Towards open access of bed forms data, standardization of its analysis, and some steps to these ends

Ronald R. Gutierrez *University of the North, Barranquilla, Colombia – rgutierrezll@uninorte.edu.co*
Alice Lefebvre *MARUM, University of Bremen, Bremen, Germany – alefebvre@marum.de*
Francisco Núñez-González, *Leichtweiß-Institut für Wasserbau, University of Braunschweig, Braunschweig, Germany – f.nunez-gonzalez@tu-braunschweig.de*
Humberto Avila *IDEHA, University of the North, Barranquilla, Colombia – havila@uninorte.edu.co*

ABSTRACT: This contribution highlights the challenges and opportunities for the community of coastal and fluvial engineers and morphologists to consolidate a paradigm of open and data-intensive scientific culture through for example data sharing, formal recognition of scientists who collect the data, free accessible code, and standardized procedures to collect and analyze data. We stress that the technical challenges can be tackled through the body of knowledge being built on the matter by some institutions such as the Federation of Earth Science Information Partners. We also underline the potential of Bedforms-ATM to standardize some bed forms data analysis techniques and enumerate the applications that could be developed and incorporated into it in the short term.

1 INTRODUCTION

Many rivers and coastal areas around the World are facing increasing demands on both land and water resources for human settlement growth, navigability, and energy. Thus, there is a necessity for improving the prediction of flow and sediment transport for a wide range of river sizes and coastal seas. Bed forms are ubiquitous features in shallow and deep-water environments, having a strong influence on flow properties and sediment transport; as such, a better understanding of their dynamics is of relevance for engineers, geomorphologists, and planers (Best, 2005; ASCE, 2002).

Despite the significant improvement on the scientific understanding of bed form dynamics from field, laboratory and numerical investigations performed in the last decades, many aspects remain obscure (Allen, 1983; Best, 2005). The availability of more sophisticated equipment (such as multibeam echo sounder) and data analysis techniques played an important role in this evolution by providing a large amount of detailed measurements which can be analyzed to help improve our knowledge on the mutual interaction of bed forms and flow. However, these data may not always be freely available and are certainly not all analyzed in a standardized manner.

In recent years, the necessity for the scientific community to move towards a paradigm for open and free sharing of scientific data and software has been underlined (McNutt et al., 2016). Although some disciplines (e.g., astronomy and oceanography) have a long tradition of sharing data, we believe that it is fair to state that it is still a pending challenge for the community of engineers and scientists devoted to the analysis of river and coastal dynamics, as well as the community dealing with bed forms dynamics, which regularly meets at the MARID conference.

The objectives of this contribution are twofold, namely: [1] to discuss on the challenges and opportunities of open and free sharing of bed form data based on similar experiences from other disciplines, and [2] to identify the applications that can be added to Bedforms-ATM that potentially would help improve our understanding of bed form dynamics and/or standardize bed form data analysis.

2 CHANGING OUR PARADIGMS

2.1 The opportunity

To date, two paradigms are becoming more prevalent in scientific research, namely: openness, and data-intensive scientific approach, which may even turn into a big-data-intensive scientific approach soon. Openness [1] enhances the productivity and efficiency of research by preventing to repeat scientific studies; [2] is essential to the validation of hypothesis, theories, data, and results; and [3] helps to promote trust among scientists by fostering cooperation and collaboration. Openness demands not only allocating information but also allocating the necessary resources to understand, validate, and apply information such as data, results, methods, and tools (Resnik, 2006; McNutt et al., 2016).

Data-intensive scientific approach has called the attention of the scientific community and has become a new opportunity and incentive to knowledge discovery because it expands the correlations among multi-disciplinary data, which subsequently triggers the discovery of new models, new rules, and new knowledge (Hey et al., 2009; Guo et al., 2017). Some researchers (e.g., Peng et al., 2016) even propose that "universities, research institutions, and funding agencies should develop new measures to evaluate a research project's success not only based on publications and other outcomes it produces but also based on the amount and quality of data it makes available for the wider user community and society."

As it is the case in many disciplines, we believe that the aforementioned paradigms are not currently fully present in the coastal and river geomorphology community, and thus, it represents a pending challenge.

