



Sensitivity Analysis of the Groundwater Risk Index in the Middle East and North Africa Region

Adam Milewski¹  · Khalil Lezzaik^{1,2}  · Rachel Rotz¹ 

Received: 25 September 2019 / Accepted: 30 December 2019 / Published online: 8 January 2019
© Springer Nature Switzerland AG 2020

Abstract

Composite indicators are popular tools for assessing and comparing multidimensional phenomena in different countries. This paper tests the methodology of the Groundwater Risk Index (GRI), a water vulnerability composite index developed to evaluate groundwater depletion in the Middle East and North Africa (MENA) region (Lezzaik et al. 2018). A sensitivity analysis using a one-factor-at-a-time (OFAT) approach is used to determine the impact of alternate methodological choices on country-level GRI scores and ranks. The analysis focuses on the GRI sensitivity to: (1) the selection of constituent indicators; (2) the choice of normalization scheme; and (3) the choice of aggregation method. Results show that the GRI scores are not impacted by the selection of indicators and the choice of alternative normalization schemes. Conversely, the GRI was sensitive to arithmetic and multiplicative aggregation methods. High-income, oil-rich gulf countries exhibited decreases in rankings due to an imbalance between water resource allocations and adaptive capacity parameters, whereas countries with balanced conditions exhibited increases in the GRI rankings. The sensitivity analysis provides useful insights into the development of GRI for end-users to assess the effects of groundwater vulnerability against potential changes or stressors in semi-arid to hyper-arid regions.

Keywords Composite indicators · Robustness testing · Sensitivity analysis · Groundwater risk

Highlights

- A sensitivity analysis on the Groundwater Risk Index (GRI) is presented
 - Outputs highlight different perspectives on groundwater depletion risk in the MENA region
 - The structural flexibility of the GRI allows for its use in regions with a high dependency on groundwater resources
-

✉ Adam Milewski
milewski@uga.edu

Khalil Lezzaik
khalil.lezzaik@owp.csus.edu

Rachel Rotz
rrotz@uga.edu

Extended author information available on the last page of the article

1 Introduction

Composite indicators, also referred to as indices, have expanded in the past decade due to their increased use by governments, think tanks, national, and international organizations in assessing multidimensional phenomena such as human development, governance, and environmental degradation (Foa and Tanner 2012; Saisana and Tarantola 2002). Water resources trends in the Middle East and North Africa (MENA) region indicate increasing vulnerability due to socio-economic factors (Droogers et al. 2012; Sowers et al. 2011). In this paper, the Groundwater Risk Index (GRI), a composite index developed to evaluate groundwater depletion in the MENA region (Lezzaik et al. 2018) is considered as a case study (Fig. 1). The conceptual framework of the GRI assumes that groundwater risk, defined as the probability of an entity experiencing groundwater depletion, cannot be solely assessed by models of hydrological systems, mechanics, and limitations. Instead, the GRI assesses groundwater risk by combining hydrogeological assessments of groundwater reserves and storage changes, with governance, food security, and energy costs into a composite index.

Composite indices are central tools in policymaking given their ability to comparatively analyze and benchmark country performance on elusive issues (Cherchye et al. 2008). A survey of inter-country composite indices by Bandura (2008) found that of the approximately 180 indices developed, 50% were in the previous five years alone. Composite indicators are calculated by aggregating indicators into a single index based on a conceptual framework dictated by what is being measured, under what conditions, and for what intended purpose. Constructing an index and computing its scores requires normalizing, weighting, and aggregating constituent indicators into a final index using an aggregative function.

The creation and review of composite indices for groundwater and other environmental factors has been on the rise in the recent literature. Pires et al. (2017) performed a thorough review of over

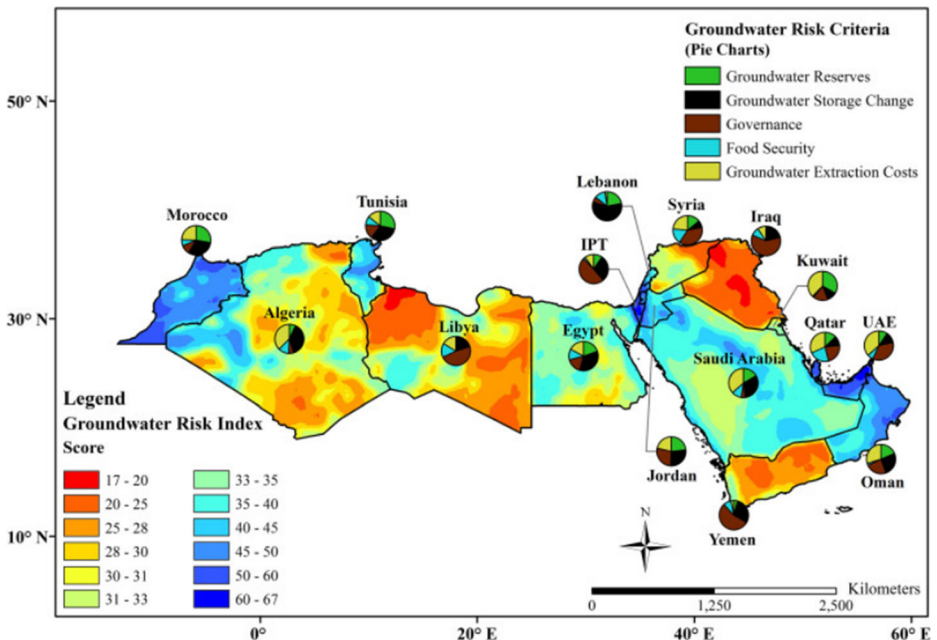


Fig. 1 The GRI assessment for the MENA region between 2003 and 2014 by Lezzaik et al. (2018)

170 current sustainability indicators to assess their quality and reliability. Indicators were reviewed according to the DPSIR (Driving Forces, Pressures, States, Impacts, and Responses) framework and concluded that 24 indicators met their objective criteria. The GRI, the authors suggest, would satisfy at least three of the five driving indicators deeming it a useful indicator. Jeet et al. (2019) developed a composite hydrologic index for a semi-arid region in India using principal component analysis and sub-basins for their weighting schemes. Hosseini et al. (2019) developed an integrated environmentally sustainable groundwater management index (ESGMI) based on weighted aggregation of thirteen adopted indicators. Lastly, Azizi et al. (2019) developed their own coastal vulnerability index using a weighting scheme based on expert knowledge of influential parameters. Additional index development similar to this study can be found in the literature (Cameron and White 2004; Gleeson et al. 2012; Godfrey et al. 2002; Lewis et al. 2012; Mattas et al. 2014; Preda et al. 2013, among others).

Composite indicators are contentious with its supporters and critics (Saisana and Tarantola 2002). In addition to summarizing complex multidimensional phenomena, composite indices provide a single score of what is measured, which enables ease of interpretation vis-à-vis using multiple benchmarks (Foa and Tanner 2012). Consequently, indices support decision makers and facilitate communication with the general public (López-Claros 2010). However, critics argue that overreliance on composite indices amplifies the risk of extrapolating erroneous policy recommendations based on simplistic big-picture summarizations of complex phenomena (Saisana et al. 2005). Critics primarily cite the underlying developmental process of composite indices as the fundamental reason behind their unreliability as policy-informing tools (Cherchye et al. 2008; López-Claros 2010; Saisana and Tarantola 2002).

