A multi-chemistry modelling framework to enable flexible and reproducible water quality simulations in existing hydro-models: 1. The OpenWQ concept and the water quality modelling lab

Diogo Costa[1], Kyle Klenk[2], Wouter Johannes Maria Knoben[2], Andrew Ireson[2], Raymond J Spiteri[2], and Martyn P. Clark[3]

[1]University of Évora
[2]University of Saskatchewan
[3]Centre for Hydrology, University of Saskacthewan Coldwater Laboratory

June 19, 2023

Abstract

This work advances the incorporation and cross-model deployment of multi-biogeochemistry and ecological simulations in existing process-based hydro-modelling tools. It aims to transform the current practice of water quality modelling as an isolated research effort into a more integrated and collaborative activity between science communities. Our approach, which we call "Open Water Quality" (OpenWQ), enables existing hydrological, hydrodynamic, and groundwater models to extend their capabilities to water quality simulations, which can be set up to examine a variety of water-related pollution problems. OpenWQ's objective is to provide a flexible biogeochemical model representation that can be used to test different modelling hypotheses in a multi-disciplinary co-creative process. In this paper, we introduce the general approach used in OpenWQ. We detail aspects of its architecture that enable its coupling with existing models. This integration enables water quality models to benefit from advances made by hydrologic- and hydrodynamic-focused groups, strengthening collaboration between the hydrological, biogeochemistry, and soil science communities. We also detail innovative aspects of OpenWQ's modules that enable biogeochemistry lab-like capabilities, where modellers can define the pollution problem(s) of interest, the appropriate complexity of the biogeochemistry routines, and test different modelling hypotheses. In a companion paper, we demonstrate how OpenWQ has been coupled to two hydrological models, the "Structure for Unifying Multiple Modelling Alternatives" (SUMMA) and the "Cold Regions Hydrological Model" (CRHM), demonstrating the innovative aspects of OpenWQ, the flexibility of its couplers and internal spatiotemporal data structures, and the versatile eco-modelling lab capabilities that can be used to study different pollution problems.

A multi-chemistry modelling framework to enable flexible and reproducible water quality simulations in existing hydro-models: 1. The OpenWQ concept and the water quality modelling lab

Diogo Costa[1,2], Kyle Klenk[1], Wouter Knoben[1], Andrew Ireson[1], Raymond J. Spiteri[1], Martyn Clark[1]

[1]Mediterranean Institute for Agriculture, Environment and Development & Global Change and Sustainability Institute, Department of Geosciences, Universidade de Évora, Pólo da Mitra, Ap. 94, 7006-554 Évora, Portugal
[2]University of Saskatchewan, 105 Administration Pl, Saskatoon, SK S7N 5A2, Canada

Abstract

This work advances the incorporation and cross-model deployment of multi-biogeochemistry and ecological simulations in existing process-based hydro-modelling tools. It aims to transform the current practice of water quality modelling as an isolated research effort into a more integrated and collaborative activity between science communities. Our approach, which we call "Open Water Quality" (OpenWQ), enables existing hydrological, hydrodynamic, and groundwater models to extend their capabilities to water quality simulations, which can be set up to examine a variety of water-related pollution problems. OpenWQ's objective is to provide a flexible biogeochemical model representation that can be used to test different modelling hypotheses in a multi-disciplinary co-creative process. In this paper, we introduce the general approach used in OpenWQ. We detail aspects of its architecture that enable its coupling with existing models. This integration enables water quality models to benefit from advances made by hydrologic- and hydrodynamic-focused groups, strengthening collaboration between the hydrological, biogeochemistry, and soil science communities. We also detail innovative aspects of OpenWQ's modules that enable biogeochemistry lab-like capabilities, where modellers can define the pollution problem(s) of interest, the appropriate complexity of the biogeochemistry routines, and test different modelling hypotheses. In a companion paper, we demonstrate how OpenWQ has been coupled to two hydrological models, the "Structure for Unifying Multiple Modelling Alternatives" (SUMMA) and the "Cold Regions Hydrological Model" (CRHM), demonstrating the innovative aspects of OpenWQ, the flexibility of its couplers and internal spatiotemporal data structures, and the versatile eco-modelling lab capabilities that can be used to study different pollution problems.

1 Introduction

Societies are becoming increasingly aware of the importance of harmonizing economic development with thriving aquatic ecosystems (Keith et al., 2022; Frank & Schlenker, 2016). This has led to an increase in the use of ecological and water quality models to study management solutions to pollution problems such as nutrient excess due to the use of fertilizers or microbiological contamination due to the discharge of untreated wastewater in rivers.

Over the past decades, substantial progress has been made in computational hydrology, producing many powerful generic and fit-for-purpose modeling tools. Some of these models have been particularly tailored to specific environments, such as the Cold Regions Hydrological Model (J. Pomeroy et al., 2022; J. W. Pomeroy et al., 2007), the Wetland DEM Ponding Model (Shook et al., 2013), and the FLUXOS-Overland model (Costa, Shook, et al., 2020), that have been developed to deal with the specific hydrological challenges of simulating the Canadian Prairie pothole region due to the lack of well-defined river networks. Unfortunately, the science and modelling progress achieved by such hydrological model development communities rarely transfers to (and translates into) water quality modeling improvements because hydrological and water quality modeling tools are typically developed in isolation, a problem that becomes clearly noticeable when examining the hydrological calculations embedded in popular water quality modeling tools, such as SWAT (Arnold et al., 1998), HYPE (Lindström et al., 2010; Arheimer et al., 2012), and INCA (Whitehead et al., 1998; A. J. Wade et al., 2002; Jackson-Blake et al., 2016), that often rely on simplifications that may be limiting for some regions and applications (Costa, Baulch, et al., 2020a).

Besides the issues with the communication and transfer of knowledge between these two scientific communities, other scientific and technical biogeochemistry-specific challenges remain critical for reliable simulations across climate regions, landscapes, and seasons (Costa, Baulch, et al., 2020a; Fu et al., 2019; Wellen et al., 2015). Some key challenges include (1) the adequate representation of the complexity and heterogeneity of biogeochemical processes and their dependency on weather, soil, and sediment characteristics; (2) the dependency on accurate hydrological drivers to reliably track the vertical and lateral movement of chemical

constituents (being at the end of the climate-hydrology-ecology modeling chain invariably propagates uncertainties into water quality simulations); and (3) the often limited range of applicability of models to specific pollution problems and landscape conditions due to their rigid and heavily parameterized representation of water quality processes.

Underlying all of these issues are foundational questions of scientific reproducibility in computational hydrology (C. Hutton et al., 2016), as well as considerable challenges in defining the appropriate complexity, scale, and scope of water quality models (Costa, Baulch, et al., 2020a; Moore et al., 2006; Mekonnen, 2016; Shoemaker, 1997). This challenge is in part due to uncertainties associated with the prediction of many hydrological and biogeochemical responses at different spatial scales and geographical settings, which are aggravated by the often sparse water quality data available for testing and validation of the models. This raises critical questions for the design, application, and benefit of such modeling tools (Beck, 1987; Moore et al., 2006; A. Wade et al., 2008; Costa, Baulch, et al., 2020a).

Improving process-based hydro-biogeochemical models for terrestrial hydrological simulations requires effective and continuous integration of progress across several research areas, including hydrology, biogeochemistry, and soil science. Models must combine practical aspects related to model application with theoretical scientific insights at various degrees of scientific maturity and geographical applications. Such cross-disciplinary modeling efforts require modelers to make model development decisions based on several considerations common to hydrological models (Clark et al., 2011) that include (1) model fidelity, complexity and practicality, (2) scientific reproducibility and transparency, and (3) data availability.

It has been recognized that more flexible water quality modelling approaches are needed to address these issues (Yang et al., 2022; Costa, Shook, et al., 2020). Some approaches have emerged that address some of the challenges and improve transparency and engagement among hydrologists, biogeochemists, soil scientists, and decision-makers, such as the STELLA (Structural Thinking and Experiential Learning Laboratory with Animation) software (Richmond, 2003) and the MIKE Eco Lab, which is a water quality modelling addition to the MIKE tools (Refsgaard & Storm, 1995). However, such tools are not suitable for integration into modelling platforms developed by various hydrological communities around the world. In the case of STELLA, the tool is standalone and more suitable for applications where simpler batch-reactor assumptions are applicable, such as for small reservoirs and wastewater treatment plants. The MIKE tools from DHI now provide an Eco-Lab module for more flexible water quality and ecological simulations, but this module is a proprietary tool integrated into the MIKE ecosystem of tools, so it does not benefit the wider hydrology and modelling communities, and it is constrained by the chemical constituents and types of biogeochemical processes that are explicitly introduced by the MIKE model developers. More recently, Yang et al. (2022) proposed a new model named HiWaQ for flexible catchment water quality assessments with compatibility for multiple hydrological model structures. This is a valuable contribution toward a more unified hydrological-water quality modelling approach, but the current version is limited to the nitrogen cycle.

2 Current modelling capabilities and needs

Widely used process-based catchment nutrient models such as SWAT, HYPE, and INCA have been at the forefront of innovation in water quality modeling, paving a way for researchers and decision-makers around the world to investigate solutions to a variety of pollution problems, particularly related to nutrient pollution. These models have become increasingly complex and heavily parameterized, but they remain a limited representation of reality because hydro-biogeochemical processes are highly complex in natural environments (A. Wade et al., 2008; Beck, 1987; Costa, Baulch, et al., 2020b). These models typically simulate a series of biogeochemical processes conceptualized to address particular pollutions problems, and processes are represented through a combination of empirical and physico-chemically based methods, often leading to many calibration parameters and thus increasing

the risk of parameter equifinality (Costa, Baulch, et al., 2020b). Sparse and sporadic water quality measurements, which are common problems for most water bodies, limit the further application and development of these water quality models.

The combination of process-representation methods with varying degrees of empiricism is in part due to knowledge gaps in understanding the drivers and controls of hydrological and biogeochemical responses at various spatial scales and across different landscapes and climate zones. In regions where relatively uncommon processes may play an important role in the overall water quality dynamics, the use of these models becomes problematic because there is little flexibility for adjusting conceptual models, adding or removing processes, or testing different modeling hypotheses. For example, research has shown that in some cold regions, processes such as preferential infiltration of hydrochemical into frozen soils (Lilbaek & Pomeroy, 2007), preferential elution of hydrochemicals from melting snowpacks (Davies et al., 1987; Marsh & Pomeroy, 1999; Costa & Pomeroy, 2019), microbial uptake and fixing of nitrogen in melting snowpack (Jones, 1999), and volatilization of nitrogen during snow redistribution and sublimation (J. W. Pomeroy et al., 1991; J. Pomeroy et al., 1999) may affect water quality, but they are not represented in most popular models. This lack of representation may compromise their use in such regions and calls for a more flexible approach to water quality modeling that enables a systematic and controlled approach for the addition and removal of processes as needed to reflect (1) regional and climate characteristics, (2) data limitations, and (3) objectives of the study.

3 The OpenWQ concept

3.1 Overview

OpenWQ is a coupler-modelling framework designed to provide portable and customizable multi-chemistry modelling capabilities to existing hydro-models. The vision is to create a tool that could plug into existing process-based hydrological, hydrodynamic, and groundwater models to extend their capabilities to environmental and ecological studies. The approach allows tailoring chemical-microbiological constituents and biogeochemistry-cycling processes to enable the representation of different pollution problems and landscapes, as well as to compare modelling hypotheses. This framework is a much-needed effort to bring the hydrology and biogeochemistry communities together, optimizing research and investment efforts. The work stems from previous model developments by Environment and Climate Change Canada and the University of Saskatchewan, particularly CRHM-WQ (Cold Regions Hydrological Model - Water Quality) (Costa et al., 2021) that extends the original CRHM platform (hydrology) model to nitrogen and phosphorus simulations (J. W. Pomeroy et al., 2007), the WINTRA framework (Costa et al., 2017), the multiphase multilayer PULSE snow hydrochemistry model (Costa et al., 2018), and the FLUXOS-OVERLAND model for watershed hydrodynamic-water quality simulations suitable for Prairie regions (Costa, Shook, et al., 2020).

OpenWQ aims to address three main challenges with existing water quality models: (1) structural rigidity in the representation of chemical constituents and biogeochemical processes, (2) over-simplification and limitations of hydro-flux calculations, and (3) inadequacy for testing different modelling hypotheses for proper quantification of structural uncertainty. Structural rigidity is perhaps the key factor that hinders the effective use of models across landscapes and in complex, diverse environments (e.g., permafrost, peatlands, variable contributing areas) that require more investigative, open-ended, and interactive simulation approaches. A flexible environment that enables integrating new methods and concepts from complementary disciplines and experts (e.g., limnology, soil science, biogeochemistry) is critical to advance science and promote meaningful and impactful cross-disciplinary collaborations. The static, hard-coded implementation of biogeochemical reaction-network transformations limits their suitability for a wide range of environmental problems. It also hinders multi-disciplinary co-creation efforts because models provide little flexibility for

changing, expanding, and testing different strategies for biogeochemical cycling representation. Finally, the hydro-flux calculations that are embedded in water quality modelling tools are often outdated or limited compared to dedicated, disciplinary hydro-models (e.g., hydrological, hydrodynamic, and groundwater modelling tools). Uncertainty in model structure, process representation, and future scenarios (e.g., climate change) cannot be adequately quantified without flexible and transparent modelling structures.

3.2 General design

This section describes how OpenWQ was designed to address the key challenges in water quality modelling described before. First, focus is given to describe how OpenWQ tackles the need for more robust hydro-flux and associated solute transport calculations through integration within existing hydro-models. This includes details on how OpenWQ can be linked to existing hydro-models considering key practical aspects of version control (Section 3.2.1), portability (Section 3.2.2), code integration (Sections 3.2.3), and state variables and spatiotemporal discretization (Section 3.2.4) . Second, focus is given to describe how OpenWQ tackles the need for more flexible representation of physical and biochemical processes, which includes aspects of model input (Section 3.3), model structure to enable testing modelling hypotheses and quantifying structural uncertainty (Sections 3.4 and 3.5).

3.2.1 Obtaining OpenWQ

OpenWQ can be obtained from the official GitHub repository: `https://github.com/ue-hydro/openwq`. The recommended method for obtaining OpenWQ is to clone the repository with "git clone". Downloading the repository is also possible, but it is less advisable because OpenWQ is continuously updated and improved. OpenWQ is designed for the user to update their copy quickly and efficiently with "git pull" instead of requiring them to download each new version. The cloning (or download) should be performed inside the host hydro-model code directory as shown in Fig. 1. It will create a new folder named "openwq" with the entire repository. The "main" branch should be used because it will contain the latest official release version. Compiling OpenWQ is carried out with CMake (Kitware, 2022) and the provided "CMakeLists" file. When new versions of OpenWQ are made available, they will only require the user to recompile using the same "CMakeLists" file. Once the coupling of OpenWQ to a hydro-model is completed, carrying out updates to OpenWQ's source code will not break OpenWQ's API calls and interface/coupler functions in the host hydro-model.

3.2.2 Portability via internal dynamic coupler

In order to optimize the implementation of OpenWQ in hydro-models, the model has been designed as an internal coupler module (Fig. 2). External coupling (i.e., OpenWQ reading output files from the host-model and running standalone) was not a viable option because hydro-models, particularly hydrological models, often deal with many water fluxes moving around vertically and horizontally within and across hydrological compartments (e.g., snow, soil), which are not always possible to export and disentangle. Even in cases where models allow exporting all water fluxes separately, mapping those in OpenWQ and harmonizing units to correctly compute the corresponding solute mass transport would be extremely difficult for most applications. Focusing on a flexible full-coupling approach with minimal code re-engineering allowed us to address this problem, with OpenWQ being specifically designed to adapt its internal structure (e.g., spatial domain, temporal resolution) to that of the host hydro-model with the support of interface and coupler functions that establish one-way, plug-in-type communications between the two simulation systems. Although these interface and coupler functions may need adjustments to harmonize particular aspects of the host hydro-models, a generic "hydro-link" file is provided with a template for these

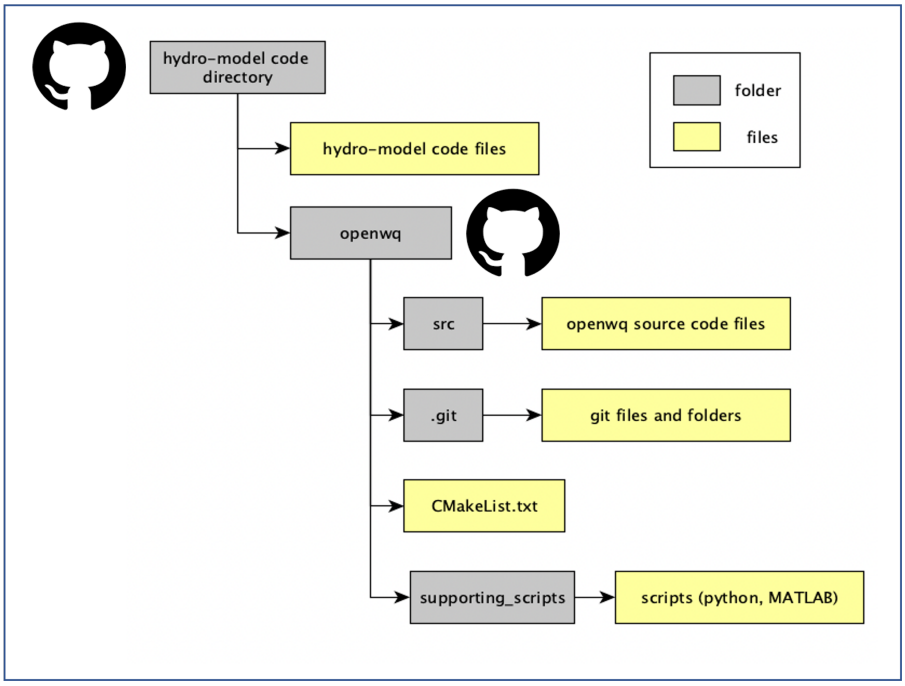


Figure 1: General model structure used in hydrological model

functions. It is possible to create such a template because most hydro-models follow the general model architecture depicted in Fig. 3.

