# Application of Data Integrity Algorithm for Geotechnical Data Quality Management

https://doi.org/10.3991/ijim.v12i8.9569

Tarik Chafiq(✉)
University Hassan II, Casablanca, Morocco
Tarik.chafiq1@gmail.com

Mohammed Ouadoud
Abdelmalek Essaadi University (UAE), Tetouan, Morocco

Hassane Jarar Oulidi
Hassania School of Public Works, Casablanca, Morocco

Ahmed Fekri
University Hassan II, Casablanca, Morocco

**Abstract—**The aim of this research work is to ensure the integrity and correction of the geotechnical database which contains anomalies. These anomalies occurred mainly in the phase of inputting and/or transferring of data. The algorithm created in the framework of this paper was tested on a dataset of 70 core drillings. In fact, it is based on a multi-criteria analysis qualifying the geotechnical data integrity using the sequential approach. The implementation of this algorithm has given a relevant set of values in terms of output; which will minimalize processing time and manual verification. The application of the methodology used in this paper could be useful to define the type of foundation adapted to the nature of the subsoil, and thus, foresee the adequate budget.

**Keywords—**GIS, algorithm, geotechnical data, engineering, drilling

## 1 Introduction

The importance of geotechnical database in the decision-making for the installation of various works (dams, railway, subway, sewage networks, telephone relays, etc.) are obvious. These rich databases would constitute a specific knowledge of subsoil features [1]. This information is essential for private and public organizations, universities and companies for the development of architectural plans of the different infrastructure and the development of the territory [2,3]. In fact, this type of information, which is collected across national laboratories operating in geotechnical studies, will provide a better understanding of the risk of differential underground movement or for the construction of works (pavilions, buildings, roads, structures, etc.). These data are obtained in a traditional way (manually) from investigation reports [4] in digital format. Knowing that 75% of the data quality problems occurs when entering or/and feeding of the information by the collaborators [5]. This can lead to the decrease of the visualization, interrogation and develop geotechnical models for yielding the maximum value and understanding of the data, including purview, which eventually supplies the process of the decision making. Some recent studies have demonstrated the applicability of multivariate data analyses in various fields of engineering geology such as rock engineering [6,7], soil liquefaction [8], Sensitivity analysis to landslides [9,10] and even in investigating The correlation between clay mineralogy and soil shear strength [11]. Therefore, these previous studies were ignored the anomalies (anomaly related to layer overlap, anomaly related to the gap between the layers and anomaly related to lack of information or value) that can happen during the transferring the information to databases. The dataset under investigation of this work was obtained from 70 core drillings collected at the study site. Each point (sample) represents multiple layers superimposed. The present paper attempts to point out a new approach to resolve the problem of geotechnical's data anomalies by using an algorithm to ensure the integrity and correction.

## 2 Geotechnical data and anomalies

### 2.1 Anomaly related to layer overlap

The layers will be in a state of overlap [12], i.e. the movement of the layers, one in the other (Fig. 1) (Table 1).

**Table 1.** Attribute table represents the overlap of data

| Id_Borehole (Id Samples) | Cot_Superior (Upper side) | Cot_Inferior (Under side) | Epa (Thickness) | Lithology (Lithology) |
|---|---|---|---|---|
| 997 | 0 | 0.5 | 0.5 | Topsoil |
| 997 | 0.5 | 0.65 | 0.15 | Lime stone |
| 997 | 0.65 | 2.65 | 2 | Sand Pack |
| 997 | 2.5 | 3 | 0.5 | Sand Stone |

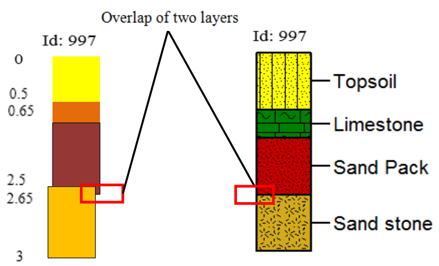


**Fig. 1.** The header image of online-journals.org

The geotechnical data table consists of rows and columns. Each line corresponds to an entity. The columns carry the attributes of the entities. These attributes in this table are:

- *Id_Borehole*: Represent a unique number that represents a sample. Each sample contains layers superimposed on each other.
- *Cot_Superior*: Represent the upper side as compared to soil of each layer in a sample.
- *Cot_Inferior*: Represent the underside as compared to soil of each layer in a sample.
- *Epa*: Represent the thickness of the layer.
- *Lithology*: Represent a description of its physical characteristics visible at outcrop.

