



Commentary: *Aedes albopictus* and *Aedes japonicus*—two invasive mosquito species with different temperature niches in Europe

Konstantinos Demertzis^{1*}, Lazaros Iliadis¹ and Vardis-Dimitrios Anezakis²

¹ Department of Civil Engineering, Faculty of Mathematics Programming and General Courses, School of Engineering, Democritus University of Thrace, Xanthi, Greece, ² Department of Forestry and Management of the Environment and Natural Resources, Democritus University of Thrace, Orestiada, Greece

Keywords: Asian bush mosquito, Asian tiger mosquito, climate change, invasive species, species distribution modeling, ensemble learning, machine learning

A commentary on

Aedes albopictus and *Aedes japonicus*—two invasive mosquito species with different temperature niches in Europe

by Cunze, S., Koch, L. K., Kochmann, J., and Klimpel, S. *Parasit. Vectors* (2016). 9:573. doi: 10.1186/s13071-016-1853-2

OPEN ACCESS

Edited by:

Aristides (Aris) Moustakas,
Universiti Brunei Darussalam, Brunei

Reviewed by:

Dimitris Poursanidis,
Foundation for Research and
Technology Hellas, Greece
Stelios Katsanevakis,
University of the Aegean, Greece

*Correspondence:

Konstantinos Demertzis
kdemertz@fmenr.duth.gr

Specialty section:

This article was submitted to
Environmental Informatics,
a section of the journal
Frontiers in Environmental Science

Received: 27 September 2017

Accepted: 23 November 2017

Published: 11 December 2017

Citation:

Demertzis K, Iliadis L and
Anezakis V-D (2017) Commentary:
Aedes albopictus and *Aedes*
japonicus—two invasive mosquito
species with different temperature
niches in Europe.
Front. Environ. Sci. 5:85.
doi: 10.3389/fenvs.2017.00085

INTRODUCTION

In this interesting and original study, the authors present an ensemble Machine Learning (ML) model for the prediction of the habitats' suitability, which is affected by the complex interactions between living conditions and survival-spreading climate factors. The research focuses in two of the most dangerous invasive mosquito species in Europe with the requirements' identification in temperature and rainfall conditions. Though it is an interesting approach, the ensemble ML model is not presented and discussed in sufficient detail and thus its performance and value as a tool for modeling the distribution of invasive species cannot be adequately evaluated.

METHODOLOGY USED

The authors use an Ensemble Approach (ENAP) based on 10 timely ML algorithms, aiming to draw up the habitats' maps for both species of mosquitoes. Ensemble methods are meta-algorithms that combine several techniques into a unique predictive model to decrease variance. For example, in Bagging different training data subsets are randomly drawn—with replacement—from the entire training dataset, to train a different classifier. In Boosting, resampling is strategically geared to provide the most informative training data for each consecutive classifier, or to improve predictions. Stacking, involves training to combine the predictions of several other learning algorithms (Zhou, 2012).

Unlike a statistical ensemble in statistical mechanics which is usually infinite, a ML ensemble consists of only a concrete finite set of alternative models, but typically allows for much more flexible structures to exist among those alternatives. Perhaps one of the earliest works on ensemble systems is the paper by Dasarathy and Sheela (1979). They first introduced an ENAP for partitioning the feature space, using two or more classifiers, in a divide-and-conquer fashion. Over a decade later, Hansen and Salamon (1990) showed the variance reduction property of an ENAP.

They managed to improve the generalization performance of an ANN by using an ensemble of similarly configured ANN. But it was Schapire's work that has put the ENAP at the center of ML research, as he has proven that a strong classifier can be generated by combining weak classifiers (Schapire, 1990). Finally, Buisson et al. suggested that attention should be paid to the use of predictions ensembles resulting from the application of several statistical methods. Forecasted impacts should always be provided with an assessment of their uncertainty (Buisson et al., 2010).

Unfortunately, the authors of this interesting paper, do not offer a deep description of the proposed ENAP and it is not clear if their approach can cover the main points of the ensemble techniques. For example, the proposed ENAP convert species' probability of occurrence into binary presence-absence data using a predefined threshold. Assessing models based on presence only data, it is difficult to learn the overall species occurrence probability, based on false or misleading information or unjustified simplifying assumptions, because there is typically no validation data with true presences and absences (Hastie and Fithian, 2013). The ENAP that was proposed cannot surmount this problem, it only makes it more hidden.

