Response to Referee 1

Manuscript: egusphere-2026-3475

Title: A High-Resolution Water Resources Assessment Model (HyDRA version 1.0) for Industrial and Domestic Water Use: A regional application for the Gulf Cooperation Council (GCC) Countries and Jordan

Authors: Amr Al-Zu'bi and Yoshihide Wada

We sincerely thank the referee for the detailed and constructive review. The questions raised led to substantial changes and improvement: 1) a single-sector benchmark, 2) a decomposition of the industrial change, 3) quantitative evaluation statistics with the reference year excluded, 4) an independent evaluation of the TEF in 119 countries, 5) separate reporting of oil-extraction water, 6) a section on industrial consumption, and 7) a table of representative-year totals (Table S3). We also corrected the inconsistencies the referee found in the parameters, and removed or qualified the claims the results did not support.

Referee comments are shown in blue italics, followed by our response. Page numbers in the referee comments refer to the preprint; the page and line numbers given after each response refer to the revised manuscript and supplement. Section, figure, table and equation numbers refer to the revised manuscript unless marked "preprint". Two new figures were added (Fig. 6, production by subsector; Fig. 9, industrial consumption), so the domestic figures are now Figs. 10–13 (Figs. 8–11 of the preprint). The supplement has been reorganised to follow the order of the main text; supplementary sections, figures and tables are numbered as in the revised supplement.

General assessment

> *This paper builds a subsector-resolved model of industrial and domestic water use and runs it for the Gulf states and Jordan over 1990 to 2025. The region is a blind spot for the global models, and the authors have gathered more national data for it than I have seen in one place. So let me say up front that I think there is a useful model and dataset here. My problem is the distance between what the paper shows and what it says it shows. Several of the big claims are never tested, and parts of the accounting do not add up yet. I would recommend major revision. Most of what follows are questions.*

Response:

We thank the referee for this critical yet positive assessment. We agree that several claims went beyond what the preprint tested. In the revised manuscript, each claim is either supported by a test or removed. Our point-by-point responses follow.

Main comments

Comment R1.1 – Source of oil-extraction water

What water is being counted in the oil module? Section 2.1.8 says the external supply can be "seawater, freshwater, or desalinated water." Section 3.1.3 then calls the result "industrial freshwater withdrawal." Which is it, country by country? If the split isn't known, why show the total as freshwater at all? For a region that runs on desalination this may be the most consequential number in the industrial module.

Response:

We agree, and thank the referee for this correction. HyDRA does not resolve where the water injected into oil fields comes from, and the split by country is not known. In the study region this water is mainly treated seawater and saline groundwater, with smaller amounts of fresh and desalinated water, so presenting the total as freshwater was wrong.

The revised manuscript reports oil-extraction withdrawal separately from industrial withdrawal of fresh or desalinated water throughout. Section 2.1.8 states that the source is not resolved and that the total should not be read as freshwater. The Fig. 5 caption identifies oil extraction as water of unresolved source, and Table S3 lists it in its own column.

Location of changes. Revised manuscript: Sect. 2.1.8, p. 18, ll. 9–11; Fig. 5 caption, p. 30, ll. 20–25. Supplement: Table S3 caption, p. 12, ll. 10–17.

Comment R1.2 – Oil-extraction intensities and the definition of consumption

The oil intensities do not reproduce. Putting the page 17 parameters into Eqs. (11) and (12), I get roughly 1.84 and 0.69 m^3 per m^3; the paper reports 1.96 and 0.83. Which set went into the runs? And a question of convention while we are here: water that is injected and stays in the reservoir never returns to the sea or the aquifer it came from, so why isn't it consumed under your own definition?

Response:

The referee is correct. With the parameters of Table 4, the withdrawal and consumption equations (Eqs. 12 and 13; Eqs. 11 and 12 of the preprint) give 1.84 and 0.69 m^3 per m^3 of oil. The values of 1.96 and 0.83 $m^3\ m^{-3}$ in the preprint were inconsistent with these parameters. The text, Table 3 and the model inputs use 1.84 $m^3\ m^{-3}$ for withdrawal, and all results have been recomputed with it. Following the referee's second point, consumption now equals withdrawal (see below).

On the convention, we agree with the referee and have adopted the source-based definition. Water injected to replace the produced oil stays in the reservoir, and produced water that is not recycled is disposed of, mostly by reinjection into deep formations, so none of the external withdrawal returns to its source. Oil-extraction consumption therefore now equals withdrawal, 1.84 m^3 per m^3 of oil (Sect. 2.1.8 and Table 3). Section 2.1.8 keeps the surface losses of the produced-water stream (Eq. 13; 0.69 $m^3\ m^{-3}$) as a narrower alternative, and Sect. 3.1.4 states that using it would lower industrial consumption by about a third. With the source-based definition, oil extraction accounts for two thirds of industrial consumption in 1990 and for half in 2025.

Location of changes. Revised manuscript: Table 3 caption, p. 12, ll. 2–3; Sect. 2.1.8, p. 18, ll. 16–25; Sect. 2.1.8, p. 19, ll. 5–6; Sect. 3.1.4, p. 34, ll. 17–20.

