Supporting Information for

Lifting China's Water Spell

Dabo Guan[1,2,3], Klaus Hubacek[4], Martin Tillotson[5], Hongyan Zhao[3], Weidong Liu[6], Zhu Liu[7], Sai Liang[8]*

[1.] Water Security Research Centre, School of International Development, University of East Anglia, Norwich, NR4 7TJ, UK

[2.] water@leeds, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK

[3.] Ministry of Education Key Laboratory for Earth System Modelling, Centre for Earth System Science, Tsinghua University, Beijing 100084, China

[4.] Department of Geographical Sciences, University of Maryland, College Park, Maryland, 20742, USA

[5.] water@leeds, School of Civil Engineering, University of Leeds, Leeds, LS2 9JT, UK

[6.] Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China

[7.] Sustainability Science Program, John F. Kennedy School of Government, Harvard University, Cambridge, USA

[8.] School of Natural Resources and Environment, University of Michigan, Michigan 48109-1041, USA

[*]To whom correspondence should be addressed. E-mail: dabo.guan@uea.ac.uk

This document contains Supporting Information (SI) for the manuscript "Lifting China's Water Spell". The first section details the various models used. The second section provides a detailed data description. The SI contains 18 pages, two tables and four figures.

1. Method

1.1 Defining water consumption, pollution discharge and unavailable water

In the main text, we employ three water consumption indicators – water consumption, pollution discharge and cumulative unavailable water. We aim to further explain the meaning of those three terms here.

Water Consumption: we use the definition for 'quantity of water consumption' in China's "code of practice for water resources bulletin" by General Administration of Quality Supervision, Inspection and Quarantine and National Emblem of the People's Republic of China [1]. The definition is: water consumed up and not be able to return to surface- or ground- water bodies during water transport and utilisation via evaporation, soil filtration, production processes, people and animal daily usage. Please note the agricultural water consumption does not include rain water (or green water in water footprint context).

Pollution discharge: the definition for 'quantity of wastewater discharge' in China's "code of practice for water resources bulletin is the secondary, tertiary and urban households wastewater discharge, excluding the cooling water in electricity production and unaccountable wastewater in during mining process. In this paper, we separate the wastewater from production process and household consumption. The production related wastewater discharge includes agriculture, secondary, and tertiary wastewater discharge. Data of secondary and tertiary wastewater is from Chinese official statistics (see data section), and agriculture wastewater is estimated by using the proxy of amount of fertiliser used.

Unavailable water: Traditionally, the term 'water demand' comprises the amount of net water consumed for economic production and domestic usage; however the flows of polluted water resources after economic activities back into the ecosystem are usually not accounted for. The quality of the return flows usually changes; the water quality being lower than when it entered the production process initially. The entered pollutants would mix and spread in the water bodies to develop a dynamic process causing pollution in the same and sometimes other economic regions. For example, the polluted wastewater infiltrates into the groundwater or mixes with surface water and flows downstream where it contaminates other freshwater resources. Furthermore, the pollutants may stay in the hydro-ecosystem for a long time span, and contaminates the return freshwater from new water cycles. The hydro-ecosystem has the ability to self-purify some waste, but this ability is determined by hydro- or geo- conditions and biological, physical or chemical characters of the pollutants. If the cumulative pollutants research a certain level in an ecosphere, the ecosystem would lose their purification ability. To recover such natural ability, large amount of freshwater would be require diluting the pollution concentration level to a standard level. The water requirement by the ecosystem can refer as the 'unavailable' water to any economic system. The cumulative unavailable water refers to total amount of water is required to dilute all remaining pollution in water bodies over years.

1.2 Hydro-economic inventories

In the main text section 3.1, we explain the construction of 7 annual hydro-economic inventories for 1992 – 2007. Each inventory is constructed based on our previous constructed method published in Guan and Hubacek [2]. The framework of the method is shown in Table 1 at the main text, which illustrates the economic-ecological exchange by input-output model.

Matrix **A** describes the economic system. The key part of the static Leontief's input-output model is the following algebraic equation:

$$\mathbf{y} = (\mathbf{I} - \mathbf{A})\mathbf{x}; \quad (1)$$

where **x** represents sectoral production output; **y** represents total final demand; **I** is the $n \times n$ identity matrix; and **A** is a $n \times n$ matrix of technical coefficients. The **A** represents the components of intermediate demand, where the coefficients a_{ij} refer to the amount of input from sector i required by the sector j for each unit of output.

