

Predicting Coastal Dissolved Inorganic Nitrogen Levels by Applying Data-Driven Modelling: The Case Study of Cyprus (Eastern Mediterranean Sea)

Ekaterini Hadjisolomou^{1,2,}, Konstantinos Antoniadis³, Maria Rousou³, Lavrentios Vasiliades³, Rana Abu-Alhaija⁴, Herodotos Herodotou¹, Michalis Michaelides¹, and Ioannis Kyriakides^{2,5}*

¹Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, 3036 Limassol, Cyprus

² University of Nicosia Research Foundation, 46 Makedonitissas Avenue, CY-2417 P.O. Box 24005, CY-1700 Nicosia, Cyprus.

³ Department of Fisheries and Marine Research, Ministry of Agriculture, Rural Development and the Environment, 2033 Nicosia, Cyprus

⁴ Cyprus Subsea Consulting and Services (CSCS) Ltd, 34A Paragotitissas Street, 2326 Lakatamia, Nicosia, Cyprus

⁵ Cyprus Marine & Maritime Institute, CMMI House, Vasileos Pavlou Square, P.O.Box 40930, 6023 Larnaca, Cyprus

Abstract. A surfeit of Dissolved Inorganic Nitrogen (DIN), which is defined as the total amount of nitrite, nitrate, and ammonium levels in water, may cause negative effects to the marine environment. For example, elevated levels of DIN may promote surplus production of algae and possible depletion of oxygen in the water column. The DIN in the marine water column is monitored as part of the Water Framework Directive (WFD), the Nitrates Directive and the EU Marine Strategy Framework Directive (MSFD). Data-driven models have been proved to be an excellent management tool for environmental issues related to coastal water quality protection and management. Based on data-drive models, and specifically the Artificial Neural Networks (ANNs), the DIN levels from coastal stations in Cyprus were predicted. To do so, three different ANNs models were created, each of them calculating nitrite, nitrate, and ammonium levels respectively with high accuracy ($r > 0.95$). The results derived from these models can be used to identify hot-spot areas with increased DIN levels and to evaluate management scenarios and measures to be implemented in order to maintain the good Environmental Status and quality of the coastal waters.

1 introduction

Coastal eutrophication caused by anthropogenic nutrient inputs is one of the most serious environmental issues and is responsible for the degradation of marine ecosystems worldwide [1]. Increased levels of human-derived inputs of nutrients (mainly nitrogen (N) and phosphorus (P) compounds) are associated with the eutrophication phenomenon [2]. Usually, the coastal waters are extensively polluted by the nutrients in areas affected by the urbanization and agriculture activities [3]. In that case, the elevated levels of nutrients originated from anthropogenic activities trigger the eutrophication process into the coastal water ecosystem [4], causing unpleasant consequences. These eutrophication-originated side effects are mainly hypoxia (dissolved oxygen (DO) depletion in the water column); and harmful algal blooms [5] which are associated among others to aesthetic degradation of the water body (scum formation), harmful cyanotoxins production,

biodiversity decline, massive fish kills and seagrass loss [5- 6].

The marine eutrophication is mainly affected/controlled by the N parameter. As stated by the Hudon et al. [7] the excess increase of N concentration into the coastal water because of cultural eutrophication, has as a result the marine phytoplankton production to suffer since it is mainly N-limited in temperate areas. Interestingly, according to Howarth and Marino [3] the N is the primary nutrient controlling coastal eutrophication instead of the P (which is considered the primary nutrient regarding freshwater eutrophication), something that was realized/accepted by the scientific community only during the recent decades. However, it must be noted that the previous authors [3], also stated that the optimal management of coastal eutrophication requires controlling both N and P, since primary production might be P-limited in some systems. In their study Zhang et al. [4] are evaluating the impact of dissolved inorganic nitrogen (DIN) pollution in coastal waters adjacent to Hainan Island (China) and reported