Comment R1.3 – Non-revenue water as consumption; the UAE trajectory

Eq. (8) puts all of non-revenue water into consumption, while water that reaches urban users is consumed at 10 percent. You note yourselves that the Qatar NRW figures include apparent losses as well as real ones. What is in those apparent losses, and why would any of it be consumptive? Then the UAE: its NRW was estimated "assuming a trajectory similar to Qatar's improvements." If that decline is prescribed, how can the falling UAE consumption-to-withdrawal ratio be read as evidence of better network efficiency?

Response:

Apparent losses are water that reaches users but is not billed: metering errors, unauthorised or illegal use and data-handling errors. We agree that they are not consumptive in themselves. They are delivered water, of which only the fraction CF is consumed. Real losses are leakage from the network. Part of this water percolates to shallow urban groundwater, which in the study region is often saline or contaminated and is rarely recovered, and part evaporates. Counting real losses as consumed is therefore closer to reality there, although it is still an upper bound. The study countries do not report real and apparent losses separately in a consistent way, so we cannot split NRW.

Section 2.2 now states that counting all NRW as consumption (Eq. 17; Eq. 8 of the preprint) gives an upper bound, and that domestic consumption is therefore overestimated where reported NRW includes apparent losses, as in Qatar. The Discussion section discusses this limitation.

On the UAE, the referee is right. The NRW decline is prescribed from Qatar's trajectory, so the resulting fall in the UAE consumption-to-withdrawal ratio is an assumption, not evidence. We have removed the statement that it indicates better network efficiency. Section 2.2 now states that the change is an assumption, not a result. Section 3.2 attributes the decline in Qatar to the reported fall in NRW (from over 35 % to 15 %) and the decline in the UAE to the prescribed trajectory.

Location of changes. Revised manuscript: Sect. 2.2.1, p. 24, ll. 16–17; Sect. 2.2.1, p. 24, ll. 25–28; Sect. 3.2.2, p. 36, ll. 12–14; Sect. 4, p. 42, ll. 36–37.

Comment R1.4 – Benchmark against a simpler formulation

Better than what? The abstract says the model "advances beyond conventional top-down scaling approaches," and the Discussion ends by saying it can "substantially improve" on aggregated approaches. I looked for the conventional model that was run alongside and could not find one. Would a constant-intensity reconstruction, a single lumped industrial sector, or a plain GDP regression actually do worse against the same observations? Even a simple baseline would tell us what the extra complexity buys.

Response:

We agree that the claim needed a benchmark. Section 3.1.3 and Fig. 5 now compare HyDRA with a single-sector reconstruction driven by the same variables. It divides each country's 2020 HyDRA total by the national TEF of each year, i.e. it uses the variables of the TEF (GDP, electricity use and energy intensity) directly as drivers of total withdrawal, with one sector and no production structure. Because the TEF falls as GDP and electricity use grow, its inverse acts as a single activity index built from the same variables, in the way single-sector formulations scale industrial withdrawal with such indicators, so the comparison isolates what the production structure adds.

In 1990 the single-sector series is 57–132 % higher than HyDRA in Saudi Arabia, Kuwait, Oman, Bahrain and the UAE, where water-using industries expanded faster than these drivers, and 31 % lower in Jordan. We did not use the national records of total industrial withdrawal to rank the two reconstructions. They cover few years, contain the breaks discussed under R1.5, and AQUASTAT industrial withdrawal gets accounted in a different way, so a closer fit to these records would not show which structure is right. The model is tested against observations where the records are consistent: the TEF against the industrial water intensity of the screened countries (Supplementary Sect. S5) and against the intensity records of four subsectors (Sect. 3.1.2, Fig. 4). The benchmark answers the second part of the question, what the extra structure buys.

The decomposition described under R1.15 shows where this difference comes from. The production mix raised withdrawal in Oman, the UAE, Bahrain and Kuwait and lowered it in Saudi Arabia, Qatar and Jordan, which a single sector driven by the same variables cannot represent. The difference matters most for scenarios in which the industrial structure changes, such as the planned expansion of petrochemicals and refining in Saudi Arabia.

Location of changes. Revised manuscript: Sect. 3.1.3, p. 29, ll. 17–23; Sect. 3.1.3, p. 29, l. 30 – p. 30, l. 5; Fig. 5 caption, p. 30, ll. 20–25.

Comment R1.5 – Quantitative evaluation and what a failed test looks like

The evaluation is visual from start to finish. No bias, no RMSE, no slope, no R^2. Figure 3 is normalized to 2020, Figure 4 to a reference year the caption does not name, and the domestic series are pinned to the observed 2020 value, so agreement near that year is guaranteed. Why not report skill with the anchor year left out, with absolute values next to the normalized ones? And what would a failed test look like? Section 3.1.1 keeps explaining disagreement as reporting problems. That may well be true in places, but if agreement counts for the model and disagreement counts against the data, the model cannot lose.

Figure 3 is the clearest case. The paragraph beneath it opens with the comparison "highlights the model's ability to capture long-term efficiency improvements." Then come the panels: Saudi Arabia and Oman do not move the same way in the two series, and the UAE observations fall far more than the model does. Why open with the success statement? Say where it fits and where it doesn't, then draw whatever conclusion the panels support.