Despite some similarities with mathematical or computational models that have universally accepted scientific rules, composite indices rely on subjective judgments (Cherchye et al. 2008). Key decisions, informed by subjective judgments, are made at the different stages of index development, ranging from the selection of constituent indicators to the selection of a normalization method, weighting scheme, and aggregation approach for the composite index. Due to their increased influence, composite indicator creators are being asked to address the aforementioned critiques. Indeed, according to Saisana and Tarantola (2002), sensitivity analyses are included in only a few composite index studies. This paper analyses the robustness of the GRI by assessing the sensitivity of its country-level scores and its ranking to the selection of constituent indicators, normalization, and aggregation methods.

2 Review of the GRI Development

2.1 Theoretical Framework and Indicator Selection

Groundwater risk extends beyond hydrogeologic models of groundwater systems to include societal adaptive capacity criteria. The following section summarizes the rationale and methodology of the GRI development as in Lezzaik et al. (2018). The GRI was developed to help countries assess the hotspots of groundwater depletion risk in arid environments. Moreover, it enables countries to compare and benchmark groundwater risk and its determinants by allowing them to assess their position relative to others. The GRI is a national or regional screening tool that provides easily interpretable assessments based on a single composite measure. The GRI results can then be followed-up by an individual country-level analysis of the factors driving groundwater risk. The design of the index reflects two key concerns. The

first is the determination of the dominant factors underlying groundwater risk (is it physical groundwater endowments, politico-economic adaptive capacity, or both?). The second is the comparison of groundwater risk levels across all countries, and the possible identification and extrapolation of policy recommendations, especially to poorly ranked, high-risk countries. To accomplish this, the GRI focuses on five dimensions:

- a) *Groundwater Storage Reserves*: Also known as groundwater supply availability, groundwater reserves constitute the principle and objective natural framework on which groundwater risk can be conceptualized and measured. An accurate quantitative assessment of groundwater reserves was achieved by integrating distributed saturated aquifer thickness estimates with gridded effective porosity values (Lezzaik and Milewski 2018).
- b) *Groundwater Storage Changes*: Due to human consumption, groundwater is overburdened with 30% of the world's largest aquifers currently overstressed and undergoing little to no natural replenishment (Richey et al. 2015a; Richey et al. 2015b). Groundwater storage change estimates, which inherently included the indirect effects of climate change over the study period, were calculated by disaggregating GRACE-derived terrestrial water storage data, using GLDAS-generated land surface parameters (Lezzaik and Milewski 2018; Voss et al. 2013).
- c) *Governance*: Decentralized pumping, caused by poor governance, is one of the primary drivers of groundwater depletion. Srinivasan et al. (2012) argue that poor governance translates into a lack of water reallocation mechanisms and an ineffective control over water rights and flows, which consequently drives decentralized pumping by rural and urban dwellers. The World Bank's Worldwide Governance indicators (Kaufmann et al. 2011) were used to generate aggregate governance scores based on five different dimensions (e.g., the rule of law, government effectiveness).
- d) *Food Security*: The globalization and internationalization of trade results in reduced demands and pressures on local, national, and regional groundwater resources. The food security indicator is a proxy measure of countries' capacities to rely on exogenous virtual water trade to meet their population's caloric requirements. This would incidentally include the role of human dependence on groundwater for domestic, agricultural and industrial use in each country. The three pillars of food security (i.e., affordability, availability, and dietary diversity) were calculated using data from the World Bank (2016), FAO (2016), and independent studies, followed by their integration into an overall measure of country-level food security.
- e) *Groundwater Extraction Cost*: Within the framework of the energy-water nexus, energy costs influence groundwater extraction rates and affect associated groundwater risk. Numerous studies have established a negative correlation between increasing energy prices and groundwater extraction rates (Pfeiffer and Lin 2014; Zhu et al. 2007). The indicator measures groundwater extraction costs as a function of country-level diesel energy prices provided by (World Bank 2016) and groundwater table depths modeled by Fan et al. (2013).

2.2 Normalization

The GRI is composed of indicators with different data classifications and measurement units. Therefore, aggregating indicators into a composite index requires normalizing the original data

using a linear transformation method that re-expresses the original value for each indicator on a unitless scale from 0 to 100, with the following formula:

$$N_{q,c} = \frac{x_{q,c} - \min(x_q)}{\text{range}(x_q)} \times 100 \quad (1)$$

where $N_{q,c}$ denotes the normalized value of the indicator q for country c , and $x_{q,c}$ denotes the raw value of the indicator q for country c .

2.3 Weighting

Assigning the relative importance to indicators that determine the phenomenon being measured is a difficult and subjective design decision undertaken in composing composite indices. The GRI attaches equal weights to each dimension of groundwater risk. The equal weighting scheme is indicated for its simplicity in building the model and usefulness for the exchange of relevant variables (OECD/European Union/JRC 2008).

Alternative weighting indices for sustainability related composite indices are increasing in use but face criticism regarding bias depending on the quantity, quality, and the collection methods of the data (Nardo et al. 2005; Sharpe and Andrews 2012). Additional details regarding the GRI equal weighting scheme is discussed within the methodology section of this paper and in detail by Lezzaik et al. (2018).

2.4 Aggregation

An additive arithmetic mean model was selected to aggregate the indicators into a final index:

$$CI_c = \sum_{i=1}^n w_q N_{q,c} \quad (2)$$

where CI_c denotes the composite score of country c , $N_{q,c}$ denotes the normalized value of indicator q for country c , and Wq denotes the weight of the indicator q .

The aggregation method was selected based on the GRI's theoretical framework that allows for compensability and trade-offs between the constituent indicators (Aguna and Kovacevic 2010). Moreover, for the purpose of facilitating ease of interpretation, additive aggregation methods are superior to multiplicative methods, insofar as rendering the GRI effective and practical for use by different end-users.

3 The GRI Sensitivity Analysis

The GRI was developed as a distributed composite index to assess and evaluate groundwater depletion risk by combining different environmental and socioeconomic datasets and models. The GRI was designed as a multi-criteria diagnostic tool to identify and determine the severity and probability of an area experiencing the adverse effects of groundwater changes. In the preceding section, we briefly outlined the original GRI model (GRI_{original}), as one that calculated groundwater risk from a set of indicators that were rescaled using a min-max normalization, assigned an equal weighting scheme, and aggregated into a final composite index score using a simple additive arithmetic mean

function. This section focuses on describing and discussing the steps undertaken to assess the robustness of the design behind GRI_{original}, by testing the sensitivity of GRI output values to the index design decision-making and considerations. The GRI design choices, including adopting an equal weighting scheme and a linear additive aggregation approach, promote structural flexibility that enables the modification, application, and implementation of the index in other semi- to hyper-arid regions with a high level of dependency on groundwater resources.

