Table 1
Types of anomalies likely encountered in in situ sensor-generated water-quality time series, along with the importance ranking of each type with respect to the priority end-user goal in this case study (i.e. time series visualization), and relevance to potential end-users.

Anomaly type	Type code (Class[b])	Description	Examples in the literature and/or alternative terminology	Importance ranking (with respect to time series visualization in this case study)	Potential end-users and applications[c]
Large sudden spike	A[a] (1)	Anomalous value is isolated and 'much' higher or lower than surrounding data, and the spike occurs in a very short window of time (e.g. only one data point is anomalously high or low).	Point or collective anomaly (Goldstein and Uchida, 2016)	First priority (at any point in the time series)	Management, monitoring and compliance; Policy and decision makers; Public; Data managers; Sensor maintenance technicians
Low variability/ persistent values	B (3)	Values constant though time or with very minimal variation compared with that expected	Data value persistence (Horsburgh et al., 2015); collective anomaly (Chandola et al., 2009)	Second priority (especially during event flow)	Data managers; Sensor maintenance technicians
Constant offset (e.g. calibration error)	C (3)	Values are in error by some constant. Likely related to/seen before and/or after sudden shifts	Incorrect offset or calibration (Horsburgh et al., 2015)	Third priority	Data managers; Sensor maintenance technicians
Sudden shifts	D (1)	Values suddenly shift to a new range (higher or lower than previous range)	Level shifts (Tsay, 1988)	Equal third priority (especially when shift is considered large)	Management, monitoring and compliance; Policy and decision makers; Public; Data managers; Sensor maintenance technicians
High variability	E (3)	Values oscillate considerably over short time periods (more than expected during natural daily cycles or events)	Variance change (Tsay, 1988); collective anomaly (Chandola et al., 2009)	Fourth priority	Data managers; Sensor maintenance technicians
Impossible values	F (2)	Values impossible or highly unlikely for that water-quality variable (e.g. negative values for all, conductivity values nearing or at zero ('too fresh'))	Out of range values (Horsburgh et al., 2015)	Important, but should be detected easily (e.g. using a simple rule)	Sensor manufacturers; Statisticians; Data managers; Sensor maintenance technicians
Out-of-sensor-range values	G (2)	Values that the sensors are incapable of detecting (outside of their detection range). Some of these anomalies may be first captured under type F (impossible values)	Not distinguished from type F by Horsburgh et al. (2015)	Important, but should be detected easily (e.g. using a simple rule)	Sensor manufacturers; Statisticians; Data managers; Sensor maintenance technicians
Drift	H (3)	Gradual change in values in positive or negative direction	Sensor drift (Horsburgh et al., 2015); collective anomaly (Chandola et al., 2009)	Comparatively low priority (most likely observed in turbidity), but important to flag as being a possible occurrence of an anomaly e.g. when gradual increase or decrease occurs before a sudden shift	Sensor manufacturers; Data managers; Sensor maintenance technicians
Clusters of spikes	I[a] (1)	Multiple spikes in a short period of time	Micro cluster (Goldstein and Uchida, 2016); collective anomaly (Chandola et al., 2009)	Low priority (isolated spikes much more important to detect)	Management, monitoring and compliance; Policy and decision makers; Public; Data managers; Sensor maintenance technicians
Small sudden spike	J[a] (1)	Anomalous value is 'somewhat' higher or lower than surrounding data, and the spike occurs in a very short window of time (e.g. only one data point is anomalously high or low)	Point anomaly (Goldstein and Uchida, 2016)	Very low priority	Data managers; Sensor maintenance technicians
Missing values	K (2)	Gaps in time series (i.e. greater than the set frequency of measurement)	Skipped or no-data values (Horsburgh et al., 2015)	Undetermined	Data managers; Sensor maintenance technicians; Sensor manufacturers; Statisticians; Policy and decision makers

[a] Spikes may be in the positive or negative direction with respect to surrounding data (i.e. can include a sudden isolated decrease and/or a sudden isolated increase in value).
[b] Classes of anomalies, as defined in this paper: (1) involve a sudden change in value from the previous observation, (2) are detectable by automated classification rules, (3) likely require user intervention to identify observations as anomalous.
[c] Monitoring, management and compliance: agencies, industries and landholders etc. concerned with water quality monitoring, management and compliance checking – summary statistics such as means are strongly influenced by such anomalies; Policy and decision makers – to limit use of incorrect data and for reporting purposes; Public – to avoid false warning of water quality breaches; Data managers – for quality control and assurance and to increase confidence in the data by reporting the presence of such anomalies; Sensor maintenance technicians – to ensure timely and correct calibration and maintenance of equipment; Sensor manufacturers – to improve performance, e.g. extend battery life, improve wiper quality to further minimize biofouling; Statisticians – for AD methods to better detect other non-trivial anomaly types and/or for methods requiring regular and frequent observations.