Response:

We agree with each point. The revised manuscript reports skill with the anchor year left out, and absolute values next to the normalized ones:

- All evaluation statistics exclude the reference year, so agreement imposed by the normalization or by the domestic anchor is not counted.
- Each panel of Fig. 3 reports Pearson r, the concordance correlation coefficient (CCC; Lin, 1989) and the median absolute percentage error (MdAPE).
- For the subsector test (Fig. 4), Sect. 3.1.2 reports the median bias (−3 %) and the MdAPE for each subsector: 9 % for petro-refining, 14 % for paper and pulp, 17 % for basic metals and 31 % for food and beverage. Each panel of Fig. 4 now shows the MdAPE, the median bias and the share of points within ±20 %.

- For the domestic sector, Sect. 3.2.1 reports skill excluding 2020. At national level (94 country-years), MdAPE is 7 %, r = 0.91 and the median bias is −5 %. For the 13 Saudi regions (156 region-years), MdAPE is 15 %, r = 0.80 and the median bias is +2 %.
- Each caption states its reference year: 2020 for Fig. 3, and for Fig. 4 a series-specific year between 2003 and 2020.
- Table S3 gives absolute values for 1990, 2010 and 2025.

We also agree that a test the model cannot fail is not a test. We define a failed test as a country with consistent records whose observed intensity moves against the TEF. Section 3.1.1 now opens with where the TEF agrees with the observations: Jordan, Kuwait, Bahrain and the UAE (r = 0.71–0.96). It then states where it does not. In Saudi Arabia and Oman the reported intensity rises until about 2007 and 2016, respectively, and in Qatar it changes abruptly, whereas the TEF declines steadily in all three. Possible causes are discussed only after the disagreement has been stated.

So that the data are not judged by the model, breaks in the reported records are now identified from the observations alone. The screen flags a series whose value departs more than five-fold from its 2020 value, or changes more than two-fold between consecutive values. We applied it to the 119 countries with at least 15 years of records (Supplementary Sect. S5). It flags 37 countries, including Qatar and the UAE. In the remaining 82, the TEF reproduces the direction of change in 77 % of countries and predicts a significant decline in 50 of the 53 countries where one is observed. It also outperforms a no-change benchmark in 65 % of country-years (Supplementary Sect. S5).

On the paragraph below Fig. 3: the opening success statement has been removed. The paragraph now states where the panels agree and where they do not, and draws its conclusion from that.

Location of changes. Revised manuscript: Sect. 3.1.1, p. 25, l. 1; Fig. 3 caption, p. 25, ll. 6–10; Sect. 3.1.1, p. 25, l. 11 – p. 26, l. 12; Sect. 3.1.2, p. 26, ll. 22–23; Fig. 4 caption, p. 27, ll. 4–7; Sect. 3.1.2, p. 27, ll. 8–14; Sect. 3.2.1, p. 35, ll. 8–17. Supplement: Sect. S5, p. 9, ll. 6–9.

Comment R1.6 – Functional form of the TEF

> *TEF carries most of the industrial time series, so its form matters. Why average total GDP with two per capita quantities, in countries whose populations grew several times over? Why equal weights, and why 0.5 on the GDP term? These are choices, but the text presents them as properties of the system. It also says all three terms are "designed to exhibit a saturating, plateau-like response," when only the GDP term is transformed. And one thing I kept wondering: when GDP contracts, does the model read that as intensity going up?*

Response:

We agree that these are choices. Section 2.1 now presents them as such: the functional form, the exponent and the weights are modelling assumptions, not fitted relationships, and they are evaluated in Sects. 3.1.1 and 3.1.2.

On mixing total and per-capita quantities, only one of the three terms is per capita. The GDP term uses total GDP as a measure of the scale of the economy, so it does respond to population growth. The exponent of 0.5 expresses the assumption that efficiency gains grow less than proportionally with economic scale. Energy intensity is total energy use per unit of GDP, so population cancels (Eq. 4). Only the electricity term is per capita; it represents the development of electricity-dependent activities. The terms are averaged with equal weights because we had no basis for weighting one above another. We

have deleted the sentence stating that all three terms are designed to saturate, since only the GDP term is transformed.

On a GDP contraction: yes. Each term is a ratio to 2020, so a fall in GDP raises the GDP term, and with it the TEF and the modelled intensity. A 10 % fall in GDP raises the GDP term by 5.4 % and, at unchanged energy use, the energy-intensity term by 11 %, so that the TEF rises by between 1.8 % (energy use falling in proportion to GDP) and 5.5 % (energy use unchanged).

Location of changes. Revised manuscript: Sect. 2.1, p. 10, ll. 16–22.

Comment R1.7 – Which subsectors use the TEF

> *Which subsectors use TEF? Eq. (2) gives it a subsector index; Eq. (3) does not. Page 10 says it applies "for all the subsectors," page 18 says it "was not considered" for oil, and the power-plant equation has no TEF in it. If it is a single national multiplier, what is validated in a section called "Validation of Sub-Sector Efficiency Changes"? Subsector-specific baseline intensities and subsector-specific trends are different things, and only the first seems to exist.*

Response:

The TEF is a single national multiplier, and we have removed the subsector index from Eq. (2). It is applied identically to every subsector in a country except oil extraction and electricity generation, which use fixed coefficients. For oil extraction, gains in drilling and recovery are offset by the rising water cut of maturing fields, with no clear net trend (Sect. 2.1.8). For electricity generation, water intensity depends on the fuel type and cooling technology of each plant, which the model represents directly; changes over time arise from changes in the plant fleet and in total generation (Sect. 2.1.1). The text that said the TEF applies to all subsectors has been corrected.

