Water Quality Digital Twin Survey

Dick, J. Dobel A., Fry M., Harrison S., Qu Y., Khamis D, Keller V., Thackeray S. J.

Client Ref: NE/R016429/1
Issue number v1
01/11/2023

UK Centre for Ecology & Hydrology

Contents

Executive Summary 3

1. Background to survey 4

2. Method 5

3. Results 8

3.1 Response time and respondent type 8
3.2 Feasible spatial scale of a water quality digital twin 10
3.3 Feasible temporal scale of a water quality digital twin 15
3.4 Major obstacles to creating a digital twin of running and standing water quality 20
3.5 Most important determinand in a water quality digital twin 28
3.6 Output variables considered as most important water quality 36
3.7 Respondents' views on output from a digital twin of water quality 41
3.8 Additional comment from respondents 47

4. References 49

Annex 1 Survey questions 50

Annex 2 Participant information 51

Annex 3 Survey results 52

Executive Summary

Stakeholders' views on the creation of a Water Quality Digital Twin, for assessing the impacts of multiple stressors on standing and flowing freshwaters, were sought through an anonymous survey composed of closed and open questions. The work was funded as part of the UKSCAPE integration fund (https://uk-scape.ceh.ac.uk/about) and sought to co-develop a blueprint for a UKCEH water quality digital twin. Fifty-nine participants responded and provided a wealth of viewpoints, from academia, industry, regulators, and policy makers' perspectives.

In general, the catchment scale was considered the most feasible, useful, realistic, and deliverable scale for a Water Quality Digital Twin. Respondents considered the desired temporal scale to be dependent on the use envisaged for the digital twin. However, sub-daily or daily scales emerged as providing the most actionable knowledge if data were available and it was computationally feasible.

Overall, there was consensus that nutrient concentrations were the most important determinands to include in a water quality digital twin (100% of respondents scoring these as one of their three most important determinand categories). However, several additional abiotic and biotic determinands were also scored highly, partly depending on the required use of the resultant digital twin.

Predictions of a combination of physico-chemical and biological variables were considered the most important output for a water quality digital twin but, if forced to select only one output type, then physico-chemical variables were judged the most important (52% of respondents). This was in part because of a view that other outputs could then be calculated or inferred. While human wellbeing and financial impact as output variables were considered very important for decision making, they were recognised as more difficult to model.

An interactive web portal with visualisations of predictions was selected by the majority (92%) as a suitable means for end users to interact with, and receive output from, a digital twin of water quality. However, most respondents favoured several formats. From the written rationale provided for their selection, some respondents wished to have the data interpretated for them while others simply wanted access to the data so they could conduct further analysis themselves.

Almost twice as many respondents considered a lack of real time data and models on water quality to be more important obstacles to digital twin implementation than the equivalent for water quantity. The need for real time data on both water quality and quantity was considered by the majority of respondents to be a major obstacle to creating a fully functional digital twin of running and standing water quality.

1. Background to survey

Fresh waters supply several essential services to society, such as water supply for various sectors (including drinking water), agriculture, manufacturing, the leisure industry (e.g. bathing water), and habitats for fauna and flora (Millenium Ecosystem Assessment, 2005). Despite their overwhelming importance, fresh waters around the world are under increasing pressure due to the interacting effects of climate change, pollution, overexploitation, and socio-economic change (Tickner et al., 2020). As a result, in Europe, only 40% of surface water bodies are achieving good ecological status, as required under the Water Framework Directive (European Environment Agency, 2018). In the UK, this figure drops down to only 14% (House of Commons, 2022). It is therefore vitally important that we increase our understanding of freshwater ecosystem dynamics under current conditions and apply this enhanced understanding to make projections of likely future change and scenarios. Armed with such understanding, we would greatly facilitate adaptive management of freshwater resources and biodiversity.

Over recent decades, scientists across the UK and beyond have collected a wealth of environmental data using an ever-increasing array of approaches and technologies (Thackeray & Hampton, 2020). Long-term and high-frequency monitoring networks have been established and maintained, gathering invaluable evidence on the changing state of fresh waters over time. Such data not only represent direct observations of processes and states in fresh waters but are also essential to inform and drive a variety of process-based models.

Process-based models are valuable tools in ecological research and management, providing digital representations of real-world processes that allow the investigation of impact scenarios in “virtual experiments”. However, such models also have recognised limitations. Typically, models only capture specific aspects of the freshwater environment (e.g. rivers or lakes, droughts or floods, hydrology or water quality), or they focus on only specific aspects of environmental stress (e.g. point vs. diffuse sources of pollutants, macronutrients, metals). Furthermore, the structure and parametrisation of process-based models is typically static in time, limiting their ability to capture the constantly changing nature of the environment. Finally, the use of data for the single purpose of process-based modelling (driving data, calibration, validation) is not an optimal use of the wealth of data available to scientists; new “big data” technologies can also unlock new knowledge from this information and facilitate efforts towards increasing the resilience of freshwater ecosystems.

Over recent years, scientists have explored a new technological paradigm, the digital twin concept, to provide improved modelling capabilities that would enable

more accurate forecasting, for potential use in decision making. The commonly accepted definition of a digital twin is a *virtual representation of a system that is constantly updated to accurately represent the current state and behaviour of the system*. These systems allow the integration of the plethora of information and technology available to scientists including monitoring observations, remote sensing data, process-based and data-driven models. Crucially, digital twins include feedbacks on the way that we interact with the real environment (Siddorn et al., 2022).

To maximise the benefits to, and uptake of, the digital twin application by the wider community, we have undertaken work to engage and identify the needs of stakeholders. This work was conducted under the *Develop a blueprint for a UKCEH water quality digital twin (WaDiTi)* project, funded under the NERC funded UK-SCAPE project (grant number NE/R016429/1). The project aims to deliver through three inter-linked work packages (Figure 1), which will bring together stakeholder community needs with in-depth understanding of digital twin architectures, data, and models.

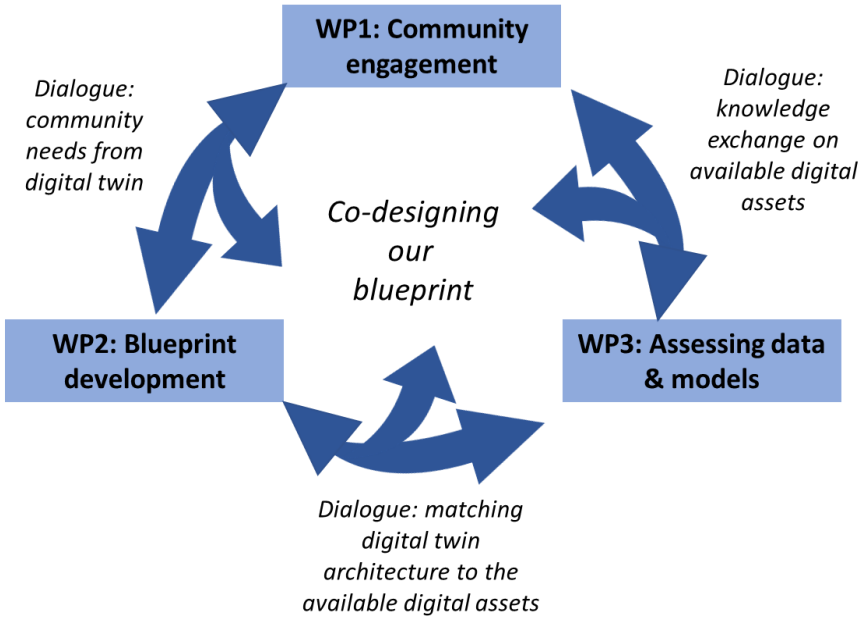


Figure 1. The WaDiTi project structure

This report details the work conducted in the initial stakeholder consultation in work package 1 (WP1) of the project.

2. Method

The questionnaire was designed by all members of the team to capture opinions on fundamental considerations when creating a blueprint for a water quality digital twin

including spatial and temporal scale, determinands to consider, target audience, outputs in terms of content and format, foreseen risks and limitations (see Annex 1 for full questionnaire). The questionnaire was created using UKCEH approved software JISC, the online survey tool designed for academic research, education, and public sector organisations. The questionnaire was distributed UK wide, using a snowball approach whereby the team encouraged respondents to share it with peers that might be interested in the topic, i.e., invitees invite others. The link to the questionnaire was embedded in the invitation. Box 1 contains the standard email invitation which was tailored by team members depending on their relationship with the network or individual invited.

A link to the questionnaire was shared through personal contacts of UKCEH staff and via learned societies and other groups as appropriate. Consequently, no database of invitees or respondents was stored.

Box 1 Example invitation text.

Dear freshwater colleague,

Water quality and freshwater biodiversity are challenged by a diverse array of human-induced stressors, and the ways in which fresh waters respond to these threats are highly complex, impacting their status, management, and restoration.

To allow better management and restoration of freshwaters a new approach called digital twinning is progressing. Put simply, a **Digital Twin** is a virtual representation of a real system that is constantly updated to accurately represent the current state and behaviour of that system.

UKCEH have been asked to create a blueprint and development plan for a surface water quality Digital Twin and we are looking for **your input** to help us understand the needs and requirements of the stakeholder community for such a system.

To gather your thoughts and needs, we have put together a small ~10 min survey, which is open until mid-August can be accessed here.

It is anonymous; however, we do ask three non-obligatory questions about survey respondents to enable reporting the range of responds who contributed to the development of the blueprint (I.e. type of organisation/institute, age group, gender). **No individual contact details are sought**.

We have also attached the **Participant Information Sheet**, which provides full details of the survey and how your data will be stored and used.

Please pass on the link to all relevant colleagues or contacts so we can report the communities needs and aspirations in the exciting development of a Water Quality Digital Twin.

It is not possible to determine the response rate, but it is estimated that over 8,000 invitations were distributed. We accept that many people may have been invited more than once as they belong to several relevant networks (Table 1).

The nature of the survey was briefly explained in the Participant Information Sheet, along with the purposes for which the data would be used (Annex 2).

Table 1. Organizations and networks invited to complete the questionnaire, using a snowball approach.

Personal message to groups/lists with request to forward to interested colleagues	**Estimated reach**
Digital Twin Hub (https://digitaltwinhub.co.uk/)	4500
Alan Turing Institute internal Slack Channel	1832
British Hydrological Society (email to distribution list)	800
Aquatic Ecology Special Interest Group, BES (email to distribution list)	744
Turing Environment and Sustainability Slack (public)	379
CEEDS (email to distribution list)	239
Rivers Trust CaSTCo Slack channel (Includes all water companies)	149
NERC Constructing a Digital Environment Expert Network	79
Water Resources Science Area UKCEH Teams channel (personal message to list)	74
Lake Ecosystem Group, UKCEH (personal message to group)	16
Freshwater Restoration & Sustainability Group (incl. students) UKCEH Edinburgh	10
Rivers Trust CaSTCo project (personal message to group)	8
Personal message to individuals with request to forward as appropriate	
Deputy Chief Scientific Adviser Defra,	1
Chief Scientific Adviser, Environment Agency	1
James Hutton, Head of groups	3
WaterAware Collective	2
Hydronation and Stirling University	1
Floods & Water Research, Defra	1
United Utilities	1
Severn Trent Water	1
Thames Water	1
UK & Ireland Lakes Network	1
Floods and Droughts Research Initiative	1
Scottish Freshwater Group	1
CREW	1
Broads Authority	1
Natural England	1
Freshwater Biological Association	1
Environment Agency	1
Scottish Water	1
Leeds PhD student	1

3. Results

3.1 Response time and respondent type

A total of 59 responses were delivered between 09/06/2023 and 26/07/2023 (Figure 2). One participant started the survey one evening and apparently did not complete until 02:18 the following morning, some 5.5 hours later. The responses appear valid, so the answers were accepted as useable, but omitted from the response time calculations. On average the participants took 22 min to complete the survey with one participant taking only 4:36 min and another leaving the survey open for 1:34 min.

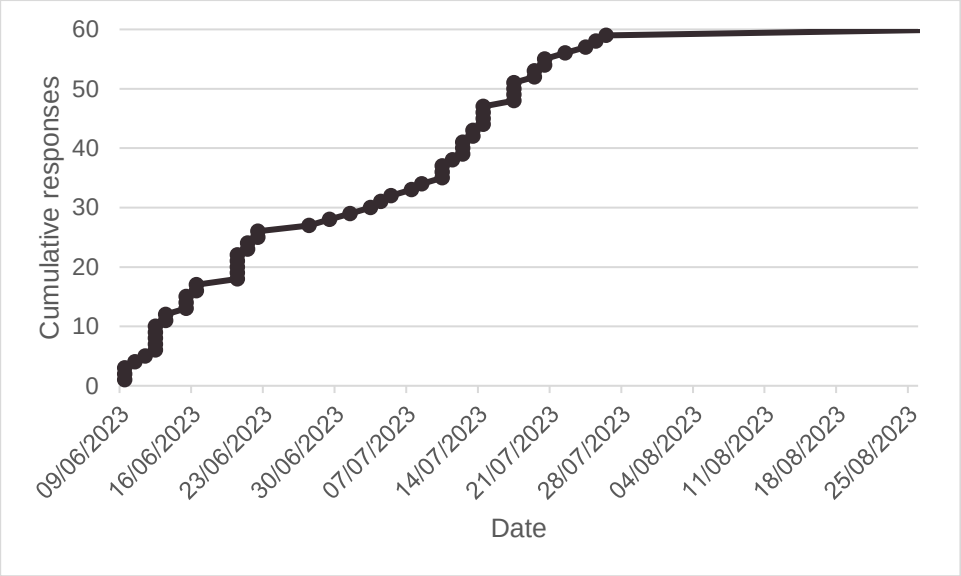


Figure 2. Cumulative response to survey between June and August 2023

Some 40% of the respondents declared affiliation with the research institute and/or university sector (Figure 3). Non-Governmental public bodies (NGOs), water companies and government agencies were roughly equally represented, with only five respondents declaring an affiliation with a commercial business, and only two representatives of government responding to the survey. Six respondents declared that they worked for more than one type of organization, with three stating they were affiliated with an NGO and a research institute and/or university, while one respondent stated that, in addition, he was also affiliated with a government agency. One respondent declared he was affiliated with an NGO and was a consultant. Finally, another respondent declared multiple affiliations, reporting connections with a water company, a commercial business, and a research institute and/or university.

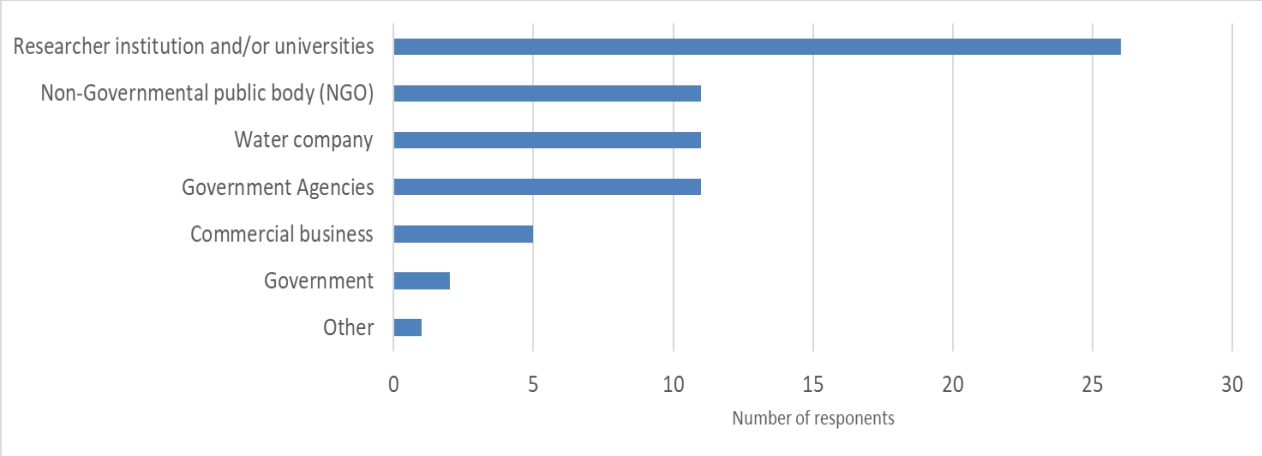


Figure 3. Affiliation of respondents

The respondents represented the typical bell-shaped distribution regarding age, with no respondents below 20 years of age or over 70 years (Figure 4). Respondents identifying as male and female were almost equally represented in the cohort of respondents who provided data (25 and 26 male and females respectively). Two responds declined to answer the questions concerning age and gender.

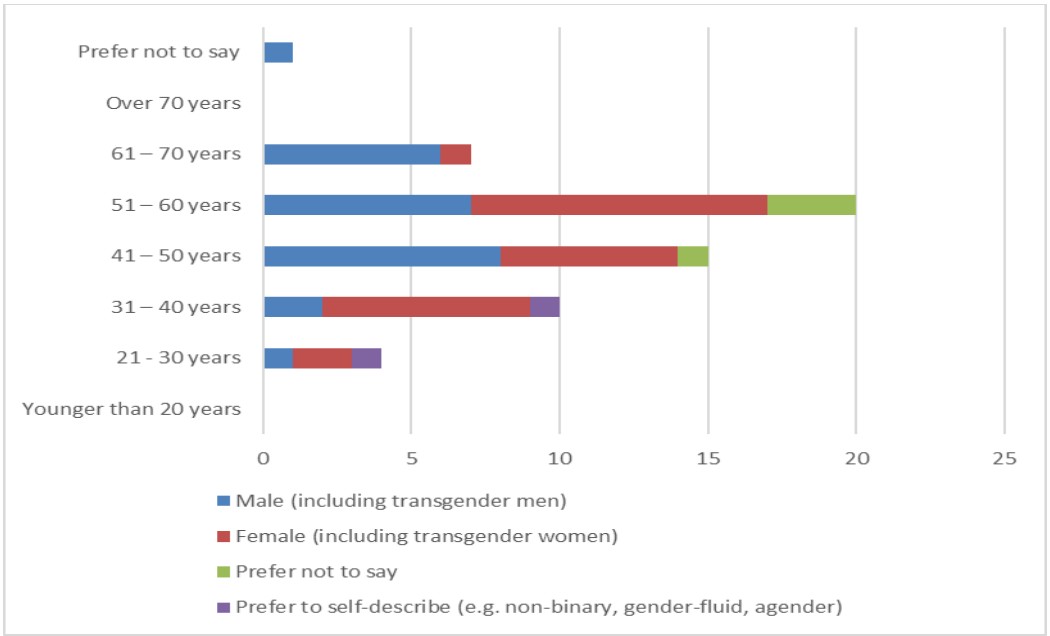


Figure 4. Age and gender of respondents

It is concluded that, in terms of age and gender, the cohort of respondents are likely to be representative of those working in the sectors relevant to the creation of a water quality digital twin. A full summary of the results is provided in Annex 3.

3.2 Feasible spatial scale of a water quality digital twin

The majority (75%) of respondents who expressed a single scale preference considered the catchment scale was the most feasible, useful, realistic, and deliverable scale for a Water Quality Digital Twin. Only 4% and 6% selected the local or regional scale respectively, while 16% selected national scale. However, eight respondents (14%) considered that there was not a single feasible, useful, realistic, and deliverable scale for a Water Quality Digital Twin (Figure 5), instead considering delivery at multiple scales desirable.

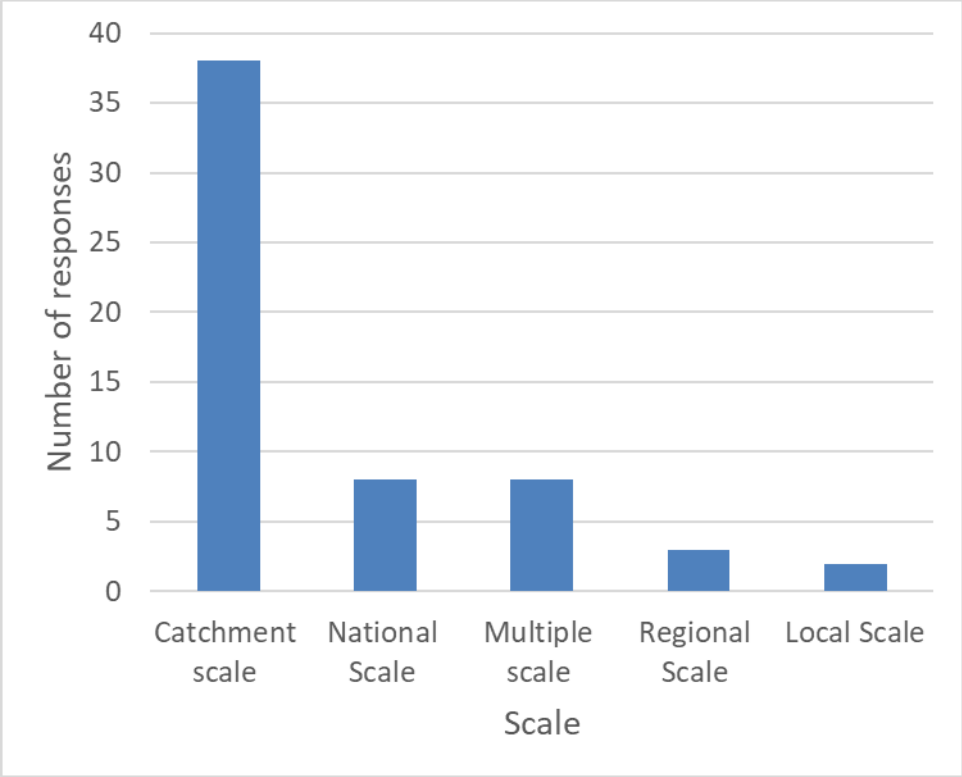


Figure 5. Respondents' views on the most feasible, useful, realistic, and deliverable spatial scale for a Water Quality Digital Twin

The foremost rationales for preferring the catchment scale reflected that it was "*the usable spatial scale"* and *"more achievable to deliver",* reflecting the "*specific nature of each catchment"* and "*provides the highest resolution of data/information upon which to act".* In addition, respondents considered that "*Catchment-scale mirrors the currently active river basin management planning process"* and it was recognised that "*catchments often transcend national or even regional boundaries",* noting *"management that doesn't recognise this is insufficient"* (Table 2).

Respondents who selected national as the single most relevant scale tended to focus on policy objectives e.g. "*Needs to be appropriate for government policy at that scale"* and *"Decision making is needed at national scale".* While others also

commented "*It might have more impact if you can do nationally"* and *"It will be crude so a broad national picture might be most useful".*

Only three people selected regional scale as the single most appropriate, considering it "*a good middle ground; balance between usefulness and feasibility"* and *"probably more relevant to a wider audience … likely more manageable than catchment scale but will still provide enough detail that potentially would be lost if this were at a national scale. A happy medium".*

Two individuals wrote that they considered local scale most appropriate. One considered that "*we need to be nowcasting at a scale relevant to a public thinking about recreation, aesthetics, and environmental impacts on a very local scale"*. The other respondent considered that the scale of delivery should be "*Local, individual rivers"* because they "*frankly would rather have real time monitoring in place than a model approach".*

Those who selected multiple scales observed that "*a mix of scales is required because sources (and therefore potential mitigations) are often small scale (especially point source) while effects can be regional".* One respondent suggested that "*a digital twin should be scalable and appropriate across all the different scales mentioned. only answering one would, in my opinion, go against the definition of a digital twin".*

In conclusion there was a desire for scalable Water Quality Digital Twin which would provide local mitigation knowledge while simultaneously enable national policy formulation. However, if this was not possible then catchment scale was considered the most suitable.