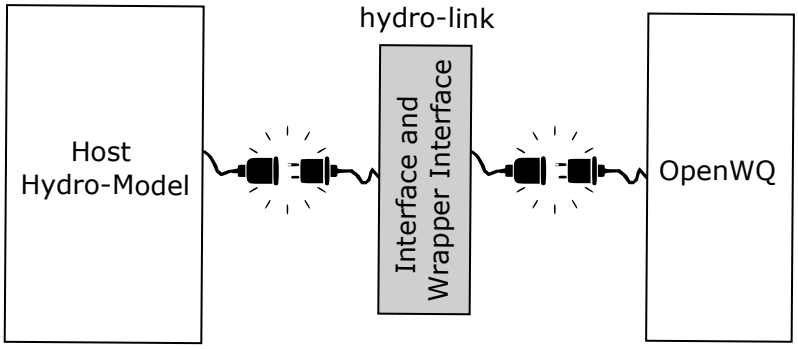


Figure 2: OpenWQ concept as a coupler

The integration of OpenWQ into existing hydro-models is carried out through four coupler functions that are responsible for (1) converting datatypes and data structures between OpenWQ and the "host" hydro-model, (2) passing hydro-fluxes into OpenWQ, and (3) calling OpenWQ's APIs (Fig. 4). These four coupler functions are invoked through interface routines contained within a C++ file with the default name "*OpenWQ_hydrolink.cpp*". This is a one-way communication from the host hydro-model to OpenWQ, so no information is returned to the host model. The interface functions are generic and serve as templates (coupling recipes) that have been optimized to streamline the coupling procedure. The adjustments needed in the interface functions are to ensure that the data types and structure

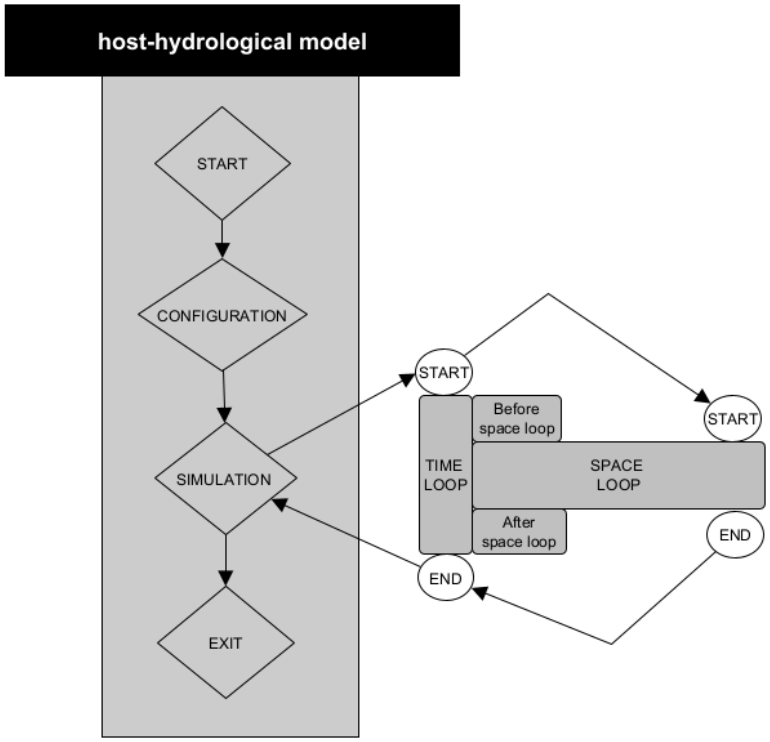


Figure 3: General model structure commonly used in dynamic hydrological, hydrodynamic, and groundwater model

used in the hydro-model are properly translated into OpenWQ's own data structure and conventions so that OpenWQ's APIs can perform adequately.

The following general steps can be performed to couple OpenWQ to a hydro-model: (STEP 1) identify critical elements of hydro-model structure (see Fig. 3), (STEP 2) git-clone OpenWQ, (STEP 3) create the C++ OpenWQ classes and objects to be loaded as modules or libraries in the hydro-model, (STEP 4) identify the appropriate places in the hydro-model to call OpenWQ's coupler functions and APIs (see Fig. 4), (STEP 5) materialize such calls, (STEP 6) adjust OpenWQ's coupler function to adapt to hydro-model data structures (referred to "COUPLER CODE", see below), and (STEP 7) compile the new coupled model. Fig. 5 shows the general structure of each of the four coupler functions, where the location and purpose of the "COUPLER CODE" block are also explained:

1. Coupler function 1 (*openwq::decl*) invokes a series of API calls that handle tasks associated with the initial configuration of the model, initialization of variables, and pre-processing of the input data;
2. Coupler function 2 (*openwq::run_time_start*) invokes a series of API calls that handle tasks required at the start of each time step;
3. Coupler function 3 (*openwq::run_space*) contains a series of API calls that handle tasks related to the spatial domain of the model; and
4. Coupler function 4 (*openwq::run_time_end*) contains a series of API calls that handle tasks required at the end of each time step.

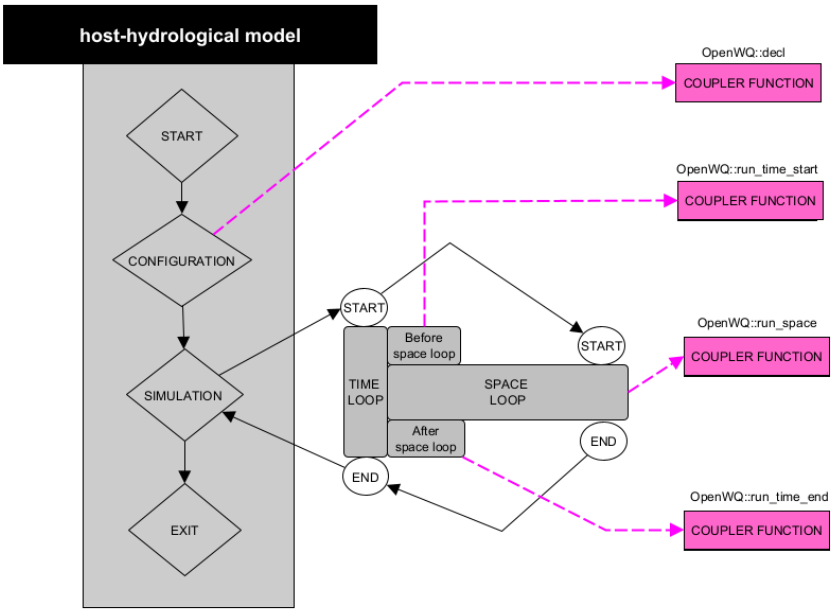
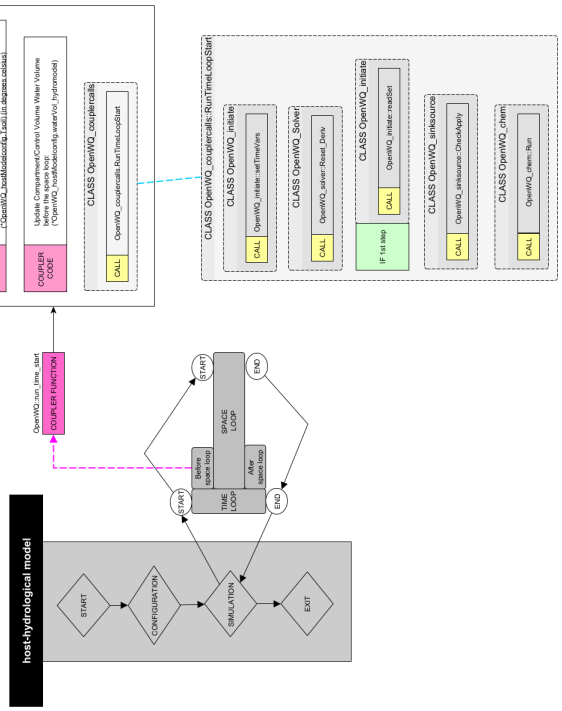


Figure 4: General placement of the calls for OpenWQ's coupler functions

3.2.3 General architecture

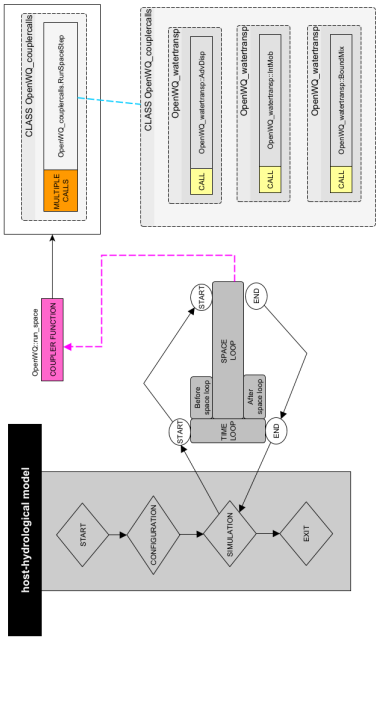
OpenWQ has been created in a way that separates the physics and biochemistry calculations from the numerical implementation. Such an approach was implemented in the SUMMA model (Clark et al., 2015a,b), and we adopted it here to the development of OpenWQ's core structure to improve scalability. This approach addresses a major problem with many hydrological, hydrodynamic, and water quality models where the specification of the model equations is intertwined with their numerical solution (Clark & Kavetski, 2010). This specification complicates the selection and assessment of different model representations (hypotheses) and makes introducing and evaluating alternative numerical methods challenging. As such, the state variables of OpenWQ are only updated inside OpenWQ's own numerical solver based on rates of chemical mass changes (time and space derivatives) caused by different physical and biogeochemical processes. These processes and "rates of change" are computed separately, in process-specific routines, and then passed into the numerical solver for a controlled and contained update of OpenWQ's state variables, which is currently performed as a finite volume problem solved via a simple forward Euler method. The separation of the numerical solver from the physics-biogeochemistry calculations will enable improving the robustness of the numerical implementation in a contained and controlled manner in the future.

The modules in OpenWQ are divided into four groups. Each of these groups is responsible for the calculation of a chemical mass change (time derivative) driven by a particular process (or phenomenon) or group of processes (or phenomena), specifically, (1) initial conditions (*dm_ic*), (2) sinks and sources of chemical load (i.e., chemical mass entering or exiting the model domain) (*dm_ss*), (3) biogeochemical processes (*dm_dt_chemistry*), and (4) physical transport of chemical constituents with water flow (*dm_dt_transport*) (Fig. 6). Each of the modules, in turn, enables a series of modeling options that can be explored, but only one module can be activated for each of the four spatiotemporal-derivative calculations. In other words, the first layer of decision in representing a process pertains to the selection of the key modules that will be responsible for computing each of these four derivatives.

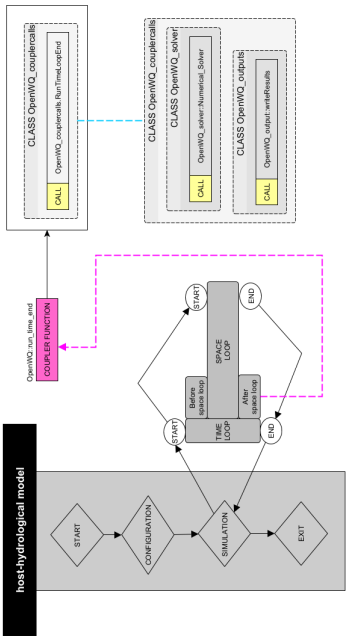


(a) Coupler Function 1: *OpenWQ :: decl*

(b) Coupler Function 2: *OpenWQ :: run_time_start*



(c) Coupler Function 3: *OpenWQ :: run_space*



(d) Coupler Function 4: *OpenWQ :: run_time_end*

Figure 5: Generic coupler functions to integrate OpenWQ into a hydro-model.

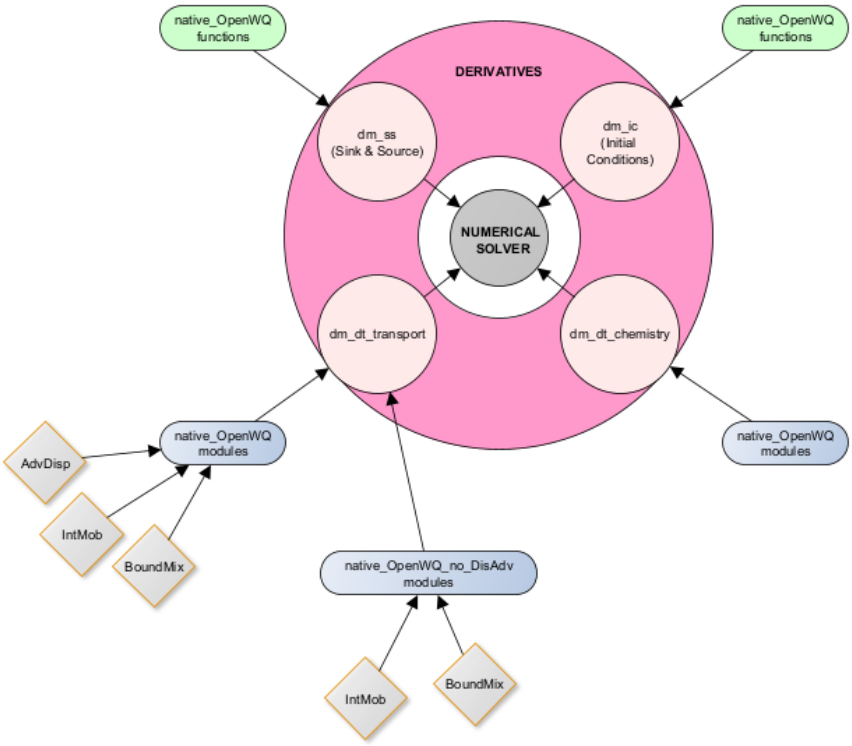


Figure 6: General modular architecture of OpenWQ that separates the different modules and methods available

3.2.4 State variables & Model-adaptive spatial and temporal discretization

OpenWQ was designed to automatically adapt to the hydro-model spatial and temporal discretization structure, which can include Hydrological Response Units (HRUs), 1D–3D spatial distributed meshes, structured or unstructured meshes, and multiple domains or hydrological compartments (e.g., snow, soil, groundwater). OpenWQ's state variables (column 1 in Fig. 7), such as solute mass, are stored in a hierarchical data structure organized with the following nested fields: (1) domain that refers to hydrological compartments (column 2 in Fig. 7), (2) chemical species (column 3 in Fig. 7), and (3) sub-domain that refers to the internal 1D–3D spatial discretization of the domains or hydrological compartments (column 4 in Fig. 7). The hierarchical data structures are built via the Armadillo C++ library for linear algebra and scientific computing (Sanderson & Curtin, 2016, 2018).

The model interface and coupler are responsible for passing information about the spatiotemporal configuration of the host hydro-model into OpenWQ. Such domain configuration options are specific to each host hydro-model and depend on its particular domain discretization scheme and model decisions specified by the users of the hydro-model. Once that information is digested by the couplers and passed on to OpenWQ, a series of API calls dynamically create the corresponding hierarchical data structures and sub-structures that match those in the host model. Each of the data structures stores information about a state variable or a supporting non-state variable (e.g., water fluxes, time derivatives). The state-variable data structures record the spatiotemporal evolution of the mass of the different chemical species tracked in each model domain or hydrological compartment (e.g., snow, soil, lake).

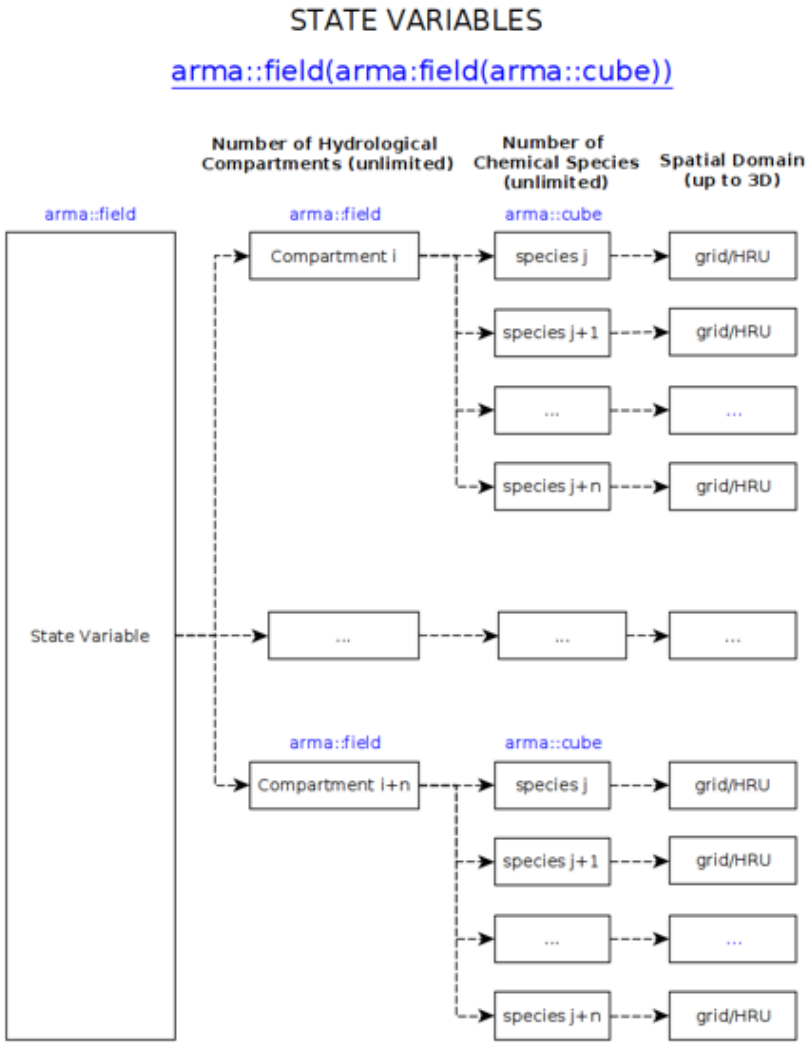


Figure 7: State variables are stored in OpenWQ via dynamic hierarchical data structures

3.3 Model setup and configuration: Inputs and Outputs

The configuration of OpenWQ is provided via JavaScript Object Notation (JSON) files. JSON is an ideal format for large data inputs that have a hierarchically structured relationship. JSON files are composed of key/value pairs, e.g., "RUN_MODE_DEBUG: TRUE". OpenWQ requires four JSON files to run. Each of these files deals with a particular aspect of the model setup and is given one of the following designations (1) "runManagement", (2) "Biogeochemistry", (3) "Configuration", and (4) "Source/Sink".