Hence, a_{ij} (unit: Yuan/Yuan) are technological coefficients, defined as Equation (2),

$$a_{ij} = \frac{x_{ij}}{x_j} \qquad (2)$$

where the x_{ij} (unit: Yuan/year) represents the monetary flows from i^{th} sector to j^{th} sector. x_j (unit: Yuan/year) is the total economic output of j^{th} sector in monetary term. If the **(I-A)** is not singular, we can convert Equation (1) as,

$$\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{y} \qquad (3)$$

The matrix $(\mathbf{I}\text{-}\mathbf{A})^{-1}$ gives the so-called Leontief multiplier matrix which accounts for the total accumulative effect on sectoral output **(x)** by the changes in final demand **(y)**.

Matrix **F** describes water inputs to economic sectors. Water is a primary input involved in production of goods and services. This connection can be captured in the $m\times n$ **F** matrix. The water input for production consists of three sources, surface water, groundwater and rainfall. The direct water consumption coefficient, f_{kj} (unit: m^3/Yuan) is defined in Equation (4):

$$f_{kj} = \frac{g_{kj}}{x_j} \qquad (4)$$

where g_{kj} (unit: m^3/year) is the amount of water supplied from the k hydrological sectors consumed in economic sector j; x_j (unit: Yuan/year) is the total input of the j^{th} sector. This coefficient represents the direct or the first round effects of the sectoral interaction in the economy. However, water is not only consumed directly but also indirectly. In order to combine both direct and indirect water consumption, we have to generate the total water consumption multipliers matrix (**F**) by multiplying the diagonalized matrix of direct water consumption coefficients $\hat{f}_k$ (k = surface, ground or rainfall water) with the Leontief multiplier matrix $(\mathbf{I}\text{-}\mathbf{A})^{-1}$, which represents an indicator of the total amount of water used up throughout the production chain for each sector. Equation (5) describing the direct and indirect effects of water inputs by increasing a unit of final consumption, named as 'Net water consumption':

$$\textbf{Net Water Consumption} = \hat{f}_k (\mathbf{I} - \mathbf{A})^{-1}\mathbf{y} \qquad (5)$$

Matrix **R** describes flows from the economy to the hydro-ecosystem. The output (COD emissions) of each production sector to the water supply sources shall be captured in the **R** matrix with dimensions $n\times m$. Similarly to the process of determining net water consumption, Equation (6) representing the total amount of wastewater generated in an economy by final consumption, referred to as 'Discharged wastewater':

$$\textbf{Discharged COD} = \hat{r}_l (\mathbf{I} - \mathbf{A})^{-1}\mathbf{y} \qquad (6)$$

Matrix **B** identifies the water flows within the hydrological ecosystem and the impacts of discharged wastewater to the freshwater resources. In other words, it measures the natural water losses (e.g. evaporation loss), and also quantifies the amount of freshwater sources necessary to dilute the pollutants in the discharged wastewater to a respective standard rate (that is e.g. stated in the regulation of water quality and management). For illustrate purpose, we use COD (Chemical Oxygen Demand) as water quality indicator measured in *gram/m^3*. We employ the following water quality model which is constructed based on a mass balance approach (Equation 7). It calculates the concentration of pollutants in the water body after the mixing processes of the discharged wastewater from economic sectors into the original water resources. Equation 7 is a simple water quality model, however for a large watershed, the effectiveness is similar to the sophisticated model in estimating water pollution dispersion [3].

$$c_{mixed} = \frac{1}{1 + k_1 \frac{v}{q}} \left(k_2 \frac{q_0}{q} c_0 + \frac{q_p}{q} c_p \right) \qquad (7)$$

$$q = q_0 + q_p$$

Parameters:

c_{mixed} – pollutant concentration after mixture processes

c_0 – initial pollutant concentration in the water body

c_p – pollutant concentration in wastewater

q – runoff rate after completion of the mixing process[1]

q_0 – initial runoff rate

q_p – wastewater discharge rate

v – volume of freshwater in the water bodies

k_1 – Total reaction rate of pollutants after entering the water bodies

k_2 – pollution purification rate before entering to the water bodies (e.g. filler effect of soils)