First, we identified and assessed the sources of sensitivity. Generally, when developing a composite index, sensitivities arise from some or all of the following steps involving subjective judgment and decision-making for the design process:

- i. indicator selection
- ii. data selection and editing
- iii. data normalization
- iv. indicator weighting scheme
- v. index aggregation function

In this work, we focus on three sources of sensitivity within the GRI framework: (1) indicator selection; (2) normalization scheme; and (3) aggregation scheme. The indicator weighting scheme was not selected because an equal weighting scheme was deemed the most appropriate based on the lack of access to subjective weightings and the conceptual impediments to multivariate analysis. As noted in Lezzaik and Milewski (2018), determining the relative importance and the proper balancing of the plurality of perspectives through the weighting of different indicators is the most contentious problem in building composite indices. Conventional participatory schemes, in which weights are assigned on the basis of expert consultation, are criticized for their bias and subjectivity or motivated to reflect the pre-existing beliefs of the stakeholders (Booyesen 2002; Nardo et al. 2005). The Analytic Hierarchy Process (AHP) is one such participatory method where results depend on evaluator selection and the experimental setting (Saaty 1987). The AHP and other participatory methods may not be ideal in cases where indicators span a disparate set of disciplines, a lack of resources, or a lack of consensus exists on alternative solutions (Sharpe and Andrews 2012).

In regard to weighting approaches based on statistical models, complex multivariate statistical analyses sacrifice the functionality of composite indices by imposing a conceptual rigidity on the selection and weighting of indicators and by inhibiting the ease of the index results interpretation (Cox et al. 1992; Ginsburg et al. 1986). Beyond the limitations of normative and statistical weighting methods, the GRI was limited by a lack of resources to access societal and expert viewpoints, necessary to assign weights using participatory approaches such as the AHP.

Hagerty and Land (2007) recommended adopting an equal weighting scheme for composite indices that do not qualify for subjective or statistical weighting schemes. While an equal weighting scheme is the norm for many composite indices (OECD/European Union/JRC 2008), as well as the one chosen for the GRI, the authors acknowledge that any weighting scheme is a value judgment. In the case of the equal weighting scheme, the fidelity and correlation of the variable data for each indicator was neither rewarded nor punished. As a result, the sensitivity analysis of the indicator weighting scheme is considered beyond the scope of this paper.

3.1 Indicator Selection

Composite indices are developed, and their indicators selected, based on the underlying theoretical framework and empirical observations of the phenomenon being assessed. Consequently, the GRI indicator choices are debatable and present a source of sensitivity to the underlying phenomenon being assessed.

To assess the sensitivity associated due to the choice of indicators, the authors exercised the inclusion/exclusion of individual indicators. The process involves a one-at-a-time exclusion of an individual indicator, followed by an execution of the composite index, and an examination of the differences in the resultant index scores between the original (baseline) and modified GRI scores for each country:

$$\Delta score_c = score_{original, c} - score_{excl, q, c} \quad (3)$$

where $\Delta score_c$ denotes the change in score for country c , $score_{original, c}$ denotes the original GRI score for country c , and $score_{excl, q, c}$ denotes the modified GRI score for country c after the exclusion of indicator q .

Additionally, shifts in country ranks will be explored, following the exclusion of individual indicators:

$$\Delta rank_c = rank_{original, c} - rank_{excl, q, c} \quad (4)$$

where $\Delta rank_c$ denotes the change in rank for country c , $rank_{original, c}$ denotes the original GRI rank for country c , and $score_{excl, q, c}$ denotes the modified GRI rank for country c after the exclusion of indicator q .

The investigation of $\Delta score_c$ and $\Delta rank_c$ will be the scope of the sensitivity analysis targeting GRI robustness against its indicator selection.

3.2 Normalization

In $GRI_{original}$, a min-max re-scaling (Eq. 1) was adopted as the method of normalization to transform indicator datasets into a common scale. The choice was based on multiple considerations. First, as a linear transformation, min-max rescaling preserves the data in the data original values (OECD/European Union/JRC 2008). Second, the ease of communicating index data and outputs lying within an identical and bounded range [0,100] facilitates GRI's functionality by enabling easy interpretation of index results. However, since data normalization involves the transformation of data, the use of different normalization schemes could influence index outputs. Therefore, an examination of the sensitivity posed by the choice of normalization schemes is necessary.

In addition to min-max rescaling, there are two main normalization methods (Mazziotta 2013): standardization (Z-score) and indicization (distance to a reference). The standardization method converts the indicators to a common scale of mean zero and a standard deviation of one. Consequently, the Z-score method rewards exceptionally higher than average scores:

$$S_{q,c} = \frac{x_{q,c} - \text{mean}(x_q)}{\text{std}(x_q)} \quad (5)$$

where $S_{q,c}$ denotes the standardized value of the indicator q for country c , and $x_{q,c}$ denotes the raw value of the indicator q for country c .

Indicization takes the ratio of an indicator for a specific country $x_{q,c}$ with respect to a reference country. In our case, the highest scoring country will be taken as a reference, such that an indicator score for each country will be divided by the indicator score of the best ranking country. The ‘distance-to-largest-value’ (DLV) is calculated using the following formula:

$$DLV_{q,c} = \frac{x_{q,c}}{x_{q,ref}} \quad (6)$$

where $DLV_{q,c}$ denotes the normalized value of the indicator x_q for country c , $x_{q,c}$ denotes the raw value of the indicator x_q for country c , and $x_{q,ref}$ denotes the raw value of the indicator x_q for the highest ranking reference country.

GRI was executed separately with different normalization schemes to assess normalization-related sensitivities in terms of shifts in country rank (Eq. 7):

$$\Delta rank_c = rank_{min-max,c} - rank_{z-score/DLV,c} \quad (7)$$

where $\Delta rank_c$ denotes a rank change in country c , $rank_{min-max,c}$ denotes the original GRI rank for country c using a min-max rescaling scheme, $rank_{z-score/DLV,c}$ denotes the modified GRI rank for country c using alternative normalization schemes.

3.3 Aggregation

The two main aggregation methods are additive arithmetic mean and multiplicative geometric mean. The methodological choice between the two models rests upon an index theoretical framing of how indicator performances are rewarded and punished (Nardo et al. 2005). The GRI aggregated its indicators using an additive arithmetic mean model, based on the assumption of perfect substitutability and the superiority of additive models insofar their effectiveness in facilitating easy interpretation and adoption by experts and the public alike.

A geometric mean aggregation, however, reflects and awards trade-offs between different indicators by allowing for partial or imperfect substitutability, where a negative performance by one indicator cannot fully compensate for a positive performance by another. Moreover, geometric aggregation rewards balanced indicator scores and severely penalizes poor performance in one or more indicators (Aguna and Kovacevic 2010).

To account for the sensitivity associated with aggregation methods, the authors executed GRI using a geometric mean formula:

$$CI_c = \prod_{i=1}^n [q_c]^{1/n} \quad (8)$$

where CI_c denotes the composite score of country c , and q_c denotes indicator q score for country c .

Sensitivities inherent to the choice of aggregation function were evaluated through shifts in country-level groundwater risk score and rank.