The "RunManagement" file is the entry point to OpenWQ. It provides the basic instructions, simulation and model output decisions, and the full or relative paths to the other JSON files needed to run a model. Currently, OpenWQ supports HD5F and CSV output file formats. The "Biogeochemistry" file is where the chemical species and biogeochemical cycling frameworks are created and characterized. It uses a standard template structure for the characterization transformations and the chemical species involved. Different "Biogeochemistry" files can be prepared to simulate different pollution problems, which can be readily loaded simultaneously (or used separately or swapped as needed) into a simulation

via the "RunManagement" file. The "Biogeochemistry" file is also where the contaminant species available for the simulation are defined and given a "tag" that can be invoked in other input files. All cycling frameworks and their respective transformations are also provided with a unique "tag" that can be used to flexibly set up the model as desired. These cycling framework tags are particularly important in the "Configuration" file to assign the desired transformations to each domain or hydrological compartment of the model (e.g., snow, soil, groundwater). The "Configuration" file also includes information about the initial conditions. A "Sink/Source" file provides information about the mass exchange with regions outside the model domain. This information can include external chemical mass loading into (source) or loss from (sink), the model domain that can be associated with particular external water fluxes, as in the case of precipitation, or mass inputs that can be independent of the hydrological cycle. Typical examples of such mass inputs can include fertilizer application, which is a relatively localized source, or atmospheric deposition, which is a relatively distributed source. There can be multiple "Sink/Source" files in a given OpenWQ configuration.

3.4 Native Modules and Process representation

OpenWQ contains four main groups of modules, each producing rates of chemical mass change (time derivatives) associated with specific processes, phenomena, and model aspects, which are passed into the numerical solver for updating the state-variables throughout the simulation (see Fig. 6). The modules deal with (1) initial conditions, (2) sinks and sources, (3) transport with water, and (4) biogeochemistry.

The first two groups of modules deal with initial conditions and sinks and external sources with a focus on translating user inputs into the simulation. For example, the sink and source module applies chemical load as prescribed by the users that can include (1) continuous load arising from sources like atmospheric deposition or the outlet of a wastewater treatment plant or (2) episodic/instantaneous loads arising from sources such as fertilizer and chemical spills. The remaining two groups of modules focus respectively on (1) the physical transport of chemical constituents as water moves through the system and (2) biogeochemical transformations.

3.4.1 Transport Module

Currently, there are two options available for computing the physical transport of dissolved solutes and fully suspended sediments. The first option accounts for both advection and dispersion and solves the hyperbolic-parabolic advection-diffusion partial differential equation (PDE) in up to 3 dimensions depending on the host model spatial discretization scheme (Equation 1); the second option only accounts for advection (Equation 2). In both cases, the PDEs are solved inside OpenWQ's numerical solver as described in Section 3.2.3 using the specific hydro-fluxes and corresponding source and recipient domain cells prescribed by the host models. This means that for each water flux computed by the host-model (at each grid cell and timestep), OpenWQ calculates the corresponding solute mass transported, which will be intimately linked to the internal spatio-temporal water-flux representations of the host-model (i.e., if the host model represents snow as a 1 vertical layer domain (lumped), OpenWQ will compute solute concentrations at that spatial resolution; if the host model runs at daily time steps, OpenWQ will calculate solute concentrations at that temporal resolution). The physical transport can be between computation elements (e.g., HRUs, grid cells) and across domains or hydrological compartments (e.g., runoff, soil saturated, soil unsaturated, canopy), as prescribed by the host model.

$$\frac{\partial(Vc_s)}{\partial t} + \nabla \cdot (V\vec{u}c_s) = \nabla \cdot (V\mathbf{E} \cdot \nabla c_s) + S, \tag{1}$$

$$\frac{\partial(Vc_s)}{\partial t} + \nabla \cdot (V\vec{v}c_s) = S, \tag{2}$$

where V is the volume of the computational element/cell [L^3], c_s is the concentration of a given dissolved substance [ML^{-3}]; $\vec{u} = (u_x, u_y, u_z)$ is the velocity in the x, y, and z directions [LT^{-1}]; $\mathbf{E}$ is the (diagonal) diffusivity tensor [L^2T^{-1}]; and S is a source term [MT^{-1}] that is linked to the chemistry modules. The $\mathbf{E}$ term accounts for the combined effect of different mixing phenomena:

$$\mathbf{E} = \mathbf{E}^{turb} + \mathbf{E}^{tsd} + \mathbf{E}^{sgt} + \mathbf{E}^{d}, \tag{3}$$

where $\mathbf{E}^{turb}$ is the turbulent or eddy diffusivity, which is a complex phenomenon with multi-fractal behaviour dominated by friction forces, $\mathbf{E}^{tsd}$ is the so-called Taylor shear dispersion that arises from the unresolved vertical variation of the horizontal flow, $\mathbf{E}^{sgt}$ accounts for the sub-grid eddy viscosity and diffusivity arising from unresolved mixing occurring at sub-grid scale, and $\mathbf{E}^{d}$ is the background molecular diffusion resulting from the probabilistic Brownian motion concept occurring at particle scales. $\mathbf{E}$ is approximated to the dominant turbulent dispersion based on the eddy viscosity concept: $\mathbf{E}^{turb} = \boldsymbol{\nu}_t/\sigma$, where σ is the Prandtl–Schmidt number. Turbulent viscosity ($\boldsymbol{\nu}_t$) depends on shear velocity ($\vec{u}^*$) and on a turbulent length scale ($l_t = 0.07h$). In this model, this relationship is approximated by an algebraic expression: $\boldsymbol{\nu}_t \sim k\vec{u}^* l_t$, where k is a user-defined scaling factor to account for sub-grid scale eddies (Costa et al., 2016).

3.4.2 Biogeochemistry module

Biogeochemical cycling frameworks are characterized in the appropriate biogeochemistry JSON file. Each cycling framework is given a "tag" that is used to load it into simulations. The characterization of each of these cycling frameworks involves the definition of the associated biogeochemical transformations and the chemical species involved. In turn, the characterization of each of these transformations requires information about the chemical species consumed and produced and the reaction kinetics. It allows for the creation of reaction parameters that can be loaded into the kinetics solver, which relies on the comprehensive C++ Mathematical Expression Toolkit Library (*ExprTk*) developed by Arash Partow (1999–2020) (Partow, 1999). The implementation of *ExprTk* in OpenWQ is simple to use and provides an efficient run-time mathematical expression parser and evaluation engine. *ExprTk* supports numerous forms of functional, logical, and vector processing semantics and is easily extendible. The equations can be written with (1) multiple chemical species, (2) user-defined parameters, and (3) built-in hydro-model dependencies. These model dependencies are tailored to each hydro-model, but they usually include variables like soil moisture and air and soil temperature.

The biogeochemistry JSON file can be prepared manually following the appropriate OpenWQ JSON structure (i.e., key-value pairs). Alternatively, the cycling frameworks can be characterized via a diagram drawn using a *GraphML* editor. *GraphML* is an XML-based file format for graphs, and there are several free *GraphML* editors, such as *yEd*. After the diagrams have been drawn in *GraphML* format, they can be converted into OpenWQ's JSON format using a Python script available on OpenWQ's GitHub repository and copied locally during the coupling step defined in Section 3.2.1. This graphical option has been developed to enable collaboration and co-creation of water quality models through a more visual interaction, which can be particularly helpful for water quality modeling activities that often involve interaction between different disciplines that may be less familiar with JSON files.

The expressions used to represent the different reaction kinetics in the model via the biogeochemistry JSON files can take many forms and are deployed in OpenWQ via the *ExprTk* integration. However, these expressions often take the form of sequences of reaction networks involving single or multiple chemical species governed by first-, second-, or third-order kinetics (Eq. 4, Eq. 5, Eq. 6), respectively:

$$\frac{dc_A}{dt} = -k\lambda c_A, \tag{4}$$

$$\frac{dc_A}{dt} = -k\lambda c_A^2, \tag{5}$$

$$\frac{dc_A}{dt} = -k\lambda c_A^2 c_B, \tag{6}$$

where c_A and c_B [ML^{-3}] are the concentrations of chemical species A and B, parameter/variable λ represents weather/hydrological dependencies (such as soil moisture and temperature), and k is the reaction rate [$ML^{-3}T^{-1}$]. The reaction rate k can be provided as the reaction rate using standard maximum at a reference temperature (often 20°C) or using expressions that can include relationships with the hydrological/weather dependent variables/parameters.

3.5 Eco-modelling lab and Cross-model deplyoment: Benefits and Innovation

OpenWQ's concept as a coupler and customizable biogeochemistry modeling framework allows for cross-model deployment of eco-lab, co-creation modeling capabilities, providing a long-awaited transformative direction for innovation in water quality modeling that tackles the inherent challenges of being at the intersection between several scientific fields, including biogeochemistry, soil science, hydrology, hydrodynamics, and hydrogeology. This deployment is an essential step to enhancing collaborative efforts and streamlining knowledge/innovation transfer between the different disciplines involved, ultimately benefiting the entire environmental and ecological research and management community. The ability to deploy OpenWQ across different hydro-models provides the following transformational changes in the current paradigm of water quality modeling.

1. Optimization of investment because focus is given to a single biogeochemical tool, thus reducing code maintenance,
2. Maximization of benefit because it allows (a) any existing hydro-tool to extend its capabilities to water quality and ecological studies, (b) progress achieved by the biogeochemistry and soil-science research community to be transferred into OpenWQ and automatically benefits multiple hydro-models via update installs, (c) OpenWQ's input files to be transferred across hydro-models, providing experience and knowledge transfer between environmental projects and research communities, and
3. Reproducibility is reinforced because OpenWQ's input files are transferable across hydro-models, allowing for more rigorous cross-model comparisons.

Once different biogeochemical modeling hypotheses have been identified and set up for testing in one hydro-model-OpenWQ coupled model, they can be easily transferred for simulation in any other hydro-amodel that has been coupled to OpenWQ (Fig. 8).

4 Discussion

4.1 Unifying Different Approaches to Water Quality Modelling

OpenWQ provides a unifying modeling framework to deploy different approaches to physiochemically based soil and water quality modeling across existing hydrological, hydro-

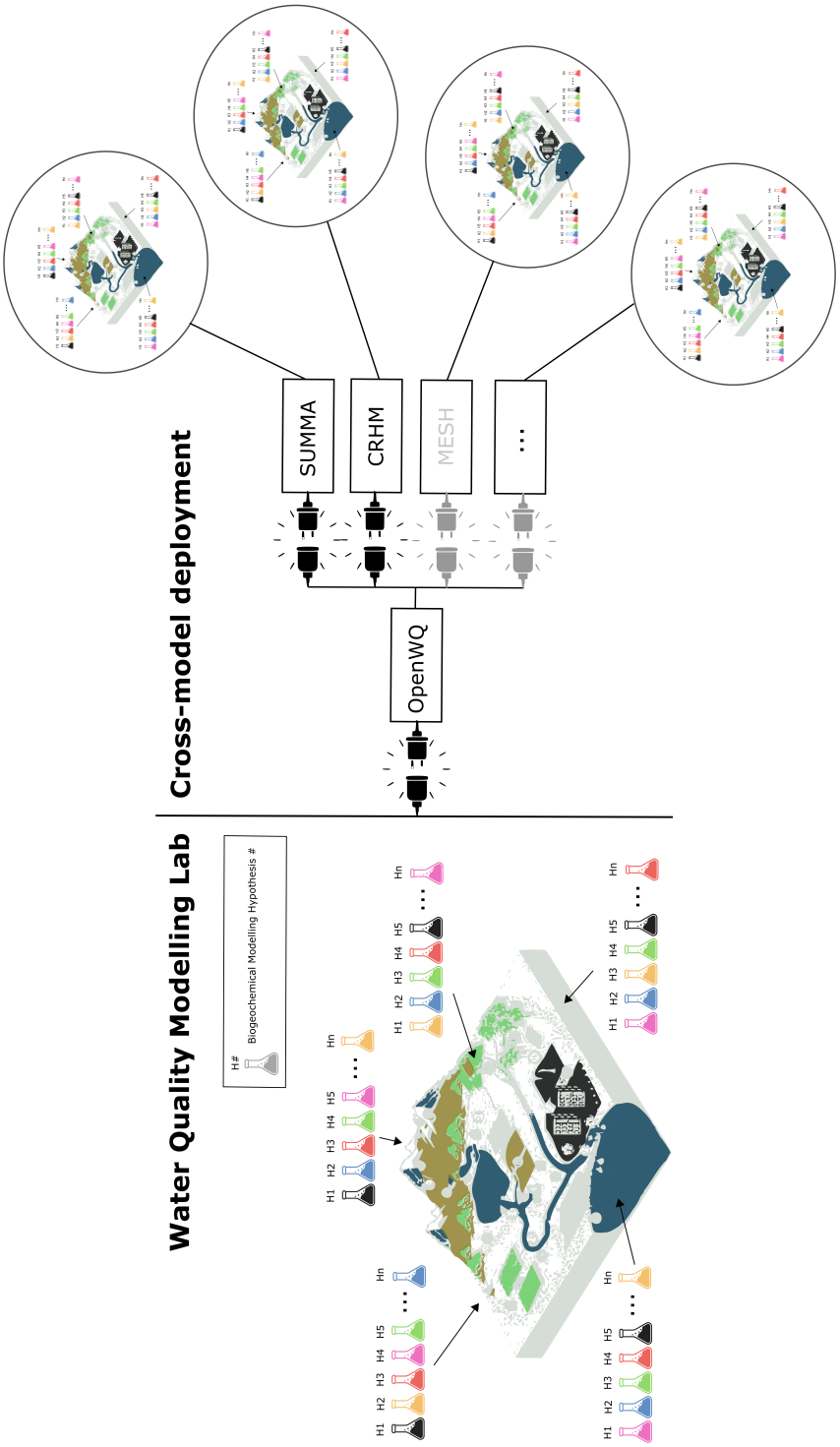


Figure 8: Cross-model deployment of biogeochemistry modelling setup, hypothesis and scenários

dynamic, and hydrogeological models. Here, we emphasize how OpenWQ can help unify different modeling approaches to simulate different landscapes and pollution problems.

1. *Sub-catchment versus catchment-scale modeling.* OpenWQ provides a unique opportunity to harmonize spatial scales in water quality modeling. Water quality models usually belong to one of two scales regarding the spatial domain, (a) sub-catchment and (b) catchment. Sub-catchment water quality models, including river-reach models such as WASP (Wool et al., 2020; Di Toro et al., 1983) and QUAL2E (Brown & Barnwell, 1987), lake models such as MyLake (Saloranta & Andersen, 2007) and Delft3D (Lesser et al., 2004), and aquifer models such as MODFLOW-MT3D (Harbaugh, 2005; Bedekar et al., 2016) and FEFLOW (Trefry & Muffels, 2007), require the characterization of both horizontal and vertical boundary conditions because they do not represent entire closed systems such a river basin. Instead, they focus on subregions within such closed systems. On the other hand, catchment models such as INCA (Jackson-Blake et al., 2016; A. J. Wade et al., 2002; Whitehead et al., 1998) and SWAT (Arnold et al., 2012) only require vertical boundary conditions (e.g., precipitation) because they focus on the larger closed system that contains the entire basin area. Accordingly, these two scales of models tend to focus on different pollution problems. Sub-catchment models tend to focus more on point sources, such as wastewater discharge, and catchment models often look at diffuse pollution (e.g., agriculture nutrients and fertilizer use). These differences also result in different chemical species and biogeochemical cycles of focus, e.g., river models often address problems related to low Dissolved Oxygen (DO) levels caused by biomass decomposition and BOD (Biological Oxygen Demand) arising from wastewater discharges and other point sources discharging directly in river and lakes.
   OpenWQ enables multi-scale chemistry simulations that can be integrated into hydro-models of sub-catchment and catchment scales; thus, it can help bridge the gap between these two approaches, as well as allow for better integration of cross-dependent biogeochemical cycles. For example, whereas the DO and nitrogen cycles are often simulated in river-reach models, catchment models tend to focus more on the nitrogen cycle and assume that there is an unlimited amount of DO available. Although this assumption could be considered valid in many cases for surface runoff, it is certainly erroneous for water bodies such as rivers, lakes, and groundwater.

2. *Simple versus detailed biogeochemistry representation.* Water quality models have been developed with varying degrees of detail in the representation of biogeochemical processes (Costa, Baulch, et al., 2020a). For example, while INCA and HYPE (Lindström et al., 2010) provide simpler biogeochemical modeling approaches for the nitrogen and phosphorous cycles, requiring a smaller number of reaction-kinetic parameters to calibrate them may be more suitable for data-scarce regions; more complex biogeochemical models like HSPF ("Hydrological Simulation Program–Fortran, User's manual for version 11: U.S. Environmental Protection Agency", 1997) arguably provide higher model fidelity but may only be applicable in data-rich environments. However, when selecting a model for a particular region, modelers often have to make compromises with process representation because some models have more detailed physics-based coupled water-energy balance computations such as AnnAGNPS (Bosch et al., 1998) but may offer more limited biogeochemical capabilities (Costa, Baulch, et al., 2020a). OpenWQ enables addressing this issue through its flexible eco-modelling lab, which allows testing different biogeochemical conceptual models and modelling hypotheses, from simple biogeochemical cycles and transformations to more complex and intertwined reaction networks involving dozens or hundreds of chemical species.

3. *Unconnected versus interconnected biogeochemical cycling representation.* Most catchment models simulate biogeochemical cycles in isolation. For example, popular

models such as SWAT and HYPE simulate the phosphorous and nitrogen cycles without integrating them with the DO-BOD cycle, despite them being strongly interrelated because DO is used in both cycles. In the nitrogen cycle, DO is used in nitrification, where ammonia is oxidated into nitrite (NO_3) and nitrate (NO_2). In the DO-BOD cycle, DO is used in the biological, aerobic decomposition of organic matter. OpenWQ aims to allow for a biogeochemistry representation that is less compartmentalized, making interactions between cycling frameworks more fluid and closer to reality (i.e., model fidelity). In reality, DO dynamics affect the cycling of many chemical constituents, from nutrients to heavy metals. OpenWQ's eco-modelling lab addresses this issue through its open reaction-network solver, which provides the flexibility to deploy any number of biogeochemical cycling representations of any number of chemical species simultaneously, which can be connected or unconnected.