Table 2. Rational, as expressed by respondents, for their desired spatial scale for a Water Quality Digital Twin. Note only spelling/typographical errors corrected, speech and grammatical integrity maintained as delivered by respondents.

Catchment scale
Catchment scale work is what most agencies are now working to so this would seem appropriate
An ecosystem is a complex system: to depict it with enough accuracy, there is the necessity to medium-high resolution data and models
I feel it is important to start small before scaling up
Sufficient driving data could be collected at that scale
That is the usable spatial scale for water quality
It seems the most sensible.
Because you would want to tweak things such as catchment management practice and so on. Sub-catchment scale would be even better

National would be ideal but realistically due to severe data limitations I think there would be a large amount of work required to gather the necessary data to create an accurate and useful digital twin. I think it should be built upon catchment by catchment up to regional and national scale.
catchments often transcend national or even regional boundaries, management that doesn't recognise this is insufficient
Water quantity and quality in catchment-based scale is generally homogeneous.
Due to the specific nature of each catchment
This is a highly ambitious project. I selected catchment scale as I felt it would be challenging but more achievable to deliver for pilot catchments. This would then enable user feedback and also assist scaling up to national scale- which would be the ultimate objective.
I generally work on individual developments, having a catchment scale model could lead to the ability to define site specific pollutant treatment approaches.
Full pathway needed
1) technically feasible - connecting models and sensors required for these systems will be complex and starting small is probably a good idea 2) Predictions can be readily implemented into decision support for everyday management of systems e.g. reservoir or lake manager changing abstraction depth or providing warning for harmful algae or risk to fish from DO decline 3) cost - starting at a small scale will help to ensure the full digital twin can be implemented within a realistic budget - we should learn lessons from previous projects e.g. Demonstration Test Catchments where funding only covered equipment and not interventions initially
Think it provides the highest resolution of data/information upon which to act.
In an ideal world, a Water Quality Digital Twin would exist for every waterbody in the UK so National scale would be useful. However, without knowing what resource is available to deliver the Water Quality Digital Twin, I can only assume it is going to be done with the resource of a typical research group. In this case, catchment-scale seems most feasible, realistic and deliverable to gather the input data required for modelling, update forecasts and generate outputs.
This is the spatial scale for catchment planning and the majority of stakeholders we work with would recognise the catchment as a physical space in which their interest is most tangible
For immediate action and impact using bottom up approach
It needs to be realistic and national scale will require many compromises and assumptions which could be more refined at a local level. Catchment is the building block of hydrological units.
Water exists and (mostly) interacts with other elements of biota, including humans at a catchment scale
This fits with RBMP spatial scale and also makes good scientific sense as processes and pressures operate at a catchment scale
More actionable, good to test approach at relatively small scale first
Catchment is the most tractable to be able to integrate the range of variables/stressors required for accuracy
Water quality is impact by a range of geographical, land use and water management issues which manifest at a local level. Consequently the digital twin needs to be

granular enough to measure and predict changes in a way that allows plans and actions to have an impact
Much of WQ challenge and issues are catchment related and on that scale. Would increase ability to identify issues and address root causes
Monitoring benefits and data
Huge variation in river types and pressures on them across national and regional scales, but catchment scale allows for useful twinning
Catchment-scale mirrors the currently active river basin management planning process.
I was erring between regional or catchment scale. It would be useful to understands trends across the whole of a region to put trends into context, i.e. are they specific to that location or indicative of wider environmental change. But catchment helps to show how/where interventions have been beneficial.
Each catchment is different in terms of characteristics and biodiversity. Much of the work we do is on a catchment scale rather than regional or national.
There are too many regional differences in geology, land use, human activity to be meaningful at wider scale.
Regional and National would be insufficiently detailed to be useful.
National and regional scale models would not reflect local differences sufficiently to understand what needs to be done to improve or maintain water quality and ecology in the catchment.
Catchments are unique and though imperfect control volumes (eg GW-shed <> SW-shed), is the best way we have to account for fluxes & storages, where budgets for constituents are a key aspect for further understanding & better management
Catchment scale would be most useful but data could be lacking, depending on catchment size
Catchments are large enough to be considered as a whole system but also mean something to people locally.
Our organisation works at a catchment scale
National Scale
Needs to be appropriate for government policy at that scale.
It might have more impact if you can do nationally. This might show for example that the Clyde is good and the Forth is bad and make actions happen. Doing on one catchment-scale while more practical might not have an impact.
The entire British Isles should be covered. Information must be available though at a local scale, e.g. river reach scale.
It will be crude so a broad national picture might be most useful - if it gives water temperature and nutrient loads for say bathing waters
It would give the clearest picture across the full range of watercourses where regional or local may show a bias in areas with fewer issues.
At the outset, catchment level would be sensible to begin with. The ultimate goal should be national scale.
Decision making is needed at national scale - for the Environment Agency this effectively means England, but GB level could also be useful

because there is limited value in modelling a single lake, and why restrict coverage to only regional if national is likely to be almost as feasible?
Regional Scale
This scale is probably more relevant to a wider audience and likely more manageable than catchment scale but will still provide enough detail that potentially would be lost if this were at a national scale. A happy medium.
It is a balance between usefulness and feasibility. Waters in a region experience similar weather and major pressures that might make the twinning feasible while being at a scale that is useful
national is too generalised to be of use. Catchment level could be too detailed to be feasible and transferable. Regional is a good middle ground
Local Scale
Local, individual rivers. Frankly, would rather have real time monitoring in place than a model approach. Comes across as wanting to monitor water quality on the cheap. Complex range of pollutants today & actions should be taken on real world data.
Local scale, albeit allowing people to see information at the other stated scales. With heightened public interest and demands for improvement, we need to be nowcasting at a scale relevant to a public thinking about recreation, aesthetics and environmental impacts on a very local scale.
Multiple scale
A digital twin should be scalable and appropriate across all the different scales mentioned. only answering one would, in my opinion, go against the definition of a digital twin
It depends on the use and the precision of WQ info you can produce. e.g. I might be interested in DO at a national scale to see where hot weather incidents might occur.
Local (point source specific). Most water management currently happens on a catchment scale and there is little joined up work between neighbouring catchments. A digital twin provides an opportunity to understand how specific water quality issues affect both within- and between- catchment ecologies. National is a big ask so perhaps not realistic? But a mix of scales is required because sources (and therefore potential mitigations) are often small scale (especially point source) while effects can be regional.
Landscape type e.g. chalk, greensand, etc, land use, size, vegetation type/habitat, abstraction and any tidal influences. A mix of the bigger picture versus local impact and influences. To understand where pressures and their impacts alter the norm and how comparable they might be across different catchments for reproducibility of data/data certainty.
National to allow view of national changes and prediction and driving policy: catchment scale to allow management decisions and planning
If the regulators are to use this tool to manage water companies, the model must be the same across the country. It is probably more feasible to look at issues at catchment level as conditions vary greatly between catchments before we add human input
Needs to be able to look at all scales! and flow of pollutants between catchments.
The digital twin should ideally be available at multiple scales, to allow cross-scale comparisons etc.

3.3 Feasible temporal scale of a water quality digital twin

There was no consensus on the most suitable temporal scale to deliver output from a Water Quality Digital Twin. Only 35 respondents (59%) selected a single temporal scale that they considered the most feasible, useful, realistic, and deliverable for a Water Quality Digital Twin while the others reported either a combination of the suggested scales or an alternative temporal scale i.e., weekly or seasonally (Figure 6).

Overall, 76% and 80% of respondents marked ‘sub-daily plus daily’ or ‘daily plus monthly’ alone or in combination with other temporal scales as the most suitable temporal scale to deliver output from a Water Quality Digital Twin.

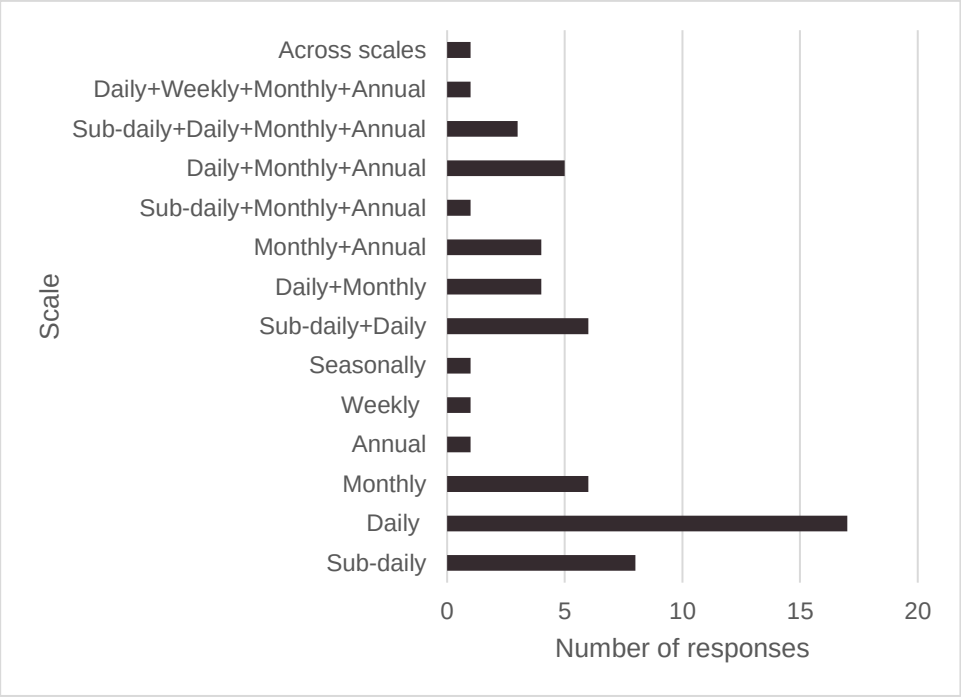


Figure 6. Respondents’ views on the most feasible, useful, realistic, and deliverable temporal scale for a Water Quality Digital Twin.

Respondents rational for their selection was also captured, to facilitate the interpretation of the results (Table 3). This lack of agreement on temporal scale appears to result from respondents’ concerns on (i) the use of the digital twin output, (ii) computational feasibility, and (iii) availability of data.

Several respondents commented that the most appropriate temporal scale depended on the question asked of a Water Quality Digital Twin, e.g. “*I think the potential uses need to be mapped to the most relevant scale to answer this - who wants the outputs of this model and what for?”* Another commented similarly; “*The*

question asks for most feasible, realistic and deliverable. This is tricky because feasible may be difficult to do on a too high frequency if data storage and cost are prohibitive, but realistically for maximum benefit you need regular monitoring more frequent than annual and monthly. Weather patterns are changing so seasonal shifts/months may change. Daily cycles will also be missed if daily is chosen as this may be quoted as a daily average or a fixed time and not reflect changes throughout the day".

Several others agreed with these sentiments but often answered this question on temporal scale of delivery based on their assumptions on the use of the digital twin. For example, "*Event impacts are often most important for response and recovery and are needed on sub daily scale. Whereas seasonal and long term, monthly or weekly* [may be] *enough. Daily doesn't quite answer either";* and *"If this twin is going to influence drinking water treatment decisions daily updates on algal blooms would be very beneficial. However long term trends are also important for planning purposes and investment decisions".*

Respondents generally considered that sub-daily and daily were useful if technically feasible e.g. "*Reflects data availability; If digital twin outputs were sub-daily they would be most useful for control of water quality impacts, and they would be able to utilise 15-minute interval update river water quality monitoring being installed by water companies from 2025 onwards; This is the only way to capture diurnal patterns and the impact pressures have on those patterns".*

Other respondents were aware of the computing requirements of resolving a digital twin at fine temporal scales, and the influence this may have for the use of the digital twin e.g. "*Daily and sub-daily can be implemented but may threaten the computability of the model and arise a lot of noise in the output (e.g., photosynthesis has an oscillatory trend due to diurnal fluctuations, but on a larger time scale such fluctuations must be analysed with a time-series analysis approach to filter out the oscillations and find whether a trend is present behind the data on a larger time scale)."*

Others feared that useful data and inferences would be lost if the digital twin did not deliver at a sub-daily or daily temporal scale e.g. one respondent mentioned the need for a "*Balance between temporal scales of variability across different processes";*

Some respondents commented that, if delivered at sub-daily or daily scales, the data could be integrated to longer temporal scales e.g. "*Annual can then be inferred from monthly anyway".*

Several respondents requested a temporal scale linked to real time monitoring e.g. "*ideally near real-time if possible"; "Realtime data gives the best results".*

Table 3. Rational as expressed by respondents for their desired temporal scale for a Water Quality Digital Twin. Note only spelling/typographical errors corrected, speech and grammatical integrity maintained as delivered by respondents.

Sub-daily
This will allow us to understand diurnal cycles of eg nutrients
Water quality is highly variable, this was demonstrated by data used for the development of the WRc WQ Monitoring Manual back in the 1980s. Continuous water quality monitoring demonstrates large diurnal variations in WQ data. E.g. dissolved oxygen, temperature and un-ionised ammonia.
Again, the public want to know the situation right now, albeit allowing people to see information at the other stated scales
tbf different answers for feasible than useful! Event impacts are often most important for response and recovery and are needed on sub daily scale. Whereas seasonal and long term, monthly or weekly enough. Daily doesn't quite answer either.
Fw ecosystems are highly variable, so several times during the day is of more value than less frequently (ie smaller scale, more depth).
Most appropriate for the likes of extreme rainfall and flooding events and the rapid pressures these pose on CSOs
If digital twin outputs were sub-daily they would be most useful for control of water quality impacts, and they would be able to utilise 15-minute interval update river water quality monitoring being installed by water companies from 2025 onwards.
Upper catchments are key systems to protect and restore, and being small, their response time is short
Daily
Because changes quickly, which is why a mix of water chemistry & biomonitoring best. Chemical data is a snap shot in time, with biota who are living there get a more realistic account if pollution levels.
This is the only way to capture diurnal patterns and the impact pressures have on those patterns. Realtime data gives the best results
Daily corresponds to a major criteria of daily rainfall in most parts of the world
Daily should be sufficient, longer term would not help with catchment management
Reflects data availability
Daily would be required to inform tactical decision-making, for example on Bathing Water information to the public, or say on smarter abstraction licenses linked to quality as well as flow.
1) This scale is probably a good balance between requirements for short-term management decisions and effective/ accurate sensor resolution 2) This does depend quite a lot on the question you are trying to answer though - so this should be identified once the purpose of the digital twin is defined 3) It will also be a trade-off between the processes captured by the models and the achievable data resolution of sensors

I'm not sure daily is necessarily most feasible but I think it would be suitably useful when balancing the technological power required to generate forecasts (and the carbon footprint required to supply this). Sub-daily would be very powerful but it would be a trade-off between whether it would be used vs the power needed to generate these models.
Sub-daily (hourly or 15minutes) is likely to require too much computing power and produce lots of data, easier to work on a daily timestep.
Otherwise, either too late or we drown in data.....
Parameters vary
With the new requirement for River quality monitoring in the Environment Act should allow meaningful daily or sub daily models to be used
Sub-daily might be too difficult for the beginning, but should be looked at for future iterations.
Daily provides the potential to represent event-scale behaviour - probably the most useful for stakeholders. Sub-daily seems unrealistic.
Monthly not resolute enough, particularly with climate change, impact of summer storms, CSO events etc
Water quality can change significantly overtime, capturing sub-daily data would create significant amount of data. Monthly is too little.
Most realistic for outputs and inputs
Monthly
Availability of data would limit temporal scale. Monthly would still provide a valuable output.
This is a policy tool, not a hazard forecast tool I think. At this stage at least.
I think the potential uses need to be mapped to the most relevant scale to answer this - who wants the outputs of this model and what for?
I think sub-daily and daily outputs may be unrealistic if the Water Quality Digital Twin is to be implemented beyond catchment-scale due to the resource and time required to process input data, update models, and generate outputs.
Maybe impractical more often though fortnightly a better option?
Manageable levels of data that can provide useful outputs
Annual
In current practice we tend to design for the 1 in 1 years storm event.
Weekly
Weekly. Monthly doesn't feel like a good match for a digital twin but daily and sub-daily you could be lost in a swamp of data. Weekly (at least at first) gives a good useful goal. Could come down to sub-daily for predictive modelling in time
Seasonally
I do not comprehend the scale of your project. However, I would imagine daily would be too much data to deal with and annual would not take into consideration the extremes that bring heavy rain in autumn and heat/low rain some summers.
Sub-daily + Daily

The temporal scale needs to be fine enough to see whether any intervention is working. Another critical aspect is how far away in the future the Digital Twin can predict (given some assumed inputs).
Most appropriate for both quantity and quality assessment.
To be relevant it needs to be sub-daily - if its to be anything close to real-time. sensor data should be relatively continuous? needs to link to that...
Balance between temporal scales of variability across different processes; ideally near real-time if possible
The question asks for most feasible, realistic and deliverable. This is tricky because feasible may be difficult to do on a too high frequency if data storage and cost are prohibitive, but realistically for maximum benefit you need regular monitoring more frequent than annual and monthly. Weather patterns are changing so seasonal shifts/months may change. Daily cycles will also be missed if daily is chosen as this may be quoted as a daily average or a fixed time and not reflect changes throughout the day.
Problems with water quality can easily be missed if done only at the daily scale, e.g. due to diurnal cycle in DO, sewage inputs
Daily + Monthly
Daily and monthly scales are what can be monitored with good certainty.
To relate closely with realities
Daily would be ideal, but if that's not possible then monthly would be OK. Annual can then be inferred from monthly anyway
Daily would be very interesting and the most useful but could prove difficult to deliver so monthly would be the next best.
Monthly + Annual
Daily and sub-daily can be implemented but may threaten the computability of the model and arise a lot of noise in the output (e.g., photosynthesis has an oscillatory trend due to diurnal fluctuations, but on a larger time scale such fluctuations must be analysed with a time-series analysis approach to filter out the oscillations and find whether a trend is present behind the data on a larger time scale)
This period are ok to account for changes
Daily response would be valuable but probably not currently feasible. Monthly would be useful while annual provides summary over the seasonal cycle.
Overall annual assessment of water quality would be useful for monitoring purposes but monthly timescale would allow seasonal pressures etc to be detected/modelled
Sub-daily + Monthly + Annual
Many changes in water courses happen slowly eg due to climate change or delivery of measures. eg AMP measures tend to happen on 5 year cycles. However, seasonal and daily variation in water quality would also help to identify limiting factors in a catchment.
Daily + Monthly + Annual
Driving data and uncertainty scale with those temporal scales. Annual would be the least demanding, daily the most.

Daily data would allow it to be used to guide leisure activities. Monthly and annual would be useful for research purposes.
I think sub-daily is probably too detailed for most people to warrant it being useful. Daily, monthly and annually allow some flexibility for the user to choose how they want to look at the information (given often different temporal scales are more appropriate for specific tasks) and I think would be more useful and more realistic to achieve.
Annual or slightly longer, to reflect the response of river systems to chronic issues. Short term (daily for initial response and monthly for recovery response) more appropriate for individual incidents e.g. a pollution spill
Again it depends on the use but for the DO example I would want daily - interested in the need to deploy bubblers and or prepare for fish kills
Sub-daily + Daily + Monthly + Annual
Real time monitoring - sub daily Different time scales to allow different monitoring and modelling.
If this twin is going to influence drinking water treatment decisions daily updates on algal blooms would be very beneficial. However long term trends are also important for planning purposes and investment decisions.
With real or near real time monitoring capabilities, sub daily should be possible and less detailed temporal data can be gained from that dataset
Daily + Weekly + Monthly + Annual
Upland catchments respond quickly to weather (hence daily output), but we also need to know longer term climatic effects (annual output)
Across scales
See answer to 1b ...a digital twin should be scalable and appropriate across all the different scales mentioned. only answering one would, in my opinion, go against the definition of a digital twin

In conclusion, the desired temporal scale depends on the use envisaged for the digital twin but sub-daily or daily was considered to provide the most practical knowledge if data was available and it was computationally feasible.

3.4 Major obstacles to creating a digital twin of running and standing water quality.

Respondents were asked what they considered to be the major obstacles to creating a fully functional digital twin of running and standing water quality and were offered four choices related to lack of real time data or models on water quantity and quality (with an additional ‘other’ category). Only ten responds selected a single category with the majority selecting multiple choices. However, almost twice as many respondents considered a lack of real time data and models on water quality to be more important obstacles than the equivalent for water quantity (Figure 7). It

is possible that the latter observation may reflect the cohort of individuals who completed the survey.

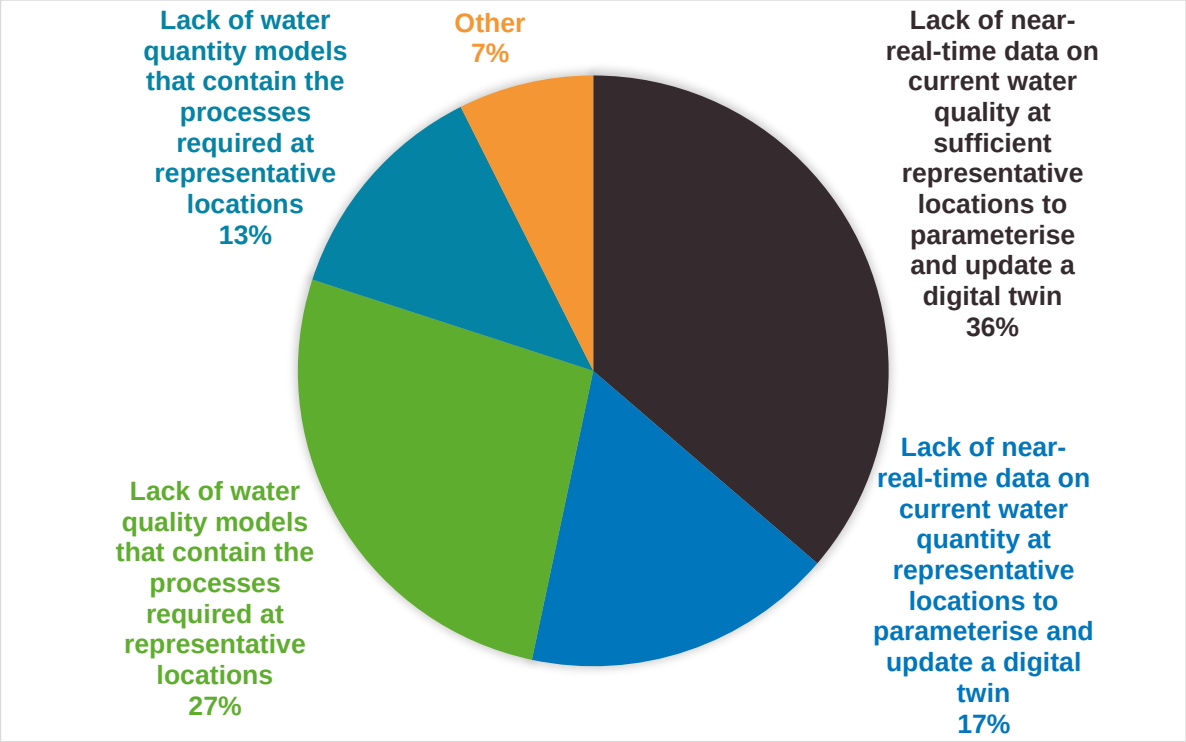


Figure 7. Respondents' views on the lack of real time data or models on water quantity and quality (with an additional 'other' category) as obstacles to creating a fully functional digital twin of running and standing water quality.