The referee is right that the model has subsector-specific baseline intensities but not subsector-specific trends, and Sect. 2.1 now says so. Section 3.1.2 tests whether this single national trend describes the temporal change of intensity within individual subsectors, using the subsectors and countries for which records exist. We have retitled the section "Validation of Temporal Intensity Changes at the Subsector Level", and it now states this purpose at the start.

Location of changes. Revised manuscript: Sect. 2.1, p. 10, ll. 11–12; Sect. 2.1, p. 10, ll. 25–26; Sect. 2.1.1, p. 14, ll. 7–9; Sect. 2.1.8, p. 19, ll. 24–25; Sect. 3.1.2 title, p. 26, l. 14; Sect. 3.1.2, p. 26, ll. 19–21.

Comment R1.8 – Coverage of the subsector test

> *Figure 4 tests four subsectors on countries outside the region. Fair enough, that is where the data are. But how do the ranges of the GDP, electricity and energy-intensity ratios in those countries compare with what the model meets in the Gulf? If they hardly overlap, the test checks interpolation in one set of economies while the model extrapolates in another. And why extend the conclusion to subsectors as a whole when four of ten were tested, and two of the untested ones don't use TEF anyway?*

Response:

We agree that this needs to be addressed. The subsector test examines how the water intensity of a sub-sector develops over time in countries at different stages of economic development. Technological development and industrial practices in these sub-sectors are largely shared regionally and globally (Keller, 2004; Comin and Hobijn, 2010), because international companies supply, license and operate the

process technology and set common standards; the first petrochemical plants in Saudi Arabia, for example, were joint ventures in which the foreign partners supplied the technology (OTA, 1984). Local conditions such as regulation and water pricing also play a role (Reynaud, 2003), but the evolution of water intensity is partly shaped by this shared development, which the test examines across a wide range of economies. Section 3.1.2 and Supplementary Sect. S6 now state this.

Most of the test countries are at an economic level similar to the GCC countries, and the subsector test performs best for these countries: the median absolute error is 18 %, against 31 % at lower income levels, and the median bias is −1 %, against −10 % (Supplementary Sect. S6). This demonstrates the relevance of the test to the economies to which the model is applied. Linking sub-sectoral efficiency changes to GDP provides a means of representing technological and economic development over time and follows previous global water-use modelling studies (Alcamo et al., 2003; Flörke et al., 2013; Wada et al., 2016) and the observed dependence of water use on income (Duarte et al., 2013).

The comparison in the new Figs. S6 and S7 shows that the test also covers the range in which the TEF is applied: since 2000, the TEF of the study countries lies within the range covered by the test in 94 % of country-years. However, the relative contributions of the three terms differ between the two groups. GDP is the dominant contributor to TEF variation in the study countries, whereas energy intensity dominates in the test countries. This reflects their different energy-intensity trajectories, which declined at almost all test points but increased in part of the Gulf country-years. Because no subsector intensity records exist for the study countries, applying the TEF there assumes that the relation found in the test countries also holds under these conditions. We state this assumption and the condition of increasing energy intensity as a limitation in Sect. 3.1.2 and Supplementary Sect. S6, and note that subsector intensity records for the region would allow it to be tested.

Sect. 3.1.2 now draws the conclusion only for the four tested subsectors, which account for 58–68 % of the TEF-adjusted withdrawal in the study countries. Sect. 2.1 states that oil extraction and electricity generation do not use the TEF.

Location of changes. Revised manuscript: Sect. 3.1.2, p. 27, l. 15 – p. 28, l. 8. Supplement: Sect. S6, p. 11, l. 9 – p. 12, l. 4.

Comment R1.9 – Choice of 2020 as reference year

> *Why 2020? It was a pandemic year, and everything is scaled from it. Do the trajectories change with 2019, or an average of a few years, as the reference? If not, say so. If they do, the reference year is a real uncertainty and should be treated as one.*

Response:

We chose 2020 because it was the most recent year with complete inputs for all countries when the model was set up (Sect. 2.1). We tested 2019 as the reference year, as the referee suggests. Because the TEF terms are ratios to the reference year, using 2019 rescales each country's industrial series by about 1/TEF(2019), between −1.9 % (Jordan) and +2.7 % (Bahrain), without changing its shape. For the domestic sector, anchoring on the observed 2019 value shifts each national series by the 2019 model–observation difference, between −7.2 % (Bahrain) and +3.2 % (UAE), again without changing its shape. In the 119-country evaluation, using 2019 or 2010 as the normalisation year changes the median per-country correlation by less than 0.02 (Supplementary Sect. S5).

The trajectories therefore do not depend on the pandemic year, and Sect. 2.1 now says so. The 2020 fall in GDP enters through the inputs, while household size, connection rates and the other structural inputs change slowly and are not affected. We kept a single reference year, rather than a multi-year average, so that the baseline intensities and the reference domestic withdrawal refer to the same year and the same set of inputs.