4. *Background chemical transport driver.* Typically, hydro-models are developed and maintained by research communities with specific research motivations driven by a regional context. These modeling tools often become highly specialized for particular environments and applications, and modelers commonly find it hard to find a water quality modeling tool that integrates such regionally important hydro-transport modeling capabilities. There is also in-house expertise in such communities and research groups that is passed on over the years between elements of the research community, which can make them reluctant to switch to other modeling tools because it may involve steep learning curves. OpenWQ aims to address this issue by enabling extension to water quality modeling capabilities directly on those models so that modelers can continue using the hydro-modeling tool that they consider more suitable to the environments on which they focus.

4.2 Advancing Current Model Development Paradigms

Our focus on providing a flexible biogeochemical model representation framework that is deployable on any process-based hydro-models and offers a unique opportunity to transform the way environmental studies involving water quality modeling are carried out. It also provides a concrete way for streamlining collaboration, co-creation, and knowledge and innovation transfer between the hydrological and biogeochemistry/soil science communities. Such streamlining is materialized by OpenWQ's coupler interface, which links developments/progress in the hydro-modeling tool (hydro-modeling community) with developments/progress in OpenwQ (biogeochemistry communities). Another major issue with the current water quality modeling paradigm is that most models offer little flexibility to adapt the representation of biogeochemistry processes to (1) local/region context, (2) data availability, and (3) application(s) of interest, resulting in the need to make difficult choices when selecting a modeling tool because compromises between the representation of hydrofluxes and biogeochemistry are often required (Costa, Baulch, et al., 2020a).

4.3 Limitations

In OpenWQ, the physical transport of chemical constituents relies on the quality of the water fluxes passed by the host hydro-model. Although this may be an advantage because it allows the user to choose the hydro-model that best suits the modeling needs, it can become an issue if the calculated fluxes in the hydro-model are not accurate. Modelers should select the hydro-model coupled to OpenWQ that is the most suitable for the application at hand or pursue the coupling of OpenWQ to another hydro-model.

The native biogeochemistry module of OpenWQ that provides water quality-lab capabilities assumes that biogeochemical cycling can be represented via a series of sequential and parallel reaction networks. Although this representation is true for most chemical constituents, pollution problems, and environmental studies, some biogeochemistry may involve

formulations that rely on variable dependencies that may not be available in a particular coupled model. For example, simulating microbiological pollution in lakes, rivers, and beaches, such as contamination with fecal coliforms originating from leaking septic tanks and wastewater discharges, requires the simulation of die-off rates due to exposure to solar radiation. In the case of hydrodynamic models, such information may not be available. In the case of hydrological models, which typically deal with such data, that variable dependency may not have been passed into OpenWQ during the development of the coupler interface, so updates to the coupler may be needed. OpenWQ's coupler functions have been designed to make the addition of new dependency variables straightforward.

The portability of OpenWQ is materialized through a series of coupler functions and wrapper interface functions that enable its coupling to hydro-models written in C++ or Fortran. However, the use of a more standardized model interface framework, such as the Basic Model Interface (BMI Peckham et al., 2013; E. W. Hutton et al., 2020), could help streamline further the coupling process through the use of standard control and query functions. This could make that model both easier to learn and easier to couple with other software elements. BMI, for example, currently supports five languages: C, C++, Fortran, Java, and Python.

5 Conclusions

This paper describes a unified framework for enabling multi-biogeochemical modeling capabilities in existing hydro-models. The framework, which we call Open Water Quality (OpenWQ), was designed with both coupler and water quality-lab modeling capabilities to enable (1) flexible co-creation and testing of biogeochemistry modeling representations, (2) systematic implementation and evaluation of alternative modeling approaches for process representation, and (3) identification of specific causes of model weaknesses.

This work arises from the recognition that the hydrological transport of contaminants in the environment strongly affects their concentrations in aquatic ecosystems, but the interaction between the biogeochemical and hydrological-hydrodynamic-hydrogeological communities lacks the appropriate mechanisms for an efficient transfer of knowledge and innovation between the two communities. Many hydro-models are developed and maintained by specific research communities addressing particular regions and climate zones. The investment in such tools creates invaluable in-house expertise, making them less likely to switch to other modeling tools and hindering their expansion to water quality and ecological studies. OpenWQ aims to address this problem and provide a concrete direction for innovation in connecting communities through an optimized plug-in-like water quality model that can be coupled to existing hydro-modeling tools, extending the value of these tools for water quality and ecological studies in their region of focus. Different process representations and different spatial configurations can be integrated into the structural model core, which enables users to decompose the modeling problem into the individual decisions made as part of model development and evaluate different "fine-grained" model development decisions in a systematic and controlled way.

OpenWQ can provide the necessary model flexibility to progress toward answering the following fundamental modeling questions and challenges: (1) which hydro-biogeochemical processes should be represented explicitly in different environmental settings, and, correspondingly, which processes can be ignored or greatly simplified; (2) what modeling approaches should be used to represent the dominant biogeochemistry at the spatial scale of the model discretization; (3) how should heterogeneity in pollution and biogeochemical processes be represented across spatial scales, including the complexity of transport across landscapes; and (4) how can we provide insights into the sources of model uncertainty. The companion paper describes the integration of OpenWQ into two hydro-models, SUMMA and CRHM, describing how coupling interfaces between the two models have not only enabled water quality simulation capacities in these host hydro-models but, even more importantly,

established a direct and permanent link for the transfer of innovation between the associated modeling communities, promoting cooperation and co-creation. Example applications of pollution studies enabled by the coupling of the tools are also provided to begin to address some of these fundamental modeling challenges.

Acknowledgments

This work was supported by the NSERC-Alliance Grant ALLRP 571910-22. We acknowledge support from Environment and Climate Change Canada, the Global Water Futures programme, and the Natural Sciences and Engineering Research Council of Canada under Discovery Grant RGPN-2020-04467 (RJS).

References

Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., & Strömqvist, J. (2012). Water and nutrient simulations using the HYPE model for Sweden vs. the Baltic Sea basin – influence of input-data quality and scale. *Hydrology Research*, *43*(4), 315-329. doi: 10.2166/nh.2012.010

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., ... Jha, M. K. (2012). SWAT: Model Use, Calibration, and Validation. *Trans. ASABE*, *55*(4), 1491–1508.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment Part I: Model Development. *JAWRA Journal of the American Water Resources Association*, *34*(1), 73-89. doi: https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water Resources Research*, *23*(8), 1393–1442. doi: 10.1029/WR023i008p01393

Bedekar, V., Morway, E. D., Langevin, C. D., & Tonkin, M. J. (2016). *MT3D-USGS version 1: A U.S. Geological Survey release of MT3DMS updated with new and expanded transport capabilities for use with MODFLOW* (Tech. Rep.). Reston, VA. doi: 10.3133/tm6A53

Bosch, D., Theurer, F., Bingner, R., & Felton, G. (1998). Evaluation of the AnnAGNPS water quality model. *Agricultural Non-Point Source Water Quality Models: Their Use and Application; John, EP, Daniel, LT, Rodney, LH, Eds*, 45–54.

Brown, L., & Barnwell, T. (1987). The enhanced stream water quality models qual2e and qual2e-uncas: documentation and user manual. *Environmental Protection Agency*.

Clark, M. P., & Kavetski, D. (2010). Ancient numerical daemons of conceptual hydrological modeling: 1. fidelity and efficiency of time stepping schemes. *Water Resources Research*, *46*(10). doi: https://doi.org/10.1029/2009WR008894

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9). doi: https://doi.org/10.1029/2010WR009827

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Rasmussen, R. M. (2015a). A unified approach for process-based hydrologic modeling: 1. modeling concept. *Water Resources Research*, *51*(4), 2498-2514. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017198` doi: https://doi.org/10.1002/2015WR017198

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Marks, D. G. (2015b). A unified approach for process-based hydrologic modeling: 2. model implementation and case studies. *Water Resources Research*, *51*(4), 2515-2542. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017200` doi: https://doi.org/10.1002/2015WR017200

Costa, D., Baulch, H., Elliott, J., Pomeroy, J., & Wheater, H. (2020a). Modelling nutrient dynamics in cold agricultural catchments: A review. *Environmental Modelling & Software*,

124, 104586. doi: https://doi.org/10.1016/j.envsoft.2019.104586

Costa, D., Baulch, H., Elliott, J., Pomeroy, J., & Wheater, H. (2020b). Modelling nutrient dynamics in cold agricultural catchments: A review. *Environmental Modelling & Software*, *124*, 104586. doi: https://doi.org/10.1016/j.envsoft.2019.104586

Costa, D., Burlando, P., & Liong, S.-Y. (2016). Coupling spatially distributed river and groundwater transport models to investigate contaminant dynamics at river corridor scales. *Environmental Modelling & Software*, *86*, 91-110. doi: https://doi.org/10.1016/j.envsoft.2016.09.009

Costa, D., Pomeroy, J., & Wheater, H. (2018). A numerical model for the simulation of snowpack solute dynamics to capture runoff ionic pulses during snowmelt: The pulse model. *Advances in Water Resources*, *122*, 37-48. doi: https://doi.org/10.1016/j.advwatres.2018.09.008

Costa, D., & Pomeroy, J. W. (2019). Preferential meltwater flowpaths as a driver of preferential elution of chemicals from melting snowpacks. *Science of The Total Environment*, *662*, 110–120. doi: 10.1016/J.SCITOTENV.2019.01.091

Costa, D., Pomeroy, J. W., Brown, T., Baulch, H., Elliott, J., & Macrae, M. (2021). Advances in the simulation of nutrient dynamics in cold climate agricultural basins: Developing new nitrogen and phosphorus modules for the Cold Regions Hydrological Modelling Platform. *Journal of Hydrology*, *603*, 126901. doi: https://doi.org/10.1016/j.jhydrol.2021.126901

Costa, D., Roste, J., Pomeroy, J., Baulch, H., Elliott, J., Wheater, H., & Westbrook, C. (2017). A modelling framework to simulate field-scale nitrate response and transport during snowmelt: The wintra model. *Hydrological Processes*, *31*(24), 4250-4268. doi: https://doi.org/10.1002/hyp.11346

Costa, D., Shook, K., Spence, C., Elliott, J., Baulch, H., Wilson, H., & Pomeroy, J. W. (2020). Predicting variable contributing areas, hydrological connectivity, and solute transport pathways for a canadian prairie basin. *Water Resources Research*, *56*(12), e2020WR027984. doi: https://doi.org/10.1029/2020WR027984

Davies, T. D., Brimblecombe, P., Tranter, M., Tsiouris, S., Vincent, C. E., Abrahams, P., & Blackwood, I. L. (1987). *The Removal of Soluble Ions from Melting Snowpacks* (H. G. Jones & W. J. Orville-Thomas, Eds.). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-009-3947-9_20

Di Toro, D. M., Fitzpatrick, J. J., Thomann, R. V., & Hydroscience, I. (1983). *Documentation For Water Quality Analysis Simulation Program (WASP) And Model Verification Program (MVP)* (Tech. Rep.).

Frank, E. G., & Schlenker, W. (2016). Balancing economic and ecological goals. *Science*, *353*(6300), 651-652. doi: 10.1126/science.aaf9697

Fu, B., Merritt, W. S., Croke, B. F., Weber, T. R., & Jakeman, A. J. (2019). A review of catchment-scale water quality and erosion models and a synthesis of future prospects. *Environmental Modelling & Software*, *114*, 75–97. doi: 10.1016/J.ENVSOFT.2018.12.008

Harbaugh, A. W. (2005). *MODFLOW-2005, the US Geological Survey modular groundwater model: the ground-water flow process* (Vol. 6). US Department of the Interior, US Geological Survey Reston, VA, USA.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016). Most computational hydrology is not reproducible, so is it really science? *Water Resources Research*, *52*(10), 7548–7555. doi: https://doi.org/10.1002/2016WR019285

Hutton, E. W., Piper, M. D., & Tucker, G. E. (2020). The basic model interface 2.0: A standard interface for coupling numerical models in the geosciences. *Journal of Open Source Software*, *5*(51), 2317. doi: 10.21105/joss.02317

Hydrological Simulation Program–Fortran, User's manual for version 11: U.S. Environmental Protection Agency. (1997). *National Exposure Research Laboratory, Athens, Ga., EPA/600/R-97/080, 755 p.*

Jackson-Blake, L. A., Wade, A. J., Futter, M. N., Butterfield, D., Couture, R.-M., Cox,

B. A., ... Whitehead, P. G. (2016). The INtegrated CAtchment model of phosphorus dynamics (INCA-P): Description and demonstration of new model structure and equations. *Environmental Modelling & Software*, *83*, 356–386. doi: 10.1016/J.ENVSOFT.2016.05.022

Jones, H. G. (1999). The ecology of snow-covered systems: a brief overview of nutrient cycling and life in the cold. *Hydrological Processes*, *13*(14-15), 2135–2147. doi: 10.1002/(SICI)1099-1085(199910)13:14/15⟨2135::AID-HYP862⟩3.0.CO;2-Y

Keith, D. A., Ferrer-Paris, J., Nicholson, E., Bishop, M. J., Polidoro, B. A., Ramirez-Llodra, E., ... Kingsford, R. T. (2022). A function-based typology for earth's ecosystems. *Nature*, *610*(7932), 513–518. doi: 10.1038/s41586-022-05318-4

Kitware. (2022). *CMake: Reference documentation.*

Lesser, G., Roelvink, J., van Kester, J., & Stelling, G. (2004). Development and validation of a three-dimensional morphological model. *Coastal Engineering*, *51*(8), 883-915. (Coastal Morphodynamic Modeling) doi: https://doi.org/10.1016/j.coastaleng.2004.07.014

Lilbaek, G., & Pomeroy, J. W. (2007). Modelling enhanced infiltration of snowmelt ions into frozen soil. *Hydrological Processes*, *21*(19), 2641–2649. doi: 10.1002/hyp.6788

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., & Arheimer, B. (2010). Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology Research*, *41*(3-4), 295. doi: 10.2166/nh.2010.007

Marsh, P., & Pomeroy, J. W. (1999). Spatial and temporal variations in snowmelt runoff chemistry, Northwest Territories, Canada. *Water Resources Research*, *35*(5), 1559–1567. doi: 10.1029/1998WR900109

Mekonnen, B. A. (2016). *Modeling and management of water quantity and quality in cold-climate Prairie Watersheds* (Unpublished doctoral dissertation). University of Saskatchewan.

Moore, D. S., Bingner, R. L., & Theurer, F. D. (2006). ANNAGNPS: Accounting for Snowpack, Snowmelt, and Soil Freeze-Thaw. In *Eighth federal interagency sedimentation conference (8thfisc),* (pp. 475–482). Reno, NV, USA.

Partow, A. (1999). *ExprTk C++ Mathematical Expression Library.*

Peckham, S. D., Hutton, E. W., & Norris, B. (2013). A component-based approach to integrated modeling in the geosciences: The design of csdms. *Computers & Geosciences*, *53*, 3–12.

Pomeroy, J., Brown, T., Fang, X., Shook, K., Pradhananga, D., Armstrong, R., ... Lopez Moreno, J. (2022). The cold regions hydrological modelling platform for hydrological diagnosis and prediction based on process understanding. *Journal of Hydrology*, *615*, 128711. doi: https://doi.org/10.1016/j.jhydrol.2022.128711

Pomeroy, J., Davies, T., Jones, H., & Marsh, P. (1999). Transformations of snow chemistry in the boreal forest: accumulation and volatilization. *Hydrological Processes*, *13*(1415), 2257–2273.

Pomeroy, J. W., Davies, T. D., & Tranter, M. (1991). The Impact of Blowing Snow on Snow Chemistry. In *Seasonal snowpacks: Processes of compositional change* (pp. 71–113). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-75112-7_4

Pomeroy, J. W., Gray, D. M., Brown, T., Hedstrom, N. R., Quinton, W. L., Granger, R. J., & Carey, S. K. (2007). The cold regions hydrological model: A platform for basing process representation and model structure on physical evidence. *Hydrological Processes*, *21*(19), 2650–2667. doi: {10.1002/hyp.6787}

Refsgaard, J., & Storm, B. (1995). Mike she.[in:] singh vp (ed.), computer models of watershed hydrology. *Water Resources Publication, Colorado*, 809–847.

Richmond, B. (2003). *STELLA: An Introduction to Systems Thinking.*

Saloranta, T. M., & Andersen, T. (2007). Mylake—a multi-year lake simulation model code suitable for uncertainty and sensitivity analysis simulations. *Ecological Modelling*, *207*(1), 45-60. (Uncertainty in Ecological Models) doi: https://doi.org/10.1016/j.ecolmodel.2007.03.018

Sanderson, C., & Curtin, R. (2016). Armadillo: a template-based c++ library for linear algebra. *Journal of Open Source Software*, *1*(2), 26. doi: 10.21105/joss.00026

Sanderson, C., & Curtin, R. (2018). A user-friendly hybrid sparse matrix class in c++. , 422–430. doi: 10.1007/978-3-319-96418-8_50

Shoemaker, L. (1997). Compendium of tools for watershed assessment and tmdl development.