Many respondents (Table 4) commented on a lack of data generally e.g. *"Drive for more data, more locations, consistent".* While others considered the lack of water quality data was greater than for water quantity e.g. "*Real-time data quantity is available from SEPA via their API. Little quality data is collected in real-time"* and *"Lack of data and models linking water quality data to ecological outcomes Real time monitoring of relevant water quality parameters (eg nutrients) is difficult! Needs investment in sensing tech. Water quantity is better quantified but needs to be linked to WQ, including monitoring at same locations".* Assessing water quality was acknowledged by several respondents as difficult, e.g. *Relatively few water quality variables are straightforward to monitor at high frequency.* Furthermore, respondents considered resource requirements e.g. *"I consider the lack of near-real-time data on current water quality at sufficient representative locations to parameterise and update a digital twin to be the biggest obstacle because of the resource required to achieve this."*

Table 4. Respondents' views on the major obstacles to creating a fully functional digital twin of running and standing water quality. Note only spelling/typographical errors corrected, speech and grammatical integrity maintained as delivered by respondents.

Data water quality	Data water quantity	Models water quality	Models water quantity	Other	Respondents' comments when selecting 'other' and their elaboration on the major obstacles to creating a fully functional digital twin
1	1	1	1	0	Model's are only as good as the data fed into them. We are deficient in this. Many of our rivers only sampled once a year, few have data loggers in place. If had real time data for DO inverts, microplastics, forever chemicals, pharmaceutical drugs, heavy metals etc wouldn't need to model. Not sure it would be effective, unless do a more in depth set of monitoring on all our rivers you have no real baseline.
1	1	1	1	0	Increase the amount of sampling done.
1	1	1	1	0	Drive for more data, more locations, consistent. Utilise industry data i.e. monitoring required by environmental permits or required for modelling for permit applications. Install permanent data monitoring equipment.
1	1	1	1	0	Limited sites monitored - water quantity is covered reasonably well. But greater spatial and temporal data and models required
1	1	1	1	0	Near-real-time data is limited although probably sufficient in case of major rivers e.g. Thames. My understanding is there are no models currently set up to do this - stochastic models may be closed to achieving this with run-times that are useful, at scale.
1	1	1	1	0	These are all potentially an issue. Particularly in remote locations where gauging infrastructure or real-time loggers are not deployed. Suggest there needs to be a prioritisation of locations to target to begin with, followed by refinements and adaptive management.
1	1	1	1	0	Resources/affordable technology to enable the data collection to be undertaken
1	1	1	1	0	?

1	1	1	1	0	It's worth starting just with a hydrology digital twin which would be worthwhile in its own right as well as a necessary first step for an exercise involving water quality
1	1	1	1	0	All of the above
1	1	1	0	1	Speed of simulation At the moment we build models for specific studies so don't have full coverage. Model build, calibration and verification is increasingly hampered by limited data as the UK environmental regulators cut back on their routine monitoring. Detailed high frequency models take time to run, but the public want immediate answers when they ask a question.
1	1	1	0	1	Need to assimilate weather, and the physical, chemical and biological components along with flow paths to model water quality. There are always missing data. Instruments, data and modelling over long time periods are required.
1	1	1	0	0	Need real time data and models to apply to this data
1	1	1	0	0	Recruiting volunteers to collect samples but you also need sufficient labs to analyse those samples.
1	1	0	1	0	The setup and resources can be expensive
1	1	0	0	0	There is probably a combination of all of those points that will affect the digital twin but near-real-time data is one of the most powerful tools we could use. To scale up to a regional level there would need to be more real-time data to enable that scale to be more representative as detail does tend to be lost the further you extrapolate. Potentially look to input more real-time monitoring instruments into more water bodies for monitoring.
1	1	0	0	0	To identify more networks, groups or teams to participate in data gathering towards digital twin modelling
1	1	0	0	0	The answers are based on limited knowledge and understanding of remote and near-real-time monitoring for fundamental water quality parameters let alone harmful chemicals. I am aware that central government funding initiatives may become available in the near future to begin to address the development and application of remote sensors and near-real-time data reporting and interrogation by AI or machine learning

1	1	0	0	0	Aquatic systems are hugely diverse in nature, and some may be so unique as to be unmodellable, so representativity can be challenging. Relatively few water quality variables are straightforward to monitor at high frequency.
1	1	0	0	0	Data share across different organisations and funding can be preventative to gaining a full picture. Also remote monitoring can be issue with maintenance and battery life. There are lots of models out there, hence I have not ticked this, although revisiting these will be necessary as data improves and validation of the models are key.
1	1	0	0	0	I'm not familiar with what models are available but I do know that real time data is extremely lacking.
1	0	1	1	0	To obtain near-real-time data: I made use of remote sensing data and the approach worked to give me weekly updates. Moreover, I think that if you have the opportunity, the deployment of probes that send measurements in the cloud through a mobile connection could be helpful and informative, ensuring frequent and real-time data
1	0	1	1	0	Depending on what we want to do - its just monitoring good chemical and ecological status there is probably plenty of information there, but to go beyond this will need sampling strategies and advanced monitoring so need to link to existing ongoing research also, and likely sharing of research plans to enable a joined up sampling / monitoring strategy in the medium term to aid parametrisation.
1	0	1	1	0	Hourly water quality measurements are rare and non existent across a whole catchment. Parameters are often v limited and "traditional".
1	0	1	1	0	.
1	0	1	0	1	Lack of data and models linking water quality data to ecological outcomes Real time monitoring of relevant water quality parameters (eg nutrients) is difficult! Needs investment in sensing tech. Water quantity is better quantified but needs to be linked to WQ, including monitoring at same locations.
1	0	1	0	0	Water quality models have improved greatly over the last couple of decades but driving variables are key to making them work
1	0	1	0	0	I have selected water quality as this main obstacle since this is complex and multi-facetted while quantity in comparison is much easier to assess.

1	0	1	0	0	Installation of more gauges.
1	0	1	0	0	Water quantity has continuous monitoring with a networked design. Water Quality sampling is not as structured. Water quantity modelling has been used for water resources and flood forecasting and is more developed than water quality
1	0	1	0	0	Better data collection and investment in WQ models
1	0	1	0	0	Monitoring many critical water quality variables, particularly nutrients has been technologically challenging and requires sufficient investment in high quality, reliable sensors and the time and technical capability of staff to maintain them - these are critical elements of the infrastructure required and should not be omitted from the budget - poor quality data production will prevent the adequate parameterization of models. Good input data at the appropriate resolution is essential for the creation predictive capabilities or forecasting whilst minimizing uncertainties from that element of the process.
1	0	1	0	0	Real-time data quantity is available from SEPA via their API. Little quality data is collected in real-time
1	0	1	0	0	I am not a modeller but my sense is that hydrological models are well developed whereas those for water quality less so
1	0	1	0	0	Models will need to be bespoke and require good data sets
1	0	1	0	0	There is real-time monitoring available, and in use, but for a limited range of parameters. Need more development on remote sensing techniques to broaden this range
1	0	1	0	0	Lack of monitoring sites is likely to be an ongoing issue, but some gaps could be filled through liaison with partner organisations and perhaps some citizen science monitoring
1	0	1	0	0	Development of new data sources based on enhanced sensor systems and data networks from LOWRAN to 5G
1	0	1	0	0	The EA has reduced their environmental monitoring when, ideally, would have been maintained or increased including more parameters
1	0	1	0	0	Water company monitoring associated up and downstream of assets should lead to an increase in some data. Models would need to include information on morphology at a particular location.

1	0	1	0	0	Deploy continuous monitoring water quality sondes if robust relationship identified between easily measured parameters (e.g. conductivity, pH, temperature) and other water quality parameters
1	0	0	0	0	The cost of real time data collection may be prohibitive, but the quality of this type of data is what is so important
1	0	0	0	0	I think we have water quantity and models. Overcome it by working at a very broad time scale to start with. Maybe seasonal. No need for sub-diurnal yet.
1	0	0	0	0	We previously had continuous water quality monitoring stations in Thames Water's area and Severn Trent Water's area (operated by the National Rivers Authority). They were expensive to maintain. New monitoring will include continuous water quality monitoring stations and therefore the data will be available. I cannot comment on water quantity data, this is outside of my area of expertise.
1	0	0	0	0	I consider the lack of near-real-time data on current water quality at sufficient representative locations to parameterise and update a digital twin to be the biggest obstacle because of the resource required to achieve this. Presumably, sondes or other automated high-frequency measurements of water chemistry parameters are going to be required as well as physical sampling for biological parameters such as algal and fish communities. This is a large undertaking for even a few waterbodies and could be difficult to scale to develop Water Quality Digital Twins even at catchment-scale with monthly outputs.
1	0	0	0	0	No comments
1	0	0	0	0	Very little near real time data currently available at representative locations in catchments.
1	0	0	0	0	Combine state agency monitoring network with local partnership low-cost sensors, citizen science and RS/EO
1	0	0	0	0	Collaborative monitoring plans produced by all partners for catchments so that representative locations can be agreed in light of the issues in the catchment and all sources of data being collected by partner organisations are included.
0	1	0	0	0	The parametrization is important.
0	1	0	0	0	No comment

0	0	1	1	1	Lack of interest by key stakeholders. make sure the results have utility and to more than academics!
0	0	1	1	0	Current pollutant runoff models are poorly integrated with hydraulic models, combining the two would simplify design processes.
0	0	0	0	1	How to join things together and update them with new data. creating digital building blocks to join data and update data
0	0	0	0	1	Main obstacle is developing the methods to appropriately integrate the real time data that is available with the models. These methods need to be far more sophisticated than simple nudging or calibration that is currently done. More methodological development is needed
0	0	0	0	1	There are plenty of models is the real time inputs of pollutants that are missing. tap in to the new CSO monitoring network if you can persuade Water Companies to share it
0	0	0	0	1	Cannot choose just one. I am not sure about that one
0	0	0	0	1	Lack of understanding of how microspatial variation in water quality and flow influences how representative water quality monitoring is of a whole water body (Environment Agency water body definition). Research into the cm by cm variation of water quality and flow with depth and distance across a water body, and into the change in response from ion-selective electrodes, conductivity and temperature monitors at different flow rates with stable parameters will allow for corrections that will deliver more accurate monitoring.
0	0	0	0	1	Probably all the above but the main issue is uncertainty and local factors that are hard to model. However models may be good enough for some purposes and this would be good to explore. I want to know real time stuff for immediate response but I also want to know how water quality will change over very long time periods in response to policy/regulatory intervention. Models that can accommodate infrastructure stress tests would be good - I don't know what the main hurdles are here.

Some suggested volunteers and working in partnership with others to be helpful in relation to monitoring *"lack of monitoring sites is likely to be an ongoing issue, but some gaps could be filled through liaison with partner organisations and perhaps some citizen science monitoring".* Others also commented on the resources needed to process the volunteer sampling e.g. *"Recruiting volunteers to collect samples but you also need sufficient labs to analyse those samples."* Lack of resources was commented on by several respondents with one offering hope *"I am aware that central government funding initiatives may become available in the near future to begin to address the development and application of remote sensors and near-real-time data reporting and interrogation by AI or machine learning".*

Some considered the lack of data more important than a lack of models. One respondent wrote "*Model's are only as good as the data fed into them. We are deficient in this. Many of our rivers only sampled once a year, few have data loggers in place. If had real time data for DO inverts, microplastics, forever chemicals, pharmaceutical drugs, heavy metals etc wouldn't need to model. Not sure it would be effective, unless do a more in depth set of monitoring on all our rivers you have no real baseline".*

Other respondents acknowledged the need for models but still considered data to be fundamentally lacking e.g. *"Water quality models have improved greatly over the last couple of decades but driving variables are key to making them work"* and *"Models will need to be bespoke and require good data sets".*

Given the diversity of river and standing water systems one respondent questions the premise that a digital twin could be created writing "*Aquatic systems are hugely diverse in nature, and some may be so unique as to be unmodellable, so representativity can be challenging".*

In conclusion, the need for real time data on both water quality and quantity was considered by the majority of respondents to be the principasl obstacle to creating a fully functional digital twin of running and standing water quality.

3.5 Most important determinand in a water quality digital twin

Nutrient concentrations were considered the most important determinands of running and standing water quality to include in a water quality digital twin, with 100% of respondents scoring these as one of the three most important categories, and 71% scoring nutrients as critically important (Figure 8).

However, many respondents also considered other determinands as critically important, with only six respondents not selecting any of the suggested determinands as critically important (Fig 8). Cyanobacteria, pesticides /herbicides and heavy metals were all selected by over 80% of respondents as being important-to-critically important to include in a water quality digital twin. Water

colour and taste and odour compounds were considered the least important, with less than 60% of respondents selecting them in the top three importance categories. Only six respondents considered any of the nine determinands as unimportant (Figure 8).

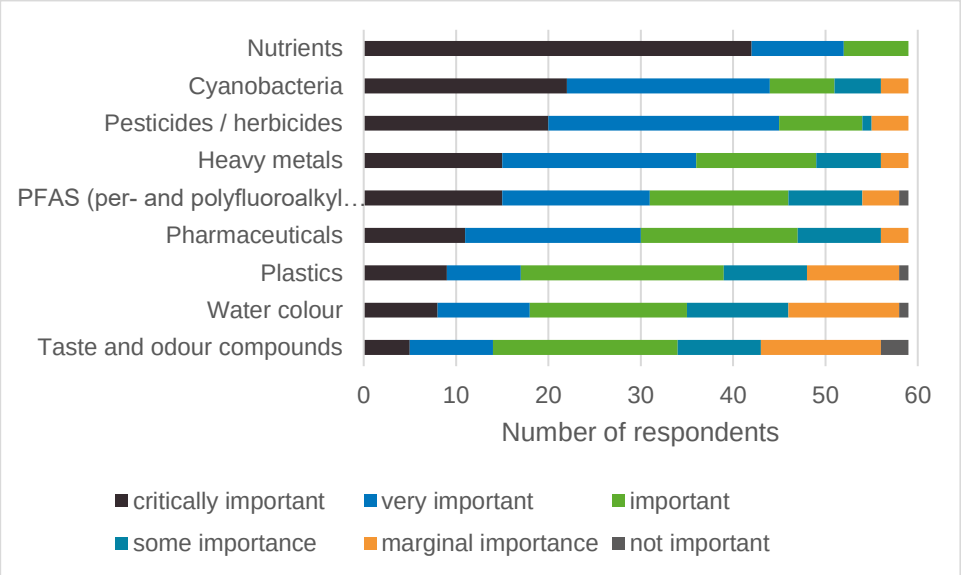


Figure 8. Respondents' views on the importance of nine determinands of running and standing water quality to include in a water quality digital twin.

Forty respondents provided details of other determinands which they considered important (Table 5). The linkage between nutrients and cyanobacteria was noted by several responds e.g. "*Algal blooms (and therefore nutrient concentrations) are of greatest concern to the general public so it would be most useful to be able to forecast these to inform management on the ground"*. The link between algal blooms and dissolved oxygen was also noted by several respondents e.g. "*Dissolved oxygen - essential for a lot of life and pretty indicative of healthy water!".*

A range of additional physico-chemical parameters were suggested e.g. "*temp, pH, BOD, COD"* and "*chlorinity and salt".* Some respondents replied with a greater degree of granularity for the "nutrients" determinand type e.g. "*phosphate, nitrate, …, and conductivity"*, "*Total ammonia",* and *"Dissolved organic carbon".* In addition, stratification and retention time were noted as important e.g. "*For standing water, information on physical structure- ie strength of stratification over time".*

Determinands related to biodiversity were considered by a few respondents as additional determinands which should be considered e.g. "*DNA for species monitoring may be health indicators, vegetation".*

The lack of a clear linkage to sewage (which can include several of the determinands) was noted by one respondent "*Have you included sewage? I can't*

see which one that is!" While other suggested "*Escherichia coli (EC) and Intestinal enterococci (IE)"* and "*Coliforms and other gut bacteria".*

One respondent who only considered nutrients, cyanobacteria and water colour as critically important remarked "*Most of the parameters listed are unimportant. We do not understand their toxicity and until we do we should not develop models to measure them - please refer to Prof Sean Comber at Plymouth University for the latest information on environmental quality standards".*

In conclusion, while there was consensus that nutrient concentrations were very important determinands to include in a water quality digital twin, several other abiotic and biotic determinands were also suggested, partly depending on the required use of the resultant digital twin.

Table 5. Individual scores for nine determinands of running and standing water quality to include in a water quality digital twin with comments plus details of other determinands respondents considered were important (i.e. would score 4 or above on importance scale). Note only spelling/typographical errors corrected, speech and grammatical integrity maintained as delivered by respondents.

Nutrients	Cyanobacteria	Pesticides herbicides	Heavy metals	PFAS	Pharmaceuticals	Plastics	Water colour	Taste and odour	Details of other determinands respondents think are important (i.e. would score 4 or above on scale provided in Q4).
1	1	1	1	1	1	1	2	2	Polyaromatic hydrocarbons (PAHs), Endocrine disruptors compounds (EDCs) Bisphenol, POPs
1	1	1	1	1	1	1	4	4	
1	1	1	1	1	1	2	4	4	Ammonium compounds, dissolved oxygen, temperature, suspended solids. pH.
1	1	1	1	3	3	3	3	3	
1	1	1	1	5	4	4	5	5	Where possible, physico-chemical parameters i.e., temp, pH, BOD, COD etc
1	1	1	2	1	2	2	2	3	Harmful algae; chlorinity and salt potentially affecting soils (if the water basin is connected to the sea)
1	1	1	2	1	2	3	2	1	Temperature and other algae types
1	1	1	3	3	3	5	2	3	Again its not whats important to the digital twin that matters but who wants the outputs and what for - how sensitive are their decisions to the model uncertainty?

1	1	2	2	2	1	1	3	3	No comment
1	1	2	2	2	4	1	3	3	
1	1	2	2	2	2	2	2	2	
1	1	2	2	2	2	3	1	1	Water pH
1	1	2	2	2	2	3	1	3	
1	1	2	2	2	2	3	4	4	UN ambient water quality indicator calls for phosphate, nitrate, oxygen, pH, and conductivity. This seems a sensible benchmark for universal core parameters.
1	1	2	2	3	3	5	2	4	Dissolved oxygen - essential for a lot of life and pretty indicative of healthy water! Might also be simpler to start with more easily measurable indicators Algal biomass (irrespective of species - since bloom decay can result in low DO and other impairments to freshwaters e.g smothering habitats)
1	1	2	2	4	2	3	5	6	Algal blooms (and therefore nutrient concentrations) are of greatest concern to the general public so it would be most useful to be able to forecast these to inform management on the ground.
1	1	2	3	3	2	3	2	2	Do you include iron and manganese with heavy metals? Dissolved oxygen is critical. Stratification and temperature.

1	1	3	5	3	5	5	1	2	You have not included dissolved oxygen, this is the number one parameter. Total ammonia and pH are also highly important. Most of the parameters listed are unimportant. We do not understand their toxicity and until we do we should not develop models to measure them - please refer to Prof Sean Comber at Plymouth University for the latest information on environmental quality standards.
1	1	5	5	5	4	5	5	5	Have you included sewage? I can't see which one that is!
1	2	1	1	1	2	2	3	3	Temperature- 1
1	2	1	2	1	2	1	5	5	
1	2	1	2	1	1	2	4	3	Presence/impact of invasive non-native species
1	2	1	2	2	2	3	3	4	DNA for species monitoring may be health indicators, vegetation, suspended solids, Dissolved Oxygen, Temperature, pH
1	2	1	2	2	3	3	4	4	
1	2	1	4	2	4	1	3	2	
1	2	2	1	2	2	3	5	5	Information on water flow
1	2	2	2	2	2	2	3	3	eDNA
1	2	2	2	3	4	3	2	4	
1	2	2	3	2	2	2	3	1	
1	2	3	2	3	3	3	3	3	
1	2	3	3	2	4	4	2	2	

1	2	5	4	5	5	5	4	3	For standing water, information on physical structure- ie strength of stratification over time.
1	2	5	4	6	4	3	3	2	Chlorophyll as a measure of total algal biomass. Temperature and retention time are also very important.
1	3	1	3	3	3	3	5	3	Sediment
1	3	2	2	2	1	3	3	3	Greenhouse gases DIC and DOC Salt (road salting - indication of all road pollution)
1	3	2	3	3	3	1	4	4	Waste material - rubbish
1	4	1	1	1	1	3	5	5	N species, P species, zinc, nickle, cypermethrin, fipronil, antibiotics, hormones, bisphenol A.
1	4	1	4	1	3	5	4	6	Potentially bacteria (FIOs) but only in urban environments or where there is high human usage.
1	4	2	3	2	3	5	5	5	Ammonia, BOD, DO, Escherichia coli (EC) and Intestinal enterococci (IE)
1	4	2	4	4	4	4	4	5	
1	5	2	1	4	1	3	2	2	Temperature, ammonia, dissolved oxygen, feacal indicator organisms.
1	5	3	3	3	3	5	5	5	Ammonium, dissolved oxygen, temperature, pH more important than many listed above
2	1	3	2	4	3	4	1	1	Dissolved organic carbon
2	1	3	3	3	3	4	1	4	
2	2	1	1	1	1	1	3	3	Suspended sediment
2	2	1	1	1	2	2	1	1	
2	2	2	1	1	1	3	5	5	Coliforms and other gut bacteria. Score 1. Not currently done for rivers unless a reported issue.