Most countries, including China, have implemented water quality regulations using standards for the quality of wastewater and for the receiving water bodies. In order to avoid water pollution, the pollutant concentration in the water body after the mixing processes needs to be less than the standard rate of the respective standard (i.e. $c_{standard} \geq c_{mixed}$). If we replace the c_{mixed} by $c_{standard}$, the Equation 7 can be re-written as follows:

$$v \geq \frac{1}{k_1 c_{\text{standard}}} (q_0 c_0 + k_2 q_p c_p - q c_{\text{standard}}) \qquad (8)$$

$$q = q_0 + q_p$$

Hereby, the scalar v is the amount of freshwater in the hydro-ecosystem that would be required to dilute pollutants in the discharged wastewater in order to reduce the pollution concentration level to the standard rate. In this study, we use 'Grade V' (e.g. usable for agricultural irrigation purpose only) as the minimum regulatory standard to calculate the 'dilution water required'.

After annual 'dilution water requirement' is calculated, cumulative dilution water (shown in below equation) is a summation of each year's dilution water during 1992 – 2007 (measured by α) after further purification processes taken (measured by η). The further purification processes can be further water treatment in water body and changes of natural processes (e.g. flooding or storms to accelerate velocity of runoffs etc). Due to research constraints, we set $\eta = 1$ for this research, which means there is no further purification processes considered during modelling. Further and more collaborative interdisciplinary research would be required to better refine the factor η. We acknowledge that the factor η can vary significantly by seasons and regions, but this is out of scope of this study.

$$v_{cumulation} = \sum_{1-\alpha} v \cdot \eta \qquad \alpha = 15$$

[1] Runoff is categorized as surface runoff and sub-surface runoff for surface and ground water respectively.

1.3 Structural Decomposition Analysis

The principal idea of Structural Decomposition Analysis (SDA) can be illustrated as shown in Equation (6.1) in the case of a two determinant multiplicative function.

$$y = x_1 \cdot x_2 \tag{9}$$

The change of *y* (Δy) can be decomposed between two time points, *t* and *t-1* into changes of the driving forces, x_1 and x_2. However there is no unique solution for the decomposition. For example, one can start the decomposition from the base year (e.g. *t-1*), which is referred to the Lasperyres index, whereas one can also begin the process from the target year (e.g. *t*), which is referred to as the Paasche index, as shown in Equation (10). Together using both Lasperyres and Paasche indices for decomposition analysis is referred to as so-called polar decompositions [4]. Heokstra and van den Bergh [5] examined that neither of the polar decompositions would lead to a complete decomposition (e.g. no residual terms), and the usual solution for this is to combine the two perspectives, so either Lasperyres- Paasche or Paasche-Lasperyres index.

$$\begin{gathered} \Delta y = y_t - y_{t-1} = \Delta(x_1 \cdot x_2) \\ \text{or,} \\ \Delta y = y_{(t-1)} - y_t = \Delta(x_1 \cdot x_2) \end{gathered} \tag{10}$$

Furthermore, the change of x_1 can be expressed as: $\Delta x_1 = x_{1(t)} - x_{1(t-1)}$; similarly, $\Delta x_2 = x_{2(t)} - x_{2(t-1)}$.

One of the possibilities is to decompose Equation (10) is by using Lasperyres- Paasche index as shown in Equation (11):

$$\begin{aligned} \Delta(x_1 \cdot x_2) &= x_{1(t)} \cdot x_{2(t)} - x_{1(t-1)} \cdot x_{2(t-1)} \\ &= (\Delta x_1 + x_{1(t-1)}) \cdot x_{2(t)} - x_{1(t-1)} \cdot x_{2(t-1)} \\ &= \Delta x_1 \cdot x_{2(t)} + x_{1(t-1)} \cdot x_{2(t)} - x_{(t-1)} \cdot x_{2(t-1)} \\ &= \Delta x_1 \cdot x_{2(t)} + x_{1(t-1)} \cdot (\Delta x_2 + x_{2(t-1)}) - x_{(t-1)} \cdot x_{2(t-1)} \\ &= \Delta x_1 \cdot x_{2(t)} + x_{1(t-1)} \cdot \Delta x_2 \end{aligned} \tag{11}$$

However, the other possibility to decompose Equation 10 is by using Paasche- Lasperyres index as shown in Equation 12:

$$\Delta(x_1 \cdot x_2) = \Delta x_1 \cdot x_{2(t-1)} + x_{1(t)} \cdot \Delta x_2 \tag{12}$$