Shook, K., Pomeroy, J. W., Spence, C., & Boychuk, L. (2013). Storage dynamics simulations in prairie wetland hydrology models: evaluation and parameterization. *Hydrological Processes*, *27*(13), 1875–1889. doi: 10.1002/hyp.9867

Trefry, M. G., & Muffels, C. (2007). Feflow: A finite-element ground water flow and transport modeling tool. *Groundwater*, *45*(5), 525-528. doi: https://doi.org/10.1111/j.1745-6584.2007.00358.x

Wade, A., Jackson, B., & Butterfield, D. (2008). Over-parameterised, uncertain 'mathematical marionettes' — How can we best use catchment water quality models? An example of an 80-year catchment-scale nutrient balance. *Science of The Total Environment*, *400*(1-3), 52–74. doi: 10.1016/j.scitotenv.2008.04.030

Wade, A. J., Durand, P., Beaujouan, V., Wessel, W. W., Raat, K. J., Whitehead, P. G., ... Lepisto, A. (2002). A nitrogen model for european catchments: Inca, new model structure and equations. *Hydrology and Earth System Sciences*, *6*(3), 559–582. doi: 10.5194/hess-6-559-2002

Wellen, C., Kamran-Disfani, A. R., & Arhonditsis, G. B. (2015). Evaluation of the current state of distributed watershed nutrient water quality modeling. *Environmental Science and Technology*, *49*(6), 3278–3290. doi: 10.1021/es5049557

Whitehead, P., Wilson, E., Butterfield, D., & Seed, K. (1998). A semi-distributed integrated flow and nitrogen model for multiple source assessment in catchments (inca): Part ii — application to large river basins in south wales and eastern england. *Science of The Total Environment*, *210-211*, 559-583. doi: https://doi.org/10.1016/S0048-9697(98)00038-2

Wool, T., Ambrose, R. B., Martin, J. L., & Comer, A. (2020). Wasp 8: The next generation in the 50-year evolution of usepa's water quality model. *Water*, *12*(5). doi: 10.3390/w12051398

Yang, X., Tetzlaff, D., Soulsby, C., & Borchardt, D. (2022). Hiwaq v1.0: A flexible catchment water quality assessment tool with compatibility for multiple hydrological model structures. *Geoscientific Model Development Discussions*, *2022*, 1–26. doi: 10.5194/gmd-2022-239

A multi-chemistry modelling framework to enable flexible and reproducible water quality simulations in existing hydro-models: 1. The OpenWQ concept and the water quality modelling lab

Diogo Costa[1,2], Kyle Klenk[1], Wouter Knoben[1], Andrew Ireson[1], Raymond J. Spiteri[1], Martyn Clark[1]

[1]Mediterranean Institute for Agriculture, Environment and Development & Global Change and Sustainability Institute, Department of Geosciences, Universidade de Évora, Pólo da Mitra, Ap. 94, 7006-554 Évora, Portugal
[2]University of Saskatchewan, 105 Administration Pl, Saskatoon, SK S7N 5A2, Canada

Abstract

This work advances the incorporation and cross-model deployment of multi-biogeochemistry and ecological simulations in existing process-based hydro-modelling tools. It aims to transform the current practice of water quality modelling as an isolated research effort into a more integrated and collaborative activity between science communities. Our approach, which we call "Open Water Quality" (OpenWQ), enables existing hydrological, hydrodynamic, and groundwater models to extend their capabilities to water quality simulations, which can be set up to examine a variety of water-related pollution problems. OpenWQ's objective is to provide a flexible biogeochemical model representation that can be used to test different modelling hypotheses in a multi-disciplinary co-creative process. In this paper, we introduce the general approach used in OpenWQ. We detail aspects of its architecture that enable its coupling with existing models. This integration enables water quality models to benefit from advances made by hydrologic- and hydrodynamic-focused groups, strengthening collaboration between the hydrological, biogeochemistry, and soil science communities. We also detail innovative aspects of OpenWQ's modules that enable biogeochemistry lab-like capabilities, where modellers can define the pollution problem(s) of interest, the appropriate complexity of the biogeochemistry routines, and test different modelling hypotheses. In a companion paper, we demonstrate how OpenWQ has been coupled to two hydrological models, the "Structure for Unifying Multiple Modelling Alternatives" (SUMMA) and the "Cold Regions Hydrological Model" (CRHM), demonstrating the innovative aspects of OpenWQ, the flexibility of its couplers and internal spatiotemporal data structures, and the versatile eco-modelling lab capabilities that can be used to study different pollution problems.

1 Introduction

Societies are becoming increasingly aware of the importance of harmonizing economic development with thriving aquatic ecosystems (Keith et al., 2022; Frank & Schlenker, 2016). This has led to an increase in the use of ecological and water quality models to study management solutions to pollution problems such as nutrient excess due to the use of fertilizers or microbiological contamination due to the discharge of untreated wastewater in rivers.

Over the past decades, substantial progress has been made in computational hydrology, producing many powerful generic and fit-for-purpose modeling tools. Some of these models have been particularly tailored to specific environments, such as the Cold Regions Hydrological Model (J. Pomeroy et al., 2022; J. W. Pomeroy et al., 2007), the Wetland DEM Ponding Model (Shook et al., 2013), and the FLUXOS-Overland model (Costa, Shook, et al., 2020), that have been developed to deal with the specific hydrological challenges of simulating the Canadian Prairie pothole region due to the lack of well-defined river networks. Unfortunately, the science and modelling progress achieved by such hydrological model development communities rarely transfers to (and translates into) water quality modeling improvements because hydrological and water quality modeling tools are typically developed in isolation, a problem that becomes clearly noticeable when examining the hydrological calculations embedded in popular water quality modeling tools, such as SWAT (Arnold et al., 1998), HYPE (Lindström et al., 2010; Arheimer et al., 2012), and INCA (Whitehead et al., 1998; A. J. Wade et al., 2002; Jackson-Blake et al., 2016), that often rely on simplifications that may be limiting for some regions and applications (Costa, Baulch, et al., 2020a).

Besides the issues with the communication and transfer of knowledge between these two scientific communities, other scientific and technical biogeochemistry-specific challenges remain critical for reliable simulations across climate regions, landscapes, and seasons (Costa, Baulch, et al., 2020a; Fu et al., 2019; Wellen et al., 2015). Some key challenges include (1) the adequate representation of the complexity and heterogeneity of biogeochemical processes and their dependency on weather, soil, and sediment characteristics; (2) the dependency on accurate hydrological drivers to reliably track the vertical and lateral movement of chemical

constituents (being at the end of the climate-hydrology-ecology modeling chain invariably propagates uncertainties into water quality simulations); and (3) the often limited range of applicability of models to specific pollution problems and landscape conditions due to their rigid and heavily parameterized representation of water quality processes.

Underlying all of these issues are foundational questions of scientific reproducibility in computational hydrology (C. Hutton et al., 2016), as well as considerable challenges in defining the appropriate complexity, scale, and scope of water quality models (Costa, Baulch, et al., 2020a; Moore et al., 2006; Mekonnen, 2016; Shoemaker, 1997). This challenge is in part due to uncertainties associated with the prediction of many hydrological and biogeochemical responses at different spatial scales and geographical settings, which are aggravated by the often sparse water quality data available for testing and validation of the models. This raises critical questions for the design, application, and benefit of such modeling tools (Beck, 1987; Moore et al., 2006; A. Wade et al., 2008; Costa, Baulch, et al., 2020a).

Improving process-based hydro-biogeochemical models for terrestrial hydrological simulations requires effective and continuous integration of progress across several research areas, including hydrology, biogeochemistry, and soil science. Models must combine practical aspects related to model application with theoretical scientific insights at various degrees of scientific maturity and geographical applications. Such cross-disciplinary modeling efforts require modelers to make model development decisions based on several considerations common to hydrological models (Clark et al., 2011) that include (1) model fidelity, complexity and practicality, (2) scientific reproducibility and transparency, and (3) data availability.

It has been recognized that more flexible water quality modelling approaches are needed to address these issues (Yang et al., 2022; Costa, Shook, et al., 2020). Some approaches have emerged that address some of the challenges and improve transparency and engagement among hydrologists, biogeochemists, soil scientists, and decision-makers, such as the STELLA (Structural Thinking and Experiential Learning Laboratory with Animation) software (Richmond, 2003) and the MIKE Eco Lab, which is a water quality modelling addition to the MIKE tools (Refsgaard & Storm, 1995). However, such tools are not suitable for integration into modelling platforms developed by various hydrological communities around the world. In the case of STELLA, the tool is standalone and more suitable for applications where simpler batch-reactor assumptions are applicable, such as for small reservoirs and wastewater treatment plants. The MIKE tools from DHI now provide an Eco-Lab module for more flexible water quality and ecological simulations, but this module is a proprietary tool integrated into the MIKE ecosystem of tools, so it does not benefit the wider hydrology and modelling communities, and it is constrained by the chemical constituents and types of biogeochemical processes that are explicitly introduced by the MIKE model developers. More recently, Yang et al. (2022) proposed a new model named HiWaQ for flexible catchment water quality assessments with compatibility for multiple hydrological model structures. This is a valuable contribution toward a more unified hydrological-water quality modelling approach, but the current version is limited to the nitrogen cycle.

2 Current modelling capabilities and needs

Widely used process-based catchment nutrient models such as SWAT, HYPE, and INCA have been at the forefront of innovation in water quality modeling, paving a way for researchers and decision-makers around the world to investigate solutions to a variety of pollution problems, particularly related to nutrient pollution. These models have become increasingly complex and heavily parameterized, but they remain a limited representation of reality because hydro-biogeochemical processes are highly complex in natural environments (A. Wade et al., 2008; Beck, 1987; Costa, Baulch, et al., 2020b). These models typically simulate a series of biogeochemical processes conceptualized to address particular pollutions problems, and processes are represented through a combination of empirical and physico-chemically based methods, often leading to many calibration parameters and thus increasing

the risk of parameter equifinality (Costa, Baulch, et al., 2020b). Sparse and sporadic water quality measurements, which are common problems for most water bodies, limit the further application and development of these water quality models.

The combination of process-representation methods with varying degrees of empiricism is in part due to knowledge gaps in understanding the drivers and controls of hydrological and biogeochemical responses at various spatial scales and across different landscapes and climate zones. In regions where relatively uncommon processes may play an important role in the overall water quality dynamics, the use of these models becomes problematic because there is little flexibility for adjusting conceptual models, adding or removing processes, or testing different modeling hypotheses. For example, research has shown that in some cold regions, processes such as preferential infiltration of hydrochemical into frozen soils (Lilbaek & Pomeroy, 2007), preferential elution of hydrochemicals from melting snowpacks (Davies et al., 1987; Marsh & Pomeroy, 1999; Costa & Pomeroy, 2019), microbial uptake and fixing of nitrogen in melting snowpack (Jones, 1999), and volatilization of nitrogen during snow redistribution and sublimation (J. W. Pomeroy et al., 1991; J. Pomeroy et al., 1999) may affect water quality, but they are not represented in most popular models. This lack of representation may compromise their use in such regions and calls for a more flexible approach to water quality modeling that enables a systematic and controlled approach for the addition and removal of processes as needed to reflect (1) regional and climate characteristics, (2) data limitations, and (3) objectives of the study.

3 The OpenWQ concept

3.1 Overview

OpenWQ is a coupler-modelling framework designed to provide portable and customizable multi-chemistry modelling capabilities to existing hydro-models. The vision is to create a tool that could plug into existing process-based hydrological, hydrodynamic, and groundwater models to extend their capabilities to environmental and ecological studies. The approach allows tailoring chemical-microbiological constituents and biogeochemistry-cycling processes to enable the representation of different pollution problems and landscapes, as well as to compare modelling hypotheses. This framework is a much-needed effort to bring the hydrology and biogeochemistry communities together, optimizing research and investment efforts. The work stems from previous model developments by Environment and Climate Change Canada and the University of Saskatchewan, particularly CRHM-WQ (Cold Regions Hydrological Model - Water Quality) (Costa et al., 2021) that extends the original CRHM platform (hydrology) model to nitrogen and phosphorus simulations (J. W. Pomeroy et al., 2007), the WINTRA framework (Costa et al., 2017), the multiphase multilayer PULSE snow hydrochemistry model (Costa et al., 2018), and the FLUXOS-OVERLAND model for watershed hydrodynamic-water quality simulations suitable for Prairie regions (Costa, Shook, et al., 2020).

OpenWQ aims to address three main challenges with existing water quality models: (1) structural rigidity in the representation of chemical constituents and biogeochemical processes, (2) over-simplification and limitations of hydro-flux calculations, and (3) inadequacy for testing different modelling hypotheses for proper quantification of structural uncertainty. Structural rigidity is perhaps the key factor that hinders the effective use of models across landscapes and in complex, diverse environments (e.g., permafrost, peatlands, variable contributing areas) that require more investigative, open-ended, and interactive simulation approaches. A flexible environment that enables integrating new methods and concepts from complementary disciplines and experts (e.g., limnology, soil science, biogeochemistry) is critical to advance science and promote meaningful and impactful cross-disciplinary collaborations. The static, hard-coded implementation of biogeochemical reaction-network transformations limits their suitability for a wide range of environmental problems. It also hinders multi-disciplinary co-creation efforts because models provide little flexibility for

changing, expanding, and testing different strategies for biogeochemical cycling representation. Finally, the hydro-flux calculations that are embedded in water quality modelling tools are often outdated or limited compared to dedicated, disciplinary hydro-models (e.g., hydrological, hydrodynamic, and groundwater modelling tools). Uncertainty in model structure, process representation, and future scenarios (e.g., climate change) cannot be adequately quantified without flexible and transparent modelling structures.

3.2 General design

This section describes how OpenWQ was designed to address the key challenges in water quality modelling described before. First, focus is given to describe how OpenWQ tackles the need for more robust hydro-flux and associated solute transport calculations through integration within existing hydro-models. This includes details on how OpenWQ can be linked to existing hydro-models considering key practical aspects of version control (Section 3.2.1), portability (Section 3.2.2), code integration (Sections 3.2.3), and state variables and spatiotemporal discretization (Section 3.2.4) . Second, focus is given to describe how OpenWQ tackles the need for more flexible representation of physical and biochemical processes, which includes aspects of model input (Section 3.3), model structure to enable testing modelling hypotheses and quantifying structural uncertainty (Sections 3.4 and 3.5).

3.2.1 Obtaining OpenWQ

OpenWQ can be obtained from the official GitHub repository: `https://github.com/ue-hydro/openwq`. The recommended method for obtaining OpenWQ is to clone the repository with "git clone". Downloading the repository is also possible, but it is less advisable because OpenWQ is continuously updated and improved. OpenWQ is designed for the user to update their copy quickly and efficiently with "git pull" instead of requiring them to download each new version. The cloning (or download) should be performed inside the host hydro-model code directory as shown in Fig. 1. It will create a new folder named "openwq" with the entire repository. The "main" branch should be used because it will contain the latest official release version. Compiling OpenWQ is carried out with CMake (Kitware, 2022) and the provided "CMakeLists" file. When new versions of OpenWQ are made available, they will only require the user to recompile using the same "CMakeLists" file. Once the coupling of OpenWQ to a hydro-model is completed, carrying out updates to OpenWQ's source code will not break OpenWQ's API calls and interface/coupler functions in the host hydro-model.

3.2.2 Portability via internal dynamic coupler

In order to optimize the implementation of OpenWQ in hydro-models, the model has been designed as an internal coupler module (Fig. 2). External coupling (i.e., OpenWQ reading output files from the host-model and running standalone) was not a viable option because hydro-models, particularly hydrological models, often deal with many water fluxes moving around vertically and horizontally within and across hydrological compartments (e.g., snow, soil), which are not always possible to export and disentangle. Even in cases where models allow exporting all water fluxes separately, mapping those in OpenWQ and harmonizing units to correctly compute the corresponding solute mass transport would be extremely difficult for most applications. Focusing on a flexible full-coupling approach with minimal code re-engineering allowed us to address this problem, with OpenWQ being specifically designed to adapt its internal structure (e.g., spatial domain, temporal resolution) to that of the host hydro-model with the support of interface and coupler functions that establish one-way, plug-in-type communications between the two simulation systems. Although these interface and coupler functions may need adjustments to harmonize particular aspects of the host hydro-models, a generic "hydro-link" file is provided with a template for these

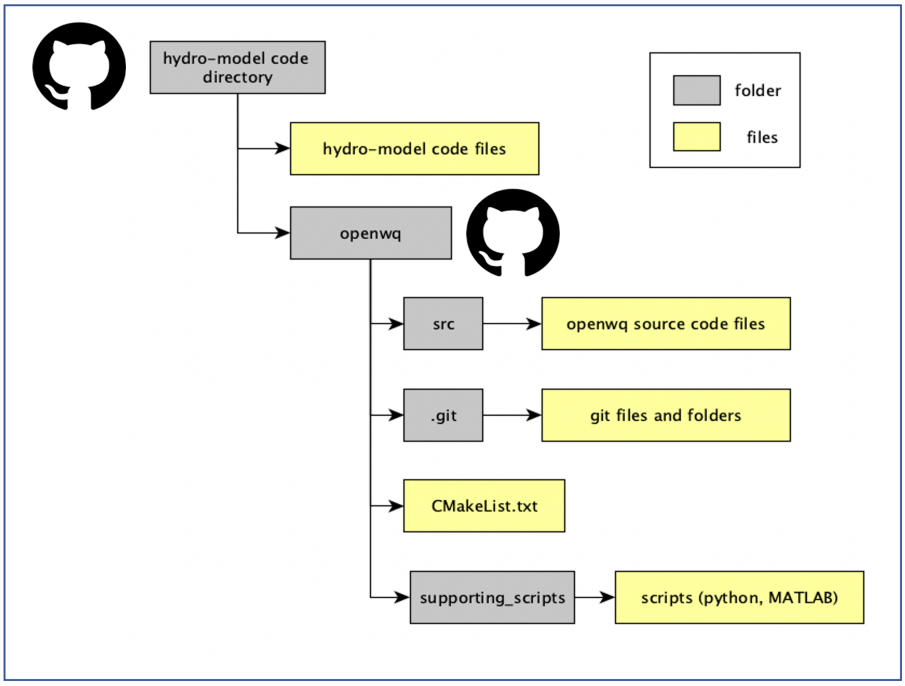


Figure 1: General model structure used in hydrological model

functions. It is possible to create such a template because most hydro-models follow the general model architecture depicted in Fig. 3.