In principle the upper side (*Cot_Superior*) as compared to the soil of each layer in a sample must have the same value as the underside (*Cot_Inferior*) of the layer that overcomes it.

## 2.2 Anomaly related to the gap between the layers:

The layers will be in a state of shift [13], this anomaly can however occur during the spacing of the layers of each other (Fig 2) (Table 2).

**Table 2.** Attribute table represents the offset of the layers

| Id_Borehole (Id Samples) | Cot_Superior (Upper side) | Cot_Inferior (Under side) | Epa (Thickness) | Lithology (Lithology) |
|---|---|---|---|---|
| 3106381 | 0 | 0.3 | 0.3 | Topsoil |
| 3106381 | 0,3 | 1,1 | 0.8 | Sand Pack |
| 3106381 | 1,3 | 3,4 | 2.1 | Sand stone |

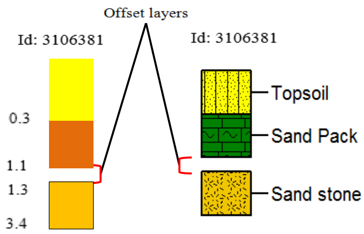


**Fig. 2.** The offset between two layers

## 2.3 Anomaly related to lack of information or value

When there is a spacing that is equal to 0.3 m or more between two layers in the same sample, they are considered as a major anomaly which should be reported (Fig.3) (Table 3).

**Table 3.** Attribute table represents the lack of information

| Id_Borehole (Id Samples) | Cot_Superior (Upper side) | Cot_Inferior (Under side) | Epa (Thickness) | Lithology (Lithology) |
|---|---|---|---|---|
| 3106021 | 0 | 0.3 | 0.3 | Topsoil |
| 3106021 | 0, 3 | 1.3 | 1 | Sand Pack |
| 3106021 | 1.7 | 3,8 | 2.6 | Sand stone |

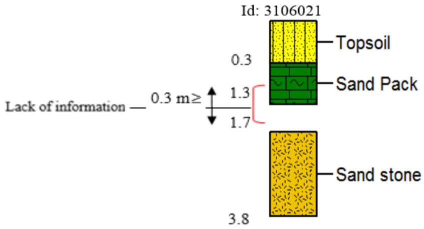


**Fig. 3.** Lack of information with a tolerance of 0.3 m or more

The main anomalies aforementioned are some errors from the samples that were detected in the geotechnical database which is the subject of this work. Knowing that, a geotechnical database of a simple study area may contain thousands of core drilling which can raise the rate of errors that can be found. Therefore, those anomalies represent an obstacle for the engineers who operate in the field of geotechnical engineering, regarding the calculation errors and the time wasted on searching for anomalies manually. To overcome these problems, this work proposes an algorithm for verification and integrity of data mentioned before efficiently.

## 3 Geographic setting

The data extracted from area of BOUSKOURA (Figure 4) is located about twenty kilometers south of Casablanca and it is known by its crust of limestone and sandstone marine dunes sporadically flushed on the surface, and the depth of the groundwater varies between 0,7 and 10 meters. The ground is relatively plain, forming a flat landscape, the altitude varies between 120 to 151 meters.