4 Results and Discussion

Sensitivity analyses are rarely reported in composite index studies. Consequently, the index robustness is questioned, and adoption by end-users is compromised. In this study, sensitivity

Table 1 Sensitivity exclusion analysis of indicators defined as groundwater reserves (GWR), groundwater storage change (GWSC), groundwater extraction cost (GWEC), groundwater storage change (GWSC), governance (GOV), food security (FS), and groundwater extraction cost (GWEC). Countries exhibiting an upward movement in rank are highlighted in green, and countries exhibiting a downward movement in rank are highlighted in red

Country	GRI _{original}		GRI _{excl. GWR}		GRI _{excl. GWSC}		GRI _{excl. GOV}		GRI _{excl. FS}		GRI _{excl. GWEC}	
	Scores	Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Israel and PT*	68.1	1	72.2	2 (-1)	73.7	1	60.8	1	67.2	1	67.2	1
Qatar	66	2	77	1 (+1)	70.2	2	58.3	2	58.8	2	66	2
UAE**	56.4	3	55.1	4 (-1)	59.4	4 (-1)	48.2	4 (-1)	58.7	3	62.8	3
Kuwait	55.8	4	58.8	3 (+1)	62.6	3 (+1)	51.8	3 (+1)	53.1	4	57.7	4
Morocco	47.3	5	52.5	5	46.5	5	47.1	6 (-1)	48.7	5	42	10 (-5)
Lebanon	46.5	6	45.5	9 (-3)	44.3	6	48.2	5 (+1)	46.2	6	45.4	7 (-1)
Jordan	44.8	7	50.2	6 (+1)	43	7	39.3	7	46	7	46	6 (+1)
Oman	44.2	8	49	7 (+1)	42.2	8	36.3	10 (-2)	44.1	8	49.3	5 (+3)
Tunisia	42.5	9	48.1	8 (+1)	40.7	9	39.3	8 (+1)	41.8	9	42.9	9
Saudi Arabia	36.9	10	40.8	11 (-1)	32.4	10	31.9	15 (-5)	34.8	10	44.7	8 (+2)
Egypt	36.2	11	40.8	10 (+1)	31.5	11	35.4	12 (-1)	34.1	11	38.6	11
Algeria	33.5	12	34.6	12	28.6	12	34.7	13 (-1)	32.4	12	36.7	12
Syria	32.6	13	31	13	27.2	13	36.6	9 (+4)	30.9	14 (-1)	36.6	13
Libya	29.8	14	30.8	14	23.6	14	35.6	11 (+3)	25.4	15 (-1)	33.6	14
Yemen	28.9	15	30.6	15	21.7	15	32.2	14 (+1)	31.7	13 (+2)	27.7	15
Iraq	25.6	16	26.7	16	18.6	16	30.5	16	24.7	16	27.6	16

*PT refers to the Palestinian Territories

**UAE refers to the United Arab Emirates

Table 2 Sensitivity analysis results for the choice of the normalization scheme. Countries exhibiting an upward movement in rank are highlighted in green and show the resultant final rank and number of rank changes in parentheses. Countries exhibiting a downward movement in the ranking are highlighted in red

Country	Minimum - Maximum Normalization (GRI _{original})			Z-Score Standardization			Indicization (divide by largest)		
	Score	Ranking	Quantile Rank	Score	Ranking	Quantile Rank	Score	Ranking	Quantile Rank
Israel and PT	68.1	1	75th or higher	2.4	1	75th or higher	47.4	1	75th or higher
Qatar	66	2	75th or higher	2.3	2	75th or higher	41.6	2	75th or higher
UAE	56.4	3	75th or higher	1.7	3	75th or higher	40	3	75th or higher
Kuwait	55.8	4	75th or higher	1.4	4	75th or higher	20.9	4	75th or higher
Oman	44.2	8	50th to 74th	0.7	5 (+3)	50th to 74th	20.3	5 (+3)	50th to 74th
Jordan	44.8	7	50th to 74th	0.6	6 (+1)	50th to 74th	20.2	6 (+1)	50th to 74th
Morocco	47.3	5	50th to 74th	0.4	7 (-2)	50th to 74th	16.1	7 (-2)	50th to 74th
Lebanon	46.5	6	50th to 74th	0.1	9 (-3)	25th to 49th	6.9	10 (-4)	25th to 49th
Tunisia	42.5	9	25th to 49th	0.3	8 (+1)	50th to 74th	13.1	8 (+1)	50th to 74th
Saudi Arabia	36.9	10	25th to 49th	0	10	25th to 49th	9	9	25th to 49th
Egypt	36.2	11	25th to 49th	-0.3	11	25th to 49th	1.7	11	25th to 49th
Algeria	33.5	12	25th to 49th	-0.6	12	25th to 49th	-1.5	12	25th to 49th
Syria	32.6	13	24th or lower	-0.8	13	24th or lower	-8.7	14 (-1)	24th or lower
Yemen	28.9	15	24th or lower	-1	14 (+1)	24th or lower	-7	13 (+2)	24th or lower
Libya	29.8	14	24th or lower	-1.1	15 (-1)	24th or lower	-12.4	15 (-1)	24th or lower
Iraq	25.6	16	24th or lower	-1.3	16	24th or lower	-16.3	16	24th or lower

analyses at different developmental stages were conducted to test GRI robustness. The paper adopts a one-factor-at-a-time (OFAT) sensitivity approach to examine the effects of the indicator selection, normalization schemes, and aggregation methods on groundwater risk scores and ranks. Sensitivity results associated with indicator choice, normalization schemes, and aggregation methods are presented in Tables 1, 2 and 3, respectively.

4.1 Indicator Sensitivity Analysis

The sensitivity analysis of GRI indicator selection on its score and rank outputs shows that the original selection of constituent indicators provides a robust measure that is not biased from potential participatory methods. This is evident by country ranks generated by different indicator combinations (Table 1). Countries within the 75th percentile or higher group with very low groundwater risk (Israel and PT, Qatar, UAE, and Kuwait), were not sensitive to any individual indicator: they consistently ranked similarly within their quantile group with the exclusion of each individual indicator. The same applies to the 25th or lower quantile group with Syria, Yemen, Libya, and Iraq slightly shifting positions within their quantile group with different indicator combinations. Examining the GRI sensitivity to each indicator confirms the aforementioned observations, with relatively small shifts in country rank, ranging between one and two positions, and little to no change in the overall ranking structure of the 16 countries. The only exception is the governance indicator, whose exclusion led to larger shifts in the country ranking than its counterparts (e.g., Saudi Arabia, Syria, and Libya). On the basis of average ranking shifts across countries per indicator exclusion, the GRI sensitivity relative to its constituent indicators is in the following decreasing order: governance (GOV) ($\Delta\overline{\text{Rank}} = 1.37$), groundwater extraction cost (GWEC) ($\Delta\overline{\text{Rank}} = 0.75$), groundwater reserves (GWR) ($\Delta\overline{\text{Rank}} = 0.75$), food security (FS) ($\Delta\overline{\text{Rank}} = 0.25$), and groundwater storage change (GWSC) ($\Delta\overline{\text{Rank}} = 0.12$). The average rank changes suggest that the relative, normalized values of each