2.2 The challenges

Some cultural, institutional, and technological constraints limit openness in many scientific disciplines and possibly, by extension, the evolution towards big-data-driven science. Based on the experience of some of the co-authors of this contribution, these limitations are more deeply present in the scientific community from developing countries. Cultural and institutional constraints will possibly be forced to change as many funding agencies and peer-reviewed journals have adopted regulations that require scientists to share data, results, methods, and tools (Resnik, 2006).

Data are often shared within a working group or with close collaborators before being published and declared to be openly available after publication (Buys et al 2015). This step of openly sharing of the data however is rarely completed. For instance, less than 1% of the collected ecological data is accessible after publication (Reichman et al., 2011), even though important discoveries were made by integrating large data sets from many systems, for example in ecology, and the potential benefits of data reuse and open data citation are widely recognized (Hey et al., 2009; Lindenmayer & Likens, 2013, Piwowar & Vision, 2013). Other disciplines, such as genomics, have shared repositories, chiefly due to the homogeneity of their data (Reichman et al., 2011). Past research has underlined the necessity for having access to large amounts of bed form data from well-documented theoretical and experimental case histories, and the need for integrated interdisciplinary studies to fully understand the morphodynamics of bed forms (Allen, 1983; Dalrymple and Rhodes 1995; Best, 2005).

Some researchers (e.g., Kwoll et al., 2014; Lefebvre & Winter, 2016; Winter et al., 2016; Gutierrez, 2017) have made bed form data publicly available. For instance, Gutierrez (2017) published synthetic bed form data, which contains ripple, dune, and bar-like features that can be used to assess bed form hierarchization techniques. Likewise, Gutierrez et al. (2018) reported the development of Bedforms-ATM, a free available software for bed forms data analysis, which also provides field bed forms data from the Parana River, Argentina.

We believe that two major technological challenges will arise in the effort to changing our paradigms, namely: [1] dealing with the heterogeneity of bed form data that results from the lack of common experimental practices, field measurement standards, and data analysis (Reichman et al., 2011; Gutierrez et al, 2013); and [2] tracking the provenance of data derived from original data sets, and the scientific outcomes stemming from them through quality control, analysis, and modeling (Reichman et al., 2011).

Despite the potential benefits of having access to massive bed form data collections, we must be aware that it might induce the proliferation of inductive science (i.e., developing research questions after having data), which, although a valid research method, may contradict the current workflow of science. It also opens the door for the existence of scientists who might hardly be motivated to gather data, because it is time and resource consuming, and who might simply take data gathered by others (Lindenmayer & Likens, 2013). This might be prevented nonetheless by following a good practice that those using open-access data sets must work in close collaboration with those who collected these data sets through co-authorship, attribution, or citation (Lindenmayer & Likens, 2013). Overall, it is recognized that data sharing increases citation rate (Piwowar et al., 2007) and we believe that it would advance bed form research in general and profit to all, those who initially collect and analyze the data and those who subsequently reuse them.

2.3 Tackling the technological challenges

Data need to be stewarded throughout the entire data lifecycle, i.e. from data collection, to management of active data sets, to long-term archive. However, most disciplines still lack the technical, institutional, and cultural frameworks to support open data access (Peng et al., 2016).

The Federation of Earth Science Information Partners (ESIP) aims at making earth science data more discoverable, accessible, and usable. In this vein, the ESIP Data Stewardship Committee has provided a set of recommendations, best practices, and guidelines allowing to influence data management carried out by government agencies and other data stewards (Downs et al., 2015). ESIP proposed a provenance and context content standard (PCCS) which lists all content items required to fully represent the provenance and context of the data products resulting from earth science missions, namely: content item name, descriptive definition, rationale (why a given item is needed), criteria (how good the content should be), priority, user community (who would most likely use the item), source, project phase capture, representation (word files, numeric files, etc.), and distribution restrictions (e.g., proprietary concerns). PCCS presented these items in a matrix that is considered a good starting point for developing a standard to offer guidance for data producers, data managers, and others (Ramapriyan et al., 2012). ESIP has stated its openness to apply PCCS standards to other types of data and has encouraged the organizational and individual membership-based participation of earth science data providers (Downs et al., 2015). ESIP has also tested the data stewardship maturity matrix developed by Peng et al. (2015), which can be applied by data centers and other data-holding organizations.

