03
	GB	0.89	0.07	0.04
ML with 15 RF selected bands	**RF**	0.77	0.17	0.05
	LR	0.78	0.16	0.05
	MLP	**0.81**	0.12	0.07
	BG	**0.81**	0.11	0.08
	SVM	0.78	0.14	0.08
ML on 8 spectral signature - guided bands	**GB**	**0.76**	0.16	0.08
	LR	0.72	0.20	0.08
	MLP	**0.75**	0.18	0.07
	SVM	0.73	0.19	0.08
ML with PCA	MLP on 3PCs	0.69	0.23	0.08
	MLP on 6PCs	**0.74**	0.17	0.09

Conclusions

This research confirmed the challenge of using VIs (both broad-bands and narrow-bands based) to assess corn nitrogen stress in early stages of corn development. VIs correctly estimated N stress on only 50% of plant samples, and caused nearly 30% of plants to be under-fertilized. On the other hand, hyperspectral sensing combined with machine learning significantly improved the assessment of corn nitrogen stress at V5, and achieved more than 90% classification accuracy when the entire spectrum was mined. Particularly, SVM, LR, MLP and GB showed promising results. These findings show the huge potential for UAV-compatible hyperspectral sensors to improve in-season corn nitrogen management. The same applies to satellite hyperspectral sensing. The approach used in this study can also guide in the selection of bands for multispectral sensor engineering. We intend to test the stability of these algorithms across different fields, growth stages and hybrids to further confirm these findings.

Acknowledgements
The authors thank the Minnesota Department of Agriculture for supporting this project.

References

Abendroth L.J., Elmore R.W, Boyer M.J., and Marlay S.K.. 2011. Corn growth and development. PMR 1009. Iowa State University Extension and Outreach, Ames, Iowa.

Breiman J. H., Friedman R. O. and Stone C. J.. 1984 ``Classification and Regression Trees,'' Belmont, CA: Wadsworth, 1984.

Chang C.C. and Lin C.J. 2001. LIBSVM: A Library for Support Vector Machines, 2001, Available online: https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf (Last accessed April 30, 2018)

Keeler B.L, Gourevitch J.S., Stephen Polasky, Forest Isbell, Chris W. Tessum, Jason D. Hill and Julian D. Marshall. 2016. The social costs of nitrogen. Science Advances Vol. 2, no. 10, e1600219

Müller A.C. and Guido S. 2016. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media. 394 pg

Pedregosa F., Varoquaux G., Gramfort A. et al. 2011. Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830., JMLR 12, pp. 2825-2830, 2011

De Veaux R. D. and Ungar L. H. 1994. Multicollinearity: a tale of two non-parametric regressions. – In: Cheeseman P. and Oldford R. W. (eds), Selecting models from data: AI and statistics IV. Springer, pp. 293–302.

Zhang H., 2004. The Optimality of Naïve Bayes. In: Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, Miami Beach, FL, USA.