Location of changes. Revised manuscript: Sect. 2.1, p. 10, ll. 7–10. Supplement: Sect. S5, p. 9, ll. 13–14.

Comment R1.10 – Parameter uncertainty

> *The uncertainty in the inputs is far wider than anything in the figures. Table 3 has intensity ranges spanning a factor of three or more, Table 4 gives WOR as 0 to 20+, and single values carry through regardless. What produces the "Predicted Range" in Figure 3 and the "median ± standard deviation" in Figure 9? Parameters, input datasets, something else? What I would want to know is whether the country rankings and the trends survive the ranges you report yourselves.*

Response:

The ranges in Fig. 3 and Fig. 11 (Fig. 9 of the preprint) are the spread across the input datasets of Table 2, i.e. the alternative sources of GDP, population, energy and household size. They do not include parameter uncertainty, and the captions now say so. Table 3 gives the range reported in the literature together with the single value used in the model, marked with an asterisk, and Sect. 2.1.8 gives the values used for the reservoir parameters of Table 4.

To answer the referee's question, we set all subsector intensities to the lower ends of the ranges in Table 3, and then to the upper ends. This scales the 2020 average regional industrial withdrawal by −40 % and +69 %. The 2020 ranking of the seven countries is unchanged, and their 1990–2025 growth changes by at most 11 %. Varying one intensity at a time changes the ranking only when the oil-extraction intensity is set to the upper end of its range, which moves Kuwait above the UAE. For oil extraction, Sect. 2.1.8 now gives the sensitivity to Bo, Bw and WOR: a WOR between 0.01 and 4 changes regional industrial withdrawal excluding electricity by about ±10 %. The Discussion states that rankings and trends are robust to the ranges in Table 3, while absolute withdrawals are not.

Location of changes. Revised manuscript: Sect. 2.1.8, p. 19, ll. 12–17; Fig. 3 caption, p. 25, ll. 6–10; Fig. 11 caption, p. 37, ll. 5–7; Sect. 4, p. 42, ll. 20–25.

Comment R1.11 – Scope promised in the Introduction

> *The Introduction faults aggregated models for masking "water policy, such as water tariffs," and promises a model that accounts for "urban form, housing structure." Eqs. (5) to (8) contain GDP per capita, household size, connection rate, population, NRW and an urban/rural factor. Where are housing and urban form? Why make tariffs the headline omission of other models and then leave them out? And what in this driver set could produce the declines in Bahrain and the UAE seen in Figure 9a? Either the model needs those variables or the Introduction needs to promise less.*

Response:

We agree that the Introduction promised more than the model can offer, and we have revised it to reflect what has been presented in the revised manuscript. "Urban form" and "housing structure" have been removed. The Introduction now names the infrastructure variables the model actually uses: household size, household connection to the water network and non-revenue water. Tariffs are no

longer presented as an omission of other models; they appear only in the review of studies that used them as a driver. Section 3.2 and the Discussion state that the model does not represent the effect of tariffs or other demand-side measures.

On the declines in Fig. 11a (Fig. 9a of the preprint): the modelled decline in the UAE between 1990 and 2010, from 354 to 286 L cap^{-1} d^{-1}, follows from the fall in real GDP per capita over that period, when population grew faster than GDP. The observed decline in Bahrain between 2000 and 2010 coincides with no comparable change in GDP per capita or household size, and the model does not reproduce it. Section 3.2 now says so and names possible causes outside the model: tariff reform, network rehabilitation and changes in reporting.

Location of changes. Revised manuscript: Sect. 1, p. 5, ll. 5–10; Sect. 3.2.2, p. 36, ll. 6–11; Sect. 4, p. 42, ll. 17–20.

Comment R1.12 – The sigmoid fit

> *On the sigmoid in Figure 2: five countries, post-2010 data only, bounds fixed at 50 and 500, and the upper end appears to rest largely on Qatar. What is the fitted equation, how many points, how good is the fit, and were the years from one country treated as independent? The reason given for dropping earlier data is that it would "obscure the GDP–DWW relationship." Why is that a reason to exclude it rather than a sign that GDP alone doesn't explain domestic use? And which countries are actually estimated from this curve?*

Response:

Section 2.2 now gives the fit in full. The curve is $DWWpc = 50 + 450 / \{1 + \exp[-k\,(GDPpc/10^5 - x_o)]\}$, with $k = 2.07$ and $x_o = 0.65$ (about 65 000 US$ per capita). It was fitted by least squares to 57 country-years from Bahrain, Jordan, Qatar, Saudi Arabia and the UAE over 2010–2022 ($R^2 = 0.81$, RMSE = 37 L cap^{-1} d^{-1}). Years from the same country were treated as independent, which overstates the information content of the fit, and the upper part of the curve rests mainly on Qatar. The text now states both points.

The curve is used for only two countries, Kuwait and Oman, which have no observed domestic withdrawal. It sets only their 2020 reference value, and their GDP per capita lies within the fitted range. Their change over time then follows Eq. (14), as for the other countries.