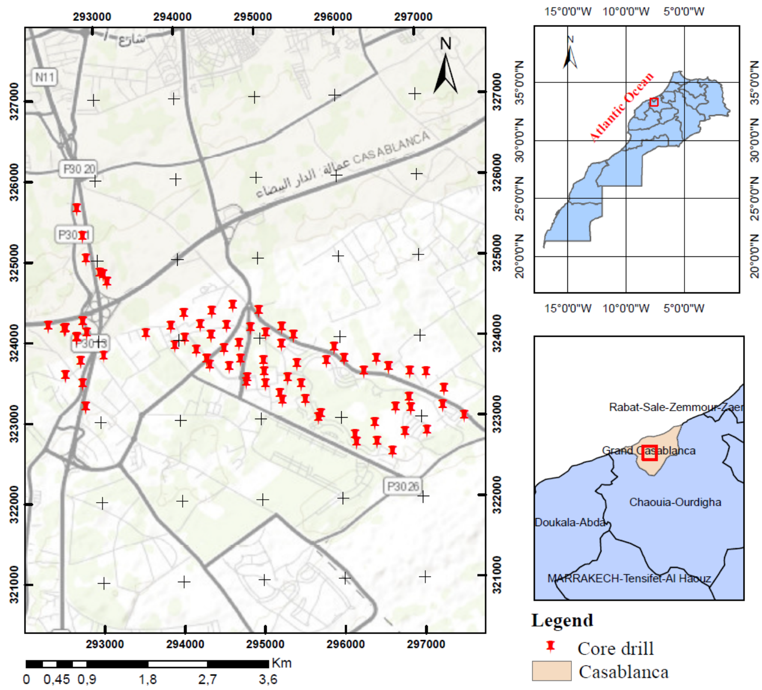


**Fig. 4.** Location map of the core drilling on the area Bouskoura Grand Casablanca region

## 4 Methodology

The approach proposed in this work is based on a multi-criteria analysis qualifying the geotechnical data integrity to maintain, ensure the accuracy and consistency of information. In fact, the integrity algorithm used in this work (Figure 5) was based on an approach which is explained as follows:

- The first step is based on two criteria:
  - Id of borehole (*Id_borehole*): Represents the ID of a core drilling which is repeated because each drilling can contain multiple layers.
  - Coast inferior of the layer (*Cot_Inferior*): Represents the underside of each layer in a core drilling.
- The second step consists of launching a query that will display geotechnical data stored in the dataset, then a Boolean variable will be created and will be given the "False" value before it enters the loop of the algorithm.
- The third step is to detect anomalies of overlapping and / or offset and / or lack of information. For this reason, the program will scan the survey points one by one, checking if the Boolean value is true, then it will access the second condition which is the verification of the "*Id_Borehole*". If they have the same ID, the program will move to the third condition that compares the parameters "*Cot_Inferior*" of the first line and "*Cot_Superior*" of the second line, and so on.

The integrity algorithm was developed as an extension of the ArcMap environment (ArcGIS; ESRI). They were created with ArcObjects, which is a developer kit for ArcGIS, based on Component Object Model (COM) [14,15], and programmed in Visual Basic using Visual Studio (Microsoft) environment [16,17].