Table 3 Sensitivity analysis results for the choice of aggregation method. Countries exhibiting an upward movement in rank are highlighted in green and show the resultant final rank and number of rank changes in parentheses. Countries exhibiting a downward movement in the ranking are highlighted in red

Country	Weighted Arithmetic Mean (GRI _{original})			Weighted Geometric Mean		
	Score	Rank	Quantile Rank	Score	Rank	Quantile Rank
Israel and PT	68.1	1	75th or higher	65.2	1	75th or higher
UAE	56.4	3	75th or higher	44.3	2 (-1)	75th or higher
Qatar	66	2	75th or higher	33.2	7 (+5)	50th to 74th
Lebanon	46.5	6	50th to 74th	43.7	3 (-3)	75th or higher
Kuwait	56	4	75th or higher	32.4	9 (+5)	25th to 49th
Morocco	47.3	5	50th to 74th	40.7	4 (-1)	75th or higher
Jordan	44.8	7	50th to 74th	36.5	5 (-2)	50th to 74th
Tunisia	42.5	9	25th to 49th	34.9	6 (-3)	50th to 74th
Oman	44.2	8	50th to 74th	33.2	8	50th to 74th
Egypt	36.2	11	25th to 49th	29.4	10 (-1)	25th to 49th
Saudi Arabia	36.9	10	25th to 49th	22.5	13 (+3)	24th or lower
Algeria	33.5	12	25th to 49th	24.9	12	25th to 49th
Syria	32.6	13	24th or lower	25.3	11 (-2)	25th to 49th
Yemen	28.9	15	24th or lower	22.5	14 (-1)	24th or lower
Libya	29.8	14	24th or lower	17.6	15 (+1)	24th or lower
Iraq	25.6	16	24th or lower	17.1	16	24th or lower

indicator do not affect the overall country rankings similar to Villholth (2013), who developed a groundwater drought risk index in sub-Saharan Africa using both natural and human indicators with a similar approach, indicators and weighting scheme.

4.2 Normalization Sensitivity Analysis

To test for the robustness of minimum-maximum rescaling schemes, GRI was executed using alternative normalization methods (i.e., Z-score standardization, Indicization methods), and alternate GRI outcomes were compared to the original results. Comparable country rank outcomes generated by the different normalization methods confirm GRI independence to normalization scheme selection and affirm the robustness of the min-max rescaling method (Table 2). Countries are more robust to the choice of normalization scheme, if they are among the best (75th or higher quantile) or worst (25th or lower quantile) performers, and relatively more sensitive to normalization schemes within the 25th to 75th quantiles. For instance, Lebanon and Morocco undergo three and two downward shifts in rank with Z-score standardization scheme, and four and two downward shifts in rank with an indicization scheme, respectively. One explanation behind the higher sensitivities exhibited in the central quantiles could be the smaller differences in country scores within these quantiles as opposed to the score of the highest and lowest performing countries. Nevertheless, average ranking shifts—relative to the original ranking with min-max rescaling scheme caused by Z-Score standardization (= 0.75) and *DLV* Indicization (= 0.93), clearly indicate the GRI overall independence and insensitivity to the choice of normalization scheme.

4.3 Aggregation Sensitivity Analysis

Finally, a sensitivity analysis on the choice of aggregation formula was conducted by executing the GRI using a geometric mean aggregation method and comparing its outcomes to those produced by the original GRI with its arithmetic mean formula. Unlike indicator selection and choice of normalization scheme, the GRI outcomes were found to be sensitive to the selection of the aggregation formula. This is not surprising given how both methods represent and simulate indicator substitutability, score balance, and performance differently (Aguna and Kovacevic 2010). Out of the 16 countries considered in the GRI, 12 experienced a shift in rank. Moreover, seven countries underwent a shift outside their quantile group into another, thus reflecting not only minor shifts in rank but an overall reconfiguration of groundwater risk ranking amongst the MENA countries. To demonstrate, Qatar and Kuwait ranked highly in the 75th or higher quantile with low groundwater risk in the original GRI. When using the geometric aggregation method, both countries experienced five downward movements in rank into lower quantile groups (Table 3). Meanwhile, Lebanon and Morocco experienced rises in rank position to three and one in the highest quartile. This significant restructuring in rank order relates to how arithmetic and geometric mean aggregations reflect tradeoffs. For example, when using the arithmetic mean aggregation, Qatar and Kuwait scored highly, despite their low scorings on the groundwater reserve and storage change indicators. This is explained by the compensability of very high scores on governance and food security that offset poor performance in groundwater reserves and storage changes. A similar outcome was realized in Jemmali and Sullivan (2014) through a sensitivity analysis for the Water Poverty Index where a lack of institutional capacity in economically poor but water rich countries increases their vulnerability risk. On the other hand, when geometric mean aggregation was

implemented, Qatar's and Kuwait's scores declined from 66 and 56 to 33 and 32, respectively, which resulted in a downward movement in rank. This is explained by the mathematical base of the geometric mean that rewards balance and penalizes differences between indicator values. Alternatively, Lebanon and Morocco experienced upward movement in rank with the implementation of a geometric mean due to their balanced indicator performance and the lack of a distinctive poor performance in one or more indicators. A general observation in Fig. 2 is that countries with well-distributed indicator performance reflected in relatively lower in-group standard deviation values, such as Lebanon, Tunisia and Morocco, which experienced upward rank movements. Inversely, countries displaying poor performance in one or more indicators with higher in-group standard deviation values are penalized, and experience decreases in rank, such as Qatar, Kuwait and Saudi Arabia.

Indeed, countries that exhibited the most downward movement in rank with geometric aggregation are high-income, oil-rich gulf countries (e.g., Saudi Arabia, Kuwait, and Qatar). Under conditions of full compensability simulated by arithmetic aggregation, these countries scored well due to the effect of wealth and governance on offsetting poor water endowments. But under conditions of partial compensability, simulated by geometric mean aggregation, oil-rich gulf countries are penalized for the imbalance between water resource allocations and adaptive capacity parameters. Similarly, countries with more balanced conditions exhibited rises in rank, as was the case with Lebanon and Tunisia. Developing a sound and coherent