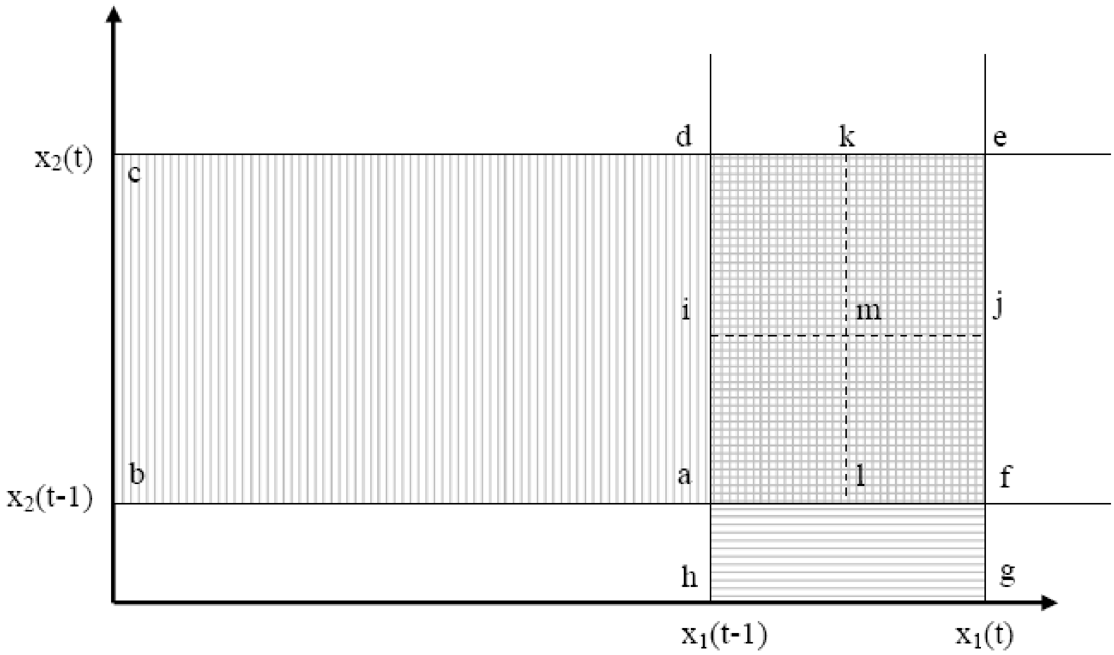
Hereby, the core question is to examine whether the above two terms are congruent with the requirements of decomposition, which comprises the following three conditions in terms of Hoekstra and van der Bergh [5] and de Boer [4].

- complete, which means there are no residual terms

- "0" robust, which means it can deal with "0" values in the calculation
- time reversal, which means the decomposition produces a reverse result if the time period has been reversed, for example $\Delta y = y_t - y_{t-1} = -(y_{t-1} - y_t)$.

The method to examine the "complete" condition can be simple illustrated in Figure 1; Equation (11) covers the areas of "hdeg + bcda" with residual term "0" while the areas of Equation (12) covering is the "hafg + bcef" with residual term of "0"., Both decompositions cover the required areas (filled with dashed lines), which means they are qualified with the requirements of 'complete' and the condition of "0" robust for implementing calculations [4, 5]. Equations (11) and (12) follow from each other by reverting base and comparison period. However neither of the above expressions satisfies the requirement of time reversal. A common approach is to take the average of these two equations to satisfy this requirement. Equation (11) and (12) is a "mirror pair" decomposition, which is the pair of permutations where the time period indication on the coefficients attached to each difference term is exactly the opposite [6].

Figure 1: Additive decomposition of $y = x_1 \cdot x_2$, discrete time



Source: [7, 8]

As mentioned previously, either Lasperyres – Paasche or Paasche – Lasperyres index would fulfil the requirement of the "complete" requirement. The two approaches are equivalent and there is no reason why one of them should be preferred to the other [7]. Therefore, the decomposition of *y* is not unique; however the result (e.g. the covered areas in Figure 1) is unique in this two determinants case. The problem is a so-called non-uniqueness which means that there exist a number of different decomposition forms and that it cannot be decided which one to prefer. Usually, the factors in SDA studies are more than three, Dietzenbacher and Los [6] proved that in the case of *n* factors, the number of possible "complete" decompositions (without any residual terms) is equal to *n!*. For example, this chapter assesses five determinants in the change of water consumption, so the possible decompositions are *5! = 120*.