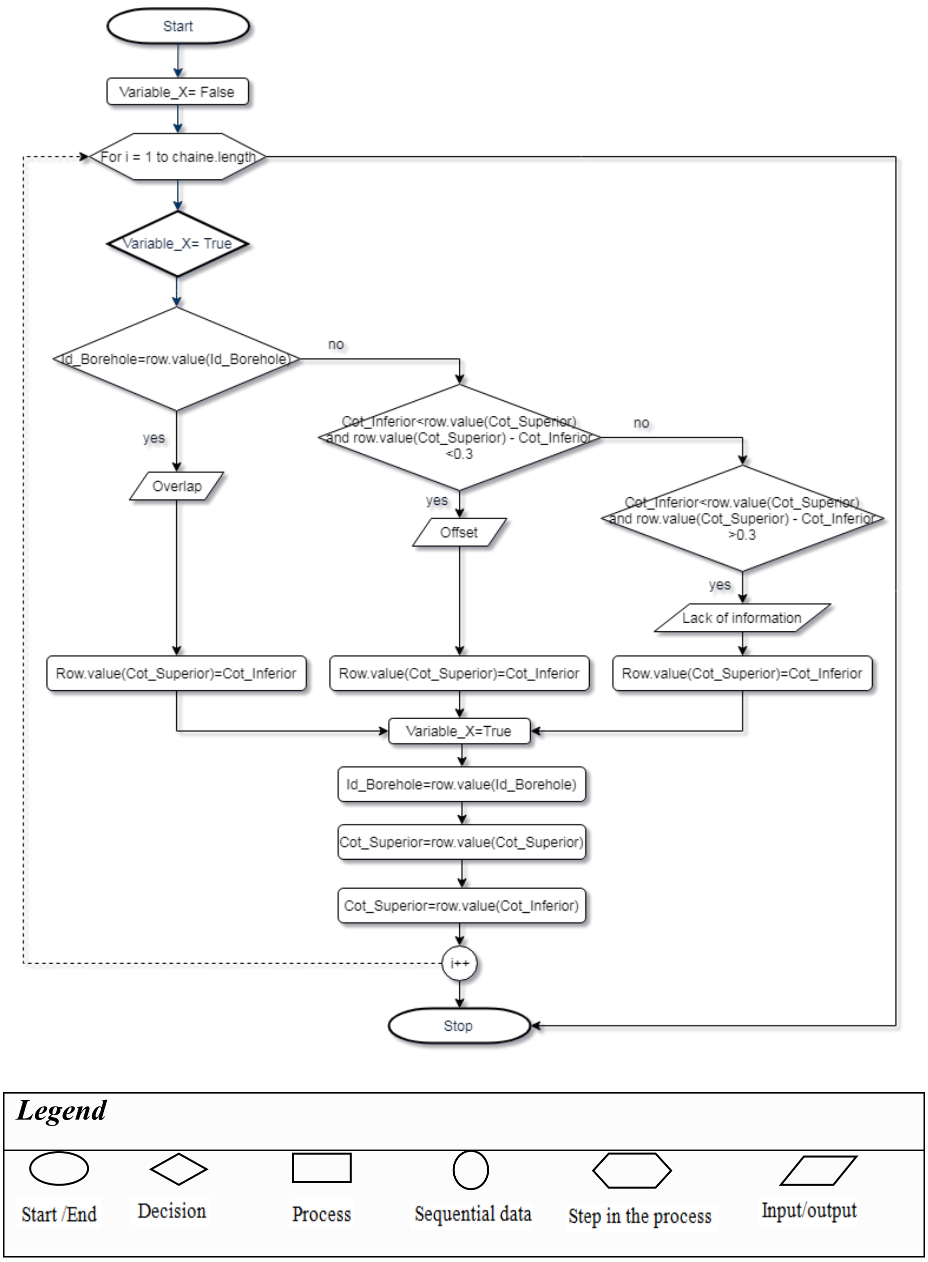


**Fig. 5.** Flowchart of the algorithm

## 5 Result

The implementation of the above algorithm in the platform of ArcObject has given satisfactory results (Table 4, Figure 6) regarding the three anomalies mentioned above. The anomalies will be detected and corrected automatically without any human intervention rapidly, which will minimalize processing time and manual verification.