* Corresponding author: e.hadjisolomou@cut.ac.cy

that when the Eutrophication Index (*EI*) was increased then the contribution of chemical oxygen demand and dissolved inorganic phosphorus may become more important than the DIN contribution to eutrophication. Nowadays, data-driven models are extensively used in almost every scientific area (e.g., medicine [8], economics [9]). Data-driven models produce relationships/associations based on machine learning methodologies, between the input parameters and the target-output parameter, without the need to consider the underlying processes that govern the modelled system. In contrast, the process-based models are based on mathematical/physical equations/principles. Therefore, data-driven models are suitable for modelling systems where process-based models cannot be built due to the lack of knowledge about the system's process [10]. According to Quetglas et al. [11] the relationships between ecological parameters are often non-linear or even unknown. Additionally, ecological data are usually noisy, non-linear, complex and affected by internal relationships between the parameters [12]. Artificial Neural Networks (ANNs) -which are data-driven models- are suitable for modelling the non-linear and complex aquatic systems, producing results highly accurate (e.g., [13]).

Several water quality modelling studies utilizing ANNs are created during the last decades (e.g., [5, 12-15]). For example, in the study of Salami-Shahid and Ehteshami [16] two ANNs were developed to predict the DO and salinity parameters using variables from a data-recording station in the San Joaquin River (USA). Similarly, in their study Huo et al. [17] developed four ANNs capable to predict the DO, Chlorophyll-a (Chl-a), total nitrogen (TN) and secchi disk depth (SD) parameters in Lake Fuxian, the deepest lake of southwest China.

In this modelling study three different ANNs were created, each of them calculating nitrite, nitrate, and ammonium levels for the coastal waters of Cyprus. Additionally, the fact that the created ANNs models are regional models, which take into consideration the unique environmental characteristics of Cyprus coastal water quality parameters [13] enables us to utilize these ANNs to investigate specialized management scenarios regarding the DIN parameter for the coastal water of Cyprus. Sensitivity analysis was calculated for each of these ANNs to investigate the impact/role of each input parameter. The results of sensitivity analysis were implemented to identify possible mechanisms/relationships between the DIN parameter and the rest of the monitored water quality parameters. In conclusion, the created ANNs may be used for DIN management/control purposes, but also provided us useful information about the DIN species interactions with the rest water quality parameters.

2 Methodology

2.1. Study Area and Data Sampling

The Republic of Cyprus is an island located in the Eastern Mediterranean, specifically in the Levantine Basin. The Levantine Basin is one of the most oligotrophic seas worldwide and exhibits ultra-oligotrophic and P-starved conditions [18]. The Levant's Sea is characterized by high mean surface water temperatures, with an annual range from 16 °C in the winter up to 26 °C in the summer period.

The following environmental parameters were measured: ammonium (NH_4^+); nitrite (NO_2^-); nitrate (NO_3^-); ortho-phosphates (PO_4^{3-}); salinity; dissolved oxygen (DO); pH; electrical conductivity (EC); water temperature (WT) and Chl-a. Totally 1552 samples were collected sporadically during the monitoring period 2000-2021, from 49 coastal stations in Cyprus (Figure 1) by the Department of Fisheries and Marine Research (DFMR) of the Cyprus Republic, as part of the implementation of Water Framework Directive (WFD) 2000/60/EC, the Marine Strategy Framework Directive (MSFD) 2008/56/EC and the Nitrates Directive 91/676/EEC as well as Regional Sea Conventions, such as the Barcelona Convention for the protection of the Mediterranean Sea.



Fig. 1. Satellite map of Cyprus, where the sampling stations are marked in green colour.

2.2 Artificial Neural Networks Development

Feed-Forward ANNs were developed for the needs of this modelling study, predicting NO_2^- , NO_3^- , and NH_4^+ levels respectively. The multilayer Feed-Forward ANNs are capable to approximate any continuous function and are characterized as “universal approximators” [19]. The multilayer Feed-Forward ANNs trained with the backpropagation algorithm are the most widely used ANNs in ecological applications [11].

Initially all the missing values in the data set was linearly interpolated, since a large proportion of the data was missing values (~15%), because by ignoring/deleting the missing values the sample size is reduced and therefore the ANN's performance is decreased. For avoiding ANN's bias, because of the different ranges of magnitudes between the variables, the data were normalized. The Min-Max normalization was used, where the data are scaled into the range (0, 1). As stated by Eesa and Arabo [20] using data normalization - like scaling into the range (0,1)- can improve the performance of the ANN.