ALIEN SPECIES DISTRIBUTION MODELING AND MACHINE LEARNING ENSEMBLES MODELS

Current practices in Alien Species Distribution Modeling (ASDM) algorithms (Lorena et al., 2011; Duan et al., 2014; Shabani et al., 2016), include Profile Methods (BIOCLIM, ENFA) (Lorena et al., 2011; Duan et al., 2014; Shabani et al., 2016), Regression-based techniques (GLM, MARS) (Lorena et al., 2011; Duan et al., 2014; Shabani et al., 2016), ML techniques (MAXENT, ANN, SVM) (Lorena et al., 2011; Duan et al., 2014; Shabani et al., 2016).

A widely used and effective method in ASDM involves creating ML ensembles' models (Duan et al., 2014). The two most important advantages of ENAP focus on the fact that they offer better prediction and more stable and robust models, as the overall behavior of a multiple model is less noisy than a corresponding single one (Kuncheva, 2004; Zhou, 2012). For example, in Zhang and Zhang (2012) the authors propose an effective ENAP to assess the impacts of predictor variables and ASDM. In Daliakopoulos et al. (2017) the Random Forest EANP has proven that it can provide a better understanding of facilitating and limiting factors of alien species presence, both for research and management purposes. Finally, Lauzeral et al. (2012) proposes an iterative ENAP to ensure noise absence and hence to improve the predictive reliability of ensemble modeling of species distributions.

Some of the most important points related to the operation and use of the ENAP that should be included and discussed thoroughly by the authors are presented below:

1. The ensemble size of the proposed model. The number of classifiers included in the creation of an ensemble model has a large impact on the accuracy of the prediction (Kuncheva, 2004; Zhou, 2012). Regarding the proposed ENAP, a 10 state-of-the-art algorithms used, nevertheless without thorough analysis and explanation. On the other hand, their theoretical framework of Ensemble Learning shows that using the same number of independent component classifiers as class labels gives the highest accuracy (Hamed and Can, 2016).
2. A detailed and complete description and justification of the classifiers selection. The choice of the proper classifiers (e.g., ANN) to be included in an ENAP (Kuncheva, 2004; Zhou, 2012) should be based on the selection of the implementation mode and on the parameters' settings which can lead to different decision boundaries, even if all other parameters remain constant (Kuncheva, 2004; Zhou, 2012). It is a fact that there is no point or advantage to combining a group of models that are identical and generalize in the same way (López et al., 2007; Bougoudis et al., 2014). In the proposed ENAP, both GLM and MAXENT were used, and there is no clear explanation on how the authors have chosen this specific architecture. As shown by Renner and Warton (2013) MAXENT is equivalent to a GLM with a Poisson error structure and differing only in the intercept term, which is scale-dependent in MAXENT. One cannot argue that MAXENT has different predictive performance than a GLM when they are equivalent.
3. A clear and sufficiently detailed discussion-explanation on the determination and handling of the weights employed by the distinct ensemble models (Kuncheva, 2004; Zhou, 2012). The weight vector is a very important parameter in the process of training an ENAP, as it is used in the determination of the classifiers' performance and of the classification confidence level (Kuncheva, 2004; Zhou, 2012). The authors do not include a detailed description of the weights employed by the distinct ensemble models, with no attempt to tie them to the problem at hand.
4. Clear description of the process that has determined the optimal model, its potential hybrid nature and justification of the proposed ensemble's architecture reliability. This can be done using inclusion of diagrams or algorithms. The variance of prediction results in a ML model is one of the most important measures for assessing the credibility of the method (Kuncheva, 2004; Zhou, 2012). The work by Yackulic et al. (2013) shows that MAXENT model outputs (i.e., maps) are presented completely casually and without providing readers with any means to critically examine modeled relationships. This fact may be hidden or masked within proposed ENAP, but the problem remains.

DISCUSSION AND CONCLUSIONS

It is worth noting that in general an ENAP can lead to much better prediction results, while offering generalization. This is one of the key issues in the field of ML, as it can reduce bias and variance and it has the potential to eliminate overfitting (Kuncheva, 2004; Zhou, 2012). Moreover, it implements robust predictive models capable of responding to high complexity

problems such as those of spreading invasive species (Demertzis and Iliadis, 2015, 2017). However, the development of these models should not be done in a black box mode research and it should be accompanied by a set of in-depth analysis regarding key training and operation decision points, thus allowing critical readers to fully and thoroughly evaluate the proposed methodology and to promote research in the broader scope. Finally, there are cases where wide variety of comparatively model-free forecasting methods outperforms the correct mechanistic data-driven model. However, according to

Moustakas (2017) “if one simply relies on data-driven science, several components of scientific methods will be made poorer.”

AUTHOR CONTRIBUTIONS

KD and LI conceived of the presented idea. KD and LI developed the theoretical formalism, performed the analytic calculations and performed the numerical simulations. KD, LI, and V-DA verified the analytical methods. All the authors contributed to the final version of the manuscript. LI supervised the project.