2	2	2	1	2	2	3	3	3	
2	3	2	3	4	3	4	6	5	Particulates/fine sediment acidity temperature
2	3	2	3	4	3	5	5	5	Temperature and dissolved oxygen. Nothing in the table about pathogens?
2	3	4	4	4	4	4	3	3	In thinking of public interest, we need to think what matters in terms of public health, recreation, aesthetics and ecology. Much harder to model, but somehow we need to link what we can measure and model more easily to what matters to people.
2	4	3	3	1	2	4	5	5	Any WFD substance.
3	1	2	4	4	2	6	4	6	Dissolved oxygen (score 1) only bioavailable nutrients, pesticides/herbicides, PFAS, pharmaceuticals and heavy metals should be measured, not totals.
3	2	1	2	1	1	1	1	3	Plastics additives - BPA, thalates etc. surfacants veterinary chemicals dyes and textile-related chemicals
3	2	2	1	3	3	3	3	2	
3	2	2	2	3	2	3	4	3	Water temperatures. Oxygen levels
3	2	3	1	3	3	4	1	5	Temperature
3	3	3	3	3	3	3	3	3	
3	5	5	5	5	5	5	3	3	None

.

3.6 Output variables considered as most important water quality

Respondents were asked to select the one most important output variable that a water quality digital twin should predict. Over half (52%) selected ‘Physico-chemical variables, including concentrations of determinands’ while less than a quarter (24%) selected ‘Biological’ variables (Figure 9).

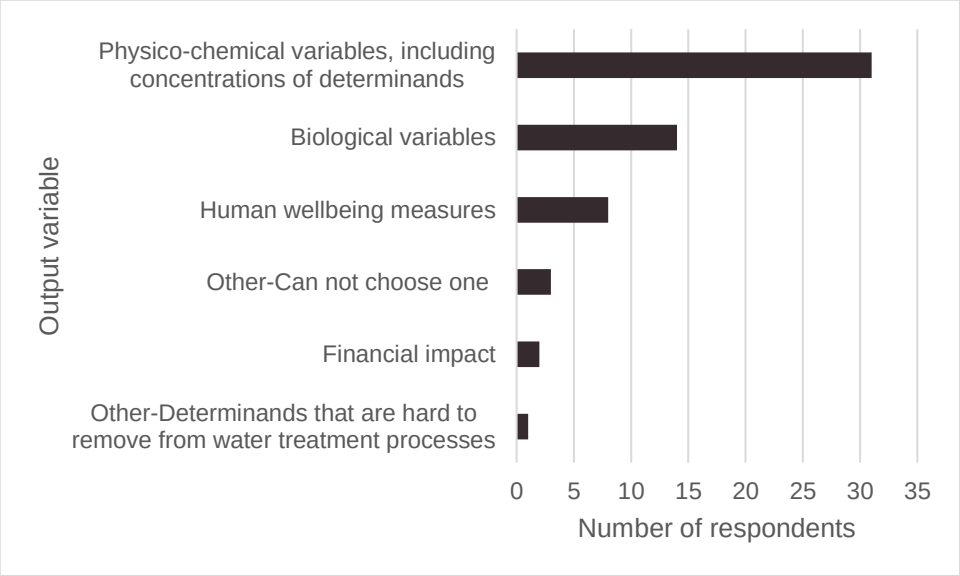


Figure 9. Respondents’ selection of output variable types that they considered to be the most important for a digital twin of running and standing water quality to predict and deliver.

Many respondents commented (Table 6) that they would have liked to select more than one but suggest that much could be inferred from the one they selected. For example, when physico-chemical variables were selected:

> *“Physico-chemical variables are at the basis of human-well being measures and financial impact. However, I would have put biological variables at the same importance of physico-chemical variables”.*
>
> *“I would have selected physico-chemical and biological variables if possible as I believe the two go hand-in-hand. However, I would prioritise physico-chemical variables as prediction of nutrient concentrations, temperature, and dissolved oxygen concentration will allow better management of algal blooms and water for other biodiversity”.*
>
> *“Once you have confidence in this then the other variables can be calculated, without this you rush to develop the d/s calculations and don't get your basics right”*

Respondents who prioritised biological variables wrote:

"As an ecologist, if I had to pick on from the above, then I would argue that information on the state of the biology gives a strong indication of the state of water body and hence can inform the other variables on the list"

"Good ecology can be linked to financial benefit and human well being. Although linked to physchem variables, it is harder to link these to financial impact and human well being. Ultimately to influence decision making the model needs to be able to inform financial impact and human wellbeing, however, if this is the model output it will be too remote from the cause of the problem. Biological variables seems like a good compromise".

Respondents who prioritised human wellbeing and financial impact as output wrote similar justification, namely, much could be inferred but these output variable types would provide a more direct measure for decision making. Below, a few examples are listed (Table 6):

"Increasing the value of water quality for human wellbeing should enable more connection and overall improvements for biological factors"

"I would have liked to also select "Biological variables" and "Human wellbeing measures" in Q5. I selected "Financial impact" because that could also reflect Human wellbeing, e.g. if drinking water supply was impacted there could be costs in providing bottled replacement water. Or if lakes were impacted by algal blooms that could reduce recreational visits and expenditure in the vicinity of the lake".

Respondents who selected 'Human wellbeing measures' also remarked that they understood these would be more difficult to model e.g.

"In thinking of public interest, we need to think what matters in terms of public health, recreation, aesthetics and ecology. Much harder to model, but somehow we need to link what we can measure and model more easily to what matters to people".

Table 6. Rational as expressed by respondents for their desired model output from a Water Quality Digital Twin. Note only spelling/typographical errors corrected, speech and grammatical integrity maintained as delivered by respondents.

Physico-chemical variables, including concentrations of determinands
These are standard for water quality monitoring - but should be done in conjunction with biological. What chemical and in what quantities feed onto wellbeing and financial (as in cost to correct)
These are the basic variables from which a host of other things can be inferred and derived (eg human well-being, financial, natural capital…)
Physico-chemical variables are at the basis of human-well being measures and financial impact. However, I would have put biological variables at the same importance of physico-chemical variables.
It is a good point for observing changes in the water
Whilst the other variables are important, the Physico-chemical variables control greatly how those other variables are expressed in the environment
The physico-chemical variables are able to provide information about what might be happening with regard other variables (i.e. biological and in some cases financial).
Physiochemical variables can be linked to the other variables by models
Because further outputs could then be based on this.
The questions people will be expecting a water quality digital twin to answer will be around physico-chemical variations, interactions, impacts
Would have like to pick several but not an option for this question. I think predicting / monitoring loads of pollutants relative to flow rates, abstraction and other human activities in the catchment is essential. The goal should be to support river management actions and to identify early accidental releases (e.g. if sub-daily readings increase beyond some threshold / usual range) then predictions of PEC relative to PNEC and when dangerous levels likely to be seen if no corrective action taken....
Physico-chemical variables could be a very powerful tool in developing solutions to point and diffuse intermittent pollution.
Day to day water quality variation for dissolved oxygen, ammonia, temperature and pH are the important elements. We infer how they impact on biological variables (e.g. fish, algae etc) from that data.
I'm not sure of the answer to this one, but it was a compulsory question.
If multiple choice I would have also chosen biological variables as physico-chemical and biological variables would be a powerful combination. I'm speaking as a biologist however and you might get higher investment from financial impact!
I would have selected physico-chemical and biological variables if possible as I believe the two go hand-in-hand. However, I would prioritise physico-chemical variables as prediction of nutrient concentrations, temperature, and dissolved oxygen concentration will allow better management of algal blooms and water for other biodiversity.
For water quality, however many stakeholders want to know is the water clean for me so knowing human wellbeing outputs is also valuable (i can only select one choice in the response)

This output would protect public health and mortality from water borne diseases
Once you have confidence in this then the other variables can be calculated, without this you rush to develop the d/s calculations and don't get your basics right
We saw biologically moribund rivers in industrial times. Leaving the eu may encourage more pollution!
Measurable and affect all others
These variables could be considered as precursors to potential impacts to biological, financial and human health elements
Just trying to be realistic, it's only worth looking at biological variables once the WQ determinands are understood
Biological variables and human well being depend on physico-chemical variables. It would be great also to understand the financial implications of the lack of WQ and also the costs of achieving better WQ
Best to get the most fundamental one done first - the others may follow depending on outcomes
Physico-chemical variables are those which we can exert most control over, by releasing reservoir water, reducing abstraction, varying treatment intensity (e.g. by using more or less ferric salt to precipitate phosphorus from wastewater), at short time intervals.
No comment
Requirement to understand how water parameters impact on ecology
All important, but physico-chemical most readily monitorable in near real time
Basic
These variables are important indicators of river health
Biological variables
It is important to understand how the determinands are impacting on the waterbody biology
Ecosystem health is very important.
These are very good for long-term predictions
Biological impact of all determinands is critical for understanding ecological health - which then has impacts on finance and wellbeing
The answer will depend on the purpose and end user. This is a scientists answer.
Water quality is the ability of life and biodiversity to thrive in the water (dependant on the type of water boy and the "natural" biodiversity expected
Biology will essentially influence a lot of other aspects
As an ecologist, if I had to pick on from the above, then I would argue that information on the state of the biology gives a strong indication of the state of water body and hence can inform the other variables on the list
The model should be aimed at predicating ecological health of the river as this is the key aim of the river basin management plans
Prediction of biological variables to provide an understanding of the influence of the measured physico-chemical variables on watercourse functionality and the biodiversity it supports
Biological responses are key aspect for considerations regarding climate change and biodiversity impacts

Includes algae and pathogens which impact both drinking water and recreational activities.
Good ecology can be linked to financial benefit and human well being. Although linked to physchem variables, it is harder to link these to financial impact and human well being. Ultimately to influence decision making the model needs to be able to inform financial impact and human wellbeing, however, if this is the model output it will be too remote from the cause of the problem. Biological variables seems like a good compromise.
We really need both phys-chem outputs and the end biological impacts but I could only tick one
Financial impact
I have selected financial as I think this will have the most influence on decision making. The data is broad in the fact it could be used to assess water use, development, habitat condition, resilience, etc
I would have liked to also select "Biological variables" and "Human wellbeing measures" in Q5. I selected "Financial impact" because that could also reflect Human wellbeing, e.g. if drinking water supply was impacted there could be costs in providing bottled replacement water. Or if lakes were impacted by algal blooms that could reduce recreational visits and expenditure in the vicinity of the lake.
Human wellbeing measures
This will be the most important for the general public.
I think that's what most users are interested--odour and taste is what people think about 'water quality', and then it's cyanobacteria and sometimes plastics and pharmaceuticals
The final deliverable is human welfare. If we can start with that, we are more likely to include all the aspects, and perhaps not focus too strongly on accuracy which might slow it down. That can come later?
Concern of public health
In thinking of public interest, we need to think what matters in terms of public health, recreation, aesthetics and ecology. Much harder to model, but somehow we need to link what we can measure and model more easily to what matters to people.
Who responds to real time outputs - swimmers, fisheries managers etc
Safeguarding human health is at the heart of what we do
Increasing the value of water quality for human wellbeing should enable more connection and overall improvements for biological factors
Other
Determinands that are hard to remove from water treatment processes. I work for a water company!
I think some kind of integrated measure of the physical-chemical-biological environment is a good first step to quantifying water quality that could then be used to generate downstream values related to policy compliance, societal costs etc.
They cannot be neatly separated to choose just one. Really hard to choose just one
I wanted to choose 1 and 4I'd like to see a water quality model that could answer questions about 'is it safe to swim/drink/fish'

In conclusion, a combination of physico-chemical variables and biological variables was considered the most important output variables that a water quality digital twin should predict. However, if forced to select only one then physico-chemical variables were judged the most important. While human wellbeing and financial impact as output variables were considered very important for decision making, they were recognised as more difficult to model.

3.7 Respondents' views on output from a digital twin of water quality

The majority of respondents selected more than one output format (76%) with approximately a quarter of respondents selecting each of 1, 2, 3, or 4 of the available options and 5% selecting all five (Figure 10and Table 7).

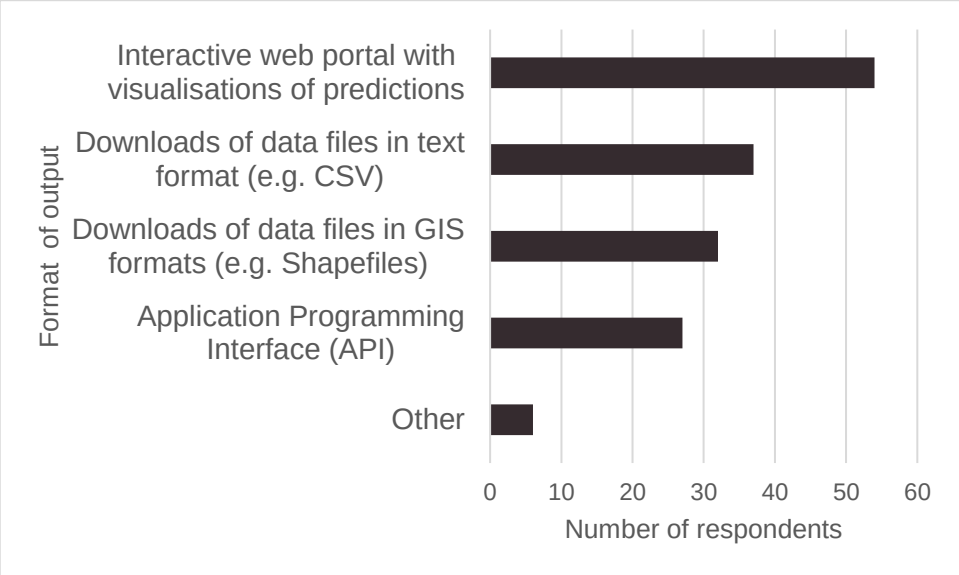


Figure 10. Respondents' selection on how end users should be able to interact with and receive output from a digital twin of water quality.

An interactive web portal with visualisations of predictions was selected by 92% of respondents. Four of the six individuals who did not select this option selected at least one of the other formats offered (Table 7) while two selected the 'Other' category and either recommended co-development i.e. *"I think this should be user defined - i.e. co-develop this with the end users in mind"* or further interpretation of the output i.e. "*Relevant knowledge that has been derived from the digital twin - don't make people have to work it our themselves…translated insights in an accessible way that doesn't involve more analysis".* These latter two observations were echoed by many respondents who elaborated on their answer to this question (Table 7).

It is clear from the output format selected and the written rationale for that selection (Table 7) that some respondents wished to have the data interpretated for them while others simply wanted access to the data so they could conduct further analysis themselves. For example the respondent who selected only 'Application Programming Interface (API) commented "*Really just worth having an API, everything else can build on top of that.".* While a respondent who selected all offered output types commented *"Having the ability to download in multiple formats is useful for lots of people. An interactive web portal with visual aids (e.g. maps, graphs etc) would be really useful particularly if they are accompanied with additional information for users to help understand what they are looking at (e.g. what is considered a high or low concentration of x, y or z)."* Another respondent who selected all output types suggested active co-development by commenting *"Possibly an open-source architecture that the community can contribute to and write interfaces and add-ons for."*

Although affiliation of respondents was collected simply to enable a description of the respondent cohort, there are clear signs that different potential users had different views on how end users should be able to interact with and receive output from a digital twin of water quality e.g. *"Most users will be satisfied with visualisations. Allowing data to be downloaded leads to confusion as some users may apply the wrong statistics leading to misuse of information. I would say that, I was a regulator!"*

Some respondents requested more innovative suggestions e.g.
"How about "what-if" scenarios, where you can ask the twin to model certain conditions and see the impacts, say discharges of certain compounds etc".

"Possibly text alerts or other notifications to relevant organisations so they can proactively undertake incident response".

"A dashboard showing when likelihood of water quality threshold exceedance"

"*There could also perhaps be an active element when information or warnings (of a cyanobacterial bloom for example) is sent to registered users."*

Other suggested linking to associated data sets e.g. *"links to data archive and other data centres where primary data available "smart" links to help allow other non-water sectors to link across to pull through the drivers and pressures on that water quality. In isolation the issues and solutions won't be visible".* One respondent expressed a desire to link to other spatially explicit datasets "*Geographically-based links to other datasets of potential interest, e.g. Met, Air Quality, Catchment Land Use, Population stats etc."* .

Table 7. Individual scores of respondents' view on how end users should be able to interact with and receive output from a digital twin of water quality and their associated comments. Note only spelling/typographical errors corrected, speech and grammatical integrity maintained as delivered by respondents.

Interactive web portal	Downloads data text -	Downloads data GIS -	API	Other	Respondents' comments when selecting 'Other' way of interacting with the digital twin and elaboration on their answer
1	1	1	1	1	Notifications of WQ changes relevant to water users would be helpful. This could also eventually be used for dynamic discharge permitting.
1	1	1	1	1	API should be linked to R package for ease of data access etc. Phone app for infield use
1	1	1	1	1	Mobile apps
1	1	1	1	0	Nil
1	1	1	1	0	Virtual reality and augmented reality
1	1	1	1	0	Interesting to look at gaming options
1	1	1	1	0	Having the ability to download in multiple formats is useful for lots of people. An interactive web portal with visual aids (e.g. maps, graphs etc) would be really useful particularly if they are accompanied with additional information for users to help understand what they are looking at (e.g. what is considered a high or low concentration of x, y or z).
1	1	1	1	0	Most important is web portal with visualisations and predictions
1	1	1	1	0	Be able to programmatically feed data into predictive models and capture outputs / predictions. e.g. how long until PEC exceeds PNEC if concentration continues to increase at current rate?
1	1	1	1	0	Possibly an open-source architecture that the community can contribute to and write interfaces and add-ons for.
1	1	1	1	0	Most users will be satisfied with visualisations. Allowing data to be downloaded leads to confusion as some users may apply the wrong statistics leading to misuse of information. I would say that, I was a regulator!

1	1	1	1	0	Different people will want different things. A simple story is essential, but some will want access to the underlying data.
1	1	1	1	0	How about "what-if" scenarios, where you can ask the twin to model certain conditions and see the impacts, say discharges of certain compounds etc
1	1	1	1	0	No
1	1	1	1	0	No
1	1	1	1	0	Possibly text alerts or other notifications to relevant organisations so they can proactively undertake incident response.
1	1	1	1	0	Not that I can think of
1	1	1	1	0	No
1	1	1	0	0	No
1	1	1	0	0	No
1	1	1	0	0	x
1	1	1	0	0	Links to data archive and other data centres where primary data available "smart" links to help allow other non-water sectors to link across to pull through the drivers and pressures on that water quality. In isolation the issues and solutions won't be visible
1	1	1	0	0	Can't think of other ways right now
1	1	1	0	0	These interaction outputs would assist with communications and monitoring for projects
1	1	1	0	0	No
1	1	1	0	0	None
1	1	1	0	0	A dashboard showing when likelihood of water quality threshold exceedance
1	1	1	0	0	No
1	1	0	1	0	There could also perhaps be an active element when information or warnings (of a cyanobacterial bloom for example) is sent to registered users.
1	1	0	1	0	Nil
1	1	0	0	0	I'm not really sure - it depends on what is created but the first two options are the most user-friendly

1	1	0	0	0	Maybe adding future predictions to the web portal
1	1	0	0	0	No
1	1	0	0	0	Geographically-based links to other datasets of potential interest, e.g. Met, Air Quality, Catchment Land Use, Population stats etc.
1	1	0	0	0	I am not familiar with GIS or API. Please make the outputs interactive intuitive and downloadable for one's own analysis.
1	0	1	0	0	No.
1	0	1	0	0	Interactive feedback from users
1	0	1	0	0	N/A
1	0	0	1	0	Have thresholds of when it is above a certain limit
1	0	0	1	0	-
1	0	0	1	0	n/a
1	0	0	1	0	Chose my answer as I think these methods will get the highest usage by the general public.
1	0	0	1	0	The outputs need to be focused on potential end users - Regulators, Water Industry and Academia
1	0	0	0	1	That is massively user specific. Water companies will want something very different to a wild swimmer in terms of info. most people likely to want translated info that is useful knowledge for their purposes
1	0	0	0	0	Web portal would be most accessible
1	0	0	0	0	No comment
1	0	0	0	0	Social media?
1	0	0	0	0	No
1	0	0	0	0	Keep it accessible and simple
1	0	0	0	0	Community engagement
1	0	0	0	0	Ability to manipulate 'inputs' to see how the system would respond if levels of a determinand are improved - predicting recovery as evidence for investment in improvement
1	0	0	0	0	Not sure

1	0	0	0	0	Ideally, I would get a graph or a pie chart to help with data visualization. A bit like SAGIS is to SIMCAT
1	0	0	0	0	No
0	1	1	1	0	A desktop version to visualize prediction would be useful, maybe more useful than the interactive web portal
0	1	0	0	0	No, and as we are talking predictions, pretty useless to many aquatic scientists who need real data.
0	0	0	1	0	Really just worth having an API, everything else can build on top of that
0	0	0	0	1	I think this should be user defined - i.e. co-develop this with the end users in mind This will depend very much on the user type - as part of the Environmental Virtual Observatory pilot project different user groups were asked about what visualisation or data needs were and as part of FREEDOM the model users were consulted about how to display that data. This should be replicated in this approach too.
0	0	0	0	1	Relevant knowledge that has been derived from the digital twin - don't make people have to work it our themselves… translated insights in an accessible was that doesn't involve more analysis.

In conclusion, if only one format were offered, an interactive web portal with visualisations of predictions was selected as the most suitable means for end users interact with and receive output from a digital twin of water quality.

3.8 Additional comment from respondents

The final question simply asked if there was anything the respondents would like to add in relation to the creation of a digital twin of running and standing water quality. Twenty-eight respondents left the question blank and another 11 answered that there was nothing else they wanted to add. Twenty respondents wrote additional comments (Table 8).

Some respondents considered it essential to co-produce the digital twin with the user community e.g. "*what will this help with beyond the water quality models we already have? The answer likely to be determined by how much co-production you employ in it".* While others focused on the impact e.g. *"Maybe impacts can be calculated as vulnerability and risk etc."* and *"How will this link to policy or catchments management - this would be an advantage".*

Other respondents suggested linking to other data sources, for example:

> . *"need to complement with water quantity (flow and volume) so as to be able to get MASS FLUXES"*
>
> *"…or models e.g. Ecoforecasting initiative in the US and Chalk Stream Digital twin https://waterinnovation.challenges.org/winners/ecological-digital-twin/".*

While others left positive remarks and wished to be kept involved in progress, one was uncertain they knew enough on the subject and one was very negative doubting the value of models.

Table 8. Respondents' additional comments related to the creation of a digital twin of running and standing water quality.

