Response to Reviewer 2 (gmd-2022-222)

In addition to (or as an emphasis of) the community comment by Jason Ke and the comments from Referee #1 I have the following suggestions for improving the paper.

Thanks for these comments. As per the other two reviewers, please find our point-by-point response below. We indicate areas where we intend to take action in purple. For the more substantial additions or changes we intend to make to the text, we have copied a draft of these changes into this document in *"italics".* Full details of these changes will be available in a manuscript with tracked changes that we will submit once the discussion period has ended.

- Model description paper in GMD should focus on a detailed description of the scientific basis of a model and the technical/ numerical implementation. In the current version of the manuscript, there is a certain imbalance between the respective contents in the main paper and the supplemental material. Most of the description of model details in the supplemental material should be directly mentioned in the main paper.

We agree with this and your subsequent comment that the balance of the manuscript can be improved, and that some more technical details should be added to the manuscript. We have summarised the information we intend to add to the manuscript below under three aspects: 1) surface water and pollutant routing; 2) pollutant loading calculations; and 3) other technical details regarding model code, data format and running time:

1. Surface water and pollutant routing

We previously referred only to the PCR-GLOBWB2 documentation (*Sutanudjaja et al., 2018)* for these descriptions to avoid too much overlap. However, we agree that specific information regarding the routing routine should also be included here, and as such we will add the following text to the manuscript:

"The routine for surface water and pollution routing follows an eight-point steepest-gradient algorithm across the terrain surface (local drainage direction) in a convergent drainage network with the lowermost cell connected to either the ocean or an endorheic basic (Sutanudjaja et al., 2018). Routing within DynQual uses the kinematic wave approximation of the Saint-Venant equations with flow described by Manning's equation, solved using a time-explicit variable sub-time stepping scheme based on the minimum Courant number (Sutanudjaja et al., 2018). In the coupled configuration, surface waters are subject to water withdrawals and return flows from the domestic, industrial, livestock and irrigation sectors calculated within the water use module of PCR-GLOBWB2."

2. Pollutant loadings

As per our response to Reviewer #1, it was a deliberate choice to keep the section on pollutant loading calculations relatively short in the manuscript given that pollutant loadings can alternatively be forced directly to DynQual. Nevertheless, we agreed with the reviewer that some pertinent information on the estimation of pollutant loadings should indeed be included in the manuscript. We will expand section 2.3 to summarise the key information with respect to pollutant loading estimates for each sector:

"Loadings from the domestic sector are estimated by multiplying gridded population numbers with region-specific per capita excretion rates (SI Section 1.1; Table S1). For the manufacturing sector, a mean effluent concentration is multiplied by location specific gridded

estimates of return flows from the manufacturing sector (SI Section 1.2; Table S2). Urban surface return flows are approximated by multiplying surface runoff (simulated by PCR-GLOBWB2) with the gridded urban fraction, which are multiplied by a region-specific mean urban surface runoff effluent concentration (SI Section 1.3; Table S3). The livestock sector is sub-divided into 'intensive' and 'extensive' production systems based on livestock densities to better account for differences in the paths by which waste enters the stream network (SI Section 1.4, Table S4). Gridded livestock numbers for buffalo, chickens, cows, ducks, goats, horses, pigs and sheep are multiplied by pollutant excretion rates per livestock type and by region (SI Section 4, Table S5 – S7). The livestock sector is sub-divided into 'intensive' and 'extensive' production systems based on livestock densities to better account for differences in the paths by which waste enters the stream network (SI Section 1.4; Table S4-S7). TDS loadings from the irrigation sector are estimated by multiplying irrigation return flows simulated by PCR-GLOBWB2 with spatially-explicit mean irrigation drainage concentrations based on salinity (as indicated by electrical conductivity) over the top- and sub-soil (SI Section 1.5). Thermal effluents (heat dumps) from thermoelectric powerplants are included as a point sources of advected heat by considering the temperature difference between the return flows and ambient surface water temperature conditions (SI Section 1.6). Pollutant loadings from the domestic, manufacturing and intensive livestock sectors, and from urban surface runoff, can be abated based on gridcell-specific wastewater practices. The proportion of pollutant loadings removed by wastewater treatment practices is estimated by multiplying the fraction of each treatment level occurring in a gridcell by the pollutant removal efficiency associated with that treatment level, as described in detail in previous work (Jones et al., 2021; Jones et al., 2022)."

As also per the request of Reviewer #1, we will also add a Table summarising the required input data for pollutant loading emissions within a DynQual run.

3. Other technical details

We will add some sentences to the manuscript to improve the description of the technical details:

"As per PCR-GLOBWB2 (Sutanudjaja et al., 2018) and DynWat (Wanders et al., 2019), DynQual is written in Python and is run using an initialization (.ini) file in which key aspects of the model run are defined (e.g. spatial extent, simulation period, paths to parameter and forcing files). Most input files required and all output are in NetCDF format. Global 5 arc-min DynQual runs that are coupled with PCR-GLOBWB2 have a wall-clock time of approximately 6 hours per year when run with parallelisation, due to the requirement to use the kinematic wave routing option for higher accuracy discharge and water temperature simulations. This is equivalent to the PCR-GLOBWB2 run times given by Sutanudjaja et al., (2018). DynQual runs performed in the stand-alone configuration are faster (~20%)."