We believe that the MARID conferences could be a unique platform to discuss on the creation, management, distribution, use, and citation of bed form data, which eventually might lead to devising an organization that can work in partnership with, for instance, ESIP. We are aware that for this end, funding, a cooperative attitude from the community of engineers and scientists, and the collaboration with public and private institutions, nongovernmental organizations, among others, will be necessary.

The bed form data set that can be potentially built does not necessarily have to be a single system. Instead, it can be made up of centralized multiple crowdsourced open-access data entities. Large complex platforms such as the Digital Earth are usually built this way (Guo et al., 2017). It is expected that stewarded bed form data would potentially encompass heterogenous, multi-source, multi-temporal, multi-scale, high dimensional, highly complex, and unstructured geospatial data sets, which are typical characteristics of many geophysical signals datasets (Nativi et al., 2015).

3 TOWARDS BED FORM DATA ANALYSIS STANDARDIZATION

With more openly accessible bed form data, scientists will require more complex processing and data analysis tools. In this regard, code used to analyze, and process data will be a fundamental requirement for transparency and reproducibility (McNutt et al., 2016). We believe that the Bedforms-ATM platform by Gutierrez et al. (2018) could potentially be used to build such code.

Bedforms Analysis Toolkit for Multiscale Modeling (Bedforms-ATM) currently comprises the following applications: [1] Bed forms wavelet analysis; [2] Power Hovmöller analysis; [3] Bed forms multiscale discrimination, which discriminates bed form fields into three scale-based hierarchies (e.g., ripples, dunes, bars); and [4] Three-dimensionality analysis, which quantifies the three-dimensionality of bed form fields. Herein we enumerate the applications that can be incorporated into Bedforms-ATM in the short term for the analysis of both marine and fluvial bed forms.

3.1 Atomization of bed form fields

The atomization of bed form fields, i.e. the identification and extraction of single bed form entities from bed form fields, and subsequent quantification of its geometric characteristics (e.g., stoss and lee side slopes, wavelengths, and amplitudes) provides information on the interactions of flow field, and suspended sediment (Best, 2005). Some researchers (e.g., van der Mark, 2008; Gutierrez et al., 2013) have already presented methodologies to perform the atomization of bed form fields. These methods are easily reproducible or openly available and can therefore be used to standardize the atomization of bed form fields. They could be incorporated into Bedforms-ATM and be expanded to include, for example, a three-dimensional atomization.

3.2 Bed form statistical analysis

Bed forms atomization is also necessary for identifying fully developed bed form fields in experimental environments and quantifying the variability of natural bed forms through statistical analysis (van der Mark, 2008; Gutierrez et al., 2013; Perillo et al., 2014; LeRoy et al., 2016). There is a body of evidences that suggest that large rivers are characterized by bed forms with leeside slope lower than the angle of repose. Thus, Best (2005) stated that the study of low-angle bed forms constitutes one of the main future research topics in the understanding of bed forms dynamics. To this end, sharing data from worldwide large rivers and standardizing bed forms atomization will be needed. A statistical analysis of bed form fields and their characteristics will enable a better characterization of their properties and provide possible explanation of their dynamics.

LeRoy et al. (2016) reported the development of SABAT (Slope-aspect bedform analysis tool), a Matlab tool that does not identify superimposed bed forms but performs a variety of statistical analyses on bed form' wavelengths and amplitudes. However, to the best of our knowledge, SABAT is not publicly available. Van der Mark et al (2008) quantified the variability in bedform geometry from a series of bathymetric measurements in the lab, in a small river (0.25 m water depth) and the Rhine (8 m water depth). Their analysis could be expanded and deepened in a variety of environment, especially in diverse large and small rivers, in order to statistically describe bed form parameters and their variability.

3.3 Characterization of experimental bed forms

Dumas et al. (2005) proposed a discrimination scheme for experimental bed forms resulting from oscillatory and combined flows. It is based on a wavelength threshold of 0.5 m for characterizing symmetrical small ripples, asymmetrical small ripples, symmetrical large ripples, and asymmetrical large ripples. This scheme has been successfully used in the past (e.g. Perillo et al., 2014). We believe that after developing a methodology for atomizing bed form fields, the scheme of Dumas et al., (2005) can be used to: [1] perform statistical analysis over single entities of bed forms, and [2], provided that enough information is available, study the time it takes for bed forms to become fully developed, which is of special concern for the understanding of bed forms morphodynamics (Allen, 1983; Best, 2005; Doré et al., 2016).