REFERENCES

- Bougoudis, I., Iliadis, L., and Papaleonidas, A. (2014). “Fuzzy inference ANN ensembles for air pollutants modeling in a major urban area: the case of Athens,” in *Proceedings of the 15th Engineering Applications of Neural Networks, Vol. 459* (Cham: Springer; Communications in Computer and Information Science), 1–14.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S., and Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. *Global Change Biol.* 16, 1145–1157. doi: 10.1111/j.1365-2486.2009.02000.x
- Daliakopoulos, I. N., Katsanevakis, S., and Moustakas, A. (2017). Spatial downscaling of alien species presences using machine learning. *Front. Earth Sci.* 5:60. doi: 10.3389/feart.2017.00060
- Dasarathy, B. V., and Sheela, B. V. (1979). Composite classifier system design: concepts and methodology. *Proc. IEEE* 67, 708–713. doi: 10.1109/PROC.1979.11321
- Demertzis, K., and Iliadis, L. (2015). “Intelligent bio-inspired detection of food borne pathogen by DNA barcodes: the case of invasive fish species *Isoetes macrospora*,” in *Engineering Applications of Neural Networks. Communications in Computer and Information Science, Vol. 517*, eds L. Iliadis and C. Jayne (Cham: Springer), 89–99.
- Demertzis, K., and Iliadis, L. (2017). “Adaptive elitist differential evolution extreme learning machines on big data: intelligent recognition of invasive species,” in *Advances in Big Data; Advances in Intelligent Systems and Computing, Vol. 529*, eds P. Angelov, Y. Manolopoulos, L. Iliadis, A. Roy, and M. Vellasco (Cham: Springer), 333–345.
- Duan, R.-Y., Kong, X.-Q., Huang, M.-Y., Fan, W.-Y., and Wang, Z.-G. (2014). The predictive performance and stability of six species distribution models. *PLoS ONE* 9:e112764. doi: 10.1371/journal.pone.0112764
- Hamed, R. B., and Can, F. (2016). “A theoretical framework on the ideal number of classifiers for online ensembles in data streams,” in *CIKM* (Indianapolis, IN: ACM), 2053.
- Hansen, L. K., and Salamon, P. (1990). Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Intellig.* 12, 993–1001. doi: 10.1109/34.58871
- Hastie, T., and Fithian, W. (2013). Inference from presence-only data; the ongoing controversy. *Ecography* 36, 864–867. doi: 10.1111/j.1600-0587.2013.00321.x
- Kuncheva, L. (2004). *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ: Wiley.
- Lauzeral, C., Grenouillet, G., and Brosse, S. (2012). Dealing with noisy absences to optimize species distribution models: an iterative ensemble modelling approach. *PLoS ONE* 7:e49508. doi: 10.1371/journal.pone.0049508
- López, M., Melin, P., and Castillo, O. (2007). A method for creating ensemble neural networks using a sampling data approach. *Ther. Adv. Appl. Fuzzy Logic* 42, 772–780. doi: 10.1007/978-3-540-72434-6_78
- Lorena, A. C., Jacintho, L. F. O., Siqueira, M. F., De Giovanni, R., Lohmann, L. G., de Carvalho, A. C. P. L. F., et al. (2011). Comparing machine learning classifiers in potential distribution modelling. *Expert Syst. Appl.* 38, 5268–5275. doi: 10.1016/j.eswa.2010.10.031
- Moustakas, A. (2017). Spatio-temporal data mining in ecological and veterinary epidemiology. *Stochast. Environ. Res. Risk Assess.* 31, 829–834. doi: 10.1007/s00477-016-1374-8
- Renner, I. W., and Warton, D. I. (2013). Equivalence of MAXENT and poisson point process models for species distribution modeling in ecology. *Biometrics* 69, 274–281. doi: 10.1111/j.1541-0420.2012.01824.x
- Schapire, R. E. (1990). The strength of weak learnability. *Mach. Learn.* 5, 197–227.
- Shabani, F., Kumar, L., and Ahmadi, M. (2016). A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecol. Evol.* 6, 5973–5986. doi: 10.1002/ece3.2332
- Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Grant, E. H. C., et al. (2013). Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.* 4, 236–243. doi: 10.1111/2041-210x.12004
- Zhang, Q., and Zhang, X. (2012). Impacts of predictor variables and species models on simulating *Tamarix ramosissima* distribution in Tarim Basin, northwestern China. *J. Plant Ecol.* 5, 337–345. doi: 10.1093/jpe/rtr049
- Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. London: CRC Press.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Demertzis, Iliadis and Anezakis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.