- Another large part of model description papers should be dedicated to the model verification. This aspect is quite underrepresented in the current version of the manuscript. Model verification is only presented in line 304 to 321 and one figure in the main paper. Some additional results are included in the supplemental material but without a thorough presentation, analysis and discussion of the results. These aspects need to gain much more space in the main paper as this is a central part of model description papers in GMD.

We agree with you and Community Comment #1 that model evaluation is under-represented in the current submission. We also agree with the comment from Reviewer #1 that “validating the global model is not easy” – comparing individual (instantaneous) observed concentrations vs. simulated daily concentrations does indeed come with challenges at large scales. Data availability, both in terms of spatial coverage and the number of observations per water quality monitoring station, also presents significant challenges. Lack of global spatial coverage impacts validation efforts for all four water quality constituents, while data availability issues is particularly limiting for BOD and FC. This is somewhat of a “catch-22” for large-scale water quality modelling efforts – poor data availability across space and time is a key motivation for developing physically-based water quality models (to fill in these data gaps), but also is a severe limitation for both model development and evaluation.

We appreciate the praise from Reviewer #1 regarding our current efforts here, yet we also agree with Community Comment #1 and Reviewer #2 that this section can (and should) be further improved and expanded. To give the validation aspects more emphasis in the revised manuscript, we intend to assign “Model evaluation” into its own sub-section and expand on this section both with additional analysis (including a new figure) and discussion.

- Along with a more detailed model verification, the authors also need to emphasize the discussion on possible model limitations.

Agreed. As raised by Reviewer #1, we will add an additional paragraph with examples of specific uncertainties and model limitations of DynQual:

“Uncertainties in surface water quality simulations arise from a combination of uncertainties associated with quantifications of pollutant loadings (e.g. pollutant excretion, emission rates and sector-specific return flows), the quality of hydrological simulations (e.g. discharge and velocities) and the representation of in-stream processes (e.g. decay coefficients). These uncertainties are amplified when modelling at large spatial extents. In-stream pollutant concentrations are highly sensitive to dilution capacity, thus the quality of the river discharge simulations. This issue contributes to uncertainties in simulated concentrations particularly in headwater streams. Fixed estimates of decay coefficients have been assumed, which contributes to uncertainties in simulations of reactive constituents such as BOD and FC. In addition, the representation of lakes and reservoirs in DynQual is rudimentary, with total (routed) loadings instantaneously averaged over the volume of the water body assuming full mixing.”

With respect to pollutant loading quantifications, spatial mismatches between the generation of pollutant loadings and the location of entry to the stream network (return flows) can result in the simulation of unrealistic concentrations, particularly in gridcells with very low water availability (i.e. headwater streams). This can occur where the drivers of point-source pollutant emissions (e.g. population) do not directly coincide with the location of wastewater treatment plant outlets. A lack of temporally-explicit input data can hinder proper representation of sectors with strong inter-annual variability - a notable limitation for the livestock sector. Here, simplified assumptions are required for aspects such as livestock population numbers (assumed to be constant across days of the year), change to livestock numbers across multi-year periods (applied annually and based on regional averages) and transportation pathways to the stream network (assumed to be a function of surface runoff excluding the representation of processes that affect pollutant retention in soils). Locally relevant sources of pollution may also be entirely excluded, such as the lack of information on TDS emissions from mining activities and road-deicing.

We also intend to add a paragraph to the discussion which will highlight the key limitations and challenges of large-scale water quality modelling more generally, and further emphasise the types of research questions these approaches can help to address.

- A much larger portion of the manuscript (lines 323-491) is dedicated to the presentation of spatial patterns and trends of (long-year global) simulation results. Section 3 is named "model demonstration" and the model evaluation part (see above) is only a small part of this section. Showing possible applications of the model is, of course, interesting but model fidelity has do be demonstrated first and in much greater detail before spatial patterns and trends can be presented. Hence, there needs to be a better balance between the "model evaluation" and the "model demonstration" part.

In line with one of the previous comments, we propose to split the sub-section "Model run set-up and validation" into two separate sections: i.e. 3.1) Model run setup and 3.2) Model evaluation – and significantly expand on these sections in line with the recommendations.

We agree that, as a model description paper, the manuscript needs to be more balanced. We propose to remove results related to the average annual fluctuations (Figure 8 + associated description) and the sector-specific time series (Figure 11 + associated description) from the manuscript, instead keeping the focus on spatial patterns and trends.

Combined, we believe these efforts will bring better balance to the manuscript in line with requirements for GMD model description papers.

- The performance of DynQual should also be discussed in comparison to other water quality models, e.g. available catchment or regional scale models. The level of detail in terms of available input data and spatio-temporal resolution is, of course, different but the modelling community and potential users need to know what quality of output they could expect from this newly developed model compared to existing ones. Although the scope of the model might be a bit different, the interpretation of results needs to rely on the closeness to observation data (i.e. the prediction capability) and this one needs to be compared to already existing modelling approaches.