**Table 4.** Attribute table represents the lack of information

| Id_Borehole (Id Samples) | Cot_Superior (Upper side) | Cot_Inferior (Under side) | Epa (Thickness) | Lithology (Lithology) |
|---|---|---|---|---|
| 997 | 0 | 0.5 | 0.5 | Topsoil |
| 997 | 0.5 | 0.65 | 0.15 | Limestone |
| 997 | 0.65 | 2.65 | 2 | Sand tufa |
| 997 | 2.65 | 3 | 0.5 | Sandstone |

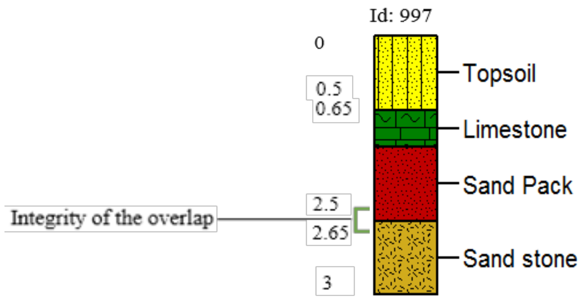


**Fig. 6.** The integrity of the anomaly caused by the overlap

To evaluate the results of the implemented algorithm, an approach of interpolation was used based on simple Kriging [18,19] as it is given in Equation (1)

$$Z(s) = \mu + \varepsilon(s) \quad (1)$$

Because the underground environment is heterogeneous, it is impossible to measure it with deterministic way [20,21]. Z(s) represents the regionalized value, (μ) is the deterministic structure for expectation function of the localization of observations and ε (s) is a Gaussian process of independent measurement errors. Each layer of core lithology (Figure 7) collected from the field was interpolated in order to have a 3D model; knowing that the three-dimensional (3D) viewpoint has always been required for geotechnical data planning purposes, and a large number of computer programs for geotechnical planning have been developed with this in mind5. Thereby, ArcScence [22] of ArcGIS were used to show the illustrative form of underground before (Figure 8a) and after (Figure 9) execution of the algorithm. Indeed, from top to bottom the lithological units are consistent with the simplified lithological column of table 1 and were integrated in the model as follow: (1) Topsoil, (2) Sand Pack and (3) sandstone.

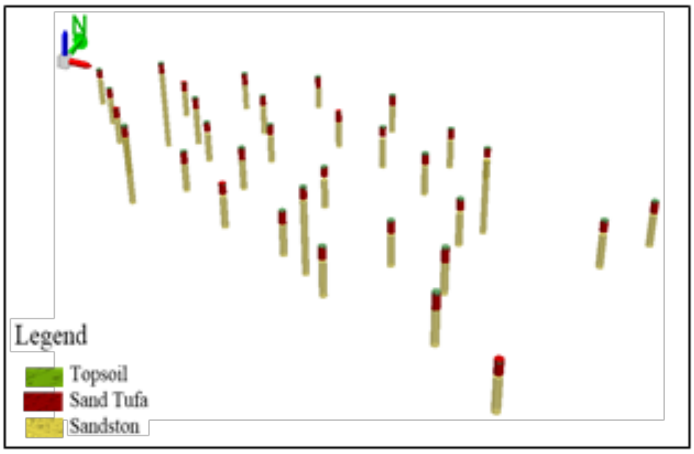


**Fig. 7.** Drilling data of study area

When the initial information was interpolated from the database of drilling points and without the use of the algorithm, the Geovolume (Fig 8) appeared with colorful spots which represent the effect of anomalies (overlap, offset, lack of information) on the 3D model. On the other hand, the integration and execution of the algorithm has positively changed the display of same Geovolume (Fig 9).