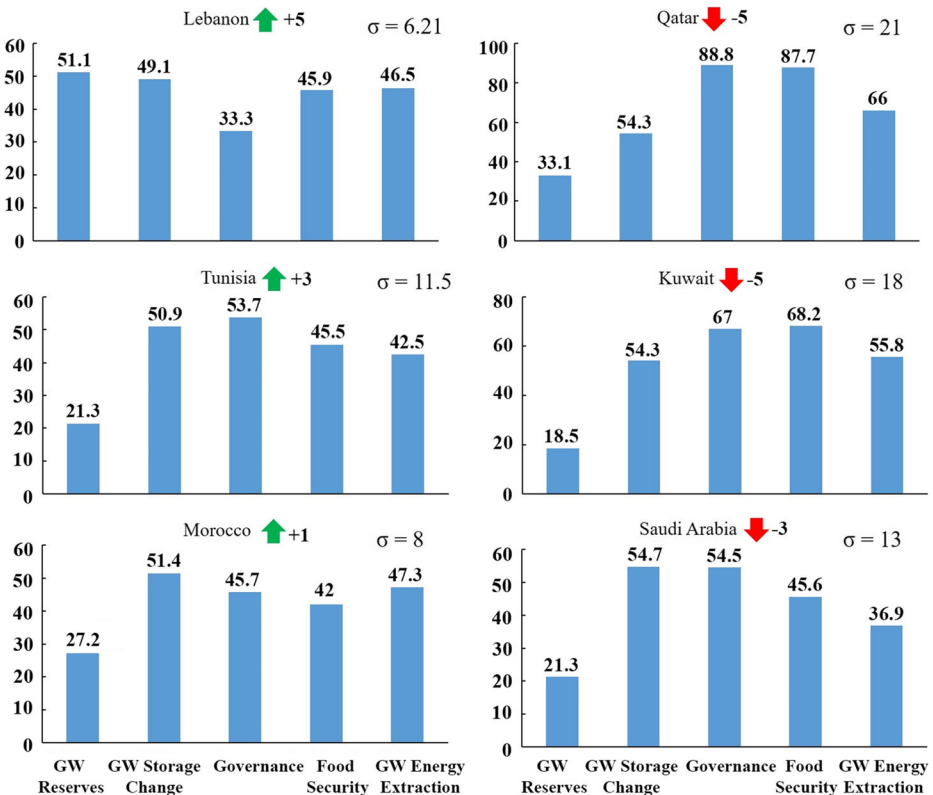


Fig. 2 Bar graphs displaying the effects of balanced indicator performance on country-level score and rank when using multiplicative geometric mean aggregation

theoretical framework that reflects the complexity of what is being measured and the interaction between the different dimensions that create it is the most pertinent step in the construction of a good composite index. The selection of an aggregation method is central to that framework, particularly as it relates to understanding and simulating the interactions between different indicators in driving a specific phenomenon. In the GRI case, executing the index using both arithmetic and geometric means resulted in outputs that are complementary to each other that highlight different perspectives on groundwater depletion risk. For instance, based on what has been discussed above, the penalizing effect of geometric aggregation on the oil-rich gulf countries stresses the reliance on oil income in mitigating groundwater risk, and raises inquiries on the sustainability of low risk scores produced by the index under full compensability conditions.

4.4 Combined Sensitivity Analyses Impacts

The preceding analysis examined the effects of different potential sources of sensitivity on the GRI output separately, with results displaying varying levels of sensitivity by indicator selection and choice of normalization and aggregations schemes. For a complete sensitivity analysis, an aggregation of the sources of sensitivity was conducted. The GRI country scores, generated with each of the sensitivity tests, were arithmetically averaged into a modified GRI output (GRI_{Modified}) reflecting country scores and ranks, as defined by alternative indicator, normalization scheme, and aggregation method selection. The results (Table 4) display the robustness of the GRI and its insensitivity to the discussed methodological alternatives. In terms of country scores, modified GRI values are negligibly different from original ones (Δ GRI Score) with the score shifts not exceeding 6 points on a [0, 100] scale, as is the case with Libya, Israel and the Palestinian Territories. In terms of country ranks, modified GRI values were also negligible, with 11 out of 16 countries experiencing no rank change (Δ GRI Rank). The remaining five countries only exhibited one shift in rank, except for Lebanon, which fell two ranks.

Moreover, the GRI modified results did not significantly alter the overall analysis, interpretations and conclusions of the original GRI results in the MENA region. Lezzaik et al. (2018) interpreted GRI results through the framework of a typological classification (Fig. 3) that grouped countries according to both their groundwater resource allotments and their governance and income levels. According to our interpretation, countries with effective governance and high incomes exhibited the lowest groundwater risk, while countries with poor governance and low incomes exhibited increased groundwater risk. Meanwhile, groundwater allotments proved inconsequential in determining risk conditions. The results of the modified GRI are consistent with the aforementioned analysis (Fig. 3). Of the five countries that experienced rank shifts, none moved outside their quantile group, thus maintaining the overall structure of country ranking order. Consequently, GRI modified results affirm the insensitivity of the index to alternative indicator selections, normalization schemes and aggregation methods.

5 Conclusions

This paper examines the sensitivity of the GRI to the methodological judgments that were made during its development by determining the effects of alternative methodological methods

Table 4 Summary of GRI sensitivity analysis score results. $GRI_{Modified}$ country scores and ranks were generated from the arithmetic mean of sensitivity analysis results to measure the combined impact of alternative methodological choices on $GRI_{Original}$ score and rank

Sensitivity Analysis		ALG*	EGY*	IRQ*	IPT*	JOR*	KWT*	LEB*	LIB*	MOR*	OMN*	QTR*	SDA*	SYR*	TUN*	UAE*	YMN*
Indicator Exclusion/Inclusion	GRI _{excl. GWR}	35	41	27	72	50	59	46	31	53	49	77	41	31	48	55	31
	GRI _{excl. GWSC}	29	32	19	74	43	63	44	24	47	42	70	32	27	41	59	22
	GRI _{excl. GOV}	35	35	31	61	39	52	48	36	47	36	58	32	37	39	48	32
	GRI _{excl. FS}	32	34	25	67	46	53	46	25	49	44	59	35	31	42	59	32
Normalization Scheme	GRI _{excl. GWEC}	37	39	28	67	46	58	45	34	42	49	66	45	37	43	63	28
	Min - Max	33	36	26	68	45	56	46	30	47	44	66	37	33	43	56	29
	Z - Score	19	26	0	100	51	72	36	4	46	54	99	35	12	43	81	7
Aggregation Method	Indicization	23	28	0	100	57	58	36	6	51	58	91	40	12	46	88	14
	Arithmetic Mean	33	36	26	68	45	56	46	30	47	44	66	37	33	43	56	29
	Geometric Mean	25	29	17	65	36	32	44	18	41	33	33	23	25	35	44	22
GRI _{Original} Score	GRI _{Original} Rank	34	36	26	68	45	56	47	30	47	44	66	37	33	43	56	29
	GRI _{Modified} Score	12	11	16	1	7	4	6	14	5	8	2	10	13	9	3	15
	GRI _{Modified} Rank	30	34	20	74	46	56	44	24	47	45	69	36	28	42	61	25
	Δ GRI Score	12	11	16	1	6	4	8	15	5	7	2	10	13	9	3	14
Δ GRI Rank	Δ GRI Score	4	2	6	-6	-1	0	3	6	0	-1	-3	1	5	0	-5	4
	Δ GRI Rank	0	0	0	0	1	0	-2	-1	0	1	0	0	0	0	0	1

* ALG (Algeria); EGY (Egypt); IRQ (Iraq); IPT (Israel and Palestinian Territories); JOR (Jordan); KWT (Kuwait); LIB (Lebanon); MOR (Morocco); OMN (Oman); QTR (Qatar); SDA (Saudi Arabia); SYR (Syria); TUN (Tunisia); UAE (United Arab Emirates); YMN (Yemen)