On the excluded years, we agree with the referee's reading. Before 2010, per-capita withdrawal changed independently of GDP per capita in some countries, for example the decline in Bahrain between 2000 and 2010. This shows that GDP per capita alone does not explain domestic withdrawal. Section 2.2 now says so, and gives this as the reason the curve is used only for the reference value. We also repeated the fit with withdrawal normalized by household size, using the same household-size scaling as the model (Fig. S3; $R^2 = 0.80$). The resulting 2020 reference values differ from those used in HyDRA by +3 % for Kuwait and −10 % for Oman, whose households are larger than those of the sample (Table S2).

Location of changes. Revised manuscript: Sect. 2.2.1, p. 22, ll. 5–22; Fig. 2 caption, p. 23, ll. 3–5. Supplement: Sect. S4, p. 5, l. 14 – p. 6, l. 5; Fig. S3 caption, p. 6, ll. 7–11; Table S2 caption, p. 6, ll. 12–18.

Comment R1.13 – The 1991 Kuwait War

The text says the model "captures" the impact of the 1991 Kuwait War. Figure 9a has almost nothing for Kuwait, and Kuwait is absent from Figure 8a. Captured against what? If there is an observation that shows it, show it. Otherwise this is the GDP input doing what the GDP input does.

Response:

We agree. No domestic observations exist for Kuwait, so the 1991 drop cannot be validated. It is the response of the income term to the 46 % fall in real GDP per capita during the Gulf War. We have replaced "captures the impact of the Kuwait War" with this statement (Sect. 3.2), and deleted the similar claim about Qatar before the 2022 FIFA World Cup. Kuwait is absent from Fig. 10a (Fig. 8a of the preprint) because there are no observations to compare with, and the caption now says so.

Location of changes. Revised manuscript: Fig. 10 caption, p. 36, ll. 1–2; Sect. 3.2.2, p. 36, ll. 6–11.

Comment R1.14 – Meaning of "high resolution"

"High-resolution" needs pinning down. The 90 m domestic fields are Eq. (7), per capita withdrawal times gridded population; industrial use is allocated to plants or spread across industrial areas by their size. So what in the model is actually known at 3 arcseconds? "Allocated to a 3-arcsecond grid" would describe it accurately. Grid spacing, information resolution and validation scale are three different things here, and the paper treats them as one.

Response:

We agree and have adopted the referee's wording. The Abstract and Sect. 2 now describe the outputs as allocated to a 3-arcsecond (~90 m) grid, and Sect. 2 states what information each module uses. Industrial withdrawal is computed nationally by subsector and assigned to plant locations in proportion to capacity, or to industrial and mining areas in proportion to their extent. Domestic withdrawal is computed per capita for each country and for each of the 13 Saudi regions, and distributed with gridded population. Section 2 also states that the model is evaluated at national and regional scales only. A new paragraph in the Discussion explains what the grid is used for and what it should not be used for, such as utility-scale planning.

Location of changes. Revised manuscript: Abstract, p. 1, ll. 19–22; Sect. 2, p. 8, ll. 4–9; Sect. 4, p. 43, ll. 1–11.

Comment R1.15 – Results that restate the model; decomposition

Several results restate the model. "Fluctuations in withdrawals closely follow economic growth and changes in household size" (p. 30) is Eq. (5) in words. Elsewhere, changes are credited to diversification policy, modernization and infrastructure development, none of which the model contains. Why not use the structure you already have and split the industrial change into production growth, composition shifts and intensity change? That would be a result nobody could get by reading the equations, and I suspect it would be the best thing in the paper.

Response:

We have followed this suggestion, and agree that it gives the most informative result in the paper. Section 3.1.3 now splits the change in industrial withdrawal between 1990 and 2025 into economic growth, production mix and water intensity (Ang, 2005). For the region, economic growth alone would

have more than tripled withdrawal excluding electricity, and declining intensity offset about a fifth of that increase.

The production-mix effect has opposite signs in different countries. It raised withdrawal in Oman, the UAE, Bahrain and Kuwait, where water-using industries grew faster than the economy. It lowered withdrawal in Saudi Arabia, Qatar and Jordan, where the economy grew faster than its water-using industries. It therefore nearly cancels for the region but not for individual countries. A new Fig. 6 shows the production changes behind these effects: each subsector is scaled to its own 2020 output, and no water data are used. Oman is the clearest case. Its industrial withdrawal grew more than threefold although its oil output changed little, because steel, petrochemicals, food and beverage, mining and aluminium were built after 2005.

For the domestic sector, Sect. 3.2 and Fig. 12 split the change in the same way. Population growth accounts for almost 90 % of the increase in regional withdrawal between 1990 and 2025, and the decline in household size for most of the rest. Income has little net effect for the region but shapes the differences between countries.

We have deleted the sentence on p. 30 that restated Eq. (5) of the preprint. We have also deleted the statements that attributed changes to diversification policy, modernisation or infrastructure development. The model takes production as an input. Where the growth of a subsector coincides with a national strategy, as in Bahrain and the UAE, the text now notes the timing without attributing the change to the strategy.

Location of changes. Revised manuscript: Sect. 3.1.3, p. 28, ll. 20–26; Sect. 3.1.3, p. 29, ll. 17–29; Fig. 6 caption, p. 31, ll. 2–8; Sect. 3.2.2, p. 36, ll. 24–30.