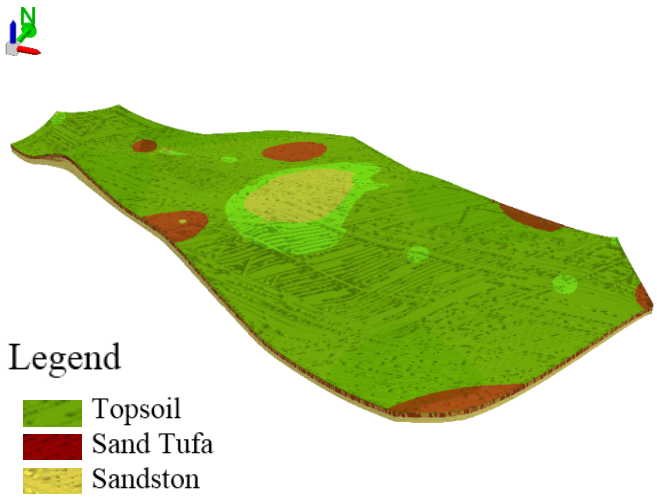


**Fig. 8.** The Geovolume before execution of the algorithm

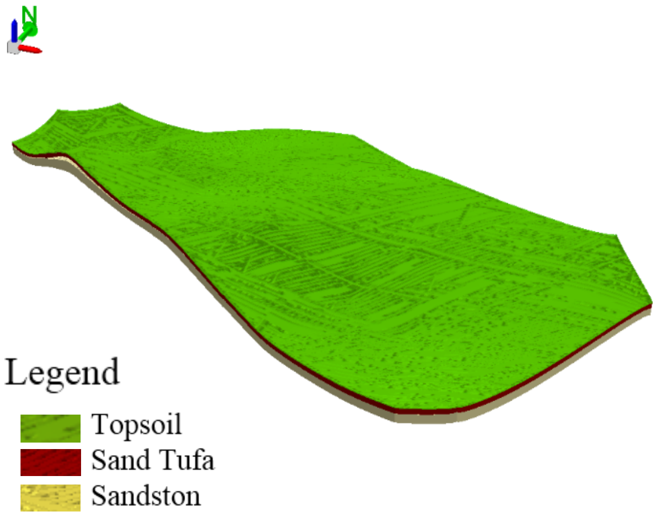


**Fig. 9.** The Geovolume after execution of the algorithm

## 6 Conclusion

The results obtained present the problem resolution related to the quality of the data. This is done by developing an algorithm for the improvement and integrity of geotechnical data in an automatic way, thus saving time for the organizations that process them. The algorithm developed and executed using ArcObject as a platform on samples, has also been tested in real working environments in order to check it. The results provide a solid starting point for contributing to the accumulation of knowledge and improving the quality of data. It is expected that the concepts developed in this study will help to pave the way for other algorithms to be developed in the field of geoscience for the aspect of data quality.

## 7 References

[1] W.H. Qazi, S. Osman, M.B. Memon, Indian Journal of Science and Technology 9 (2016).
[2] R. Zimmermann, W.-S. Ku, H. Wang, A. Zand, J.-P. Bardet, the IEEE Computer Society (2006).
[3] S. Yeniceli, M. Ozcelik, Journal of the Indian Society of Remote Sensing (2015) 1.
[4] I. Vähäaho, Tunnelling and Underground Space Technology 13 (1998) 51. https://doi.org/10.1016/S0886-7798(98)00020-0
[5] W.W. Eckerson, (2006).
[6] I. Rigopoulos, B. Tsikouras, P. Pomonis, K. Hatzipanagiotou, Construction and Building Materials 49 (2013) 747. https://doi.org/10.1016/j.conbuildmat.2013.08.065
[7] R. Ulusay, K. Türeli, M. Ider, Engineering Geology 38 (1994) 135. https://doi.org/10.1016/0013-7952(94)90029-9