1.4 Water SDA driving by five factors

In general, a country's energy demand and associated emissions change over time for a variety of reasons—population growth, increases in economic output, changes in trade structure, infrastructure investment, technical change and efficiency improvements, and changes in the production and consumption systems. This chapter will adopt five factors to assess the change of water with application of the SDA framework, and one of the *5!=120* "complete" decomposition function is shown in Equation (13) or Equation 1 in the main text:

$$
\begin{aligned}
\Delta \mathrm{Water} &= \Delta \mathrm{Water}_{(t)} - \Delta \mathrm{Water}_{(t-1)} \\
&= p_{(t)} \cdot \mathbf{F}_{(t)} \cdot \mathbf{L}_{(t)} \cdot \mathbf{y}_{\mathbf{s}(t)} \cdot \mathbf{y}_{\mathbf{v}(t)} - p_{(t-1)} \cdot \mathbf{F}_{(t-1)} \cdot \mathbf{L}_{(t-1)} \cdot \mathbf{y}_{\mathbf{s}(t-1)} \cdot y_{v(t-1)} \\
&= \Delta p \cdot \mathbf{F}_{(t)} \cdot \mathbf{L}_{(t)} \cdot \mathbf{y}_{\mathbf{s}(t)} \cdot y_{v(t)} + p_{(t-1)} \cdot \Delta \mathbf{F} \cdot \mathbf{L}_{(t)} \cdot \mathbf{y}_{\mathbf{s}(t)} \cdot y_{v(t)} \\
&+ p_{(t-1)} \cdot \mathbf{F}_{(t-1)} \cdot \Delta \mathbf{L} \cdot \mathbf{y}_{\mathbf{s}(t)} \cdot y_{v(t)} + p_{(t-1)} \cdot \mathbf{F}_{(t-1)} \cdot \mathbf{L}_{(t-1)} \cdot \Delta \mathbf{y}_{\mathbf{s}} \cdot y_{v(t)} \\
&+ p_{(t-1)} \cdot \mathbf{F}_{(t-1)} \cdot \mathbf{L}_{(t-1)} \cdot \mathbf{y}_{\mathbf{s}(t-1)} \cdot \Delta y_v
\end{aligned}
\qquad \textbf{(13)}
$$

where,

p – is a scalar, population

F – is the diagonalised water coefficient matrix

L – is the Leontief inverse matrix, $\mathbf{L=(I-A)^{-1}}$

$\mathbf{y_s}$ – is a column vector representing per capita consumption patterns

y_v – is a scalar representing the total consumption volume

Equation (13) comprises five terms in total; each term represents the contribution to change in water, water triggered by one driving force with keeping the rest of factors constant, correspondingly. The first term represents population growth, *p*; the second term represents the aggregated changes in the emission intensities (efficiency), **F**; the third term represents changes in the production structure, **L**; the fourth term represents changes in the consumption structure, $\mathbf{y_s}$; and the fifth term represents changes in the consumption volume (GDP), y_v.

1.5 "Weights" in water structural decomposition equations

As Equation (13) shows, the change of water is decomposed into five terms, and each term represents the contribution of the changing factor ("Δfactor") of the total change of water. One can perceive a logical pattern that the "Δfactor" is placed at each term in turn from left to right; and the other constant factors on the left side of the "Δfactor" are in base year value (year "*t-1*"); and the ones on the right side of the "Δfactor" are in target year value (year "*t*"). Therefore, by extracting the constant values in each term Equation (13) can be re-written as Equation (14):

$$
\Delta \mathrm{Water} = w^p \Delta p + w^{\mathbf{F}} \cdot \Delta \mathbf{F} + w^{\mathbf{L}} \cdot \Delta \mathbf{L} + w^{\mathbf{y_s}} \cdot \Delta \mathbf{y_s} + w^{y_v} \cdot \Delta y_v \qquad \textbf{(14)}
$$

where the w^p, $w^{\mathbf{F}}$, $w^{\mathbf{L}}$, $w^{\mathbf{y_s}}$ and w^{y_v} is the so-called "weight" or "coefficient" for each "Δfactor" respectively. The calculation of these weights or "coefficients" are usually done via econometric methods; alternatively, they can be generated via a more straight forward way by deriving them with the structural decomposition method [8].