Ensure Co-production
What will this help with beyond the water quality models we already have? The answer likely to be determined by how much co-production you employ in it
Make sure that you have an expectant and engaged audience who actually want this in the first place and will use it (given how many other things are also available)
For a digital twin to deliver value it must accurately represent the water bodies included, and communicate effectively enough for that representation to influence activity that impacts upon water body quality.
Ensure impact
Maybe impacts can be calculated as vulnerability and risk etc.
Some sort of hazard risk indicators?
How will this link to policy or catchments management - this would be an advantage
Priorities will be different for rural and urban catchments

Retrospective analysis of how accurate predictions were and whether the models are fixed or self learning.
Link to other projects
Whether it can link with other twins to be more holistic
Ofwat Innovation project has recently been launched by Anglian Water and Microsoft to develop a Chalk Stream Digital twin https://waterinnovation.challenges.org/winners/ecological-digital-twin/
I think it's a great initiative - there are already other projects in this space and forecasting water quality was part of a previous Water JPI project - the development of a digital twin should engage with current (e.g. Ecoforecasting initiative in the US) and past project PIs to ensure that previously gained knowledge can be built on.
Need to complement with water quantity (flow and volume) so as to be able to get MASS FLUXES
Data sources
Data should be collected upstream and downstream of all major sewage treatment works; and, upstream and downstream of major tributaries/end of catchments. The Harmonised Monitoring Suite provides a good starting point for site selection. Details should still be available showing the list of HMS sampling sites.
Flow, depth, weather and temperature should be included
There is a lack of overall monitoring of water therefore a need for more volunteers.
General Comments
Good luck and keen to hear how this progresses.
I'd love to be involved - this is very close to my heart and well-aligned with my current research.
Great idea! Uni Birmingham would love to be involved!
I felt I did not have a sufficient understanding of digital twinning to be sure of some of my responses.
Unless have real data from monitoring a large number of rivers, pretty pointless & money could be spent better. Mapping real time pollution incidents is done by others (raw sewage spills) - misconnections just not being dealt with & a big unknown. So, don't see this as a high value project because personally prefer to have real data. If a problem area, use data loggers but most wont handle the large number of pollutants in today's rivers. Being blunt here, but this is my field of research for over 30 years and have never found models to be effective or useful. You never step into the same river twice as the saying goes & you cant predict incidents in advance. Sounds as if the problem is getting bigger & don't have the resources to deal with it, so try a model.

4. References

European Environment Agency (2018). European Waters - Assessment of status and pressures 2018. EEA Report No 7/2018.

House of Common (2022). Water quality in rivers. Fourth Report of Session 2021–22. HC 74

Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-being: Wetlands and Water Synthesis. World Resources Institute, Washington, DC

Siddorn, J., Blair, G., Boot, D., Buck, J., Kingdon, A., Kloker, A., Kokkinaki, A., Moncoiffe, G., Blyth, E., Fry, M., Heaven, R., Lewis, E., Marchant, B., Napier, B., Pascoe, C., Passmore, J., Pepler, S., Townsend, P. and Watkins, J. (2022). An Information Management Framework for Environmental Digital Twins (IMFe). National Oceanography Centre, 33 pp.

Thackeray, S. J. and Hampton, S. E. (2020). The case for research integration, from genomics to remote sensing, to understand biodiversity change and functional dynamics in the world's lakes. *Global Change Biology*, 26, 3230-3240, https://doi.org/10.1111/gcb.15045

Tickner, D., Opperman, J. J., Abell, R., Acreman, M., Arthington, A. H., Bunn, S. E., Cooke, S. J., Dalton, J., Darwall, W., Edwards, G., Harrison, I., Hughes, K., Jones, T., Leclère, D., Lynch, A. J., Leonard, P., McClain, M. E., Muruven, D., Olden, J. D., Ormerod, S. J., Robinson, J., Tharme, R. E., Thieme, M., Tockner, K., Wright, M., and Young, L. (2020). Bending the Curve of Global Freshwater Biodiversity Loss: An Emergency Recovery Plan, BioScience, 70, 330–342, https://doi.org/10.1093/biosci/biaa002

Annex 1 Survey questions

Water Quality Digital Twin Distributed

Page 1: Water Quality Digital Twin

You are invited to participate in a rapid survey for an element of the UK Status, Change and Projections of the Environment Programme (UK-SCAPE). This is a NERC funded project delivered by UKCEH.

We aim to deliver a blueprint and development plan for a surface Water Quality Digital Twin, to provide now-casting and short-term forecasting of a range of environmental pollutants and ecosystem states in both rivers and lakes. The following questionnaire will inform on community needs and aspirations for a Water Quality Digital Twin.

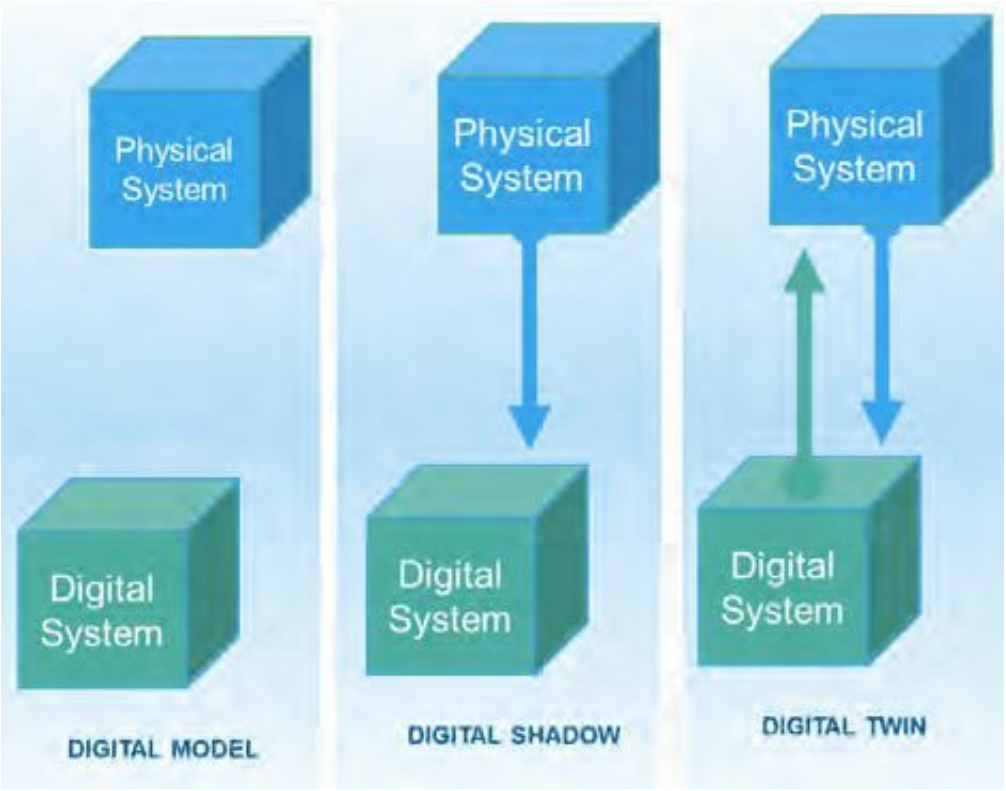
No personal data will be collected. The survey should take no more than 10 min to complete (many questions simply involve ranking or selecting options).

The information you provide will be captured electronically via this survey. The data will be stored to support analysis and any potential future publication documenting the results. We intend to archive the anonymised data for future research use; however, there will be no way for these data to be linked to project participants.

Background to the survey

Water quality and freshwater biodiversity are challenged by a diverse array of human-induced stressors, including climate change, pollution, species introduction, and land use change. The ways in which fresh waters respond to these threats are highly complex, impacting their status, management, and restoration.

A new paradigm that could improve our understanding of water quality and freshwater ecosystem change, and thus aid evidence-informed management, is that of digital twinning. Put simply, a Digital Twin is a virtual representation of a real system that is constantly updated to accurately represent the current state and behaviour of that system.



Digital Twin: a digital model with (real- or right-time) two-way information flows (Siddorn, J., et al. "An Information Management Framework for Environmental Digital Twins (IMFe)." NOC, 2022.).

Digital Twins could be used to bring together diverse data and models in a way that captures the “behaviour” of real fresh waters, allows predictions of water quality and freshwater ecosystem change, and allows us to find key unknowns in the way fresh waters “work”.

We aim to co-develop approaches to digital twinning to deliver a blueprint and development plan based on community needs for a surface Water Quality Digital Twin. The aim is to provide now-casting and short-term forecasting of a range of environmental pollutants and ecosystem states in both rivers and lakes. We are asking your help to assess community needs and aspirations for such a Water Quality Digital Twin.

We would like to hear your views on what a Water Quality Digital Twin should be able to deliver, and how such an approach would enhance your work on fresh waters.

If you consent to take part in this survey, please click on the Next button below.

Page 2: Questionnaire

1. Now that you understand our vision for a Water Quality Digital Twin, what spatial scale do you consider most feasible, useful, realistic, and deliverable for that digital twin? Should outputs be:

- [ ] National
- [ ] Regional
- [ ] Catchment-scale
- [ ] Other

1.a. If you selected Other, please specify:

1.b. Why do you consider your answer to Q1 the most appropriate spatial scale for a digital twin of water quality? ✱ *Required*

2. What temporal scale do you consider most feasible, useful, realistic, and deliverable for a Water Quality Digital Twin? Should outputs be:

Please select between 1 and 5 answers.

- [ ] Sub-daily
- [ ] Daily
- [ ] Monthly
- [ ] Annual

- [ ] Other

2.a. If you selected Other, please specify:

2.b. Why do you consider your answer to Q2 the most appropriate temporal scale for a digital twin of water quality? ✱ *Required*

3. There are many models considering aspects of water quality. What do you consider are the major obstacles to creating a fully functional digital twin of running and standing water quality?

Please select between 1 and 5 answers.

- [ ] Lack of near-real-time data on current water quality at sufficient representative locations to parameterise and update a digital twin
- [ ] Lack of near-real-time data on current water quantity at representative locations to parameterise and update a digital twin
- [ ] Lack of water quality models that contain the processes required at representative locations
- [ ] Lack of water quantity models that contain the processes required at representative locations
- [ ] Other

3.a. If you selected Other, please specify:

3.b. Please elaborate on your answer to Q3 and suggest ways to overcome the limitation(s) you highlighted. ✱ *Required*

4. Which determinand of running and standing water quality do you consider the most important to include in a water quality digital twin? 1=critically important; 2=very important; 3=important; 4=some importance; 5=marginal importance; 6=not important

Please don't select more than 1 answer(s) per row.

Please select at least 9 answer(s).

Please don't select more than 9 answer(s) in any single column.

	1	2	3	4	5	6
Nutrients	☐	☐	☐	☐	☐	☐
Pesticides / herbicides	☐	☐	☐	☐	☐	☐
PFAS (per- and polyfluoroalkyl substances)	☐	☐	☐	☐	☐	☐
Pharmaceuticals	☐	☐	☐	☐	☐	☐
Heavy metals	☐	☐	☐	☐	☐	☐
Cyanobacteria	☐	☐	☐	☐	☐	☐
Plastics	☐	☐	☐	☐	☐	☐
Water colour	☐	☐	☐	☐	☐	☐

Taste and odour compounds	☐	☐	☐	☐	☐	☐

4.a. Please provide details of other determinands which you think are important (i.e. would score 4 or above on scale provided in Q4).

5. Which output variables do you consider to be the most important for a digital twin of running and standing water quality to predict and deliver?

- ○ Physico-chemical variables, including concentrations of determinands
- ○ Biological variables
- ○ Financial impact
- ○ Human wellbeing measures
- ○ Other

5.a. If you selected Other, please specify:

5.b. Please elaborate and explain your reasoning for the answers provided in Q5. ✱
Required

6. How do you think end users should be able to interact with and receive output from a digital twin of water quality?

Please select between 1 and 5 answers.

- [ ] Interactive web portal with visualisations of predictions
- [ ] Downloads of data files in text format (e.g. CSV)
- [ ] Downloads of data files in GIS formats (e.g. Shapefiles)
- [ ] Application Programming Interface (API)
- [ ] Other

6.a. If you selected Other, please specify:

6.b. Is there any other way of interacting with the digital twin that was not included in Q6, or would you like to elaborate on your answer? ✱ *Required*

7. Is there anything else you would like to add to help us accurately report your views on creating a digital twin of running and standing water quality?

Page 3: Respondent information

To inform our analysis can you pease tell us:

8. The type of organisation/institute you are affiliated with? Please select all that apply. *Optional*

Please select between 1 and 7 answers.

- [ ] Government
- [ ] Government Agencies
- [ ] Water company
- [ ] Non-Governmental public body (NGO)
- [ ] Researcher institution and/or universities
- [ ] Commercial business
- [ ] Other

8.a. If you selected Other, please specify:

9. What is your age group? *Optional*

- ○ Younger than 20 years
- ○ 21 - 30 years
- ○ 31 – 40 years
- ○ 41 – 50 years
- ○ 51 – 60 years
- ○ 61 – 70 years

- Over 70 years
- Prefer not to say

10. How would you describe your gender?

- Male (including transgender men)
- Female (including transgender women)
- Prefer not to say
- Prefer to self-describe (e.g. non-binary, gender-fluid, agender)

Page 4: Survey complete

Thank you for completing the survey!

Annex 2 Participant information

UK-SCAPE
UK Status, Change and Projections of the Environment

Participant Information Sheet

Water Quality Digital Twin Blue-print Questionnaire

You are invited to participate in a rapid survey for UK Status, Change and Projections of the Environment Programme (UK-SCAPE).

This Participant Information explains the procedure. This will help you to understand why and how the research is being carried out and what participation will involve. Please contact Dr Jan Dick (jand@ceh.ac.uk) and/or Dr Anne Dobel (annbel79@ceh.ac.uk) if anything is unclear or you have any questions. Please feel free to share this survey with anyone you consider would like to participate.

1. **Over view of study**

1. *Who should complete this survey?*
   This survey should be completed by any stakeholders in the UK who is interested in a Water Quality Digital Twin for assessing impacts on standing and flowing freshwaters. Please feel free to share the link with whomever you consider would be interested. Please note that for GDPR reasons you must be at least 18 years old to participate in this survey.

2. *What is the purpose of this survey?*
   The survey aims to gather stakeholder opinions to create a Water Quality Digital Twin to inform research, decision making in policy, water management and conservation. Further information on the background of this project is given at the start of the survey.

3. *Who is conducting the research?*
   Representatives of UK Centre for Ecology and Hydrology (UKCEH).
   UKCEH is an independent, not-for-profit research institute and is a registered Charity in England & Wales (number 1185618) and in Scotland (number SC049849), and a registered Company Limited by Guarantee in England & Wales (number 11314957).
   The key contact from the project team is Jan Dick (jand@ceh.ac.uk), 44 (0)131 445 8578, UK Centre for Ecology and Hydrology, Bush Estate, Penicuik, EH26 0QB, Midlothian, UK

4. *Who is funding the research?*
   The UK-SCAPE programme started in 2018 and is funded by the Natural Environment Research Council as National Capability (award number NE/R016429/1).

5. *When should the survey be completed by?*
   Please complete the survey within two weeks of receiving this invitation, and no later than the 30 of June 2023.

6. *How long should the survey take?*
   It is estimated that most respondents will respond in less than 10 minutes.
   The survey consists of questions concerning your views on creating a water quality digital twin some are simple selection or ranking type questions and others are open questions

enabling you to write lengthy answers, so response time depends on the length of time it takes you to express your views.

7. *Must I complete this survey?*
No. Participation in this survey is voluntary. You may stop participating at any time and simply not submit your responses. By completing and submitting this survey, you are indicating your agreement to participate. But if after submitting you wish to withdraw your response you can contact the organisers Dr Jan Dick (jand@ceh.ac.uk) and/or Dr Anne Dobel (annbel79@ceh.ac.uk) and if they can identify your response (either by some unique statement to an open question or the time and date you submitted the questionnaire as no unique identifier like name or email is requested) it will be deleted from the database.

2. Data Protection

1. What personal data are we collecting?
The personal data (non-obligatory questions) about you that we seek to collect to describe the population of respondents includes: Type of Organisation, Age Group and Gender

2. Why are we collecting these personal data?

The information will enable us to describe the respondent population and hope to limit or at the very least be aware of any bias within our collected data.

3. How will we use your data?
The information you provide will be captured electronically. The data will be stored to support analysis and any potential future publication documenting the results. We intend to archive the anonymised data for future research use; however, there will be no way for these data to be linked to project participants.

4. How long will we keep your data for?
Any personal data collected will be retained for the duration of the project which will end on 31 March 2023. Other data will be kept for up to 3 years.

5. How secure are the data in our possession?
To collect your data securely we are using JISC to conduct our survey. The data collected will be stored in compliance with UK Data Protection Regulations.

6. Your rights and legal purpose
The GDPR clearly defines individual's rights for your data. We are collecting this information under Legitimate Interest and, as such, your rights are as follows:
Right to access, view, and edit information in a timely manner
Right to be forgotten, which means being deleted from the survey results
Right to be able to opt-out from future messages.

7. Who to contact
If you wish to complain about the use of your information please contact the UKCEH's Data Protection Officer in the first instance (email: Quentin Tucker, Data Protection Officer quetuc@ceh.ac.uk, +44 (0)1491 692427 - UK Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Wallingford, Oxon, OX10 8BB, UK). You may also wish to contact the Information Commissioner's Office (https://ico.org.uk/).

UK-SCAPE
UK Status, Change and Projections of the Environment

8. What if something goes wrong?

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions. The main point of contact is Dr Jan Dick (jand@ceh.ac.uk), 44 (0)131 445 8578, UK Centre for Ecology and Hydrology, Bush Estate, Penicuik, EH26 0QB, Midlothian, UK. If you remain unhappy and wish to complain formally, you can do this by contacting the UK Research Ethics Committee (ukcehresearchethics@ceh.ac.uk, +44(0)1491 692338).

By completing this survey, you consent to the use of your data for the purposes outlined above.

Annex 3 Survey results

Online surveys

Water Quality Digital Twin Distributed

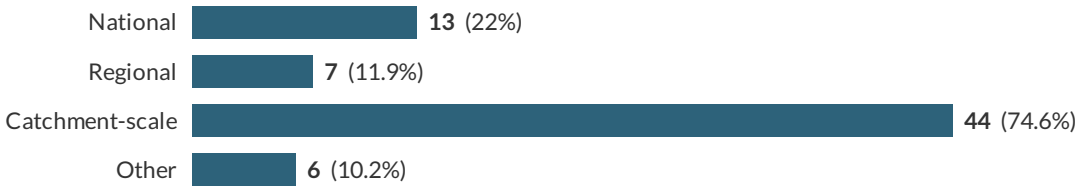
Showing 59 of 59 responses

Showing **all** responses

Showing **all** questions

Response rate: 59%

1 Now that you understand our vision for a Water Quality Digital Twin, what spatial scale do you consider most feasible, useful, realistic, and deliverable for that digital twin? Should outputs be:



Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

1.a If you selected Other, please specify:

Showing all 6 responses	
Local, individual rivers.	1064438-1064420-111750962
Local (point source specific)	1064438-1064420-111750637
a digital twin should be scalable and appropriate across all the different scales mentioned. only answering one would, in my opinion, go against the definition of a digital twin	1064438-1064420-112086292
Local scale, albeit allowing people to see information at the other stated scales	1064438-1064420-112364084
landscape type e.g. chalk, greensand, etc, land use, size, vegetation type/habitat, abstraction and any tidal influences	1064438-1064420-113562920
It depends on the use and the precision of WQ info you can produce	1064438-1064420-113894569

1.b Why do you consider your answer to Q1 the most appropriate spatial scale for a digital twin of water quality?

Showing all 59 responses

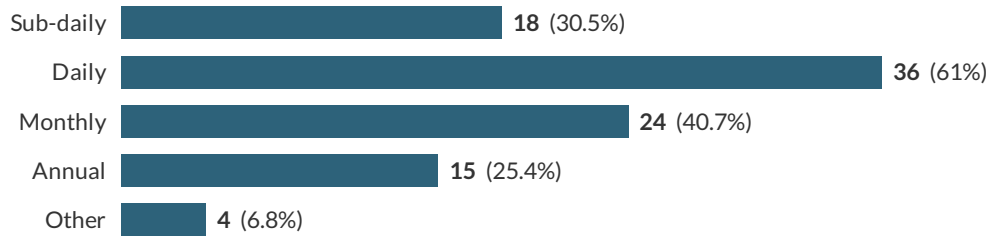
Frankly, would rather have real time monitoring in place than a model approach. Comes across as wanting to monitor water quality on the cheap. Complex range of pollutants today & actions should be taken on real world data.	1064438-1064420-111750962
Catchment scale work is what most agencies are now working to so this would seem appropriate	1064438-1064420-111756751
Most water management currently happens on a catchment scale and there is little joined up work between neighbouring catchments. A digital twin provides an opportunity to understand how specific water quality issues affect both within- and between- catchment ecologies. National is a big ask so perhaps not realistic? But a mix of scales is required because sources (and therefore potential mitigations) are often small scale (especially point source) while effects can be regional.	1064438-1064420-111750637
An ecosystem is a complex system: to depict it with enough accuracy, there is the necessity to medium-high resolution data and models	1064438-1064420-111785637
I feel it is important to start small before scaling up	1064438-1064420-111802090
Sufficient driving data could be collected at that scale	1064438-1064420-111831755
That is the usable spatial scale for water quality	1064438-1064420-111837526
It seems the most sensible.	1064438-1064420-111838425
Because you would want to tweak things such as catchment management practice and so on. Sub-catchment scale would be even better	1064438-1064420-111837340
National would be ideal but realistically due to severe data limitations I think there would be a large amount of work required to gather the necessary data to create an accurate and useful digital twin. I think it should be built upon catchment by catchment up to regional and national scale.	1064438-1064420-111852637
Needs to be appropriate for government policy at that scale.	1064438-1064420-111920414
catchments often transcend national or even regional boundaries, management that doesn't recognise this is insufficient	1064438-1064420-111934647
National to allow view of national changes and prediction and driving policy: catchment scale to allow management decisions and planning	1064438-1064420-112029029
This scale is probably more relevant to a wider audience and likely more manageable than catchment scale but will still provide enough detail that potentially would be lost if this were at a national scale. A happy medium.	1064438-1064420-112055444
It might have more impact if you can do nationally. This might show for example that the Clyde is good and the Forth is bad and make actions happen. Doing on one catchment-scale while more practical might not have an impact.	1064438-1064420-112063465
a digital twin should be scalable and appropriate across all the different scales mentioned. only answering one would, in my opinion, go against the definition of a digital twin	1064438-1064420-112086292
It is a balance between usefulness and feasibility. Waters in a region	1064438-1064420-112114441