[8] W. Zhang, A.T.C. Goh, Computers and Geotechnics 48 (2013) 82. https://doi.org/10.1016/j.compgeo.2012.09.016
[9] A. Nandi, A. Shakoor, Engineering Geology 110 (2010) 11. https://doi.org/10.1016/j.enggeo.2009.10.001
[10] P. Ramesh, J. Vivekavardhan, K. Bharathi, DESIDOC Journal of Library & Information Technology 35 (2015).
[11] A. Hajdarwish, A. Shakoor, N.A. Wells, Engineering Geology 159 (2013) 45. https://doi.org/10.1016/j.enggeo.2013.03.016
[12] J. Kovács, N. Bodnár, Á. Török, Open Geosciences 8 (2016) 52.
[13] V. Meyer, Frontiers in Offshore Geotechnics III, CRC Press, 2015. https://doi.org/10.1201/b18442
[14] R. Li, T. Sun, G. Li, Research on three-dimensional water resources information system based on ArcGIS, Water Resource and Environmental Protection (ISWREP), 2011 International Symposium on. IEEE, 2011, p. 604.
[15] W. Kim, M. Sung, IEEE Access (2018).
[16] V. Velasco, R. Gogu, E. Vázquez-Su-è, A. Garriga, E. Ramos, J. Riera, M. Alcaraz, Environmental earth sciences 68 (2013) 2145. https://doi.org/10.1007/s12665-012-1898-2
[17] V. Velasco, I. Tubau, E. Vázquez-Su-è, R. Gogu, D. Gaitanaru, M. Alcaraz, A. Serrano-Juan, D. Fernàndez-Garcia, T. Garrido, J. Fraile, Computers & Geosciences 70 (2014) 164. https://doi.org/10.1016/j.cageo.2014.04.013
[18] G. Bohling, Kansas Geological Survey, Tech. Rep (2005).
[19] D. Yang, Solar Energy 171 (2018) 876. https://doi.org/10.1016/j.solener.2018.06.055
[20] L.W. Gelhar, Stochastic subsurface hydrology, Prentice-Hall, 1993.
[21] A.E. Aboanber, A.A. Nahla, A.M. Edress, Nuclear Science and Techniques 29 (2018) 132. https://doi.org/10.1007/s41365-018-0467-0
[22] H. Kennedy, Introduction to 3D data: Modeling with ArcGIS 3D analyst and google earth, John Wiley & Sons, 2010.

## 8 Authors

**Tarik CHAFIQ** is a part-time lecturer and scientific collaborator at Hassan II University, Faculty of Sciences Ben M'Sik, who is specialised in the research field of geomatics such Geographical Information Systems (GIS) analyses and Spatial Data Infrastructures (SDI). From 2016 and 2017, he had been a research scientist at Alexandru Ioan Cuza. He got his PhD degree and master degree in geoinformatics. Furthermore, he is certified by Microsoft as an expert in C# and HTML5.

**Mohammed OUADOUD** is a PhD Candidate in Computer sciences, at the Laboratory of Informatics, Research Operational and Statistic Applied (LIROSA) at Faculty of Sciences, Abdelmalek Essaadi University. He has a Master degree in Instructional design Multimedia engineering at the *École normale supérieure* of Martil, Morocco in 2013. His current research focuses on: E-learning, Software Engineering, Man Machine Interface, Bigdata and Adaptive Hypermedia Systems.

**Hassane Jarar Oulidi** is professor of geoinformatics in the Department of Mathematics, informatics and geomatics at Hassania School for Public Works Engineering (EHTP). From 2006 and 2009, he had been a research scientist at German Research Center for Geosciences. His current research activities focus on the design, development of geospatial data infrastructure for groundwater resources management. His other research interests include geospatial sensor web, OGC standards and interoperability. He was established eLearning program course for GIS and remote sensing in Morocco. He has led and participated in many research projects and has authored and co-authored numerous academic chapter's books and articles. He was member of the program committee of several conferences in the area of geomatics and geospatial technology. He works with many companies to help them manage and implement GIS project.

**Ahmed FEKRI** is professor of a hydrogeology at Hassan II university, Faculty of Sciences Ben M'Sik. He has worked as a hydrogeological consultant at RESING consulting firm (a national pioneer firm consultancy in water resources area). He has hired at the University of Hassan II in Casablanca as an assistant professor in 1996. In 2007 Ahmed Fekri received his PhD on the impact of landfill leachate on groundwater resources at University Hassan II. In 2011 he obtained a Fulbright scholarship and he is a board member of the Moroccan Fulbright alumni association (MFAA). Since 2016 has been head of the "Applied Geology, Geomatic and Environment" laboratory.

Article submitted 20 September 2018. Resubmitted 22 November 2018. Final acceptance 23 November 2018. Final version published as submitted by the authors.