The method of backward elimination or else network trimming is applied to find the optimal set of input parameters for each of the three ANNs. As explained by Muttill and Chau [21] the network trimming starts with a set containing all inputs, and sequentially deletes the input that reduces performance. The trial-and-error procedure was applied to calculate the optimal number of neurons in the hidden layer. The ANN's training algorithm was the Levenberg-Marquardt (LM) algorithm. The data set was divided into training set (80%) and test set (20%) as required by the regularization method, which was applied to ensure that the model will not overfit the data. The ANN's outputs are evaluated for the test set data, based on performance metrics like the correlation coefficient (r), the Root Mean Square Error ($RMSE$) and the Mean Absolute Error (MAE) [22]. Finally, the sensitivity analysis was calculated with the use of the perturbation method, which calculates the effect of the inputs for a small fluctuation/ perturbation on the simulated output (more details can be found in the study of Muttill & Chau [21]). The development of ANN models and data analysis were utilized with the use of MATLAB software.

3 Results

3.1. Ammonium model's results

An ANN with the 7-6-1 topology was chosen after applying the trial-and-error procedure and the input parameters were the DO, pH, WT, EC, salinity, Chl-a and PO_4^{3-} . The inputs were determined based on the backward elimination method, from a data set containing the following candidate input parameters: DO, pH, WT, EC, salinity, Chl-a, NO_2^- , NO_3^- and PO_4^{3-} . The ANN performed well for the test set, since the $r=0.95$ and the $MAE= 0.002$. The graphical illustration between the real and the predicted data (Figure 2) verified that the created ANN is a good predictor of the NH_4^+ parameter.

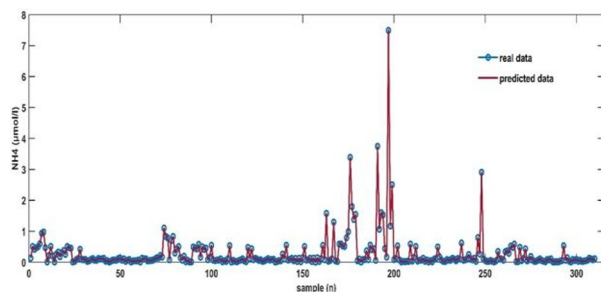


Fig. 2. Graphical illustration between the monitored NH_4^+ values against the ANN values (where the real data are illustrated with cyan colour and the predicted data with red).

Sensitivity analysis was calculated by applying the perturbation method, where the input parameters were fluctuated by a small amount of +8%. This perturbation number was chosen, in order to reflect realistic changes regarding the WT that could be anticipated over the next years due to global warming conditions [23] as indicated by global warming studies for the Eastern Mediterranean

Sea [24]; and corresponds to an increase of WT about 1-2 °C [23]. The sensitivity analysis results (Figure 3) calculated that the most influential parameter is the WT (-270.89%) followed by the DO (244.53%), pH (243.44%), EC (-159.62%), salinity (-30.35%), Chl-a (-5.01%), PO_4^{3-} (2.02%).

3.2. Nitrite model's results.

The modelling procedure applied to predict the NO_2^- levels, was similar with the one applied for the NH_4^+ model. An ANN with the 7-6-1 topology had the better performance among several tested topologies and the input parameters were the DO, pH, WT, EC, salinity, Chl-a and PO_4^{3-} . The inputs were determined based on the backward elimination method, from a data set containing the following candidate input parameters: DO, pH, WT, EC, salinity, Chl-a, NH_4^+ , NO_3^- and PO_4^{3-} .

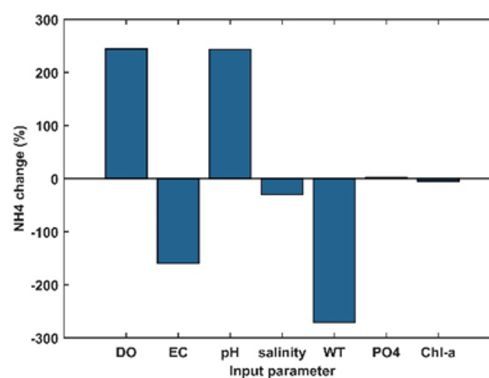


Fig. 3. ANN's sensitivity analysis results, where the input parameters were fluctuated by +8% and the associated change in NH_4^+ level is calculated for each of the inputs.

The NO_2^- model had $r=0.98$ and $MAE= 0.007$ for the test set. The measured values and the predicted values (Figure 4) are having a good match.