experience similar weather and major pressures that might make the twinning feasible while being at a scale that is useful	
water quantity and quality in Catchment-based scale is generally homogeneous.	1064438-1064420-112274258
national is too generalised to be of use. Catchment level could be too detailed to be feasible and transferable. Regional is a good middle ground	1064438-1064420-112278059
Due to the specific nature of each catchment	1064438-1064420-112276112
needs to be able to look at all scales! adn flow of pollutants between catchments.	1064438-1064420-112280299
This is a highly ambitious project. I selected catchment scale as I felt it would be challenging but more achievable to deliver for pilot catchments. This would then enable user feedback and also assist scaling up to national scale- which would be the ultimate objective.	1064438-1064420-112280475
The entire British Isles should be covered. Information must be available though at a local scale, e.g. river reach scale.	1064438-1064420-112301985
I generally work on individual developments, having a catchment scale model could lead to the ability to define site specific pollutant treatment approaches.	1064438-1064420-112337821
With heightened public interest and demands for improvement, we need to be nowcasting at a scale relevant to a public thinking about recreation, aesthetics and ennvironmental impacts on a very local scale.	1064438-1064420-112364084
It will be crude so a broad national picture might be most useful - if it gives water temperature and nutrient loads for say bathing waters	1064438-1064420-112381665
full pathway needed	1064438-1064420-112606418
1) technically feasible - connecting models and sensors required for these systems will be complex and starting small is probably a good idea 2)Predictions can be readily implemented into decision support for everyday management of systems e.g. reservoir or lake manager changing abstraction depth or providing warning for harmful algae or risk to fish from DO decline 3)cost - starting at a small scale will help to ensure the full digital twin can be implemented within a realistic budget - we should learn lessons from previous projects e.g. Demonstration Test Catchments where funding only covered equipment and not interventions initially	1064438-1064420-112757556
Think it provides the highest resolution of data/information upon which to act.	1064438-1064420-112832160
In an ideal world, a Water Quality Digital Twin would exist for every waterbody in the UK so National scale would be useful. However, without knowing what resource is available to deliver the Water Quality Digital Twin, I can only assume it is going to be done with the resource of a typical research group. In this case, catchment-scale seems most feasible, realistic and deliverable to gather the input data required for modelling, update forecasts and generate outputs.	1064438-1064420-112882234

this is the spatial scale for catchment planning and the majority of stakeholders we work with would recognise the catchment as a physical space in which their interest is most tangible	1064438-1064420-112986712
For immediate action and impact using bottom up approach	1064438-1064420-113024619
It needs to be realistic and national scale will require many compromises and assumptions which could be more refined at a local level. Catchment is the building block of hydrological units.	1064438-1064420-113159318
It would give the clearest picture across the full range of watercourses where regional or local may show a bias in areas with fewer issues.	1064438-1064420-113184308
Water exists and (mostly) interacts with other elements of biota, including humans at a catchment scale	1064438-1064420-113222578
This fits with RBMP spatial scale and also makes good scientific sense as processes and pressures operate at a catchment scale	1064438-1064420-113225723
more actionable, good to test approach at relatively small scale first	1064438-1064420-113233190
Catchment is the most tractable to be able to integrate the range of variables/stressors required for accuracy	1064438-1064420-113293314
Water quality is impact by a range of geographical, land use and water management issues which manifest at a local level. Consquently the digital twin needs to be granular enough to measure and predict changes in a way that allows plans and actions to have an impact	1064438-1064420-113327560
Much of WQ chalelnge and issues are catchment related and on that scale. Woudl increase ability to identify issues and address root causes	1064438-1064420-113328011
Monitoring benefits and data	1064438-1064420-113332353
huge variation in river types and pressures on them across national and regional scales, but catchment scale allows for useful twinning	1064438-1064420-113396447
At the outset, cachment level would be sensible to begin with. The ultimate goal should be national scale.	1064438-1064420-113411379
Decision making is needed at national scale - for the Environment Agency this effectively means England, but GB level could also be useful	1064438-1064420-113458817
The digital twin should ideally be available at multiple scales, to allow cross-scale comparisons etc.	1064438-1064420-113463697
If the regulators are to use this tool to manage water companies, the model must be the same across the country. It is probably more feasible to look at issues at catchment level as conditions vary greatly between catchments before we add human input	1064438-1064420-113466880
because there is limited value in modelling a single lake, and why restrict coverage to only regional if national is likely to be almost as feasible?	1064438-1064420-113479490
Catchment-scale mirrors the currently active river basin management planning process.	1064438-1064420-113547175
I was erring between regional or catchment scale. It would be useful to understands trends across the whole of a region to put trends into	1064438-1064420-113551659

context, i.e. are they specific to that location or indicative of wider environmental change. But catchment helps to show how/where interventions have been beneficial.	
Each catchment is different in terms of characteristics and biodiversity. Much of the work we do is on a catchment scale rather than regional or national.	1064438-1064420-113558129
A mix of the bigger picture versus local impact and influences. To understand where pressures and their impacts alter the norm and how comparablen they might be across different catchments for reproducibility of data/data certainty.	1064438-1064420-113562920
There are too many regional differences in geology, land use, human activity to be meaningful at wider scale.	1064438-1064420-113663745
Regional and National would be insufficiently detailed to be useful.	1064438-1064420-113704401
National and regional scale models would not reflect local differences sufficiently to understand what needs to be done to improve or maintian water quality and ecology in the catchment.	1064438-1064420-113741844
Catchment are unique and though imperfect control volumes (eg GW-shed <> SW-shed), is the best way we have to account for fluxes & storages, where budgets for constituents are a key aspect for further understanding & better management	1064438-1064420-113747899
Catchment scale would be most useful but data could be lacking, depending on catchment size	1064438-1064420-113817313
e.g. I might be interested in DO at a national scale to see where hot weather incidents might occur.	1064438-1064420-113894569
Catchments are large enough to be considered as a whole system but also mean something to people locally.	1064438-1064420-113941203
Our organisation works at a catchment scale	1064438-1064420-114000269

2 What temporal scale do you consider most feasible, useful, realistic, and deliverable for a Water Quality Digital Twin? Should outputs be:



Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

2.a If you selected Other, please specify:

Showing all 4 responses	
digital twin should work across scales	1064438-1064420-112086292
weekly. Monthly doesn't feel like a good match for a digital twin but daily and sub-daily you could be lost in a swamp of data. Weekly (at least at first) gives a good useful goal. Could come down to sub-daily for predictive modelling in time	1064438-1064420-112278059
weekly	1064438-1064420-112986712
I think seasonal will be more useful.	1064438-1064420-113466880

2.b Why do you consider your answer to Q2 the most appropriate temporal scale for a digital twin of water quality?

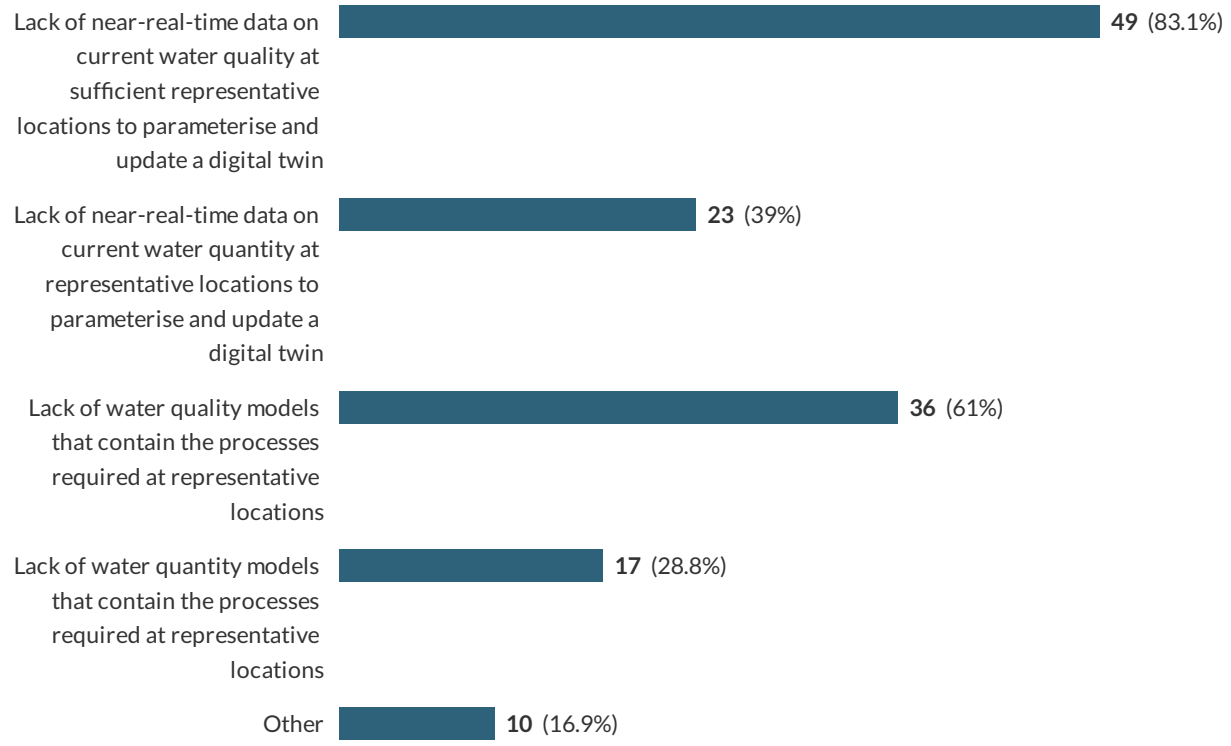
Showing all 59 responses	
Because changes quickly, which is why a mix of water chemistry & biomonitoring best. Chemical data is a snap shot in time, with biota who are living there get a more realistic account if pollution levels.	1064438-1064420-111750962
This is the only way to capture diurnal patterns and the impact pressures have on those patterns. Realtime data gives the best results	1064438-1064420-111756751
This will allow us to understand diurnal cycles of eg nutrients	1064438-1064420-111750637
Daily and sub-daily can be implemented but may threaten the computability of the model and arise a lot of noise in the output (e.g., photosynthesis has an oscillatory trend due to diurnal fluctuations, but on a larger time scale such fluctuations must be analysed with a time-series analysis approach to filter out the oscillations and find whether a trend is present behind the data on a larger time scale)	1064438-1064420-111785637
This period are ok to account for changes	1064438-1064420-111802090
Driving data and uncertainty scale with those temporal scales. Annual would be the least demanding, daily the most.	1064438-1064420-111831755
Daily and monthly scales are what can be monitored with good certainty.	1064438-1064420-111837526
Daily data would allow it to be used to guide leisure activities. Monthly and annual would be useful for research purposes.	1064438-1064420-111838425
The temporal scale needs to be fine enough to see whether any intervention is working. Another critical aspect is how far away in the future the Digital Twin can predict (given some assumed inputs).	1064438-1064420-111837340
availability of data would limit temporal scale. Monthly would still provide a valuable output.	1064438-1064420-111852637
This is a policy tool, not a hazard forecast tool I think. At this stage at least.	1064438-1064420-111920414
daily corresponds to a major criteria of daily rainfall in most parts of the world	1064438-1064420-111934647

world	
Real time monitoring - sub daily Different time scales to allow different monitoring and modelling.	1064438-1064420-112029029
I think sub-daily is probably too detailed for most people to warrant it being useful. Daily, monthly and annually allow some flexibility for the user to choose how they want to look at the information (given often different temporal scales are more appropriate for specific tasks) and I think would be more useful and more realistic to achieve.	1064438-1064420-112055444
Daily should be sufficient, longer term would not help with catchment management	1064438-1064420-112063465
see answer to 1b	1064438-1064420-112086292
Daily response would be valuable but probably not currently feasible. Monthly would be useful while annual provides summary over the seasonal cycle.	1064438-1064420-112114441
Most appropriate for both quantity and quality assessment.	1064438-1064420-112274258
weekly. Monthly doesn't feel like a good match for a digital twin but daily and sub-daily you could be lost in a swamp of data. Weekly (at least at first) gives a good useful goal. Could come down to sub-daily for predictive modelling in time	1064438-1064420-112278059
Reflects data availability	1064438-1064420-112276112
to be relevant it needs to be sub-daily - if its to be anything close to real-time. sensor data should be relatively continuous? needs to link to that...	1064438-1064420-112280299
Daily would be required to inform tactical decision-making, for example on Bathing Water information to the public, or say on smarter abstraction licenses linked to quality as well as flow.	1064438-1064420-112280475
Water quality is highly variable, this was demonstrated by data used for the development of the WRc WQ Monitoring Manual back in the 1980s. Continuous water quality monitoring demonstrates large diurnal variations in WQ data. E.g. dissolved oxygen, temperature and un-ionised ammonia.	1064438-1064420-112301985
In current practice we tend to design for the 1 in 1 years storm event.	1064438-1064420-112337821
Again, teh public want to know the situation right now, albeit allowing people to see information at the other stated scales	1064438-1064420-112364084
I think the potential uses need to be mapped to the most relevant scale to answre this - who wants the outputs of this model and what for?	1064438-1064420-112381665
tbf different answers for feasible than useful! Event impacts are often most important for response and recovery and are needed on sub daily scale. Whereas seasonal and long term, monthly or weekly enough. Daily doesn't quite answer either.	1064438-1064420-112606418
1) This scale is probably a good balance between requirements for short-term management decisions and effective/ accurate sensor resolution 2) This does depend quite a lot on the question you are trying to	1064438-1064420-112757556

answer though - so this should be identified once the purpose of the digital twin is defined 3) It will also be a trade-off between the processes captured by the models and the achievable data resolution of sensors	
I'm not sure daily is necessarily most feasible but I think it would be suitably useful when balancing the technological power required to generate forecasts (and the carbon footprint required to supply this). Sub-daily would be very powerful but it would be a trade-off between whether it would be used vs the power needed to generate these models.	1064438-1064420-112832160
I think sub-daily and daily outputs may be unrealistic if the Water Quality Digital Twin is to be implemented beyond catchment-scale due to the resource and time required to process input data, update models, and generate outputs.	1064438-1064420-112882234
upland catchments respond quickly to weather (hence daily output), but we also need to know longer term climatic effects (annual output)	1064438-1064420-112986712
to relate closely with realities	1064438-1064420-113024619
Sub-daily (hourly or 15minutes) is is likely to require too much computing power and produce lots of data, easier to work on a daily timestep.	1064438-1064420-113159318
Maybe impractical more often though fortnightly a better option?	1064438-1064420-113184308
Otherwise, either too late or we drown in data.....	1064438-1064420-113222578
Overall annual assessment of water quality would be useful for monitoring purposes but monthly timescale would allow seasonal pressures etc to be detected/modelled	1064438-1064420-113225723
parameters vary	1064438-1064420-113233190
Fw ecosystems are highly variable, so several times during the day is of more value than less frequently (ie smaller scale, more depth).	1064438-1064420-113293314
Wit the new requirement for River quality monitoring in the Environment Act should allow meaningful daily or sub daily models to be used	1064438-1064420-113327560
Sub-daily might be too difficult for the beginning, but should be looked at for future itertions.	1064438-1064420-113328011
manageable levels of data that can provide useful outputs	1064438-1064420-113332353
Annual or slightly longer, to reflect the response of river systems to chronic issues. Short term (daily for initial response and monthly for recovery response) more appropriate for individual incidents e.g. a pollution spill	1064438-1064420-113396447
Most approapriate for the likes of extreme rainfall and flooding events and the rapid pressures these pose on CSOs	1064438-1064420-113411379
Daily would be ideal, but if that's not possible then monthly would be OK. Annual can then be inferred from monthly anyway	1064438-1064420-113458817
Balance between temporal scales of variability across different processes; ideally near real-time if possible	1064438-1064420-113463697

I do not comprehend the scale of your project. However, i would imagine daily would be too much data to deal with and annual would not take into consideration the extremes that bring heavy rain in autumn and heat/low rain some summers.	1064438-1064420-113466880
Daily provides the potential to represent event-scale behaviour - probably the most useful for stakeholders. Sub-daily seems unrealistic.	1064438-1064420-113479490
If digital twin outputs were sub-daily they would be most useful for control of water quality impacts, and they would be able to utilise 15-minute interval update river water quaity monitoring being installed by water companies from 2025 onwards.	1064438-1064420-113547175
Monthly not resolute enough, particularly with climate change, impact of summer storms, CSO events etc	1064438-1064420-113551659
Water quality can change significantly overtime, capturing sub-daily data would create significant amount of data. Monthly is too little.	1064438-1064420-113558129
The question asks for most feasible, realistic and deliverable. This is tricky because feasible may be difficult to do on a too high frequency if data storage and cost are prohibitive, but realistically for maximum benefit you need regular monitoring more frequent than annual and monthly. Weather patterns are changing so seasonal shifts/months may change. Daily cycles will also be missed if daily is chosen as this may be quoted as a daily average or a fixed time and not reflect changes throughout the day.	1064438-1064420-113562920
If this twin is going to influence drinking water treatment decisions daily updates on algal blooms would be very beneficial. However long term trends are also important for planning purposes and investment decisions.	1064438-1064420-113663745
With real or near real time monitoring capabilities, sub daily should be possible and less detailed temporal data can be gained from that dataset	1064438-1064420-113704401
Many changes in water courses happen slowly eg due to climate change or delivery of measures. eg AMP measures tend to happen on 5 year cycles. However, seasonal and daily variation in water quality would also help to identify limiting factors in a catchment.	1064438-1064420-113741844
Upper catchments are key systems to protect and restore, and being small, their response time is short	1064438-1064420-113747899
Most realistic for outputs and inputs	1064438-1064420-113817313
again it depends on the use but for the DO example I would want daily - interested in eth need to deploy. bubblers and or prepare for fish kills	1064438-1064420-113894569
Problems with water quality can easily be missed if done only at the daily scale, e.g. due to diurnal cycle in DO, sewage inputs	1064438-1064420-113941203
Daily would be very interesting and the most useful but could prove difficult to deliver so monthly would be the next best.	1064438-1064420-114000269

3 There are many models considering aspects of water quality. What do you consider are the major obstacles to creating a fully functional digital twin of running and standing water quality?



Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

3.a If you selected Other, please specify:

Showing all 10 responses	
Lack of data and models linking water quality data to ecological outcomes	1064438-1064420-111750637
How to join things together and update them with new data	1064438-1064420-111837340
main obstacle is developing the methods to appropriately integrate the real time data that is available with the models. These methods need to be far more sophisticated than simple nudging or calibration that is currently done.	1064438-1064420-112086292
Speed of simulation	1064438-1064420-112364084
there are plenty of models is the real time inputs of pollutants that are missing	1064438-1064420-112381665
Lack of interest by key stakeholders	1064438-1064420-113222578
Cannot choose just one	1064438-1064420-113293314
Lack of understanding of how microspatial variation in water quality and flow influences how representative water quality monitoring is of a whole water body (Environment Agency water body definition).	1064438-1064420-113547175
Need to assimilate weather, and the physical, chemical and biological components along with flow paths to model water quality. There are always missing data.	1064438-1064420-113663745
Probably all the above but the main issue is uncertainty and local factors that are hard to model. However models may be good enough for some purposes and this woudl be good to explore	1064438-1064420-113894569

3.b Please elaborate on your answer to Q3 and suggest ways to overcome the limitation(s) you highlighted.

Showing all 59 responses	
Model's are only as good as the data fed into them. We are deficient in this. Many of our rivers only sampled once a year, few have data loggers in place. If had real time data for DO inverts, microplastics, forever chemicals, pharmaceutical drugs, heavy metals etc wouldn't need to model. Not sure it would be effective, unless do a more in depth set of monitoring on all our rivers you have no real baseline.	1064438-1064420-111750962
The cost of real time data collection may be prohibitive, but the quality of this type of data is what is so important	1064438-1064420-111756751
Real time monitoring of relevant water quality parameters (eg nutrients) is difficult! Needs investment in sensing tech. Water quantity is better quantified but needs to be linked to WQ, including monitoring at same locations.	1064438-1064420-111750637
To obtain near-real-time data; I made use of remote sensing data and	1064438-1064420-111785637

To obtain near-real-time data. I made use of remote sensing data and the approach worked to give me weekly updates. Moreover, I think that if you have the opportunity, the deployment of probes that send measurements in the cloud through a mobile connection could be helpful and informative, ensuring frequent and real-time data	1064438-1064420-111765057
the setup and resources can be expensive	1064438-1064420-111802090
Water quality models have improved greatly over the last couple of decades but driving variables are key to making them work	1064438-1064420-111831755
The parametrization is important.	1064438-1064420-111837526
Increase the amount of sampling done.	1064438-1064420-111838425
creating digital building blocks to join data and update data	1064438-1064420-111837340
Drive for more data, more locations, consistent. Utilise industry data i.e. monitoring required by environmental permits or required for modelling for permit applications. Install permanent data monitoring equipment.	1064438-1064420-111852637
I think we have water quantity and models. Overcome it by working at a very broad time scale to start with. Maybe seasonal. No need for sub-diurnal yet.	1064438-1064420-111920414
No comment	1064438-1064420-111934647
Limited sites monitored - water quantity is covered reasonably well. But greater spatial and temporal data and models required	1064438-1064420-112029029
There is probably a combination of all of those points that will affect the digital twin but near-real-time data is one of the most powerful tools we could use. To scale up to a regional level there would need to be more real-time data to enable that scale to be more representative as detail does tend to be lost the further you extrapolate. Potentially look to input more real-time monitoring instruments into more water bodies for monitoring.	1064438-1064420-112055444
Need real time data and models to apply to this data	1064438-1064420-112063465
More methodological development is needed	1064438-1064420-112086292
I have selected water quality as this main obstacle since this is complex and multi-facetted while quantity in comparison is much easier to assess.	1064438-1064420-112114441
installation of more gauges.	1064438-1064420-112274258
water quantity has continuous monitoring with a networked design. Water Quality sampling is not as structured. Water quantity modelling has been used for water resources and flood forecasting and is more developed than water quality	1064438-1064420-112278059
Better data collection and investment in WQ models	1064438-1064420-112276112
Depending on what we want to do - its its just monitoring good chemical and ecological status there is probably plenty of information there, but to go beyond this will need sampling strategies and advanced monitoring so need to link to existing ongoing research also, and likely sharing of research plans to enable a joined up sampling /	1064438-1064420-112280299