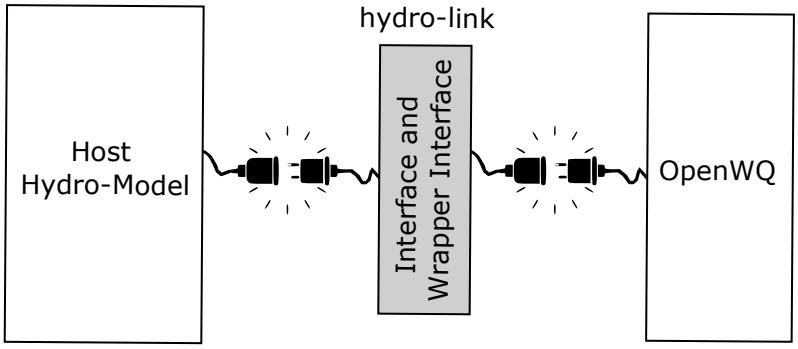


Figure 2: OpenWQ concept as a coupler

The integration of OpenWQ into existing hydro-models is carried out through four coupler functions that are responsible for (1) converting datatypes and data structures between OpenWQ and the "host" hydro-model, (2) passing hydro-fluxes into OpenWQ, and (3) calling OpenWQ's APIs (Fig. 4). These four coupler functions are invoked through interface routines contained within a C++ file with the default name "*OpenWQ_hydrolink.cpp*". This is a one-way communication from the host hydro-model to OpenWQ, so no information is returned to the host model. The interface functions are generic and serve as templates (coupling recipes) that have been optimized to streamline the coupling procedure. The adjustments needed in the interface functions are to ensure that the data types and structure

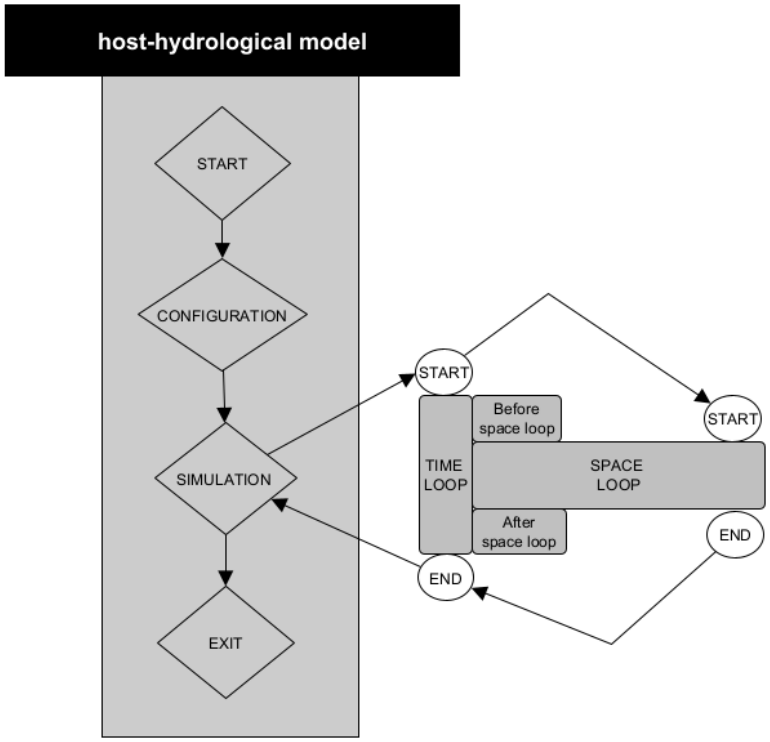


Figure 3: General model structure commonly used in dynamic hydrological, hydrodynamic, and groundwater model

used in the hydro-model are properly translated into OpenWQ's own data structure and conventions so that OpenWQ's APIs can perform adequately.

The following general steps can be performed to couple OpenWQ to a hydro-model: (STEP 1) identify critical elements of hydro-model structure (see Fig. 3), (STEP 2) git-clone OpenWQ, (STEP 3) create the C++ OpenWQ classes and objects to be loaded as modules or libraries in the hydro-model, (STEP 4) identify the appropriate places in the hydro-model to call OpenWQ's coupler functions and APIs (see Fig. 4), (STEP 5) materialize such calls, (STEP 6) adjust OpenWQ's coupler function to adapt to hydro-model data structures (referred to "COUPLER CODE", see below), and (STEP 7) compile the new coupled model. Fig. 5 shows the general structure of each of the four coupler functions, where the location and purpose of the "COUPLER CODE" block are also explained:

1. Coupler function 1 (*openwq::decl*) invokes a series of API calls that handle tasks associated with the initial configuration of the model, initialization of variables, and pre-processing of the input data;
2. Coupler function 2 (*openwq::run_time_start*) invokes a series of API calls that handle tasks required at the start of each time step;
3. Coupler function 3 (*openwq::run_space*) contains a series of API calls that handle tasks related to the spatial domain of the model; and
4. Coupler function 4 (*openwq::run_time_end*) contains a series of API calls that handle tasks required at the end of each time step.

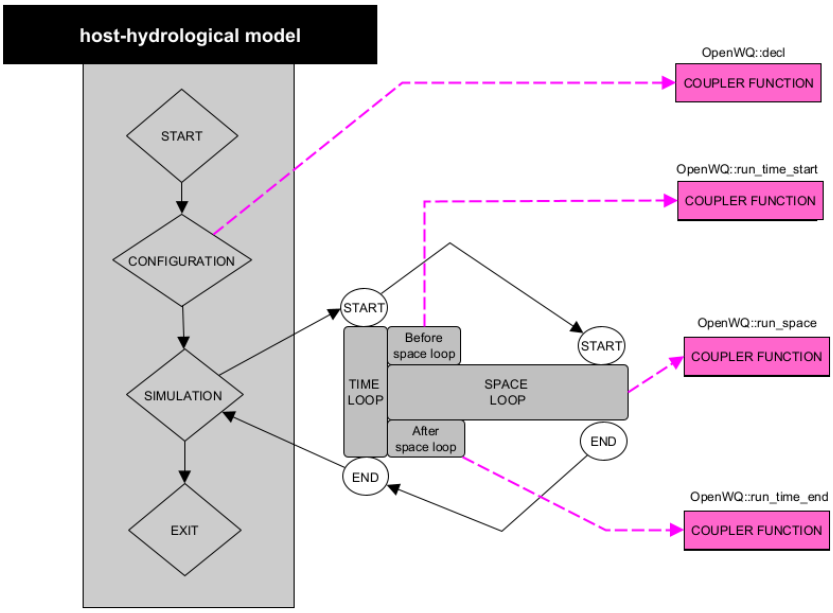
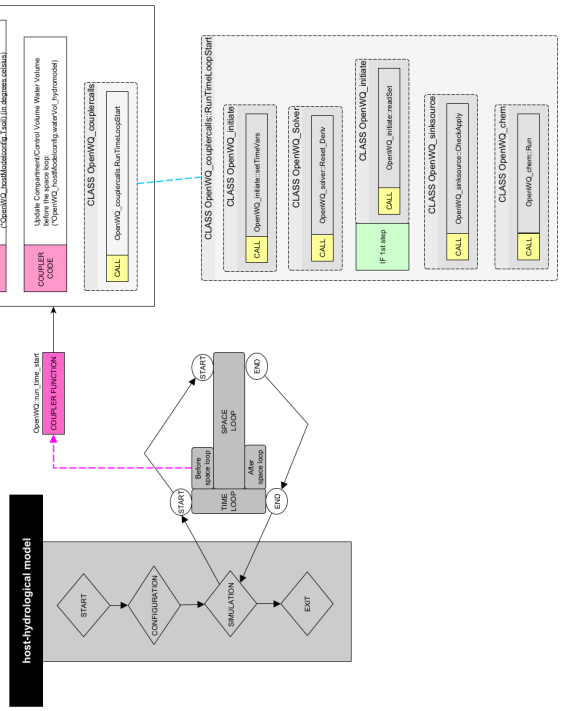


Figure 4: General placement of the calls for OpenWQ's coupler functions

3.2.3 General architecture

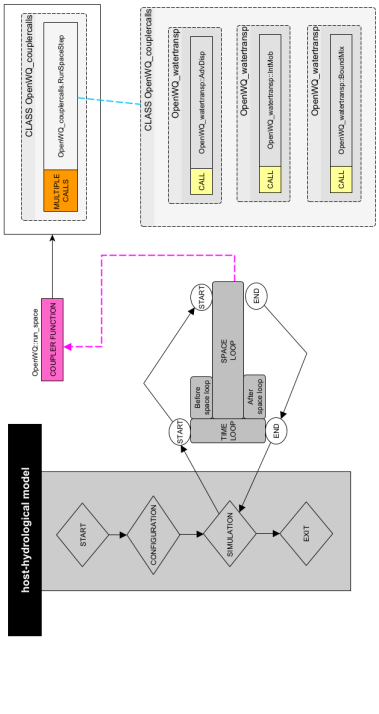
OpenWQ has been created in a way that separates the physics and biochemistry calculations from the numerical implementation. Such an approach was implemented in the SUMMA model (Clark et al., 2015a,b), and we adopted it here to the development of OpenWQ's core structure to improve scalability. This approach addresses a major problem with many hydrological, hydrodynamic, and water quality models where the specification of the model equations is intertwined with their numerical solution (Clark & Kavetski, 2010). This specification complicates the selection and assessment of different model representations (hypotheses) and makes introducing and evaluating alternative numerical methods challenging. As such, the state variables of OpenWQ are only updated inside OpenWQ's own numerical solver based on rates of chemical mass changes (time and space derivatives) caused by different physical and biogeochemical processes. These processes and "rates of change" are computed separately, in process-specific routines, and then passed into the numerical solver for a controlled and contained update of OpenWQ's state variables, which is currently performed as a finite volume problem solved via a simple forward Euler method. The separation of the numerical solver from the physics-biogeochemistry calculations will enable improving the robustness of the numerical implementation in a contained and controlled manner in the future.

The modules in OpenWQ are divided into four groups. Each of these groups is responsible for the calculation of a chemical mass change (time derivative) driven by a particular process (or phenomenon) or group of processes (or phenomena), specifically, (1) initial conditions (*dm_ic*), (2) sinks and sources of chemical load (i.e., chemical mass entering or exiting the model domain) (*dm_ss*), (3) biogeochemical processes (*dm_dt_chemistry*), and (4) physical transport of chemical constituents with water flow (*dm_dt_transport*) (Fig. 6). Each of the modules, in turn, enables a series of modeling options that can be explored, but only one module can be activated for each of the four spatiotemporal-derivative calculations. In other words, the first layer of decision in representing a process pertains to the selection of the key modules that will be responsible for computing each of these four derivatives.

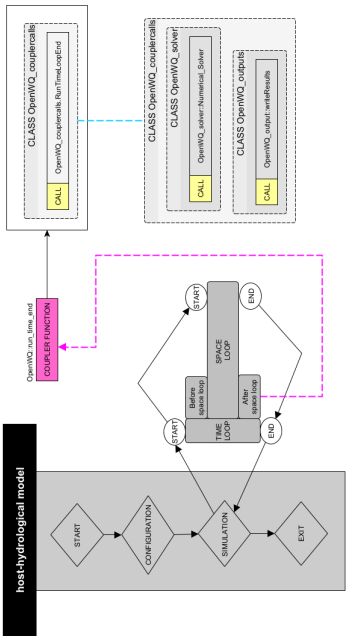


(a) Coupler Function 1: *OpenWQ :: decl*

(b) Coupler Function 2: *OpenWQ :: run_time_start*



(c) Coupler Function 3: *OpenWQ :: run_space*



(d) Coupler Function 4: *OpenWQ :: run_time_end*

Figure 5: Generic coupler functions to integrate OpenWQ into a hydro-model.

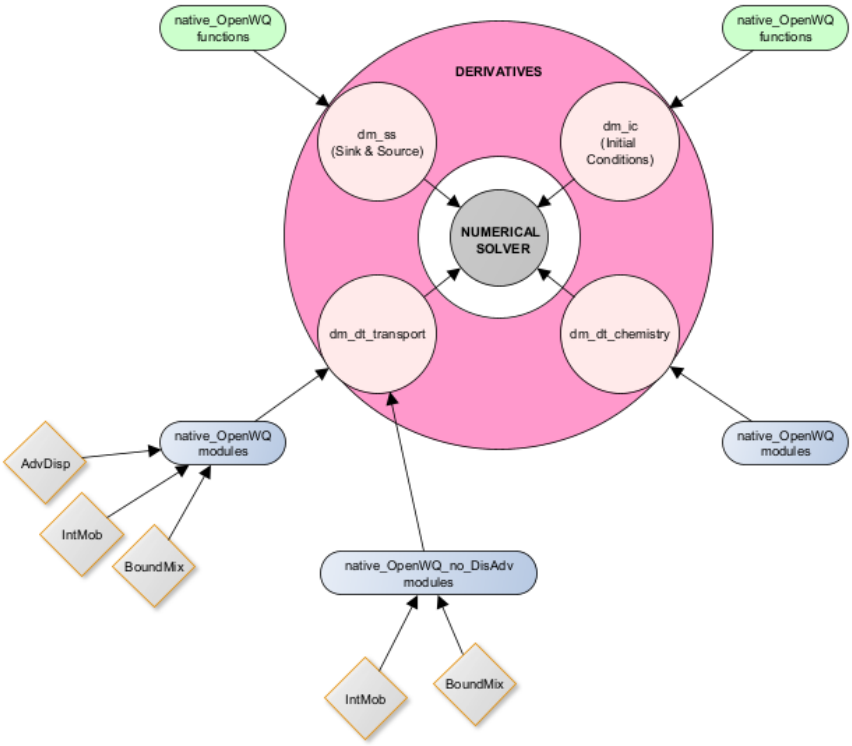


Figure 6: General modular architecture of OpenWQ that separates the different modules and methods available

3.2.4 State variables & Model-adaptive spatial and temporal discretization

OpenWQ was designed to automatically adapt to the hydro-model spatial and temporal discretization structure, which can include Hydrological Response Units (HRUs), 1D–3D spatial distributed meshes, structured or unstructured meshes, and multiple domains or hydrological compartments (e.g., snow, soil, groundwater). OpenWQ's state variables (column 1 in Fig. 7), such as solute mass, are stored in a hierarchical data structure organized with the following nested fields: (1) domain that refers to hydrological compartments (column 2 in Fig. 7), (2) chemical species (column 3 in Fig. 7), and (3) sub-domain that refers to the internal 1D–3D spatial discretization of the domains or hydrological compartments (column 4 in Fig. 7). The hierarchical data structures are built via the Armadillo C++ library for linear algebra and scientific computing (Sanderson & Curtin, 2016, 2018).

The model interface and coupler are responsible for passing information about the spatiotemporal configuration of the host hydro-model into OpenWQ. Such domain configuration options are specific to each host hydro-model and depend on its particular domain discretization scheme and model decisions specified by the users of the hydro-model. Once that information is digested by the couplers and passed on to OpenWQ, a series of API calls dynamically create the corresponding hierarchical data structures and sub-structures that match those in the host model. Each of the data structures stores information about a state variable or a supporting non-state variable (e.g., water fluxes, time derivatives). The state-variable data structures record the spatiotemporal evolution of the mass of the different chemical species tracked in each model domain or hydrological compartment (e.g., snow, soil, lake).

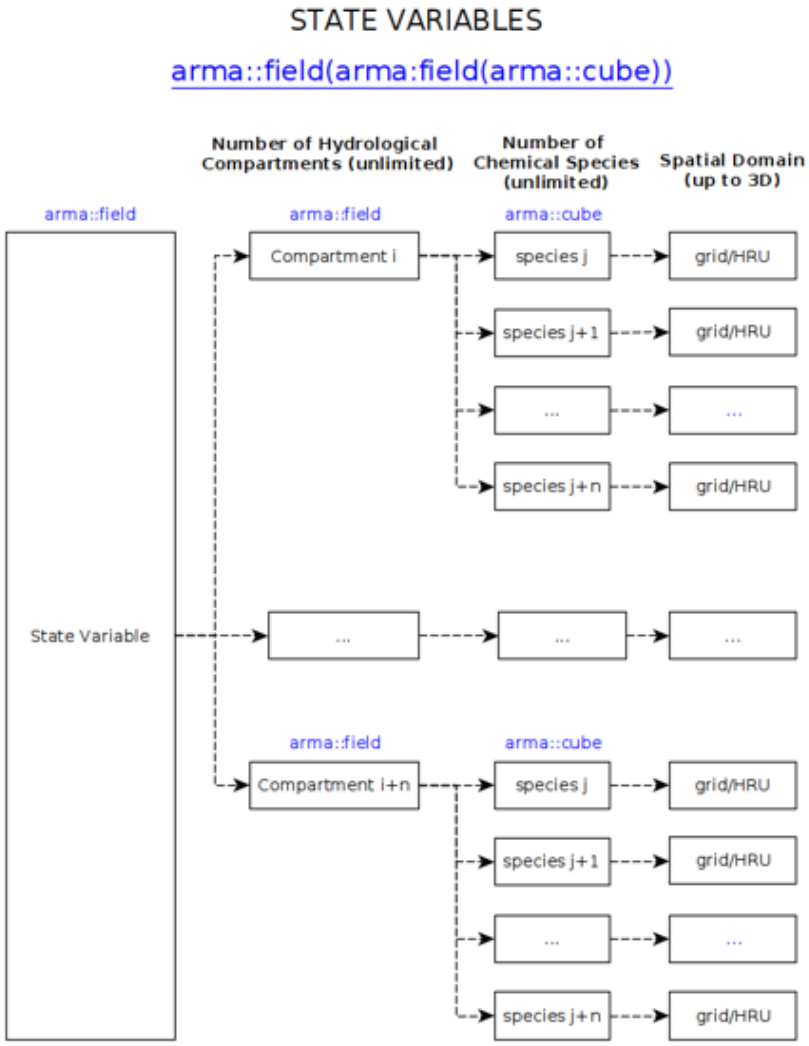


Figure 7: State variables are stored in OpenWQ via dynamic hierarchical data structures

3.3 Model setup and configuration: Inputs and Outputs

The configuration of OpenWQ is provided via JavaScript Object Notation (JSON) files. JSON is an ideal format for large data inputs that have a hierarchically structured relationship. JSON files are composed of key/value pairs, e.g., "RUN_MODE_DEBUG: TRUE". OpenWQ requires four JSON files to run. Each of these files deals with a particular aspect of the model setup and is given one of the following designations (1) "runManagement", (2) "Biogeochemistry", (3) "Configuration", and (4) "Source/Sink".

The "RunManagement" file is the entry point to OpenWQ. It provides the basic instructions, simulation and model output decisions, and the full or relative paths to the other JSON files needed to run a model. Currently, OpenWQ supports HD5F and CSV output file formats. The "Biogeochemistry" file is where the chemical species and biogeochemical cycling frameworks are created and characterized. It uses a standard template structure for the characterization transformations and the chemical species involved. Different "Biogeochemistry" files can be prepared to simulate different pollution problems, which can be readily loaded simultaneously (or used separately or swapped as needed) into a simulation

via the "RunManagement" file. The "Biogeochemistry" file is also where the contaminant species available for the simulation are defined and given a "tag" that can be invoked in other input files. All cycling frameworks and their respective transformations are also provided with a unique "tag" that can be used to flexibly set up the model as desired. These cycling framework tags are particularly important in the "Configuration" file to assign the desired transformations to each domain or hydrological compartment of the model (e.g., snow, soil, groundwater). The "Configuration" file also includes information about the initial conditions. A "Sink/Source" file provides information about the mass exchange with regions outside the model domain. This information can include external chemical mass loading into (source) or loss from (sink), the model domain that can be associated with particular external water fluxes, as in the case of precipitation, or mass inputs that can be independent of the hydrological cycle. Typical examples of such mass inputs can include fertilizer application, which is a relatively localized source, or atmospheric deposition, which is a relatively distributed source. There can be multiple "Sink/Source" files in a given OpenWQ configuration.