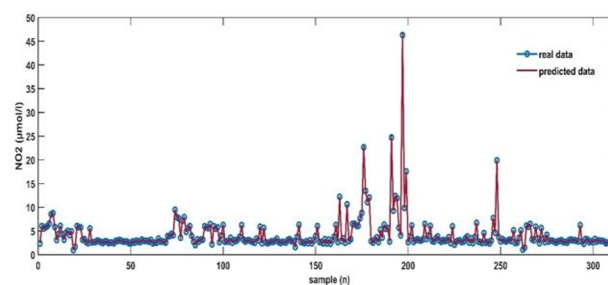


Fig. 4. Graphical illustration between the monitored NO_2^- values against the predicted values (where the real data are illustrated with cyan colour and the predicted data with red).

Sensitivity analysis was calculated by applying the perturbation method, where the input parameters were fluctuated by +8%. The sensitivity analysis (Figure 5) calculated that the most influential parameter is the WT (-16.01%) followed by the pH (15.89%), DO (12.15%), EC (-6.01%), salinity (-3.55%), PO_4^{3-} (0.04%), Chl-a (0.02%).

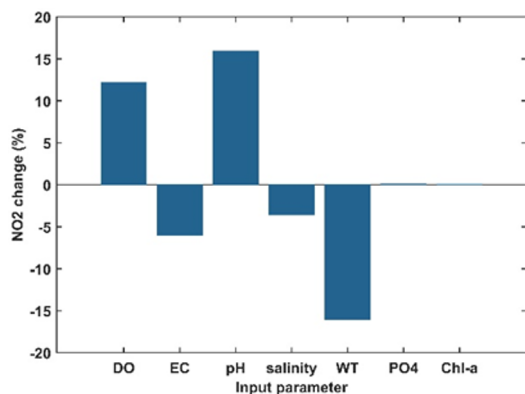


Fig. 5. ANN's sensitivity analysis results, where the input parameters were fluctuated by +8% and the associated change in NO_2^- level is calculated for each of the inputs.

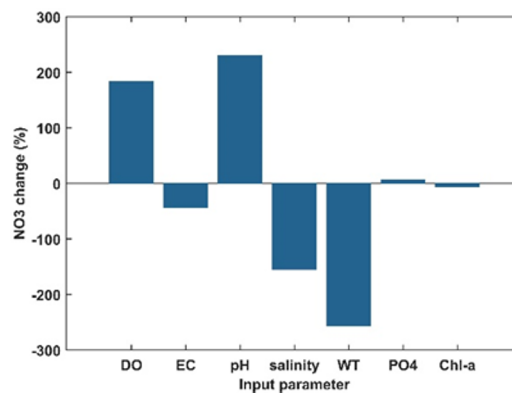


Fig. 7. ANN's sensitivity analysis results, where the input parameters were fluctuated by +8% and the associated change in NO_3^- level is calculated for each of the inputs.

3.3. Nitrate model's results.

The modelling procedure applied to calculate the NO_3^- levels, followed the same steps used for modelling the NH_4^+ and NO_2^- parameters. An ANN with the 7-6-1 topology was chosen after the trial-and-error procedure and the input parameters were the DO, pH, WT, EC, salinity, Chl-a and PO_4^{3-} . The inputs were determined based on the backward elimination method, from a data set containing the following candidate input parameters: DO, pH, WT, EC, salinity, Chl-a, NH_4^+ , NO_2^- and PO_4^{3-} . The ANN's performance was $r=0.96$ and $MAE=0.005$ for the test set. The measured values and the predicted values of the NO_3^- parameter (Figure 6) were very similar.

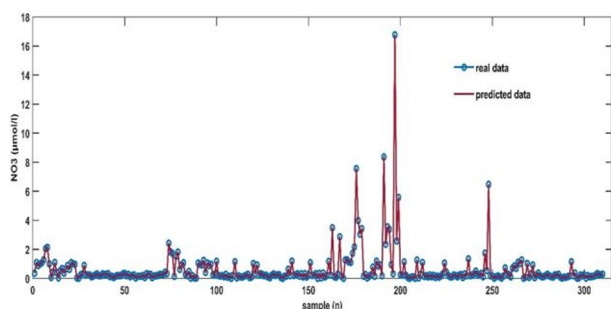


Fig. 6. Graphical illustration between the monitored NO_3^- values against the predicted values (where the real data are illustrated with cyan colour and the predicted data with red).