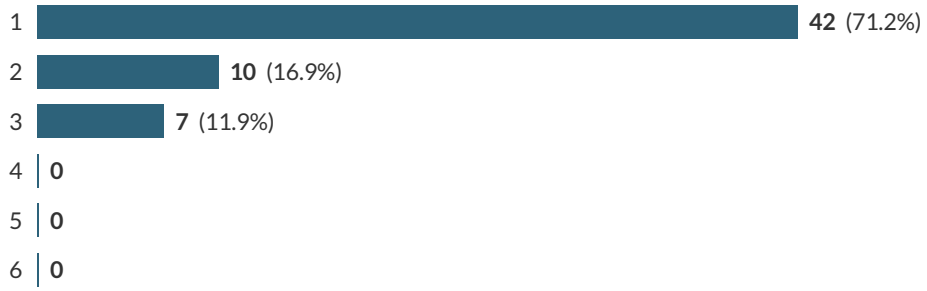
monitoring strategy in the medium term to aid parameratisation.	
Near-real-time data is limited although probably sufficient in case of major rivers e.g. Thames. My understanding is there are no models currently set up to do this - stochastic models may be closed to achieving this with run-times that are useful, at scale.	1064438-1064420-112280475
We previously had continuous water quality monitoring stations in Thames Water's area and Severn Trent Water's area (operated by the National Rivers Authority). They were expensive to maintain. New monitoring will include continuous water quality monitoring stations and therefore the data will be available. I cannot comment on water quantity data, this is outside of my area of expertise.	1064438-1064420-112301985
Current pollutant runoff models are poorly integrated with hydraulic models, combining the two would simplify design processes.	1064438-1064420-112337821
At the moment we build models for specific studies so don;t have full coverage. Model build, calibration and verification is increasingly hampered by limited data as the UK environmental regulators cut back on their routine monitoring. Detailed high frequency models take time to run, but the publci want immediate answers when they ask a question.	1064438-1064420-112364084
tap in to the new CSO monitoring network if you can persuade Water Companies to share it	1064438-1064420-112381665
Hourly water quality measurements are rare and non existent across a whole catchment. Parameters are often v limited and "traditional".	1064438-1064420-112606418
Monitoring many critical water quality variables, particularly nutrients has been technologically challenging and requires sufficient investment in high quality, reliable sensors and the time and technical capability of staff to maintain them - these are critical elements of the infrastructure required and should not be omitted from the budget - poor quality data production will prevent the adequate parameterization of models. Good input data at the appropriate resolution is essential for the creation predictive capabilities or forecasting whilst minimizing uncertainties from that element of the process.	1064438-1064420-112757556
These are all potentially an issue. Particularly in remote locations where guaging infrastructure or real-time loggers are not deployed. Suggest there needs to be a prioritisation of locations to target to begin with, followed by refinements and adaptive management.	1064438-1064420-112832160
I consider the lack of near-real-time data on current water quality at sufficient representative locations to parameterise and update a digital twin to be the biggest obstacle because of the resource required to achieve this. Presumably, sondes or other automated high-frequency measurements of water chemistry parameters are going to be required as well as physical sampling for biological parameters such as algal and fish communities. This is a large undertaking for even a few waterbodies and could be difficult to scale to develop Water Quality Digital Twins even at catchment-scale with monthly outputs.	1064438-1064420-112882234
resources/affordable technology to enable the data collection to be undertaken	1064438-1064420-112986712
To identify more networks, groups or teams to participate in data	1064438-1064420-113024619

gathering towards digital twin modeling	
Real-time data quantity is available from SEPA via their API. Little quality data is collected in real-time	1064438-1064420-113159318
Recruiting volunteers to collect samples but you also need sufficient labs to analyse those samples.	1064438-1064420-113184308
make sure the results have utility and to more than academics!	1064438-1064420-113222578
I am not a modeller but my sense is that hydrological models are well developed whereas those for water quality less so	1064438-1064420-113225723
?	1064438-1064420-113233190
I am not sure about that one	1064438-1064420-113293314
Models will need to be bespoke and require good data sets	1064438-1064420-113327560
There is real-time monitoring available, and in use, but for a limited range of parameters. Need more deveopment on remote sensing techniques to broaden this range	1064438-1064420-113328011
.	1064438-1064420-113332353
lack of monitoring sites is likely to be an ongoing issue, but some gaps could be filled through liaison with partner organisations and perhaps some citizen science monitoring	1064438-1064420-113396447
The answers are based on limited knowledge and understanding of remote and near-real-time monitoring for fundamental water quality parameters let alone harmful chemicals. . I am aware that central government funding initiatives may become available in the near future to begin to address the development and application of remote sensors and near-real-time data reporting and interrogation by AI or machine learning	1064438-1064420-113411379
It's worth starting just with a hydrology digital twin which would be worthwhile in its own right as well as a necessary first step for an exercise involving water quality	1064438-1064420-113458817
Development of new data sources based on enhanced sensor systems and data networks from LOWRAN to 5G	1064438-1064420-113463697
The EA has reduced their environmental monitoring when, ideally, would have been maintained or increased including more parameters	1064438-1064420-113466880
Aquatic systems are hugely diverse in nature, and some may be so unique as to be unmodellable, so representativity can be challenging. Relatively few water quality variables are straightforward to monitor at high frequency.	1064438-1064420-113479490
Research into the cm by cm variation of water quality and flow with depth and distance across a water body, and into the change in response from ion-selective electrodes, conductivity and temperature monitors at different flow rates with stable parameters will allow for corrections that will deliver mnore accurate monitoring.	1064438-1064420-113547175
No comments	1064438-1064420-113551659
All of the above	1064438-1064420-113558129

Data share across different organisations and funding can be preventative to gaining a full picture. Also remote monitoring can be issue with maintenance and battery life. There are lots of models out there, hence I have not ticked this, although revisiting these will be necessary as data improves and validation of the models are key.	1064438-1064420-113562920
Instruments, data and modelling over long time periods are required.	1064438-1064420-113663745
Very little near real time data currently available at representative locations in catchments.	1064438-1064420-113704401
Water company monitoring associated up and downstream of assets should lead to an increase in some data. Models would need to include information on morphology at at particular location.	1064438-1064420-113741844
Combine state agency monitoring network with local partnership low-cost sensors, citizen science and RS/EO	1064438-1064420-113747899
Deploy continuous monitoring water quality sondes if robust relationshsip identified between easily measured parameters (e.g. conductivity, pH, temperature) and other water quality parameters	1064438-1064420-113817313
I want to knwo real time stuff for immediate response but I also want to know how water quality will change over very long time periods in response to policy/regulatory intervention. Models that can accommodate infrastructure stress tests would be good - I don't know what the main hurdles are here.	1064438-1064420-113894569
Collaborative monitoring plans produced by all partners for catchments so that representative locations can be agreed in light of the issues in the catchment and all sources of data being collected by partner organisations are included.	1064438-1064420-113941203
I'm not familiar with what models are available but I do know that real time data is extremely lacking.	1064438-1064420-114000269

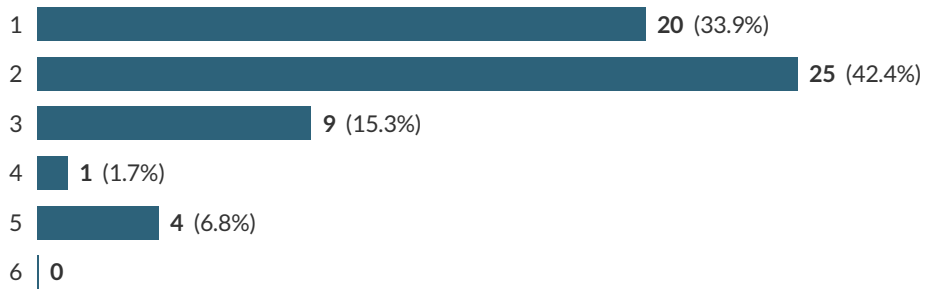
4 Which determinand of running and standing water quality do you consider the most important to include in a water quality digital twin? 1=critically important; 2=very important; 3=important; 4=some importance; 5=marginal importance; 6=not important

4.1 Nutrients



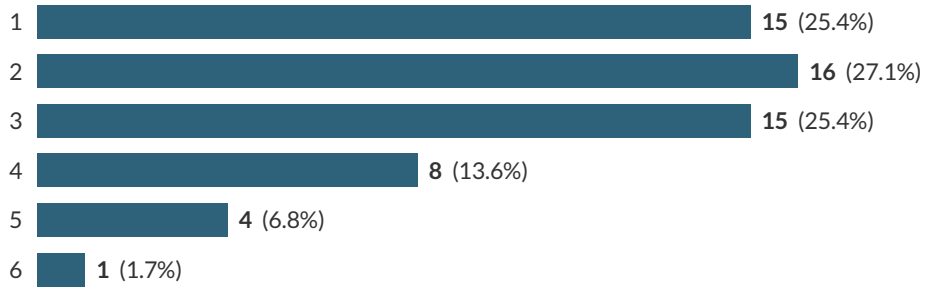
Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

4.2 Pesticides / herbicides



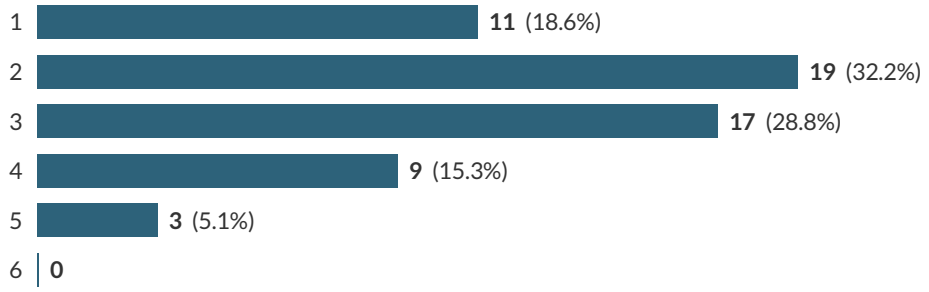
Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

4.3 PFAS (per- and polyfluoroalkyl substances)



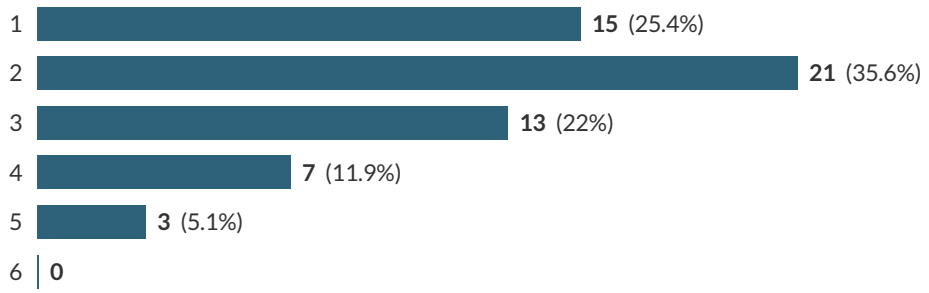
Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

4.4 Pharmaceuticals



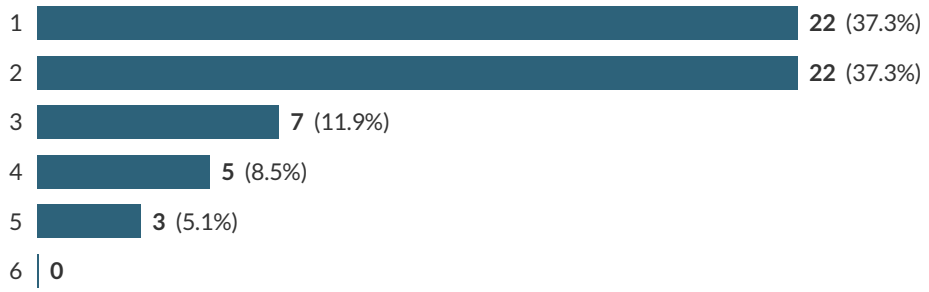
Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

4.5 Heavy metals



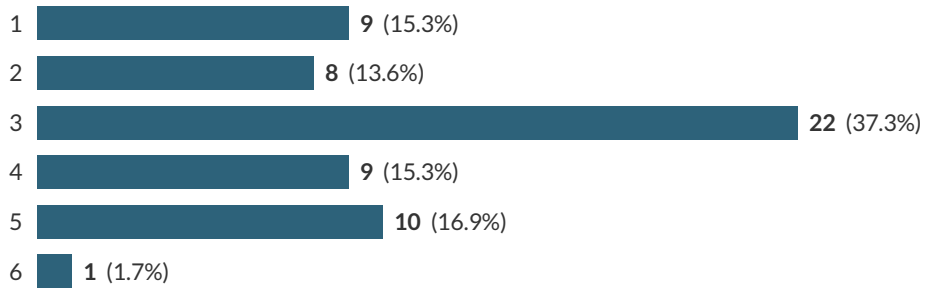
Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

4.6 Cyanobacteria



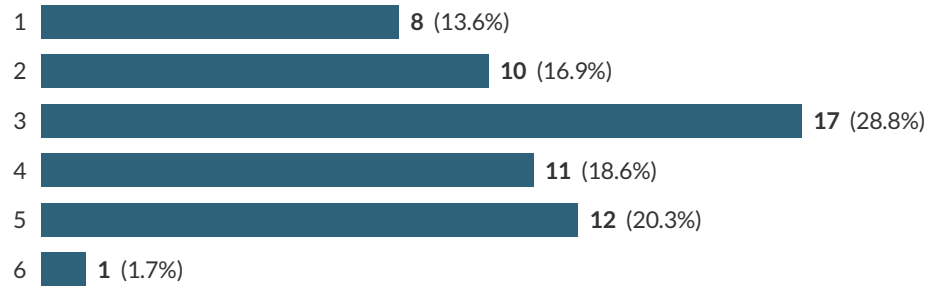
Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

4.7 Plastics



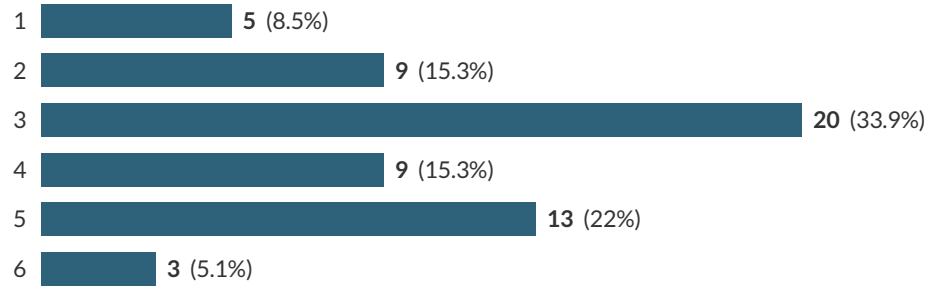
Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

4.8 Water colour



Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

4.9 Taste and odour compounds



Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

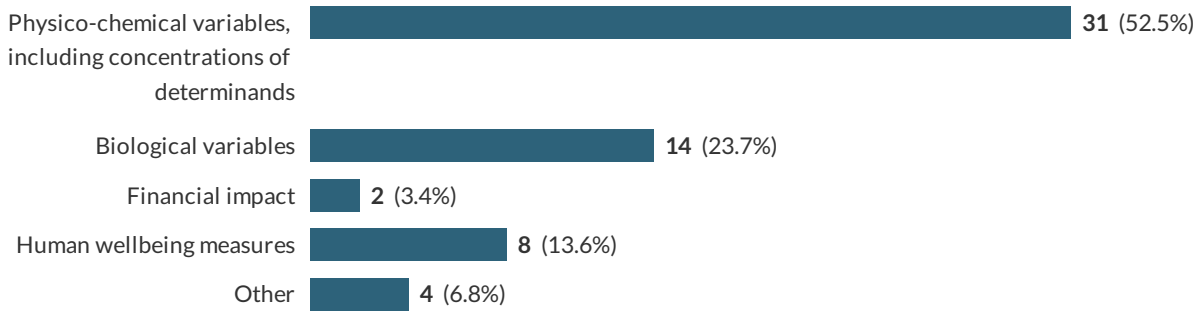
4.a Please provide details of other determinands which you think are important (i.e. would score 4 or above on scale provided in Q4).

Showing all 41 responses	
Coliforms and other gut bacteria. Score 1. Not currently done for rivers unless a reported issue.	1064438-1064420-111750962
Ammonium compounds, dissolved oxygen, temperature, suspended solids. PH.	1064438-1064420-111756751
UN ambient water quality indicator calls for phosphate, nitrate, oxygen, pH, and conductivity. This seems a sensible benchmark for universal core parameters.	1064438-1064420-111750637
Harmful algae; chlorinity and salt potentially affecting soils (if the water basin is connected to the sea)	1064438-1064420-111785637
Chlorophyll as a measure of total algal biomass. Temperature and retention time are also very important	1064438-1064420-111831755

retention time are also very important.	
Temperature- 1	1064438-1064420-111852637
Have you included sewage? I can't see which one that is!	1064438-1064420-111920414
no comment	1064438-1064420-111934647
Greenhouse gases DIC and DOC Salt (road salting - indication of all road pollution)	1064438-1064420-112063465
Information on water flow	1064438-1064420-112086292
For standing water, information on physical structure- ie strength of stratification over time.	1064438-1064420-112114441
None	1064438-1064420-112274258
temperature	1064438-1064420-112278059
plastics additives - BPA, thalates etc. surfacants veterinary chemicals dyes and textile-related chemicals	1064438-1064420-112280299
Ammonia, BOD, DO, Escherichia coli (EC) and Intestinal enterococci (IE)	1064438-1064420-112280475
You have not included dissolved oxygen, this is the number one parameter. Total ammonia and pH are also highly important. Most of the parameters listed are unimportant. We do not understand their toxicity and until we do we should not develop models to measure them - please refer to Prof Sean Comber at Plymouth University for the latest information on environmental quality standards.	1064438-1064420-112301985
In thinking of publc interest, we need to think what matters in terms of public health, recreation, aesthetics and ecology. Much harder to model, but somehow we need to link what we can measure and model more easily to what matters to people.	1064438-1064420-112364084
again its not whats important to teh digital twin that matters but who wants the outputs and what for - how sensitive are their decisions to the model uncertainty?	1064438-1064420-112381665
Dissolved oxygen - essential for a lot of life and pretty indicative of healthy water! Might also be simpler to start with more easily measurable indicators Algal biomass (irrespective of species - since bloom decay can result in low DO and other impairments to freshwaters e.g smothering habitats)	1064438-1064420-112757556
Potentially bacteria (FIOs) but only in urban environments or where there is high human usage.	1064438-1064420-112832160
Algal blooms (and therefore nutrient concentrations) are of greatest concern to the general public so it would be most useful to be able to forecast these to inform management on the ground.	1064438-1064420-112882234
Polyaromatic hydrocarbons (PAHs), Endocrine disruptors compounds(EDCs) Bisphenol, POPs	1064438-1064420-113024619

Any WFD substance.	1064438-1064420-113159318
Water temperatures. Oxygen levels	1064438-1064420-113184308
waste material - rubbish	1064438-1064420-113222578
particulates/fine sediment acidity temperature	1064438-1064420-113233190
eDNA	1064438-1064420-113327560
Temperature and other algae types	1064438-1064420-113328011
presence/impact of invasive non-native species	1064438-1064420-113396447
Where possible, physico-chemical parameters i.e., temp, pH, BOD, COD etc	1064438-1064420-113411379
N species, P species, zinc, nickle, cypermethrin, fipronil, antibiotics, hormones, bisphenol A.	1064438-1064420-113466880
water pH	1064438-1064420-113479490
dissolved oxygen (score 1) only bioavailable nutrients, pesticides/herbicides, PFAS, pharmaceuticals and heavy metals should be measured, not totals.	1064438-1064420-113547175
Sediment	1064438-1064420-113558129
DNA for species monitoring may be health indicators, vegetation, suspended solids, Dissolved Oxygen, Temperature, pH	1064438-1064420-113562920
Do you include iron and manganese with heavy metals? Dissolved oxygen is critical. Stratification and temperature.	1064438-1064420-113663745
Ammonium, dissolved oxygen, temperature, pH more important than many listed above	1064438-1064420-113704401
Temperature, ammonia, dissolved oxygen, feacal indicator organisms.	1064438-1064420-113741844
suspended sediment	1064438-1064420-113747899
Dissolved organic carbon	1064438-1064420-113817313
temperature and dissolved oxygen. Nothing in the table about pathogens? I	1064438-1064420-113894569

5 Which output variables do you consider to be the most important for a digital twin of running and standing water quality to predict and deliver?



5.a If you selected Other, please specify:

Showing all 4 responses	
Determinands that are hard to remove from water treatment processes	1064438-1064420-112276112
I couldn't select more that one - but ultimately water quality is represented by chemical and biological measures of the environment - an integrated measure is probably important	1064438-1064420-112757556
They cannot be neatly separated to choose just one.	1064438-1064420-113293314
I wanted to choose 1 and 4	1064438-1064420-113894569

5.b Please elaborate and explain your reasoning for the answers provided in Q5.

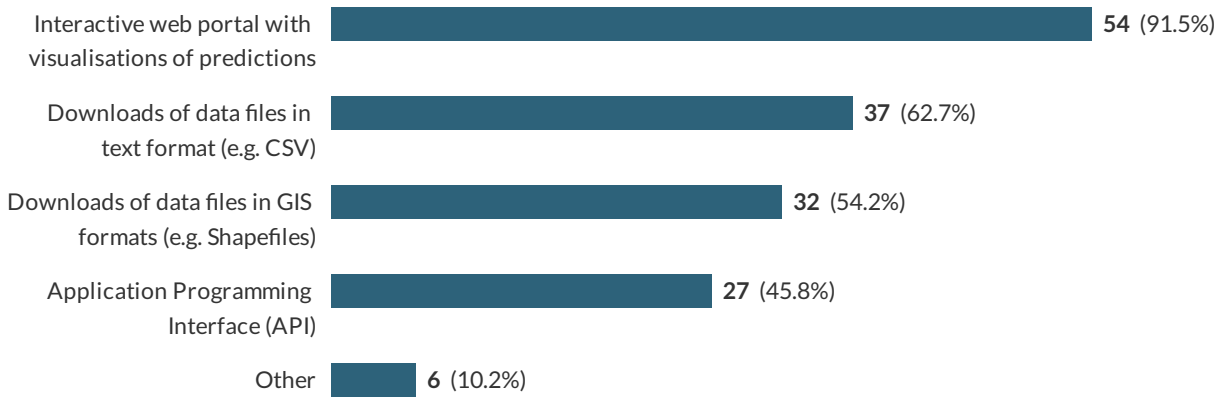
Showing all 59 responses	
These are standard for water quality monitoring - but should be done in conjunction with biological. What chemical and in what quantities feed onto wellbeing and financial (as in cost to correct)	1064438-1064420-111750962
It is important to understand how the determinands are impacting on the waterbody biology	1064438-1064420-111756751
These are the basic variables from which a host of other things can be inferred and derived (eg human well-being, financial, natural capital...)	1064438-1064420-111750637
Physico-chemical variables are at the basis of human-well being measures and financial impact. However, I would have put biological variables at the same importance of physico-chemical variables.	1064438-1064420-111785637
it is a good point for observing changes in the water	1064438-1064420-111802090
Whilst the other variables are important, the Physico-chemical variables control greatly how those other variables are expressed in the environment	1064438-1064420-111831755
Ecosystem health is very important.	1064438-1064420-111837526
This will be the most important for the general public.	1064438-1064420-111838425