3.4 Characterization of marine bed forms

Bed forms in marine environments have a great variety of dimensions and shapes, for example small-scale ripples, tidal dunes and sand waves. They are found at a wide range of depth, from the intertidal zone to the continental rise and are subjected to diverse hydrodynamic forcings (e.g., regular and storm waves, and tidal, wave-induced or contourite currents). They also form in diverse sedimentary settings such as sand, mixed sediment or sediment starved. Extensive bed form fields are now well-known and the control of their morphology by environmental parameters is better understood (e.g. Damen et al, 2018). However, Garlan et al. (2016) invites for a renewed classification of marine bed forms following the recent improvement of bed form mapping and characterization. This could be done in the framework of a large collaboration bringing marine bed form data together to be analyzed in a standardized and comprehensive way.

4 CONCLUSIONS

Scientific openness, and data-intensive science are becoming more important in today's scientific inquiry; however, they are not currently prevalent in the community of fluvial and coastal morphologists and engineers. We believe that these paradigms open a myriad of possibilities to better understand the spatio-temporal mechanisms that govern the dynamics of bed forms, which, at present, remain obscure.

A change towards scientific openness will pose some technical challenges such as handling copious data, which involves transferring, storing, managing, processing, computing, and sharing such data. We believe that MARID represents a unique platform to discuss on the opportunities and challenges related to changing our scientific paradigm.

Bedforms-ATM can potentially be used as a common platform for standardizing bed form data analysis methodologies via the contribution of river and coastal engineers and scientists. In our opinion the following applications could be developed on the Bedforms-ATM platform in the short term: [1] an application to atomize bed form fields (i.e., identifying single bed form entities), [2] an application to perform statistical analysis over single bed form entities, [3] an application to discriminate experimental bed forms based on the Dumas et al. (2005) classification scheme, and [4] an application to classify marine bed forms.

5 ACKNOWLEDGEMENT

We acknowledge the Universidad del Norte and the IDEHA research group for funding this contribution. Alice Lefebvre is founded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), project number 345915838.

6 REFERENCES

Allen, J. R. L. (1983). River bedforms: progress and problems. Modern and ancient fluvial systems, 6, 19-33.

ASCE Task Committee on Flow and Transport over Dunes (2002). Flow and transport over dunes. Journal of Hydraulic Engineering, 128(8), 726-728.

Best, J. (2005). The fluid dynamics of river dunes: A review and some future research directions. Journal of Geophysical Research: Earth Surface, 110(F4).

Buys, C. M., Shaw, P. L. (2015). Data Management Practices Across an Institution: Survey and Report. Journal of Librarianship & Scholarly Communication, 3(2).

Dalrymple, R. W., & Rhodes, R. N. (1995). Estuarine dunes and bars. In Developments in sedimentology (Vol. 53, pp. 359-422). Elsevier.

Damen, J. M., van Dijk, T. A. G. P., & Hulscher, S. J. M. H. (2018). Spatially varying environmental properties controlling observed sand wave morphology. Journal of Geophysical Research: Earth Surface,

Doré, A., Bonneton, P., Marieu, V., & Garlan, T. (2016). Numerical modeling of subaqueous sand dune morphodynamics. Journal of Geophysical Research: Earth Surface, 121(3), 565-587.

Downs, R. R., Duerr, R., Hills, D. J., & Ramapriyan, H. K. (2015). Data stewardship in the earth sciences. D-Lib Magazine, 21(7/8).

Dumas, S., Arnott, R. W. C., & Southard, J. B. (2005). Experiments on oscillatory-flow and combined-flow bed forms: implications for interpreting parts of the shallow-marine sedimentary record. Journal of Sedimentary research, 75(3), 501-513.

Garlan T., Brenon E., Marchès E., Blanpain O., 2016. From regional variability of the morphology of dunes to a new method of their classification. Proceedings of Marine and River Dune Dynamics – MARID V – 4 & 5 April 2016 – North Wales, UK, 4pp

Guo, H., Liu, Z., Jiang, H., Wang, C., Liu, J., & Liang, D. (2017). Big Earth Data: a new challenge and opportunity for Digital Earth's development. International Journal of Digital Earth, 10(1), 1-12.