Sensitivity analysis was calculated by applying the perturbation method, where the input parameters were fluctuated by +8%. The sensitivity analysis (Figure 7) produced that the most influential inputs regarding the NO_3^- parameter is the WT (-256.43%) followed by the pH (229.82%), DO (183.16%), salinity (-154.30%), EC (-43.48%), PO_4^{3-} (5.81%), Chl-a (-5.44%).

4 Discussion

Coastal eutrophication is a global environmental problem, affecting many areas in Europe, China and North America [25]. According to Yang et al. [26] understanding coastal eutrophication's mechanisms and identifying the associated influencing factors, is very important in order to addressing eutrophication problem. Nutrients inputs into the water column derived from human activities are the main cause of coastal eutrophication. The role of the DIN parameter regarding coastal eutrophication is catalytic. Therefore, data-driven models were developed for modelling the NO_2^- , NO_3^- and NH_4^+ parameters. Specifically, the algorithms developed for the needs of this modelling study were ANNs, since ANNs are considered advanced modelling tools suitable for modelling water quality parameters [27]. The created ANN models managed to simulate well the NO_2^- , NO_3^- and NH_4^+ parameters and produced output values characterized by high accuracy ($r>0.95$).

Sensitivity analysis was performed for each of the created ANNs. The ANNs sensitivity analysis revealed that the WT parameter is the most influential input for the NO_2^- , NO_3^- and NH_4^+ parameters. The DIN's concentration increase is associated with a reduction to the WT. This finding might be attributed to the winter upwelling phenomenon observed in coastal water in Cyprus [28]. During the winter upwelling (WT has lower values), nutrient rich waters emerge to the surface and the algal production is increased. Therefore, there is a positive relationship between the DIN and the Chl-a parameters.

The pH parameter was calculated as the second most influential input for the created ANNs (taking into consideration, that the sensitivity analysis for the NH_4^+ model produced very close values for the pH and DO parameters). As stated by Hansen [29] high pH values have been observed in some marine environments after the addition of nutrients. Additional, high Chl-a values are linked with high pH values [30], therefore an indirect relationship between the DIN and the pH parameters exist. The DO parameter was found to be third most contributing input. The algal photosynthesis (therefore increased levels of the Chl-a parameter) results into

oxygen production [31]. Something that might explain the positive indirect relationship between the DO and the DIN, if we consider the effect of winter upwelling. Regarding the rest inputs no clear conclusion regarding their contribution can be extracted based on the sensitivity analysis calculations.

It must be noted that the developed ANNs predicting the NO_2^- , NO_3^- and NH_4^+ parameters, only needed the DO, pH, WT, EC, salinity, Chl-a and PO_4^{3-} parameters as inputs (after applying network trimming). This is very important, particularly when only one of the NO_2^- , NO_3^- and NH_4^+ parameters is needed to be modelled. In that case, the created ANN can calculate the desired parameter without to be presented/know any of the rest DIN species parameters. In the future we aim to recalibrate the created ANNs based on a hybrid database (e.g., buoys and satellite data), aiming to forecast the DIN parameter in the wider region of Eastern Mediterranean.

5 Conclusions

Eutrophication is a serious environmental problem. It is associated mainly to anthropogenic activities, but also to climate change. The DIN parameter is considered as a key stressor to the coastal environment and promotes eutrophication. For that reason, the NO_2^- , NO_3^- , and NH_4^+ parameters were modelled utilizing ANN models. The created models predicted the NO_2^- , NO_3^- , and NH_4^+ parameters with high accuracy. Based on these ANNs several management scenarios can be examined, aiming to protect the coastal water in Cyprus.

Acknowledgements: This work was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (Open Sea Aquaculture in the Eastern Mediterranean project: INTEGRATED/0918/0046), the Cyprus University of Technology (MERMAID project: Metadidaktor POST-DOCTORAL Research Programme), and the EU H2020 Research and Innovation Programme under GA No. 857586 (CMMI-MaRITeC-X).