As mentioned earlier, Equation (13) is not a unique decomposition equation, which is only one of the 120 decomposition equations by assuming the order of the driving forces of "$p \cdot \mathbf{F} \cdot \mathbf{L} \cdot \mathbf{y_v} \cdot y_s$". However, the order can also be "$\mathbf{F} \cdot p \cdot \mathbf{L} \cdot \mathbf{y_v} \cdot y_s$" or "$\mathbf{L} \cdot \mathbf{F}\ \mathbf{y_v} \cdot y_s \cdot p$" and so on. Although each decomposition equation would produce exactly the same result for ΔWater, de Haan [9] found that the size of the contribution of each "Δfactor" significantly differs across the equations. In other words, the "coefficient" (w) of each "Δfactor" is varied in different equations.

Due to the non-uniqueness issue, Dietzenbacher and Los [6] suggested to take the average of all the $n!$ ($5!$ in this case) decomposition equations. In order to do so, all the 120 equations need to be sorted into a standard order, for example, every term in the equation needs to be re-arranged to the order of "$p \cdot \mathbf{F} \cdot \mathbf{L} \cdot \mathbf{y}_v \cdot y_s$", and the "Δfactor" is in turn placed from the first factor of "p" in the first term of the equation to the last factor of y_s in the last (fifth) term. Then, all the equations have been re-arranged in the same pattern. For example, the first term of every equation contains the information of contribution of population growth (Δp) to the change of water (ΔWater) with keeping other factors constant. The constant values are the "coefficient" for Δp. The "coefficient" $\mathbf{F}_{(t-1)} \cdot \mathbf{L}_{(t-1)} \cdot \mathbf{y}_{\mathbf{s}(t-1)} \cdot y_{v(t-1)}$ appears 24 times, and same as the "coefficient" $\mathbf{F}_{(t)} \cdot \mathbf{L}_{(t)} \cdot \mathbf{y}_{\mathbf{s}(t)} \cdot y_{v(t)}$ does. de Haan [9] and Seibel [10] found that each term in the equation always has $2^{(n-1)}$ different "coefficients" attached to the "Δfactor", $2^{(5-1)} = 16$ different "coefficients" to every "Δfactor" in this case.

Next one can calculate the "weights" of the "coefficients" which is attached to the "Δfactor". The easiest way is via observations, to count how many cases of "Δfactor" are attached to the same "coefficient". For example as mentioned previously, the "coefficient" $\mathbf{F}_{(t-1)} \cdot \mathbf{L}_{(t-1)} \cdot \mathbf{y}_{\mathbf{s}(t-1)} \cdot y_{v(t-1)}$ appears 24 times in the 120 equations, and therefore its weight is 24. However, the observation method could be difficult in large number of decomposition equations with more than 5 factors.

Seibel [10] proposed a mathematic method to deal with this. Firstly, let k represent the number of subscript "$t-1$" (base year) in a coefficient; k runs from "0" to "$n-1$"; therefore, the number of subscript "t" (target year) would be "$n-1-k$". Secondly, for each k, the number of different coefficients attached to the "Δfactor" can be calculated by Equation 15:

$$\frac{(n-1)!}{(n-1-k)! \cdot k!} \qquad \textbf{(15)}$$

In this study, n is set to 5 (five factors). So when k=0 or 4, there is only one coefficient for each case; when k=1 or 3, the number of different coefficients are 4 respectively; when k=2, there would be 6 different coefficients. Thirdly, Equation (16) calculates how many times each of these coefficients is repeated as "weights" for each "Δfactor" term in every equation of $n!$. The results for Equation (15) and (16) are shown in Table 1 for the case of n=5.

$$(n-1-k)! \cdot k! \qquad \textbf{(16)}$$

Table 1: Subscripts for the components of "Δfactor's" coefficients and their weights

k	Subscript for the components in the coefficients				**Weight**
	first	second	third	fourth	
0	*t*-1	*t*-1	*t*-1	*t*-1	**24**
1	*t*	*t*-1	*t*-1	*t*-1	**6**
	t-1	*t*	*t*-1	*t*-1	
	t-1	*t*-1	*t*	*t*-1	
	t-1	*t*-1	*t*-1	*t*	
2	*t*	*t*	*t*-1	*t*-1	**4**
	t	*t*-1	*t*	*t*-1	
	t	*t*-1	*t*-1	*t*	
	t-1	*t*	*t*	*t*-1	
	t-1	*t*	*t*-1	*t*	

	t-1	t-1	t	t	
3	t-1	t	t	t	**6**
	t	t-1	t	t	
	t	t	t-1	t	
	t	t	t	t-1	
4	t	t	t	t	**24**