I think that's what most users are interested--odour and taste is what people think about 'water quality', and then it's cyanobacteria and sometimes plastics and pharamecuticals	1064438-1064420-111837340
x	1064438-1064420-111852637
The final deliverable is human welfare. If we can start with that, we are more likely to include all the aspects, and perhaps not focus too strongly on accuracy which might slow it down. That can come later?	1064438-1064420-111920414
These are very good for long-term predictions	1064438-1064420-111934647
Biological impact of all determinands is critical for understanding ecological health - which then has impacts on finance and wellbeing	1064438-1064420-112029029
The physico-chemical variables are able to provide information about what might be happening with regard other variables (i.e. biological and in some cases financial).	1064438-1064420-112055444
Physiochemical varaibles can be linked to the other varaibles by models	1064438-1064420-112063465
because further outputs could then be based of this.	1064438-1064420-112086292
The answer will depend on the purpose and end user. This is a scientists answer.	1064438-1064420-112114441
concern of public health	1064438-1064420-112274258
the questions people will be expecting a water quality digital twin to answer will be around physico-chemical variations, interactions, impacts	1064438-1064420-112278059
I work for a water company!	1064438-1064420-112276112
Would have like to pick several but not an option for this question. I think predicting / monitoring loads of pollutants relative to flow rates, abstraction and other human activities in the catchment is essential. The goal should be to support river management actions and to identify early accidental releases (e.g. if sub-daily readings increase beyond some threshold / usual range) then predictions of PEC relative to PNEC and when dangerous levels likely to be seen if no corrective action taken....	1064438-1064420-112280299
Physico-chemical variables could be a very powerful tool in developing solutions to point and diffuse intermittent pollution.	1064438-1064420-112280475
Day to day water quality variation for dissolved oxygen, ammonia, temperature and pH are the important elements. We infer how they impact on biological variables (e.g. fish, algae etc) from that data.	1064438-1064420-112301985
I'm not sure of the answer to this one, but it was a compulsory question.	1064438-1064420-112337821
In thinking of publc interest, we need to think what matters in terms of public health, recreation, aesthetics and ecology. Much harder to model, but somehow we need to link what we can measure and model more easily to what matters to people.	1064438-1064420-112364084
who responds to real time outputs - swimmers, fisheries managers etc	1064438-1064420-112381665
water quality is the ability of life and biodiversity to thrive in the water	1064438-1064420-112606418

Water quality is the ability of life and biodiversity to thrive in the water (dependant on the type of water boy and the "natural" biodiversity expected	1064438-1064420-112...
I think some kind of integrated measure of the physical-chemical-biological environment is a good first step to quantifying water quality that could then be used to generate downstream values related to policy compliance, societal costs etc.	1064438-1064420-112757556
If multiple choice I would have also chosen biological variables as physico-chemical and biological variables would be a powerful combination. I'm speaking as a biologist however and you might get higher investment from financial impact!	1064438-1064420-112832160
I would have selected physico-chemical and biological variables if possible as I believe the two go hand-in-hand. However, I would prioritise physico-chemical variables as prediction of nutrient concentrations, temperature, and dissolved oxygen concentration will allow better management of algal blooms and water for other biodiversity.	1064438-1064420-112882234
for water quality, however many stakeholders want to know is the water clean for me so knowing human wellbeing outputs is also valuable (i can only select one choice in the response)	1064438-1064420-112986712
This output would protect public health and mortality from water borne diseases	1064438-1064420-113024619
Once you have confidence in this then the other variables can be calculated, without this you rush to develop the d/s calculations and don't get your basics right	1064438-1064420-113159318
We saw bioligically moribund rivers in industrial times. Leaving the eu may encourage more pollution!	1064438-1064420-113184308
Biology will essentially influence a lot of other aspects	1064438-1064420-113222578
As an ecologist, if I had to pick on from the above, then I would argue that information on the state of the biology gives a strong indication of the state of water body and hence can inform the other variables on the list	1064438-1064420-113225723
measurable and affect all others	1064438-1064420-113233190
Really hard to choose just one	1064438-1064420-113293314
The model should be aimed at predicating ecological health of the river as this is the key aim of the river basin management plans	1064438-1064420-113327560
Safeguarding human health is at the heart of what we do	1064438-1064420-113328011
Increasing the value of water quality for human wellbeing should enable more connection and overall improvements for biological factors	1064438-1064420-113332353
prediction of biological variables to provide an understanding of the influence of the measured physico-chemical variables on watercourse functionality and the biodiversity it supports	1064438-1064420-113396447
These variiables could be considered as precursors to potential impacts to biological, financial and human health elements	1064438-1064420-113411379
Just trying to be realistic, it's only worth looking at biological variables	1064438-1064420-113458817

just trying to be realistic, it's only worth looking at biological variables once the WQ determinands are understood	1064438-1064420-113458817
Biological responses are key aspect for considerations regarding climate change and biodiversity impacts	1064438-1064420-113463697
Biological variables and human well being depend on physico-chemical variables. It would be great also to understand the financial implications of the lack of WQ and also the costs of achieving better WQ	1064438-1064420-113466880
best to get the most fundamental one done first - the others may follow depending on outcomes	1064438-1064420-113479490
physico-chemical variables are those which we can exert most control over, by releasing reservoir water, reducing abstraction, varying treatment intensity (e.g. by using more or less ferric salt to precipitate phosphorus from wastewater), at short time intervals.	1064438-1064420-113547175
No comment	1064438-1064420-113551659
Requirement to understand how water parameters impact on ecology	1064438-1064420-113558129
I have selected financial as I think this will have the most influence on decision making. The data is broad in the fact it could be used to assess water use, development, habitat condition, resilience, etc	1064438-1064420-113562920
Includes algae and pathogens which impact both drinking water and recreational activities.	1064438-1064420-113663745
All important, but physico-chemical most readily monitorable in near real time	1064438-1064420-113704401
Good ecology can be linked to financial benefit and human well being. Although linked to physchem variables, it is harder to link these to financial impact and human well being. Ultimately to influence decision making the model needs to be able to inform financial impact and human wellbeing, however, if this is the model output it will be too remote from the cause of the problem. Biological variables seems like a good compromise.	1064438-1064420-113741844
basic	1064438-1064420-113747899
I would have liked to also select "Biological variables" and "Human wellbeing measures" in Q5. I selected "Financial impact" because that could also reflect Human wellbeing, e.g. if drinking water supply was impacted there could be costs in providing bottled replacement water. Or if lakes were impacted by algal blooms that could reduce recreational visits and expenditure in the vicinity of the lake.	1064438-1064420-113817313
'd like to see a water quality model that could answer questions about 'is it safe to swim/drink/fish'	1064438-1064420-113894569
We really need both phys-chem outputs and the end biological impacts but I coud only tick one	1064438-1064420-113941203
These variables are important indicators of river health	1064438-1064420-114000269

6 How do you think end users should be able to interact with and receive output from a digital twin of water quality?



Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

6.a If you selected Other, please specify:

Showing all 6 responses	
Notifications of important changes in WQ	1064438-1064420-111750637
that is massively user specific. Water companies will want something very different to a wild swimmer in terms of info.	1064438-1064420-112381665
I think this should be user defined - i.e. co-develop this with the end users in mind	1064438-1064420-112757556
API should be linked to R package for ease of data access etc.	1064438-1064420-113463697
Mobile apps	1064438-1064420-113562920
Relevant knowledge that has been derved from teh digital twin - don't make people have to work it our themselves	1064438-1064420-113894569

6.b Is there any other way of interacting with the digital twin that was not included in Q6, or would you like to elaborate on your answer?

Showing all 59 responses	
No, and as we are talking oreditions, oretty useless to many aquatic scientists who need real data.	1064438-1064420-111750962
Web portal would be most accessible	1064438-1064420-111756751
As above - notifications of WQ changes relevant to water users would be helpful. This could also eventually be used for dynamic discharge permitting.	1064438-1064420-111750637

A desktop version to visualize prediction would be useful, maybe more useful than the interactive web portal	1064438-1064420-111785637
nil	1064438-1064420-111802090
I'm not really sure - it depends on what is created but the first two options are the most user-friendly	1064438-1064420-111831755
No	1064438-1064420-111837526
No	1064438-1064420-111838425
virtual reality and augmented reality	1064438-1064420-111837340
x	1064438-1064420-111852637
Have thresholds of when it is above a certain limit	1064438-1064420-111920414
no comment	1064438-1064420-111934647
Interesting to look at gaming options	1064438-1064420-112029029
Having the ability to download in multiple formats is useful for lots of people. An interactive web portal with visual aids (e.g. maps, graphs etc) would be really useful particularly if they are accompanied with additional information for users to help understand what they are looking at (e.g. what is considered a high or low concentration of x, y or z).	1064438-1064420-112055444
mayeb adding future predictions to the web portal	1064438-1064420-112063465
-	1064438-1064420-112086292
There could also perhaps be an active element when information or warnings (of a cyanobacterial bloom for example) is sent to registered users.	1064438-1064420-112114441
Social media?	1064438-1064420-112274258
most important is web portal with visualisations and predictions	1064438-1064420-112278059
n/a	1064438-1064420-112276112
Be able to programatically feed data into predictive models and capture outputs / predictions. e.g. how long until PEC exceeds PNEC if concentration continues to increase at current rate?	1064438-1064420-112280299
Possibly an open-source architecture that the community can contribute to and write interfaces and add-ons for.	1064438-1064420-112280475
Most users will be satisfied with visualisations. Allowing data to be downloaded leads to confusion as some users may apply the wrong statistics leading to misuse of information. I would say that, I was a regulator!	1064438-1064420-112301985
No.	1064438-1064420-112337821
Different people will want different things. A simple story is essential, but some will want access to teh underlying data.	1064438-1064420-112364084
most people likely to want translated info that is useful knowledge for their purposes	1064438-1064420-112381665

links to data archive and other data centres where primary data available "smart" links to help allow other non-water sectors to link across to pull through the drivers and pressures on that water quality. In isolation the issues and solutions won't be visible	1064438-1064420-112606418
This will depend very much on the user type - as part of the Environmental Virtual Observatory pilot project different user groups were asked about what visualisation or data needs were and as part of FREEDOM the model users were consulted about how to display that data. This should be replicated in this approach too.	1064438-1064420-112757556
Chose my answer as I think these methods will get the highest usage by the general public.	1064438-1064420-112832160
Can't think of other ways right now	1064438-1064420-112882234
these interaction outputs would assist with communications and monitoring for projects	1064438-1064420-112986712
nill	1064438-1064420-113024619
How about "what-if" scneraios, where you can ask the twin to model certain conditions and see the impacts, say discharges of certain compounds etc	1064438-1064420-113159318
No	1064438-1064420-113184308
Interactive feedback from users	1064438-1064420-113222578
no	1064438-1064420-113225723
keep it accessible and simple	1064438-1064420-113233190
None	1064438-1064420-113293314
The outputs need to be focused on potential end users - Regulators, Water Industry and Academia	1064438-1064420-113327560
No	1064438-1064420-113328011
community engagement	1064438-1064420-113332353
ability to manipulate 'inputs' to see how the system would respond if levels of a determinand are improved - predicting recovery as evidence for investment in improvement	1064438-1064420-113396447
Not sure	1064438-1064420-113411379
Really just worth having an API, everything else can build on top of that	1064438-1064420-113458817
Phone app for infield use	1064438-1064420-113463697
Ideally, I would get a graph or a pie chart to help with data visualization. A bit like SAGIS is to SIMCAT	1064438-1064420-113466880
Geographically-based links to other datasets of potential interest, e.g. Met, Air Quality, Catchment Land Use, Population stats etc.	1064438-1064420-113479490
no	1064438-1064420-113547175

No	1064438-1064420-113551659
N/A	1064438-1064420-113558129
Nothing to add	1064438-1064420-113562920
I am not familiar with GIS or API. Please make the outputs interactive intuitive and downloadable for one's own analysis.	1064438-1064420-113663745
No	1064438-1064420-113704401
Possibly text alerts or other notifications to relevant organisations so they can proactively undertake incident response.	1064438-1064420-113741844
not that I can think of	1064438-1064420-113747899
A dashboard showing when likelihood of water quality threshold exceedance	1064438-1064420-113817313
yes, translated insights in an accessible was that doesn't involve more analysis.	1064438-1064420-113894569
No	1064438-1064420-113941203
No	1064438-1064420-114000269

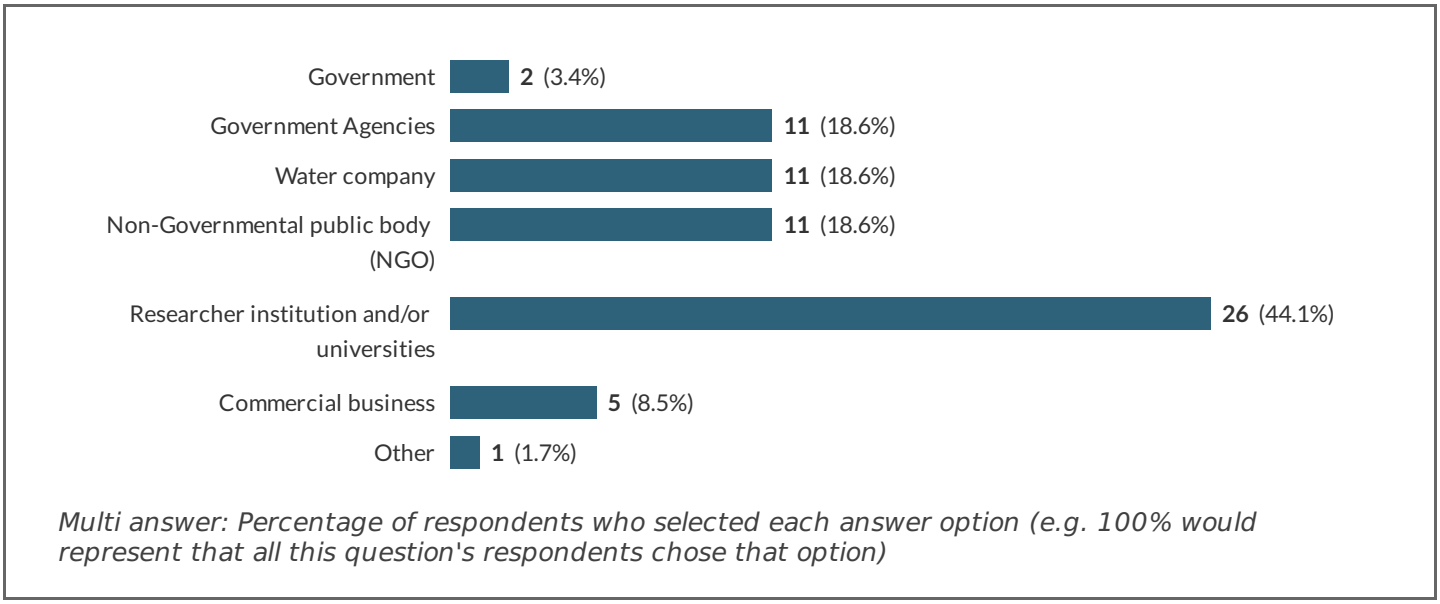
7 Is there anything else you would like to add to help us accurately report your views on creating a digital twin of running and standing water quality?

Showing all 31 responses	
Unless have real data from monitoring a large number of rivers, pretty pointless & money could be spent better. Mapping real time pollution incidents is done by others (raw sewage spills) - misconnections just not being dealt with & a big unknown. So, dont see this as a high value project because personally prefer to have real data. If a problem area, use data loggers but most wont handle the large number of pollutants in today's rivers. Being blunt here, but this is my field of research for over 30 years and have never found models to be effective or useful. You never step into the same river twice as tbe saying goes & you cant predict incidents in advance. Sounds as if the problem is getting bigger & don't have the resources to deal with it, so try a model.	1064438-1064420-111750962
I'd love to be involved - this is very close to my heart and well-aligned with my current research. Izzy Bishop (UCL) - i.bishop@ucl.ac.uk	1064438-1064420-111750637
Maybe impacts can be calculated as vulnerability and risk etc.	1064438-1064420-111837526
x	1064438-1064420-111852637
no	1064438-1064420-111934647
How will this link to policy or catchments management - this would be an advanatage	1064438-1064420-112063465
-	1064438-1064420-112086292
I felt I did not have a sufficient understanding of digital twinning to be	1064438-1064420-112114441

sure of some of my responses.	
No	1064438-1064420-112274258
whether it can link with other twins to be more holistic	1064438-1064420-112278059
great idea! Uni Birmingham would love to be involved!	1064438-1064420-112280299
Good luck and keen to hear how this progresses.	1064438-1064420-112280475
Data should be collected upstream and downstream of all major sewage treatment works; and, upstream and downstream of major tributaries/end of catchments. The Harmonised Monitoring Suite provides a good starting point for site selection. Details should still be available showing the list of HMS sampling sites.	1064438-1064420-112301985
what will this help with beyond the water quality models we already have? The answer likely to be determined by how much co-production you employ in it	1064438-1064420-112381665
I think it's a great initiative - there are already other projects in this space and forecasting water quality was part of a previous Water JPI project - the development of a digital twin should engage with current (e.g. Ecoforecasting initiative in the US) and past project PIs to ensure that previously gained knowledge can be built on.	1064438-1064420-112757556
none at the moment	1064438-1064420-113024619
There is a lack of overall monitoring of water therefore a need for more volunteers.	1064438-1064420-113184308
Make sure that you have an expectant and engaged audience who actually want this in the first place and will use it (given how many other things are also available)	1064438-1064420-113222578
no	1064438-1064420-113225723
priorities will be different for rural and urban catchments	1064438-1064420-113233190
Ofwat Innovation project has recently been launched by Anglian Water and Microsoft to develop a Chalk Stream Digital twin https://waterinnovation.challenges.org/winners/ecological-digital-twin/	1064438-1064420-113327560
No	1064438-1064420-113328011
None, thanks	1064438-1064420-113411379
For a digital twin to deliver value it must accurately represent the water bodies included, and communicate effectively enough for that representation to influence activity that impacts upon water body quality.	1064438-1064420-113547175
N/A	1064438-1064420-113558129
Flow, depth,, weather and temperature should be included	1064438-1064420-113562920
Retrospective analysis of how accurate predictions were and whether the models are fixed or self learning.	1064438-1064420-113663745
No	1064438-1064420-113704401
need to complement with water quantity (flow and volume) so as to be	1064438-1064420-113747899

able to get MASS FLUXES	
some sort of hazard risk indicators?	1064438-1064420-113894569
No	1064438-1064420-113941203

8 The type of organisation/institute you are affiliated with? Please select all that apply.

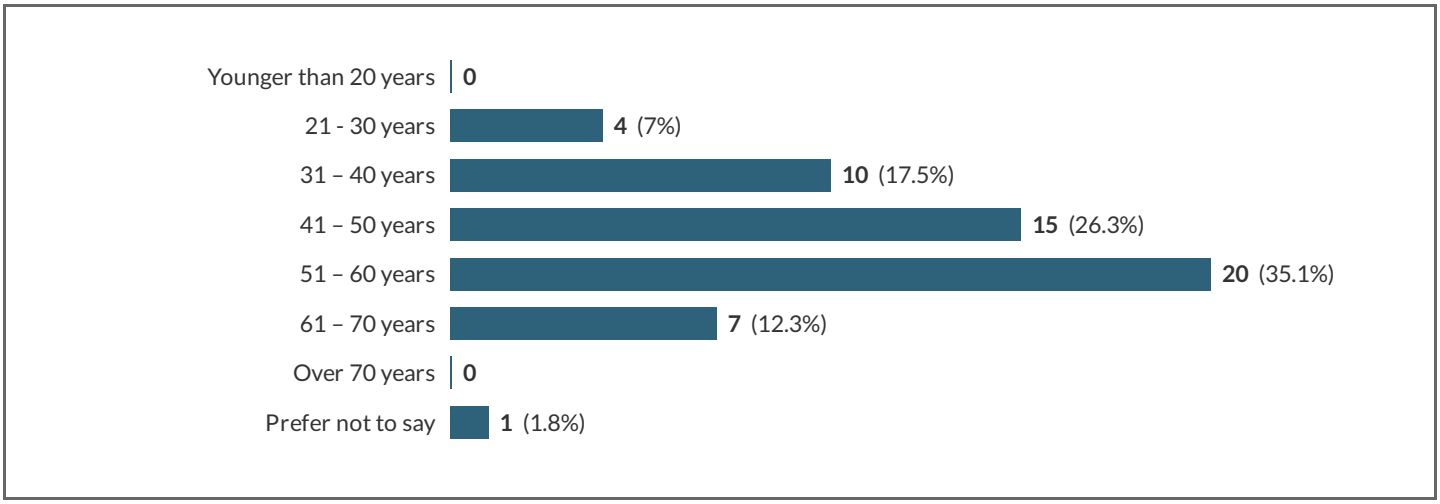


Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)

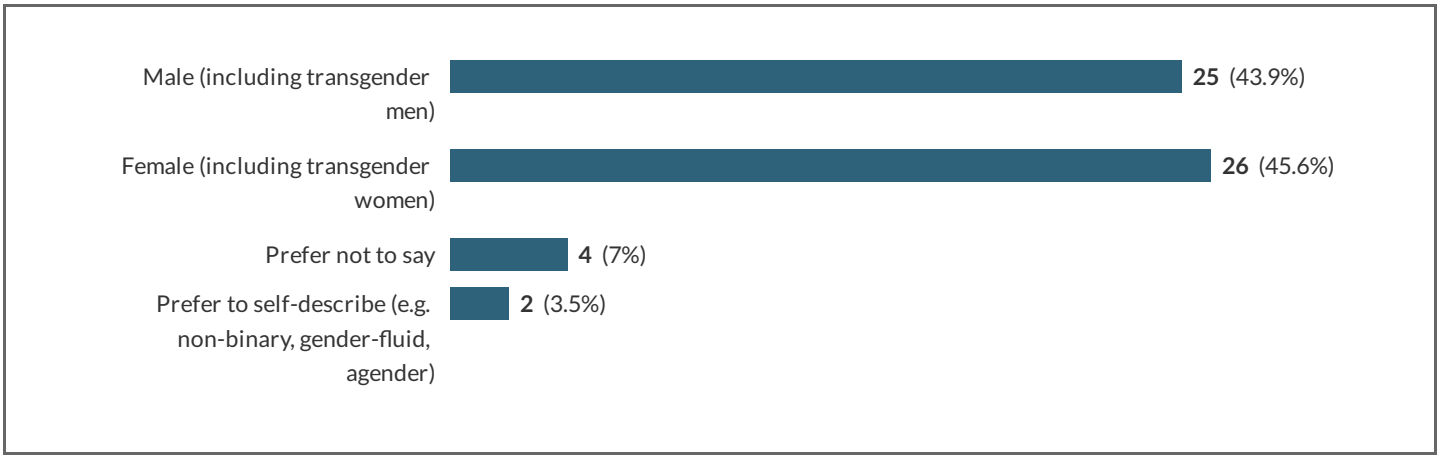
8.a If you selected Other, please specify:

Showing 1 response	
Consultant	1064438-1064420-112029029

9 What is your age group?



10 How would you describe your gender?



Contact

enquiries@ceh.ac.uk

@UK_CEH

ceh.ac.uk

———

Bangor

UK Centre for Ecology & Hydrology
Environment Centre Wales
Deiniol Road
Bangor
Gwynedd
LL57 2UW

+44 (0)1248 374500

Edinburgh

UK Centre for Ecology & Hydrology
Bush Estate
Penicuik
Midlothian
EH26 0QB

+44 (0)131 4454343

Lancaster

UK Centre for Ecology & Hydrology
Lancaster Environment Centre
Library Avenue
Bailrigg
Lancaster
LA1 4AP

+44 (0)1524 595800

Wallingford (Headquarters)

UK Centre for Ecology & Hydrology
Maclean Building
Benson Lane
Crowmarsh Gifford
Wallingford
Oxfordshire
OX10 8BB

+44 (0)1491 838800

UK Centre for Ecology & Hydrology