References

1. T. Malone, A. Newton, *Front. Mar. Sci.* **7**, 670 (2020)
2. M. Devlin, J. Brodie, *Nutrients and Eutrophication, Marine Pollution – Monitoring, Management and Mitigation* (Springer, Cham, 2023)
3. R. Howarth, R. Marino, *Limno. Oceanogr.* **51**, 18144068 (2006)
4. P. Zhang, Y. Chen, P. Conghui et al., *Reg. Stud. Mar. Sci.* **37**, 101332 (2020)
5. E. Hadjisolomou, K. Stefanidis, G. Papatheodorou, E. Papastergiadou, *Int. J. Environ. Res. Public Health* **13**, 8 (2016)
6. M. Dai, Y. Zhao, F. Chai, et al., *Cambridge Prisms: Coastal Futures* **1**, E19 (2023)
7. C. Hudon, P. Gagnon, M. Rondeau, et al., *Biogeochemistry* **135**, 251 (2017)
8. A. Woldaregay, E. Arsand, T. Botsis, et al., *J. Med. Internet. Res.* **21**, e11030 (2019)
9. J. Zhang, S. Cui, Y. Xu, et al., *Expert Syst. Appl.* **97**, 60 (2017)
10. S. Petruseva, V. Zileska-Pancovska, D. Car-Pušić, *Adv. Civ. Eng.* **2019**, 1 (2019)
11. A. Quetglas, F. Ordines, B. Guijarro, *The Use of Artificial Neural Networks (ANNs) in Aquatic Ecology*. (InTech, online, 2011)
12. Y-S. Park, P. Verdonshot, T-S. Chon, S. Lek, *Water Res.* **37**, 8 (2003)
13. E. Hadjisolomou, K. Antoniadis, L. Vasiliades, et al., *IOP Conference Series: Earth and Environ. Sci.* **1123**, 012083 (2023)
14. F. Recknagel, M. French, P. Harkonen, K-I. Yabunaka, *Ecol. Model.* **96**, 11 (1997)
15. J-T. Kuo, M-H. Hsieh, W-S. Lung, N. She, *Ecol. Model.* **200**, 171 (2007)
16. E. Salami Shahid, M. Ehteshami, *Desalin. Water Treat.* **57**, 11 (2016)
17. S. Huo, Z. He, J. Su, et al., *Procedia Environ. Sci.* **18**, 310 (2013)
18. T. Tanaka, T. Zohary, M. Krom, et al., (2007). *Deep Sea Res. Part I Oceanogr. Res. Pap.* **54**, 1721 (2007)
19. K. Hornik, M. Stinchcombe, H. White, *Neural Netw.* **2**, 359 (1989)
20. A.S. Eesa, W.K. Arabo, *Sci. J. Univ. Zakho* **5**, 319 (2017)
21. N. Muttill, K-W. Chau, *Eng. Appl. Artif. Intell.* **20**, 735 (2007)
22. A. Jierula, et al., *Appl. Sci.* **11**, 2314 (2021)
23. E. Hadjisolomou, K. Antoniadis, I. Thasitis, et al., *Exploring the Impact of Coastal Water Quality Parameters on Chlorophyll-a near Cyprus with the use of Artificial Neural Networks*, 39th IAHR World Congress (Granada, Spain 2022)
24. N. Skliris, S.Sofianos, A. Lascaratos., *J. Mar. Syst.* **65**, 400 (2007)
25. J. Garnier, G. Billen, L. Lassaletta, et al, *Environ. Res. Lett.* **16**, 023005 (2021)
26. X.E. Yang, X. Wu, H.L. Hao, Z.L. He, *J. Zhejiang Univ. Sci B.* **9**, 197 (2008)
27. E. Hadjisolomou, K. Stefanidis, G. Papatheodorou, E. Papastergiadou, *Int. J. Environ. Res. Public Health* **15**, 547 (2018)
28. G. Fyttis, S. Zervoudaki, A. Sakavara, S. Sfenthourakis, *J. Plankton Res.* **45**, 291 (2023)
29. P. Hansen, *Aquat. Microb. Ecol.* **28**, 279 (2002)
30. D. Trolle, P.A. Staehr, T.A. Davidson, et al., *Ecosystems* **15**, 1 (2012)

31. A.K. Misra, P. Chandra, V. Raghavendra, Adv. Water Resour. **34**, 1232 (2011)