3.4 Native Modules and Process representation

OpenWQ contains four main groups of modules, each producing rates of chemical mass change (time derivatives) associated with specific processes, phenomena, and model aspects, which are passed into the numerical solver for updating the state-variables throughout the simulation (see Fig. 6). The modules deal with (1) initial conditions, (2) sinks and sources, (3) transport with water, and (4) biogeochemistry.

The first two groups of modules deal with initial conditions and sinks and external sources with a focus on translating user inputs into the simulation. For example, the sink and source module applies chemical load as prescribed by the users that can include (1) continuous load arising from sources like atmospheric deposition or the outlet of a wastewater treatment plant or (2) episodic/instantaneous loads arising from sources such as fertilizer and chemical spills. The remaining two groups of modules focus respectively on (1) the physical transport of chemical constituents as water moves through the system and (2) biogeochemical transformations.

3.4.1 Transport Module

Currently, there are two options available for computing the physical transport of dissolved solutes and fully suspended sediments. The first option accounts for both advection and dispersion and solves the hyperbolic-parabolic advection-diffusion partial differential equation (PDE) in up to 3 dimensions depending on the host model spatial discretization scheme (Equation 1); the second option only accounts for advection (Equation 2). In both cases, the PDEs are solved inside OpenWQ's numerical solver as described in Section 3.2.3 using the specific hydro-fluxes and corresponding source and recipient domain cells prescribed by the host models. This means that for each water flux computed by the host-model (at each grid cell and timestep), OpenWQ calculates the corresponding solute mass transported, which will be intimately linked to the internal spatio-temporal water-flux representations of the host-model (i.e., if the host model represents snow as a 1 vertical layer domain (lumped), OpenWQ will compute solute concentrations at that spatial resolution; if the host model runs at daily time steps, OpenWQ will calculate solute concentrations at that temporal resolution). The physical transport can be between computation elements (e.g., HRUs, grid cells) and across domains or hydrological compartments (e.g., runoff, soil saturated, soil unsaturated, canopy), as prescribed by the host model.

$$\frac{\partial(Vc_s)}{\partial t} + \nabla \cdot (V\vec{u}c_s) = \nabla \cdot (V\mathbf{E} \cdot \nabla c_s) + S, \tag{1}$$

$$\frac{\partial(Vc_s)}{\partial t} + \nabla \cdot (V\vec{v}c_s) = S, \tag{2}$$

where V is the volume of the computational element/cell [L^3], c_s is the concentration of a given dissolved substance [ML^{-3}]; $\vec{u} = (u_x, u_y, u_z)$ is the velocity in the x, y, and z directions [LT^{-1}]; $\mathbf{E}$ is the (diagonal) diffusivity tensor [L^2T^{-1}]; and S is a source term [MT^{-1}] that is linked to the chemistry modules. The $\mathbf{E}$ term accounts for the combined effect of different mixing phenomena:

$$\mathbf{E} = \mathbf{E}^{turb} + \mathbf{E}^{tsd} + \mathbf{E}^{sgt} + \mathbf{E}^{d}, \tag{3}$$

where $\mathbf{E}^{turb}$ is the turbulent or eddy diffusivity, which is a complex phenomenon with multi-fractal behaviour dominated by friction forces, $\mathbf{E}^{tsd}$ is the so-called Taylor shear dispersion that arises from the unresolved vertical variation of the horizontal flow, $\mathbf{E}^{sgt}$ accounts for the sub-grid eddy viscosity and diffusivity arising from unresolved mixing occurring at sub-grid scale, and $\mathbf{E}^{d}$ is the background molecular diffusion resulting from the probabilistic Brownian motion concept occurring at particle scales. $\mathbf{E}$ is approximated to the dominant turbulent dispersion based on the eddy viscosity concept: $\mathbf{E}^{turb} = \boldsymbol{\nu}_t/\sigma$, where σ is the Prandtl–Schmidt number. Turbulent viscosity ($\boldsymbol{\nu}_t$) depends on shear velocity ($\vec{u}^*$) and on a turbulent length scale ($l_t = 0.07h$). In this model, this relationship is approximated by an algebraic expression: $\boldsymbol{\nu}_t \sim k\vec{u}^* l_t$, where k is a user-defined scaling factor to account for sub-grid scale eddies (Costa et al., 2016).

3.4.2 Biogeochemistry module

Biogeochemical cycling frameworks are characterized in the appropriate biogeochemistry JSON file. Each cycling framework is given a "tag" that is used to load it into simulations. The characterization of each of these cycling frameworks involves the definition of the associated biogeochemical transformations and the chemical species involved. In turn, the characterization of each of these transformations requires information about the chemical species consumed and produced and the reaction kinetics. It allows for the creation of reaction parameters that can be loaded into the kinetics solver, which relies on the comprehensive C++ Mathematical Expression Toolkit Library (*ExprTk*) developed by Arash Partow (1999–2020) (Partow, 1999). The implementation of *ExprTk* in OpenWQ is simple to use and provides an efficient run-time mathematical expression parser and evaluation engine. *ExprTk* supports numerous forms of functional, logical, and vector processing semantics and is easily extendible. The equations can be written with (1) multiple chemical species, (2) user-defined parameters, and (3) built-in hydro-model dependencies. These model dependencies are tailored to each hydro-model, but they usually include variables like soil moisture and air and soil temperature.

The biogeochemistry JSON file can be prepared manually following the appropriate OpenWQ JSON structure (i.e., key-value pairs). Alternatively, the cycling frameworks can be characterized via a diagram drawn using a *GraphML* editor. *GraphML* is an XML-based file format for graphs, and there are several free *GraphML* editors, such as *yEd*. After the diagrams have been drawn in *GraphML* format, they can be converted into OpenWQ's JSON format using a Python script available on OpenWQ's GitHub repository and copied locally during the coupling step defined in Section 3.2.1. This graphical option has been developed to enable collaboration and co-creation of water quality models through a more visual interaction, which can be particularly helpful for water quality modeling activities that often involve interaction between different disciplines that may be less familiar with JSON files.

The expressions used to represent the different reaction kinetics in the model via the biogeochemistry JSON files can take many forms and are deployed in OpenWQ via the *ExprTk* integration. However, these expressions often take the form of sequences of reaction networks involving single or multiple chemical species governed by first-, second-, or third-order kinetics (Eq. 4, Eq. 5, Eq. 6), respectively:

$$\frac{dc_A}{dt} = -k\lambda c_A, \tag{4}$$

$$\frac{dc_A}{dt} = -k\lambda c_A^2, \tag{5}$$

$$\frac{dc_A}{dt} = -k\lambda c_A^2 c_B, \tag{6}$$

where c_A and c_B [ML^{-3}] are the concentrations of chemical species A and B, parameter/variable λ represents weather/hydrological dependencies (such as soil moisture and temperature), and k is the reaction rate [$ML^{-3}T^{-1}$]. The reaction rate k can be provided as the reaction rate using standard maximum at a reference temperature (often 20°C) or using expressions that can include relationships with the hydrological/weather dependent variables/parameters.

3.5 Eco-modelling lab and Cross-model deplyoment: Benefits and Innovation

OpenWQ's concept as a coupler and customizable biogeochemistry modeling framework allows for cross-model deployment of eco-lab, co-creation modeling capabilities, providing a long-awaited transformative direction for innovation in water quality modeling that tackles the inherent challenges of being at the intersection between several scientific fields, including biogeochemistry, soil science, hydrology, hydrodynamics, and hydrogeology. This deployment is an essential step to enhancing collaborative efforts and streamlining knowledge/innovation transfer between the different disciplines involved, ultimately benefiting the entire environmental and ecological research and management community. The ability to deploy OpenWQ across different hydro-models provides the following transformational changes in the current paradigm of water quality modeling.

1. Optimization of investment because focus is given to a single biogeochemical tool, thus reducing code maintenance,
2. Maximization of benefit because it allows (a) any existing hydro-tool to extend its capabilities to water quality and ecological studies, (b) progress achieved by the biogeochemistry and soil-science research community to be transferred into OpenWQ and automatically benefits multiple hydro-models via update installs, (c) OpenWQ's input files to be transferred across hydro-models, providing experience and knowledge transfer between environmental projects and research communities, and
3. Reproducibility is reinforced because OpenWQ's input files are transferable across hydro-models, allowing for more rigorous cross-model comparisons.

Once different biogeochemical modeling hypotheses have been identified and set up for testing in one hydro-model-OpenWQ coupled model, they can be easily transferred for simulation in any other hydro-amodel that has been coupled to OpenWQ (Fig. 8).

4 Discussion

4.1 Unifying Different Approaches to Water Quality Modelling

OpenWQ provides a unifying modeling framework to deploy different approaches to physiochemically based soil and water quality modeling across existing hydrological, hydro-

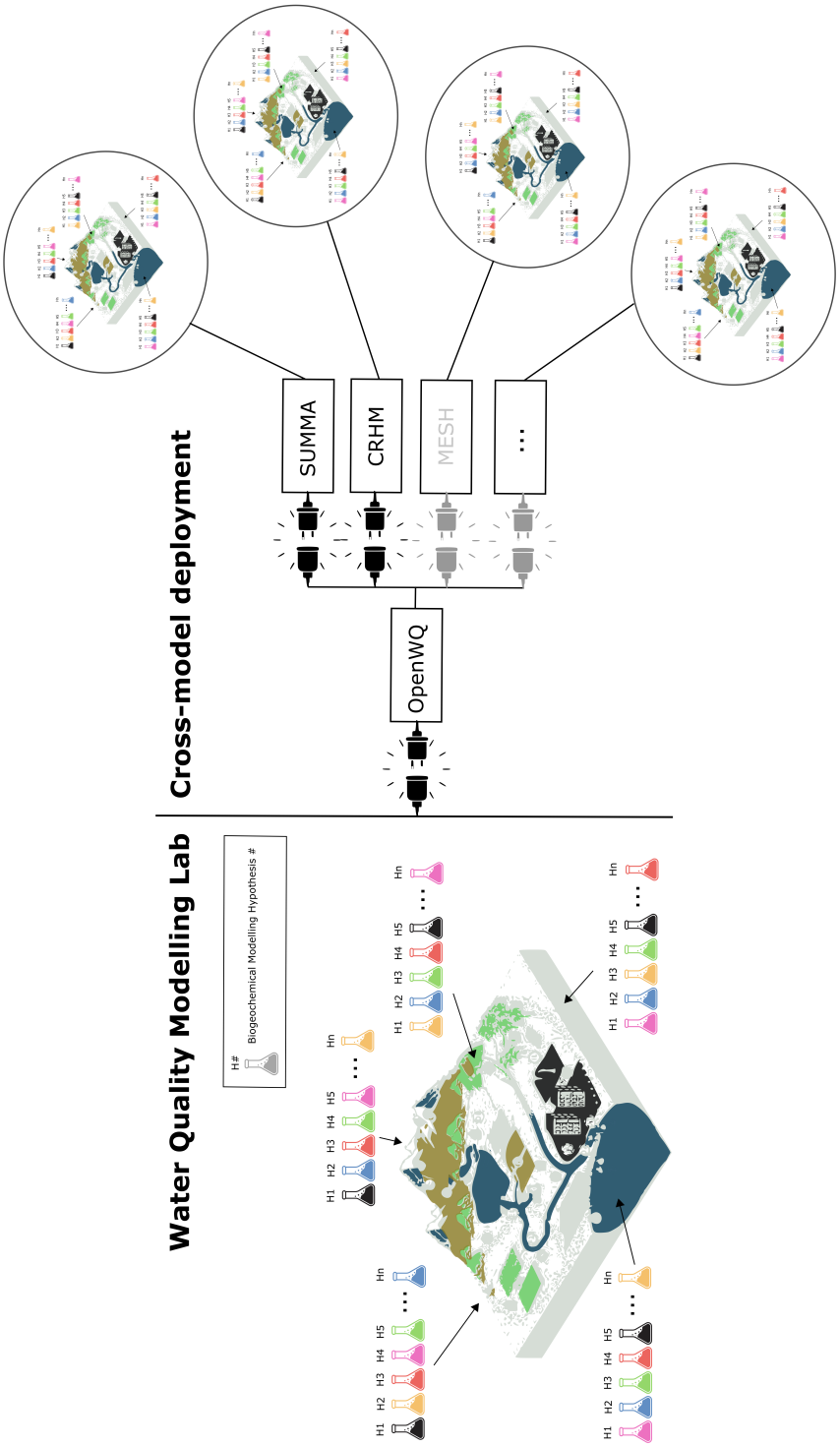


Figure 8: Cross-model deployment of biogeochemistry modelling setup, hypothesis and scenários

dynamic, and hydrogeological models. Here, we emphasize how OpenWQ can help unify different modeling approaches to simulate different landscapes and pollution problems.

1. *Sub-catchment versus catchment-scale modeling.* OpenWQ provides a unique opportunity to harmonize spatial scales in water quality modeling. Water quality models usually belong to one of two scales regarding the spatial domain, (a) sub-catchment and (b) catchment. Sub-catchment water quality models, including river-reach models such as WASP (Wool et al., 2020; Di Toro et al., 1983) and QUAL2E (Brown & Barnwell, 1987), lake models such as MyLake (Saloranta & Andersen, 2007) and Delft3D (Lesser et al., 2004), and aquifer models such as MODFLOW-MT3D (Harbaugh, 2005; Bedekar et al., 2016) and FEFLOW (Trefry & Muffels, 2007), require the characterization of both horizontal and vertical boundary conditions because they do not represent entire closed systems such a river basin. Instead, they focus on subregions within such closed systems. On the other hand, catchment models such as INCA (Jackson-Blake et al., 2016; A. J. Wade et al., 2002; Whitehead et al., 1998) and SWAT (Arnold et al., 2012) only require vertical boundary conditions (e.g., precipitation) because they focus on the larger closed system that contains the entire basin area. Accordingly, these two scales of models tend to focus on different pollution problems. Sub-catchment models tend to focus more on point sources, such as wastewater discharge, and catchment models often look at diffuse pollution (e.g., agriculture nutrients and fertilizer use). These differences also result in different chemical species and biogeochemical cycles of focus, e.g., river models often address problems related to low Dissolved Oxygen (DO) levels caused by biomass decomposition and BOD (Biological Oxygen Demand) arising from wastewater discharges and other point sources discharging directly in river and lakes.
   OpenWQ enables multi-scale chemistry simulations that can be integrated into hydro-models of sub-catchment and catchment scales; thus, it can help bridge the gap between these two approaches, as well as allow for better integration of cross-dependent biogeochemical cycles. For example, whereas the DO and nitrogen cycles are often simulated in river-reach models, catchment models tend to focus more on the nitrogen cycle and assume that there is an unlimited amount of DO available. Although this assumption could be considered valid in many cases for surface runoff, it is certainly erroneous for water bodies such as rivers, lakes, and groundwater.

2. *Simple versus detailed biogeochemistry representation.* Water quality models have been developed with varying degrees of detail in the representation of biogeochemical processes (Costa, Baulch, et al., 2020a). For example, while INCA and HYPE (Lindström et al., 2010) provide simpler biogeochemical modeling approaches for the nitrogen and phosphorous cycles, requiring a smaller number of reaction-kinetic parameters to calibrate them may be more suitable for data-scarce regions; more complex biogeochemical models like HSPF ("Hydrological Simulation Program–Fortran, User's manual for version 11: U.S. Environmental Protection Agency", 1997) arguably provide higher model fidelity but may only be applicable in data-rich environments. However, when selecting a model for a particular region, modelers often have to make compromises with process representation because some models have more detailed physics-based coupled water-energy balance computations such as AnnAGNPS (Bosch et al., 1998) but may offer more limited biogeochemical capabilities (Costa, Baulch, et al., 2020a). OpenWQ enables addressing this issue through its flexible eco-modelling lab, which allows testing different biogeochemical conceptual models and modelling hypotheses, from simple biogeochemical cycles and transformations to more complex and intertwined reaction networks involving dozens or hundreds of chemical species.

3. *Unconnected versus interconnected biogeochemical cycling representation.* Most catchment models simulate biogeochemical cycles in isolation. For example, popular

models such as SWAT and HYPE simulate the phosphorous and nitrogen cycles without integrating them with the DO-BOD cycle, despite them being strongly interrelated because DO is used in both cycles. In the nitrogen cycle, DO is used in nitrification, where ammonia is oxidated into nitrite (NO_3) and nitrate (NO_2). In the DO-BOD cycle, DO is used in the biological, aerobic decomposition of organic matter. OpenWQ aims to allow for a biogeochemistry representation that is less compartmentalized, making interactions between cycling frameworks more fluid and closer to reality (i.e., model fidelity). In reality, DO dynamics affect the cycling of many chemical constituents, from nutrients to heavy metals. OpenWQ's eco-modelling lab addresses this issue through its open reaction-network solver, which provides the flexibility to deploy any number of biogeochemical cycling representations of any number of chemical species simultaneously, which can be connected or unconnected.

4. *Background chemical transport driver.* Typically, hydro-models are developed and maintained by research communities with specific research motivations driven by a regional context. These modeling tools often become highly specialized for particular environments and applications, and modelers commonly find it hard to find a water quality modeling tool that integrates such regionally important hydro-transport modeling capabilities. There is also in-house expertise in such communities and research groups that is passed on over the years between elements of the research community, which can make them reluctant to switch to other modeling tools because it may involve steep learning curves. OpenWQ aims to address this issue by enabling extension to water quality modeling capabilities directly on those models so that modelers can continue using the hydro-modeling tool that they consider more suitable to the environments on which they focus.