We agree that the performance of DynQual should be discussed in comparison to other large-scale water quality models. We included some comparisons already to the most comparable studies (e.g. to van Vliet et al, 2021, Wen et al., 2017 and UNEP 2016). However, we will (re-)review the published large-scale water quality literature to try and expand this discussion further.

We find difficulties into statistically comparing the performance of an uncalibrated global water quality model (with global parameterisation and input data sets) to watershed specific water quality models. These two types of models have fundamentally different purposes. Watershed-scale models can incorporate locally relevant input data and processes which are impossible to meaningfully represent in global approaches given data limitations. Watershed models are parameterized for specific local conditions and typically are calibrated based upon observation data of good quality and record. This issue is also addressed in our response to Community Comment #1 Q4. We will more specifically allude to the key differences between watershed vs. global surface water quality models in the manuscript:

"Comparatively, watershed-scale surface water quality models can better incorporate locally relevant input data and processes, can be parameterized specifically for local conditions and

typically have observation data of good quality and record length for calibration and validation. This allows for higher precision and accuracy in both hydrological and water quality simulations, particularly with regards to the magnitude and timing of high and low flows and concentrations, a primary aim of watershed-scale models. However, these watershed-scale models are reliant upon detailed local knowledge which is severely lacking for many (particularly ungauged) catchments worldwide (e.g. large parts of Africa). Despite their limitations, process-based large-scale water quality models can facilitate first-order assessments of global water quality dynamics that are consistent across both space and time, such as those demonstrated in the model application section of this study."

As also requested by Reviewer #1, we will also more specifically detail the usefulness of DynQual with respect to the types of scientific questions we can use it to address:

"The presented application of DynQual allows for the investigation of research questions that only large-scale modelling efforts can address, including: global hot- and bright- spot identification (Figures 4 - 6), the relative importance of different contributing sectors to water quality status across the globe (Figure 7) and meta-trends in surface water quality dynamics (Figures 8 – 10)."

- It is unclear, why the normalized RMSE was chosen as the sole performance measure for model evaluation. Other performance measures such as the Nash-Sutcliffe efficiency (NSE) or the Kling-Gupta efficiency (KGE) are much more common in model benchmarking and allow a clearer interpretation of the model performance, e.g. NSE lower or greater than zero. In addition, KGE even combines different aspects of model performance.

Our validation approach follows the standard practice that has been adopted in evaluating comparable large-scale water quality models in terms of both 1) modelling approach; and 2) purposes (e.g. Beusen et al., 2015; UNEP, 2016; van Vliet et al., 2021). It was for this reason that we chose to focus the evaluation of DynQual using metrics such as the normalized root mean squared error (nRMSE) and by evaluating the ability to simulate concentrations within a concentration range. As discussed above, we will revisit the water quality literature to improve our model evaluation section.

Please also note that no existing large-scale water quality model has used the NSE or KGE for model evaluation, aside from for water temperature (e.g. van Vliet et al., 2011). Please also see our response to Community Comment #1 Q3&10 with regards to the evaluation of large-scale water quality model output using NSE.

- It is mentioned that DynQual can be run in two modes: (1) coupled to PCR-GLOBWB2 and (2) in offline-mode with any other hydrological model. Please describe in more detail the technical aspects of the coupling between DynQual and PCR-GLOBWB2 as well as the (technical) requirements for using DynQual with other hydrological models, e.g. what are the required input data from the hydrological model and in which form they need to be provided to DynQual (e.g. netCDF files and their format).

We will add a sentence specifying the required input data to run the stand-alone configuration (please note this information is also contained in Figure 1):

"…either: 1) in a stand-alone configuration with specific discharge (i.e. baseflow, interflow and direct runoff in m day^{-1}) fed from any land surface or hydrological model."

To clarify input data formats, and also output formats, we will add a short statement to the manuscript:

"Most input files required and all output are in NetCDF format."

To elaborate on the coupling we will include a short sentence to describe this:

"In the coupled configuration, surface waters are subject to water withdrawals and return flows from the domestic, industrial, livestock and irrigation sectors calculated within the water use module of PCR-GLOBWB2."

- Time-stepping: Some lower bounds for the sub-daily time steps are mentioned (line 184ff). Does the model ensure that numerical stability criteria (Courant number, Peclet number) are met for the reactive transport equations?

Additional information will be added to the manuscript regarding the time-steps (and, more generally, the routine for surface water and pollutant routing):

"The surface water routing routine follows an eight-point steepest-gradient algorithm across the terrain surface (local drainage direction) in a convergent drainage network with the lowermost cell connected to either the ocean or an endorheic basic (Sutanudjaja et al., 2018). Routing within DynQual uses the kinematic wave approximation of the Saint-Venant equations with flow described by Manning's equation, solved using a time-explicit variable sub-time stepping scheme based on the minimum Courant number."

- Equation 9 for nRMSE: Please double-check this equation, the square is missing

We will correct this equation in the SI.