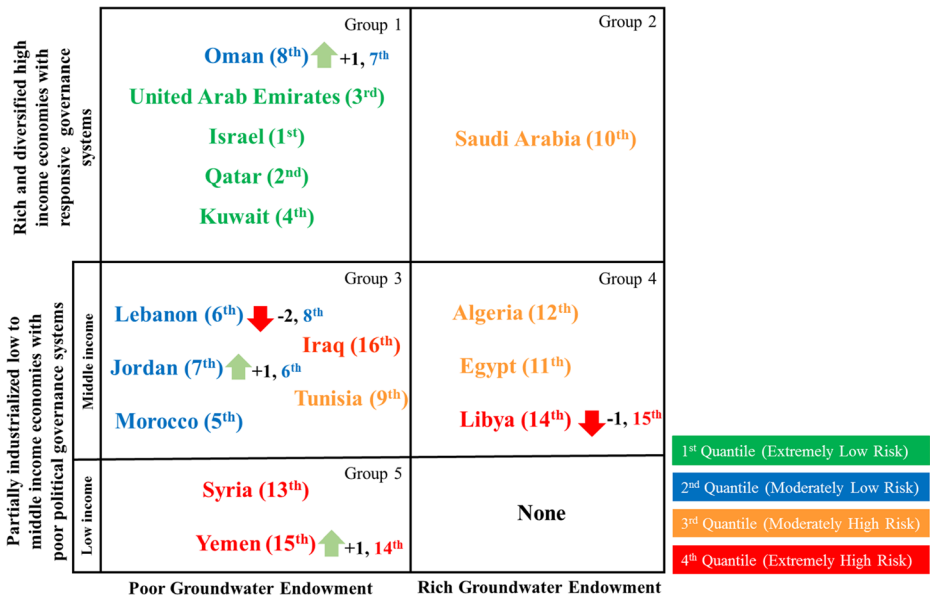


Fig. 3 A modified typology of MENA countries by hydrological systems and political economies. The figure shows $GRI_{original}$ country rank (in parenthesis). Countries experiencing a rank shift with a $GRI_{modified}$ are denoted with an arrow, followed by the number of shifts and the modified rank

on country-level groundwater risk scores and rank. A one-factor-at-a-time (OFAT) sensitivity analysis was used to measure the robustness of indicator selection, choice of normalization scheme, and choice of aggregation method. The results have shown that GRI provides a robust measure that is not biased by either the selection of the index indicators or by the choice of normalization scheme. On the other hand, the choice of aggregation methods between an additive arithmetic mean and a multiplicative geometric mean, presented a source of significant sensitivity to the GRI. This was expected given the differences in how they reflect the interactions and trade-offs between different indicators. The GRI sensitivity to the choice of aggregation method affirms the need for a representative theoretical framework that correctly simulates the interactive process between different factors contributing to groundwater risk.

Overall, however, the GRI index is explicitly insensitive to the discussed alternative methodological choices. The implication of our sensitivity analysis is significant, as it allows for the customized use of the GRI index outside the MENA region, in which alternative methodological choices, taken to fit unique regional conditions, do not significantly distort the GRI outcomes. This includes the substitution or addition of other indicators or dimensions depending on the local governance of a region as in Seward and Xu (2019), or in other semi-arid to hyper-arid transboundary aquifer systems like the MENA region.

The common thread between composite indices for groundwater and other environmental factors is the general need to evaluate groundwater or environmental sustainability, resilience, or vulnerability against potential changes or stressors. Indices may vary in their approach with different indicators, weighting schemes, or datasets, which further underscores the point that no perfect index or approach exists. Rather each one will have advantages and tradeoffs, tuned specifically for the purpose of a particular site or state variable. This paper further solidifies the

robustness of the original method (Lezzaik et al. 2018) and demonstrates a non-participatory approach that shows little sensitivity in weighting schemes. In the future, the authors recommend variance-based measures of sensitivity that explores and accounts for simultaneous variations and interactions between different input factors.

References

- Aguna C, Kovacevic M (2010) Uncertainty and sensitivity analysis of the human development index. Human development reports 11 United Nations development Programme
- Azizi F, Vadiati M, Moghaddam AA, Nazemi A, Adamowski J (2019) A hydrogeological-based multi-criteria method for assessing the vulnerability of coastal aquifers to saltwater intrusion. *Environ Earth Sci* 78:548–522. <https://doi.org/10.1007/s12665-019-8556-x>
- Bandura R (2008) A survey of composite indices measuring country performance: 2008 update. United Nations Development Programme, Office of Development Studies (UNDP/ODS Working Paper)
- Booyens F (2002) An overview and evaluation of composite indices of development. *Soc Indic Res* 59:115–151. <https://doi.org/10.1023/A:1016275505152>
- Cameron S, White P (2004) Determination of key indicators to assess groundwater quantity in New Zealand aquifers. Ministry for the Environment, Institute of Geological and Nuclear Sciences, Wellington, New Zealand
- Cherchye L, Moesen W, Rogge N, Van Puyenbroeck T, Saisana M, Saltelli A, Liska R, Tarantola S (2008) Creating composite indicators with DEA and robustness analysis: the case of the technology achievement index. *J Oper Res Soc* 59:239–251. <https://doi.org/10.1057/palgrave.jors.2602445>
- Cox DR, Fitzpatrick R, Fletcher A, Gore S, Spiegelhalter D, Jones D (1992) Quality-of-life assessment: can we keep it simple? *J R Statistic Soc Ser A (Statistic Soc)* 155:353–393. <https://doi.org/10.2307/2982889>
- Droogers P, Immerzeel W, Terink W, Hoogeveen J, Bierkens M, Van Beek L, Debele B (2012) Water resources trends in Middle East and North Africa towards 2050. *Hydrol Earth Syst Sci* 16:3101–3114
- Fan Y, Li H, Miguez-Macho G (2013) Global patterns of groundwater table depth. *Science* 339:940–943. <https://doi.org/10.1126/science.1229881>
- FAO (2016) Statistical databases. Food and Agriculture Organization of the United Nations <http://faostat.fao.org/default.htm>
- Foa R, Tanner J (2012) Methodology of the indices of social development. *ISD Working Paper Series International Institute of Social Studies of Erasmus University, Rotterdam*
- Ginsburg NS, Osborn J, Blank G (1986) Geographic perspectives on the wealth of nations. Paper no. 220. Department of Geography University of Chicago, Chicago
- Gleeson T, Alley WM, Allen DM, Sophocleous MA, Zhou Y, Taniguchi M, VanderSteen J (2012) Towards sustainable groundwater use: setting long-term goals, backcasting, and managing adaptively. *Groundwater* 50:19–26. <https://doi.org/10.1111/j.1745-6584.2011.00825.x>
- Godfrey S, Howard G, Niwagaba C, Tibatemwa S, Vairavamoorthy K (2002) Improving risk assessment and management in urban water supplies. Paper presented at the 28th WEDC Conference, Kolkata, India
- Hagerty MR, Land KC (2007) Constructing summary indices of quality of life: a model for the effect of heterogeneous importance weights. *Sociol Methods Res* 35:455–496. <https://doi.org/10.1177/0049124106292354>
- Hosseini SM, Parizi E, Ataie-Ashtiani B, Simmons CT (2019) Assessment of sustainable groundwater resources management using integrated environmental index: case studies across Iran. *Sci Total Environ* 676:792–810. <https://doi.org/10.1016/j.scitotenv.2019.04.257>
- Jeet P, Singh D, Sarangi A (2019) Development of a composite hydrologic index for semi-arid region of India. *Groundwater* 57:749–755. <https://doi.org/10.1111/gwat.12867>
- Jemmali H, Sullivan CA (2014) Multidimensional analysis of water poverty in MENA region: an empirical comparison with physical indicators. *Soc Indic Res* 115:253–277. <https://doi.org/10.1007/s11205-012-0218-2>
- Kaufmann D, Kraay A, Mastruzzi M (2011) The worldwide governance indicators: methodology and analytical issues. *Hague J Rule Law* 3:220–246. <https://doi.org/10.1017/S1876404511200046>
- Lewis AS, Reid KR, Pollock MC, Campleman SL (2012) Speciated arsenic in air: measurement methodology and risk assessment considerations. *J Air Waste Manage Assoc* 62:2–17. <https://doi.org/10.1080/10473289.2011.608620>
- Lezzaik K, Milewski A (2018) A quantitative assessment of groundwater resources in the Middle East and North Africa region. *Hydrogeol J* 26:251–266. <https://doi.org/10.1007/s10040-017-1646-5>