4.2 Advancing Current Model Development Paradigms

Our focus on providing a flexible biogeochemical model representation framework that is deployable on any process-based hydro-models and offers a unique opportunity to transform the way environmental studies involving water quality modeling are carried out. It also provides a concrete way for streamlining collaboration, co-creation, and knowledge and innovation transfer between the hydrological and biogeochemistry/soil science communities. Such streamlining is materialized by OpenWQ's coupler interface, which links developments/progress in the hydro-modeling tool (hydro-modeling community) with developments/progress in OpenwQ (biogeochemistry communities). Another major issue with the current water quality modeling paradigm is that most models offer little flexibility to adapt the representation of biogeochemistry processes to (1) local/region context, (2) data availability, and (3) application(s) of interest, resulting in the need to make difficult choices when selecting a modeling tool because compromises between the representation of hydrofluxes and biogeochemistry are often required (Costa, Baulch, et al., 2020a).

4.3 Limitations

In OpenWQ, the physical transport of chemical constituents relies on the quality of the water fluxes passed by the host hydro-model. Although this may be an advantage because it allows the user to choose the hydro-model that best suits the modeling needs, it can become an issue if the calculated fluxes in the hydro-model are not accurate. Modelers should select the hydro-model coupled to OpenWQ that is the most suitable for the application at hand or pursue the coupling of OpenWQ to another hydro-model.

The native biogeochemistry module of OpenWQ that provides water quality-lab capabilities assumes that biogeochemical cycling can be represented via a series of sequential and parallel reaction networks. Although this representation is true for most chemical constituents, pollution problems, and environmental studies, some biogeochemistry may involve

formulations that rely on variable dependencies that may not be available in a particular coupled model. For example, simulating microbiological pollution in lakes, rivers, and beaches, such as contamination with fecal coliforms originating from leaking septic tanks and wastewater discharges, requires the simulation of die-off rates due to exposure to solar radiation. In the case of hydrodynamic models, such information may not be available. In the case of hydrological models, which typically deal with such data, that variable dependency may not have been passed into OpenWQ during the development of the coupler interface, so updates to the coupler may be needed. OpenWQ's coupler functions have been designed to make the addition of new dependency variables straightforward.

The portability of OpenWQ is materialized through a series of coupler functions and wrapper interface functions that enable its coupling to hydro-models written in C++ or Fortran. However, the use of a more standardized model interface framework, such as the Basic Model Interface (BMI Peckham et al., 2013; E. W. Hutton et al., 2020), could help streamline further the coupling process through the use of standard control and query functions. This could make that model both easier to learn and easier to couple with other software elements. BMI, for example, currently supports five languages: C, C++, Fortran, Java, and Python.

5 Conclusions

This paper describes a unified framework for enabling multi-biogeochemical modeling capabilities in existing hydro-models. The framework, which we call Open Water Quality (OpenWQ), was designed with both coupler and water quality-lab modeling capabilities to enable (1) flexible co-creation and testing of biogeochemistry modeling representations, (2) systematic implementation and evaluation of alternative modeling approaches for process representation, and (3) identification of specific causes of model weaknesses.

This work arises from the recognition that the hydrological transport of contaminants in the environment strongly affects their concentrations in aquatic ecosystems, but the interaction between the biogeochemical and hydrological-hydrodynamic-hydrogeological communities lacks the appropriate mechanisms for an efficient transfer of knowledge and innovation between the two communities. Many hydro-models are developed and maintained by specific research communities addressing particular regions and climate zones. The investment in such tools creates invaluable in-house expertise, making them less likely to switch to other modeling tools and hindering their expansion to water quality and ecological studies. OpenWQ aims to address this problem and provide a concrete direction for innovation in connecting communities through an optimized plug-in-like water quality model that can be coupled to existing hydro-modeling tools, extending the value of these tools for water quality and ecological studies in their region of focus. Different process representations and different spatial configurations can be integrated into the structural model core, which enables users to decompose the modeling problem into the individual decisions made as part of model development and evaluate different "fine-grained" model development decisions in a systematic and controlled way.

OpenWQ can provide the necessary model flexibility to progress toward answering the following fundamental modeling questions and challenges: (1) which hydro-biogeochemical processes should be represented explicitly in different environmental settings, and, correspondingly, which processes can be ignored or greatly simplified; (2) what modeling approaches should be used to represent the dominant biogeochemistry at the spatial scale of the model discretization; (3) how should heterogeneity in pollution and biogeochemical processes be represented across spatial scales, including the complexity of transport across landscapes; and (4) how can we provide insights into the sources of model uncertainty. The companion paper describes the integration of OpenWQ into two hydro-models, SUMMA and CRHM, describing how coupling interfaces between the two models have not only enabled water quality simulation capacities in these host hydro-models but, even more importantly,

established a direct and permanent link for the transfer of innovation between the associated modeling communities, promoting cooperation and co-creation. Example applications of pollution studies enabled by the coupling of the tools are also provided to begin to address some of these fundamental modeling challenges.

Acknowledgments

This work was supported by the NSERC-Alliance Grant ALLRP 571910-22. We acknowledge support from Environment and Climate Change Canada, the Global Water Futures programme, and the Natural Sciences and Engineering Research Council of Canada under Discovery Grant RGPN-2020-04467 (RJS).

References

Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., & Strömqvist, J. (2012). Water and nutrient simulations using the HYPE model for Sweden vs. the Baltic Sea basin – influence of input-data quality and scale. *Hydrology Research*, *43*(4), 315-329. doi: 10.2166/nh.2012.010

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., ... Jha, M. K. (2012). SWAT: Model Use, Calibration, and Validation. *Trans. ASABE*, *55*(4), 1491–1508.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment Part I: Model Development. *JAWRA Journal of the American Water Resources Association*, *34*(1), 73-89. doi: https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Beck, M. B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Water Resources Research*, *23*(8), 1393–1442. doi: 10.1029/WR023i008p01393

Bedekar, V., Morway, E. D., Langevin, C. D., & Tonkin, M. J. (2016). *MT3D-USGS version 1: A U.S. Geological Survey release of MT3DMS updated with new and expanded transport capabilities for use with MODFLOW* (Tech. Rep.). Reston, VA. doi: 10.3133/tm6A53

Bosch, D., Theurer, F., Bingner, R., & Felton, G. (1998). Evaluation of the AnnAGNPS water quality model. *Agricultural Non-Point Source Water Quality Models: Their Use and Application; John, EP, Daniel, LT, Rodney, LH, Eds*, 45–54.

Brown, L., & Barnwell, T. (1987). The enhanced stream water quality models qual2e and qual2e-uncas: documentation and user manual. *Environmental Protection Agency*.

Clark, M. P., & Kavetski, D. (2010). Ancient numerical daemons of conceptual hydrological modeling: 1. fidelity and efficiency of time stepping schemes. *Water Resources Research*, *46*(10). doi: https://doi.org/10.1029/2009WR008894

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9). doi: https://doi.org/10.1029/2010WR009827

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Rasmussen, R. M. (2015a). A unified approach for process-based hydrologic modeling: 1. modeling concept. *Water Resources Research*, *51*(4), 2498-2514. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017198` doi: https://doi.org/10.1002/2015WR017198

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Marks, D. G. (2015b). A unified approach for process-based hydrologic modeling: 2. model implementation and case studies. *Water Resources Research*, *51*(4), 2515-2542. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017200` doi: https://doi.org/10.1002/2015WR017200

Costa, D., Baulch, H., Elliott, J., Pomeroy, J., & Wheater, H. (2020a). Modelling nutrient dynamics in cold agricultural catchments: A review. *Environmental Modelling & Software*,

124, 104586. doi: https://doi.org/10.1016/j.envsoft.2019.104586

Costa, D., Baulch, H., Elliott, J., Pomeroy, J., & Wheater, H. (2020b). Modelling nutrient dynamics in cold agricultural catchments: A review. *Environmental Modelling & Software*, *124*, 104586. doi: https://doi.org/10.1016/j.envsoft.2019.104586

Costa, D., Burlando, P., & Liong, S.-Y. (2016). Coupling spatially distributed river and groundwater transport models to investigate contaminant dynamics at river corridor scales. *Environmental Modelling & Software*, *86*, 91-110. doi: https://doi.org/10.1016/j.envsoft.2016.09.009

Costa, D., Pomeroy, J., & Wheater, H. (2018). A numerical model for the simulation of snowpack solute dynamics to capture runoff ionic pulses during snowmelt: The pulse model. *Advances in Water Resources*, *122*, 37-48. doi: https://doi.org/10.1016/j.advwatres.2018.09.008

Costa, D., & Pomeroy, J. W. (2019). Preferential meltwater flowpaths as a driver of preferential elution of chemicals from melting snowpacks. *Science of The Total Environment*, *662*, 110–120. doi: 10.1016/J.SCITOTENV.2019.01.091

Costa, D., Pomeroy, J. W., Brown, T., Baulch, H., Elliott, J., & Macrae, M. (2021). Advances in the simulation of nutrient dynamics in cold climate agricultural basins: Developing new nitrogen and phosphorus modules for the Cold Regions Hydrological Modelling Platform. *Journal of Hydrology*, *603*, 126901. doi: https://doi.org/10.1016/j.jhydrol.2021.126901

Costa, D., Roste, J., Pomeroy, J., Baulch, H., Elliott, J., Wheater, H., & Westbrook, C. (2017). A modelling framework to simulate field-scale nitrate response and transport during snowmelt: The wintra model. *Hydrological Processes*, *31*(24), 4250-4268. doi: https://doi.org/10.1002/hyp.11346

Costa, D., Shook, K., Spence, C., Elliott, J., Baulch, H., Wilson, H., & Pomeroy, J. W. (2020). Predicting variable contributing areas, hydrological connectivity, and solute transport pathways for a canadian prairie basin. *Water Resources Research*, *56*(12), e2020WR027984. doi: https://doi.org/10.1029/2020WR027984

Davies, T. D., Brimblecombe, P., Tranter, M., Tsiouris, S., Vincent, C. E., Abrahams, P., & Blackwood, I. L. (1987). *The Removal of Soluble Ions from Melting Snowpacks* (H. G. Jones & W. J. Orville-Thomas, Eds.). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-009-3947-9_20

Di Toro, D. M., Fitzpatrick, J. J., Thomann, R. V., & Hydroscience, I. (1983). *Documentation For Water Quality Analysis Simulation Program (WASP) And Model Verification Program (MVP)* (Tech. Rep.).

Frank, E. G., & Schlenker, W. (2016). Balancing economic and ecological goals. *Science*, *353*(6300), 651-652. doi: 10.1126/science.aaf9697

Fu, B., Merritt, W. S., Croke, B. F., Weber, T. R., & Jakeman, A. J. (2019). A review of catchment-scale water quality and erosion models and a synthesis of future prospects. *Environmental Modelling & Software*, *114*, 75–97. doi: 10.1016/J.ENVSOFT.2018.12.008

Harbaugh, A. W. (2005). *MODFLOW-2005, the US Geological Survey modular groundwater model: the ground-water flow process* (Vol. 6). US Department of the Interior, US Geological Survey Reston, VA, USA.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016). Most computational hydrology is not reproducible, so is it really science? *Water Resources Research*, *52*(10), 7548–7555. doi: https://doi.org/10.1002/2016WR019285

Hutton, E. W., Piper, M. D., & Tucker, G. E. (2020). The basic model interface 2.0: A standard interface for coupling numerical models in the geosciences. *Journal of Open Source Software*, *5*(51), 2317. doi: 10.21105/joss.02317

Hydrological Simulation Program–Fortran, User's manual for version 11: U.S. Environmental Protection Agency. (1997). *National Exposure Research Laboratory, Athens, Ga., EPA/600/R-97/080, 755 p.*

Jackson-Blake, L. A., Wade, A. J., Futter, M. N., Butterfield, D., Couture, R.-M., Cox,

B. A., ... Whitehead, P. G. (2016). The INtegrated CAtchment model of phosphorus dynamics (INCA-P): Description and demonstration of new model structure and equations. *Environmental Modelling & Software*, *83*, 356–386. doi: 10.1016/J.ENVSOFT.2016.05.022

Jones, H. G. (1999). The ecology of snow-covered systems: a brief overview of nutrient cycling and life in the cold. *Hydrological Processes*, *13*(14-15), 2135–2147. doi: 10.1002/(SICI)1099-1085(199910)13:14/15⟨2135::AID-HYP862⟩3.0.CO;2-Y

Keith, D. A., Ferrer-Paris, J., Nicholson, E., Bishop, M. J., Polidoro, B. A., Ramirez-Llodra, E., ... Kingsford, R. T. (2022). A function-based typology for earth's ecosystems. *Nature*, *610*(7932), 513–518. doi: 10.1038/s41586-022-05318-4

Kitware. (2022). *CMake: Reference documentation.*

Lesser, G., Roelvink, J., van Kester, J., & Stelling, G. (2004). Development and validation of a three-dimensional morphological model. *Coastal Engineering*, *51*(8), 883-915. (Coastal Morphodynamic Modeling) doi: https://doi.org/10.1016/j.coastaleng.2004.07.014

Lilbaek, G., & Pomeroy, J. W. (2007). Modelling enhanced infiltration of snowmelt ions into frozen soil. *Hydrological Processes*, *21*(19), 2641–2649. doi: 10.1002/hyp.6788

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., & Arheimer, B. (2010). Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology Research*, *41*(3-4), 295. doi: 10.2166/nh.2010.007

Marsh, P., & Pomeroy, J. W. (1999). Spatial and temporal variations in snowmelt runoff chemistry, Northwest Territories, Canada. *Water Resources Research*, *35*(5), 1559–1567. doi: 10.1029/1998WR900109

Mekonnen, B. A. (2016). *Modeling and management of water quantity and quality in cold-climate Prairie Watersheds* (Unpublished doctoral dissertation). University of Saskatchewan.

Moore, D. S., Bingner, R. L., & Theurer, F. D. (2006). ANNAGNPS: Accounting for Snowpack, Snowmelt, and Soil Freeze-Thaw. In *Eighth federal interagency sedimentation conference (8thfisc),* (pp. 475–482). Reno, NV, USA.

Partow, A. (1999). *ExprTk C++ Mathematical Expression Library.*

Peckham, S. D., Hutton, E. W., & Norris, B. (2013). A component-based approach to integrated modeling in the geosciences: The design of csdms. *Computers & Geosciences*, *53*, 3–12.

Pomeroy, J., Brown, T., Fang, X., Shook, K., Pradhananga, D., Armstrong, R., ... Lopez Moreno, J. (2022). The cold regions hydrological modelling platform for hydrological diagnosis and prediction based on process understanding. *Journal of Hydrology*, *615*, 128711. doi: https://doi.org/10.1016/j.jhydrol.2022.128711

Pomeroy, J., Davies, T., Jones, H., & Marsh, P. (1999). Transformations of snow chemistry in the boreal forest: accumulation and volatilization. *Hydrological Processes*, *13*(1415), 2257–2273.

Pomeroy, J. W., Davies, T. D., & Tranter, M. (1991). The Impact of Blowing Snow on Snow Chemistry. In *Seasonal snowpacks: Processes of compositional change* (pp. 71–113). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-75112-7_4

Pomeroy, J. W., Gray, D. M., Brown, T., Hedstrom, N. R., Quinton, W. L., Granger, R. J., & Carey, S. K. (2007). The cold regions hydrological model: A platform for basing process representation and model structure on physical evidence. *Hydrological Processes*, *21*(19), 2650–2667. doi: {10.1002/hyp.6787}

Refsgaard, J., & Storm, B. (1995). Mike she.[in:] singh vp (ed.), computer models of watershed hydrology. *Water Resources Publication, Colorado*, 809–847.

Richmond, B. (2003). *STELLA: An Introduction to Systems Thinking.*

Saloranta, T. M., & Andersen, T. (2007). Mylake—a multi-year lake simulation model code suitable for uncertainty and sensitivity analysis simulations. *Ecological Modelling*, *207*(1), 45-60. (Uncertainty in Ecological Models) doi: https://doi.org/10.1016/j.ecolmodel.2007.03.018

Sanderson, C., & Curtin, R. (2016). Armadillo: a template-based c++ library for linear algebra. *Journal of Open Source Software*, *1*(2), 26. doi: 10.21105/joss.00026

Sanderson, C., & Curtin, R. (2018). A user-friendly hybrid sparse matrix class in c++. , 422–430. doi: 10.1007/978-3-319-96418-8_50

Shoemaker, L. (1997). Compendium of tools for watershed assessment and tmdl development.

Shook, K., Pomeroy, J. W., Spence, C., & Boychuk, L. (2013). Storage dynamics simulations in prairie wetland hydrology models: evaluation and parameterization. *Hydrological Processes*, *27*(13), 1875–1889. doi: 10.1002/hyp.9867

Trefry, M. G., & Muffels, C. (2007). Feflow: A finite-element ground water flow and transport modeling tool. *Groundwater*, *45*(5), 525-528. doi: https://doi.org/10.1111/j.1745-6584.2007.00358.x

Wade, A., Jackson, B., & Butterfield, D. (2008). Over-parameterised, uncertain 'mathematical marionettes' — How can we best use catchment water quality models? An example of an 80-year catchment-scale nutrient balance. *Science of The Total Environment*, *400*(1-3), 52–74. doi: 10.1016/j.scitotenv.2008.04.030

Wade, A. J., Durand, P., Beaujouan, V., Wessel, W. W., Raat, K. J., Whitehead, P. G., ... Lepisto, A. (2002). A nitrogen model for european catchments: Inca, new model structure and equations. *Hydrology and Earth System Sciences*, *6*(3), 559–582. doi: 10.5194/hess-6-559-2002

Wellen, C., Kamran-Disfani, A. R., & Arhonditsis, G. B. (2015). Evaluation of the current state of distributed watershed nutrient water quality modeling. *Environmental Science and Technology*, *49*(6), 3278–3290. doi: 10.1021/es5049557

Whitehead, P., Wilson, E., Butterfield, D., & Seed, K. (1998). A semi-distributed integrated flow and nitrogen model for multiple source assessment in catchments (inca): Part ii — application to large river basins in south wales and eastern england. *Science of The Total Environment*, *210-211*, 559-583. doi: https://doi.org/10.1016/S0048-9697(98)00038-2

Wool, T., Ambrose, R. B., Martin, J. L., & Comer, A. (2020). Wasp 8: The next generation in the 50-year evolution of usepa's water quality model. *Water*, *12*(5). doi: 10.3390/w12051398

Yang, X., Tetzlaff, D., Soulsby, C., & Borchardt, D. (2022). Hiwaq v1.0: A flexible catchment water quality assessment tool with compatibility for multiple hydrological model structures. *Geoscientific Model Development Discussions*, *2022*, 1–26. doi: 10.5194/gmd-2022-239