Comment R1.16 – National totals, sector shares and consumption

> *There are very few numbers. The figures carry the trajectories, but I could find almost no national totals or sector shares stated in the text or in a table, and industrial consumption gets far less attention than withdrawal. One compact table with representative-year totals, sector shares and the fresh/saline split would make the reconstruction much easier to judge and to reuse.*

Response:

We agree. The new Table S3 gives the following for each country and for the region in 1990, 2010 and 2025:

- industrial withdrawal of fresh or desalinated water;
- withdrawal for oil extraction (source not resolved);
- industrial consumption;
- power-plant withdrawal of freshwater and of seawater;
- seawater used to cool refineries, petrochemical plants and metal smelters on the coast;
- domestic withdrawal, and the domestic share of freshwater withdrawal.

Section 3.1.3 now gives totals and subsector shares in the text and refers to the table. For example, regional industrial withdrawal excluding electricity almost tripled between 1990 and 2025, from 1.4 to 3.8 $km^3\ yr^{-1}$, with Saudi Arabia accounting for about half of it throughout. The share of oil extraction fell from half to under a third. The industrial sector as a whole withdrew about 24 times as much seawater as freshwater in 2025.

Industrial consumption now has its own section (Sect. 3.1.4) and figure (Fig. 9). Consumption excluding electricity rose from 1.04 to 2.27 $km^3\ yr^{-1}$ between 1990 and 2025. Each subsector consumes a fixed fraction of what it withdraws: about 16–17 % for steel, aluminium and paper and pulp, 58–65 % for cement, refining and potash, and all of it for oil extraction (R1.2); other manufacturing consumes 42 % of its withdrawal (Sect. 2.1.11). The share consumed therefore follows the subsector composition of withdrawal. It fell from 0.75 to 0.60 for the region, from 0.89 to 0.53 in Oman, where steel, petrochemicals and food and beverage reduced the share of oil extraction, and from 0.55 to 0.37 in Bahrain, where metals and food and beverage replaced refining. It rose from 0.47 to 0.54 in Qatar, where the share of steel declined. Power plants withdrew about 30 km^3 of seawater in 2025 but consume less than 1 % of it, and the seawater used to cool coastal refineries, petrochemical plants and smelters is assumed to be returned to the sea. The consumption series of each subsector are in Fig. S8.

Location of changes. Revised manuscript: Sect. 3.1.3, p. 28, ll. 11–17; Sect. 3.1.4, p. 33, ll. 7–15; Sect. 3.1.4, p. 34, ll. 1–21; Fig. 9 caption, p. 34, ll. 23–26. Supplement: Sect. S7, p. 12, ll. 6–9; Table S3 caption, p. 12, ll. 10–17; Fig. S8 caption, p. 13, ll. 3–6.

Comment R1.17 – Food and beverage intensity

> *The food and beverage intensity is given as "1 – 17 m^3/$ value added." Taken literally, even the low end gives absurd withdrawals for a modest sector. Is a scaling factor missing, or is the denominator in thousands of dollars? And why electricity per capita as the predictor for the countries without data? How well does that relationship actually work?*

Response:

The referee is right about the unit. The denominator is thousand US$ of value added, the unit of GLORIA, so the range is 0.7–17 m^3 per 1000 US$ (excluding Croatia). In the study countries the intensities are 2.9 (Jordan) to 13.3 (Bahrain) m^3 per 1000 US$. The text, Table 3 and Fig. S2 have been corrected.

Food and beverage is the most data-scarce subsector, and the relationship is used only for countries without Eurostat data. Electricity per capita, used as an indicator of technological development, is the predictor. Supplementary Sect. S3 now gives the fitted relationship, WWI = 3.91 E^0.194 − 14.0 (20 countries, 2010–2020, excluding Croatia; $R^2 = 0.48$), and lists the other predictors tested (value added per capita, the share of electricity in energy use, logistic and multilinear forms, and the shares of food, beverage and dairy production). Section 2.1.6 states that the relationship is used only for countries with missing values, and Supplementary Sect. S3 reports its fit statistics. Fig. S2 has been redrawn with the axis in m^3 per 1000 US$.

Location of changes. Revised manuscript: Table 3, pp. 12–13; Sect. 2.1.6, p. 16, ll. 29–31. Supplement: Fig. S2 caption, p. 5, ll. 4–5; Sect. S3, p. 5, ll. 6–12.

Comment R1.18 – Table 3 as the single source of truth

Table 3 should be the single source of truth, and right now it isn't. The refining text adopts a value that does not sit inside the range shown for the crude class it names, and the basic-metals section gives a mean and a median without saying which is used. Could one table list, per subsector, the value used, its unit, the period it represents, the source and where it applies? Given these mismatches, I would also want confirmation that the archived code reproduces the figures from the equations as written.

Response:

We agree. Table 3 now gives, for each subsector, the range reported in the literature, the value used in HyDRA (marked with an asterisk), the unit and the source. Where a value applies only to some countries or conditions, the table says so: the crude class for refining, the pulp share for paper and pulp, and the Dead Sea case for potash. We corrected the inconsistencies the referee found:

- The refining values (1.5, 0.96 and 40 m^3 per m^3 of oil) belong to the light-to-medium crude class, which dominates in the region. The text now names this class, and the value lies within its range (Sect. 2.1.2).
- For basic metals, the text now states that the median (24 $m^3\ t^{-1}$) is used.
- We also corrected the cement consumption value (0.163 $m^3\ t^{-1}$), the food-and-beverage unit (R1.17) and the oil-extraction values (R1.2).