- Lezzaik K, Milewski A, Mullen J (2018) The groundwater risk index: development and application in the Middle East and North Africa region. *Sci Total Environ* 628:1149–1164. <https://doi.org/10.1016/j.scitotenv.2018.02.066>
- López-Claros A (2010) The innovation for development report 2010-2011: innovation as a driver of productivity and economic growth. Palgrave Macmillan, London. <https://doi.org/10.1057/9780230299269>
- Mattas C, Kazakis N, Soulios G (2014) Groundwater vulnerability and risk assessment using drastic model in a gis environment. A case study from the gallikos river basin, northern greece. In: The 10th International Hydrogeological Congress of Greece, Thessaloniki, Greece
- Mazziotta M (2013) A non-compensatory composite index for measuring well-being over time. *Cogito Multidiscip Res J* 5:93–94
- Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman A, Giovannini E (2005) Handbook on constructing composite indicators. OECD Statistics Working Papers 3 OECD Publishing Paris, Paris
- OECD/European Union/JRC (2008) Handbook on constructing composite indicators: methodology and user guide. OECD Publishing Paris, Paris. <https://doi.org/10.1787/9789264043466-en>
- Pfeiffer L, Lin C-YC (2014) The effects of energy prices on agricultural groundwater extraction from the High Plains aquifer. *Am J Agric Econ* 96:1349–1362. <https://doi.org/10.1093/ajae/aa020>
- Pires A, Morato J, Peixoto H, Botero V, Zuluaga L, Figueroa A (2017) Sustainability assessment of indicators for integrated water resources management. *Sci Total Environ* 578:139–147. <https://doi.org/10.1016/j.scitotenv.2016.10.217>
- Preda E, Kløve B, Kværner J, Lundberg A, Siergieiev D, Boukalova Z, Postawa A, Witczak S, Balderacchi M, Trevisan M (2013) New indicators for assessing GDE vulnerability. Wageningen University & Research Netherlands <http://www.diva-portal.org/smash/get/diva2:995565/FULLTEXT01.pdf>
- Richey AS, Thomas BF, Lo MH, Famiglietti JS, Swenson S, Rodell M (2015a) Uncertainty in global groundwater storage estimates in a total groundwater stress framework. *Water Resour Res* 51:5198–5216. <https://doi.org/10.1002/2015WR017351>
- Richey AS, Thomas BF, Lo MH, Reager JT, Famiglietti JS, Voss K, Swenson S, Rodell M (2015b) Quantifying renewable groundwater stress with GRACE. *Water Resour Res* 51:5217–5238. <https://doi.org/10.1002/2015WR017349>
- Saaty TL (1987) The analytic hierarchy process—what it is and how it is used. *Math Model* 9:161–176. [https://doi.org/10.1016/0270-0255\(87\)90473-8](https://doi.org/10.1016/0270-0255(87)90473-8)
- Saisana M, Tarantola S (2002) State-of-the-art report on current methodologies and practices for composite indicator development. Joint Research Centre Ispra, Ispra
- Saisana M, Saltelli A, Tarantola S (2005) Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *J R Statistic Soc: Ser A (Statistic Soc)* 168:307–323. <https://doi.org/10.1111/j.1467-985X.2005.00350.x>
- Seward P, Xu Y (2019) The case for making more use of the Ostrom design principles in groundwater governance research: a south African perspective. *Hydrogeol J* 27:1017–1030. <https://doi.org/10.1007/s10040-018-1899-7>
- Sharpe A, Andrews B (2012) An assessment of weighting methodologies for composite indicators: the case of the index of economic well-being. Center for the study of living standards, CCLS Research Report Centre for the Study of Living Standards Ottawa, Ontario. <http://www.csls.ca/reports/csls2012-10.pdf>
- Sowers J, Vengosh A, Weinthal E (2011) Climate change, water resources, and the politics of adaptation in the Middle East and North Africa. *Clim Chang* 104:599–627. <https://doi.org/10.1007/s10584-010-9835-4>
- Srinivasan V, Lambin E, Gorelick S, Thompson B, Rozelle S (2012) The nature and causes of the global water crisis: Syndromes from a meta-analysis of coupled human-water studies. *Water Resour Res* 48. <https://doi.org/10.1029/2011WR011087>
- Villholth KG (2013) Groundwater irrigation for smallholders in sub-Saharan Africa—a synthesis of current knowledge to guide sustainable outcomes. *Water Int* 38:369–391. <https://doi.org/10.1080/02508060.2013.821644>
- Voss KA, Famiglietti JS, Lo M, De Linage C, Rodell M, Swenson SC (2013) Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resour Res* 49:904–914. <https://doi.org/10.1002/wrcr.20078>
- World Bank (2016) World development indicators 2016. World Bank © World Bank Washington, DC. <https://openknowledge.worldbank.org/handle/10986/23969>
- Zhu T, Ringer C, Cai X (2007) Energy price and groundwater extraction for agriculture: exploring the energy-water-food nexus at the global and basin levels. Linkages between Energy and Water Management for Agriculture in Developing Countries, Hyderabad

Affiliations

Adam Milewski¹ · Khalil Lezzaik^{1,2} · Rachel Rotz¹

¹ Department of Geology, University of Georgia, Water Resources & Remote Sensing Group (WRRS), 210 Field Street, Geography-Geology Building 306, Athens, GA 30602, USA

² Office of Water Programs, Sacramento State University, 6000 J Street, Sacramento, CA 95819-6025, USA