Gutierrez, R. R., Abad, J. D., Parsons, D. R., & Best, J. L. (2013). Discrimination of bed form scales using robust spline filters and wavelet transforms: Methods and application to synthetic signals and bed forms of the Río Paraná, Argentina. Journal of Geophysical Research: Earth Surface, 118(3), 1400-1418.

Gutierrez, R. R. (2017): Synthetic data for the Bedforms Analysis Toolkit for Multiscale Modeling (Bedforms-ATM). PANGAEA, https://doi.org/10.1594/PANGAEA.873304.

Gutierrez, R. R., Mallma, J. A., Núñez-González, F., Link, O., & Abad, J. D. (2018). Bedforms-ATM, an open source software to analyze the scale-based hierarchies and dimensionality of natural bed forms. SoftwareX, 7, 184-189.

Hey, T., Tansley, S., Tolle, K. M. (2009). The fourth paradigm: data-intensive scientific discovery (Vol. 1). Redmond, WA. Microsoft research.

Huang, N. E., & Wu, Z. (2008). A review on Hilbert Huang transform: Method and its applications to geophysical studies. Reviews of geophysics, 46(2).

Kwoll, E., Becker, M. and Winter, C. (2014). With or against the tide: The influence of bed form asymmetry on the formation of macroturbulence and suspended sediment patterns. Water Resources Research, 50(10), pp.7800-7815.

Lefebvre, A. and Winter, C. (2016). Predicting bed form roughness: the influence of lee side angle. Geo-Marine Letters, 36(2), pp.121-133.

LeRoy, J.Z., Rhoads, B.L., Best, J.L., Cisneros, J. 2016. Bed morphology and sedimentary dynamics at chute cutoffs: A case study of Mckey Bend, lower Wabash River, IL-IN. River Flow 2016: Iowa City, USA, July 11-14, 2016, 469-481.

Lindenmayer, D., & Likens, G. E. (2013). Benchmarking open access science against good science. The Bulletin of the Ecological Society of America, 94(4), 338-340.

Masselink, G., et al. (2007). Geometry and dynamics of wave ripples in the nearshore zone of a coarse sandy beach. Journal of Geophysical Research: Oceans 112.C10 (2007).

McNutt, M., Lehnert, K., Hanson, B., Nosek, B., Ellison, A., and King, L. (2016). Liberating field sciences samples and data. Science, 351 (6277). Doi: 10.1126/science.aad7048.

Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M., & Ochiai, O. (2015). Big data challenges in building the global earth observation system of systems. Environmental Modelling & Software, 68, 1-26.

Peng, C., Song, X., Jiang, H., Zhu, Q., Chen, H., Chen, J. M., ... & Zhou, X. (2016). Towards a paradigm for open and free sharing of scientific data on global change science in China. Ecosystem Health and Sustainability, 2(5), e01225.

Peng, G., Privette, J. L., Kearns, E. J., Ritchey, N. A., & Ansari, S. (2015). A unified framework for measuring stewardship practices applied to digital environmental datasets. Data Science Journal, 13, 231-253.

Perillo, M. M., Best, J. L., Yokokawa, M., Sekiguchi, T., Takagawa, T., & Garcia, M. H. (2014). A unified model for bedform development and equilibrium under unidirectional, oscillatory and combined flows. Sedimentology, 61(7), 2063-2085.

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. PloS one, 2(3), e308.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. PeerJ, 1, e175.

Ramapriyan, H., Moses, J., & Duerr, R. (2012). Preservation of data for Earth system science-Towards a content standard. In Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International (pp. 5304-5307). IEEE.

Reichman, O. J., Jones, M. B., Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. Science, 331(6018), 703-705.

Resnik, D. B. (2006). Openness versus secrecy in scientific research. Episteme, 2(3), 135-147.

van der Mark, C. F., Blom, A. and Hulscher, J.S. (2008). Quantification of variability in bedform geometry. J. Geophys. Res., 113, doi:10.1029/2007JF000940.

Winter, Christian; Lefebvre, Alice; Benninghoff, M; Ernstsen, Verner Brandbyge (2016): German Bight bedform reconstructions. PANGAEA, https://doi.org/10.1594/PANGAEA.858716