In the code archive, the intensities of each subsector are set in a single input table, so the values in Table 3 are the values the model uses. We have confirmed that the archived code reproduces the figures from the equations and parameter values given in Sect. 2 and Table 3.

Location of changes. Revised manuscript: Sect. 2.1, p. 10, l. 34 – p. 11, l. 1; Table 3, pp. 12–13; Sect. 2.1.2, p. 14, ll. 28–30; Sect. 2.1.5, p. 16, l. 7.

Comment R1.19 – Novelty relative to the closest studies

Where is the novelty relative to the nearest work? Table 1 itself lists Fujimori et al. (2017) with ten industrial subsectors, and two high-resolution industrial datasets for China. What does HyDRA do that they don't? Is it the pairing of industrial detail with domestic infrastructure variables, the regional harmonization, the spatial allocation? Say it in the Introduction rather than leaving the reader to work it out from the table.

Response:

We agree that the Introduction should state this directly. Fujimori et al. (2017) resolve ten industrial subsectors, but in a CGE-based integrated assessment model with 17 world regions. In that model the GCC countries and Jordan are not represented individually, and domestic use is not modelled. Hou et al. (2024) map industrial withdrawal in China at high resolution, and Jiao et al. (2024) model the industrial water use of China under technology-specific transformation pathways. None of these represents domestic use, and none has been applied outside China.

HyDRA's contribution is therefore not subsector detail or fine spatial allocation on its own. It is the combination of the two with a domestic module driven by infrastructure variables (household connection and non-revenue water), in one consistent framework applied country by country to a

region that global models aggregate. The Introduction now states this contribution relative to these studies.

Location of changes. Revised manuscript: Sect. 1, p. 4, ll. 14–20.

Smaller points

> *Figure 6 is never referred to in the text; why include it? The Figure 9 caption says "industrial" for a domestic figure, and the y-axes of Figures 5 and 6 say m³ where the values must be millions. Equation numbers repeat. Figure 7 has a placeholder date in the caption and no key for circle size; can the Google imagery be reused under the journal's license? Page 8 says "water use" will stand for both withdrawal and consumption, a strange choice when telling them apart is the point. How sensitive are the industrial totals to the 5 percent "other manufacturing" term? What did the Google Earth Engine "verification" of cooling types consist of? The Introduction ends with a list of advantages rather than a question, the Discussion's limitations mention none of the issues above, and there is no Conclusions section. And why "Novel" in the title?*

Response:

We thank the referee for these points. They have been addressed as follows.

- **Figure 6 of the preprint (electricity):** this is now Fig. 7, and Sect. 3.1.3 refers to it. Its y-axis, like that of Fig. 5, is in $Mm^3\ yr^{-1}$. (Sect. 3.1.3, p. 30, ll. 11–14; Fig. 7 caption, p. 32, ll. 2–3)
- **Figure 9 caption of the preprint:** this is now Fig. 11, and its caption says "domestic". (Fig. 11 caption, p. 37, ll. 2–3)
- **Equation numbers:** the equations are now numbered consecutively, 1 to 17.
- **Figure 7 of the preprint (map):** this is now Fig. 8. The placeholder date has been removed. The Google basemap has been replaced with the public-domain Natural Earth basemap, and a key for circle size has been added. (Fig. 8 caption, p. 33, ll. 4–5)
- **"Water use":** the convention has been dropped. Withdrawal and consumption are reported separately throughout, and "water use" is kept only for statements that apply to both (Sect. 2). (Sect. 2, p. 8, ll. 10–11)
- **Other manufacturing:** the term is a fixed proportion of the other subsectors, so halving or doubling the 5 % share changes industrial withdrawal excluding electricity by −2.4 % or +4.8 % in every country and year. Trends and rankings do not change (Sect. 2.1.11). (Sect. 2.1.11, p. 21, ll. 9–14)
- **Google Earth Engine check:** Sect. 2.1.1 now describes it. The cooling type reported by Byers et al. (2018) was checked plant by plant on high-resolution imagery, using visible features such as cooling towers, seawater intake and outfall structures, or air-cooled condensers. (Sect. 2.1.1, p. 14, ll. 13–15)
- **End of the Introduction:** the Introduction now ends with the question the paper addresses. (Sect. 1, p. 5, ll. 7–10)
- **Limitations:** the Discussion now covers the issues raised in this review. These are the unresolved source of oil-extraction water, the upper bound on domestic consumption from NRW, the single national TEF and its evaluation, demand-side measures that the TEF does not represent, input-data versus parameter uncertainty, and what the 3-arcsecond grid can and cannot be used for. (Sect. 4, p. 41, l. 15 – p. 42, l. 37)

- **Conclusions:** the section is now titled “Discussion and conclusions” and ends with the conclusions of the study. (Sect. 4, p. 43, ll. 22–28)
- **“Novel”:** removed from the title.

References cited in the responses and not listed in the manuscript

Duarte, R., Pinilla, V., and Serrano, A.: Is there an environmental Kuznets curve for water use? A panel smooth transition regression approach, Econ. Model., 31, 518–527, https://doi.org/10.1016/j.econmod.2012.12.010, 2013.