Source: Modified from [7]

Therefore, each "w" attached to the "Δfactor" in Equation 14 can be defined, for example,

$$
\begin{aligned}
w^{p}\Delta p = \frac{1}{120}[& (24\cdot\Delta p\cdot\mathbf{F}_{(t-1)}\cdot\mathbf{L}_{(t-1)}\cdot\mathbf{y}_{\mathbf{s}(t-1)}\cdot y_{v(t-1)})+\\
& (6\cdot\Delta p\cdot\mathbf{F}_{(t)}\cdot\mathbf{L}_{(t-1)}\cdot\mathbf{y}_{\mathbf{s}(t-1)}\cdot y_{v(t-1)})+\\
& (6\cdot\Delta p\cdot\mathbf{F}_{(t-1)}\cdot\mathbf{L}_{(t)}\cdot\mathbf{y}_{\mathbf{s}(t-1)}\cdot y_{v(t-1)})+\\
& (6\cdot\Delta p\cdot\mathbf{F}_{(t-1)}\cdot\mathbf{L}_{(t-1)}\cdot\mathbf{y}_{\mathbf{s}(t)}\cdot y_{v(t-1)})+\\
& (6\cdot\Delta p\cdot\mathbf{F}_{(t-1)}\cdot\mathbf{L}_{(t-1)}\cdot\mathbf{y}_{\mathbf{s}(t-1)}\cdot y_{v(t)})+\\
& (4\cdot\Delta p\cdot\mathbf{F}_{(t)}\cdot\mathbf{L}_{(t)}\cdot\mathbf{y}_{\mathbf{s}(t-1)}\cdot y_{v(t-1)})+\\
& (4\cdot\Delta p\cdot\mathbf{F}_{(t)}\cdot\mathbf{L}_{(t-1)}\cdot\mathbf{y}_{\mathbf{s}(t)}\cdot y_{v(t-1)})+\\
& (4\cdot\Delta p\cdot\mathbf{F}_{(t)}\cdot\mathbf{L}_{(t-1)}\cdot\mathbf{y}_{\mathbf{s}(t-1)}\cdot y_{v(t)})+\\
& (4\cdot\Delta p\cdot\mathbf{F}_{(t-1)}\cdot\mathbf{L}_{(t)}\cdot\mathbf{y}_{\mathbf{s}(t)}\cdot y_{v(t-1)})+\\
& (4\cdot\Delta p\cdot\mathbf{F}_{(t-1)}\cdot\mathbf{L}_{(t)}\cdot\mathbf{y}_{\mathbf{s}(t-1)}\cdot y_{v(t)})+\\
& (4\cdot\Delta p\cdot\mathbf{F}_{(t-1)}\cdot\mathbf{L}_{(t-1)}\cdot\mathbf{y}_{\mathbf{s}(t)}\cdot y_{v(t)})+\\
& (6\cdot\Delta p\cdot\mathbf{F}_{(t-1)}\cdot\mathbf{L}_{(t)}\cdot\mathbf{y}_{\mathbf{s}(t)}\cdot y_{v(t)})+\\
& (6\cdot\Delta p\cdot\mathbf{F}_{(t)}\cdot\mathbf{L}_{(t-1)}\cdot\mathbf{y}_{\mathbf{s}(t)}\cdot y_{v(t)})+\\
& (6\cdot\Delta p\cdot\mathbf{F}_{(t)}\cdot\mathbf{L}_{(t)}\cdot\mathbf{y}_{\mathbf{s}(t-1)}\cdot y_{v(t)})+\\
& (6\cdot\Delta p\cdot\mathbf{F}_{(t)}\cdot\mathbf{L}_{(t)}\cdot\mathbf{y}_{\mathbf{s}(t)}\cdot y_{v(t-1)})+\\
& (24\cdot\Delta p\cdot\mathbf{F}_{(t)}\cdot\mathbf{L}_{(t)}\cdot\mathbf{y}_{\mathbf{s}(t)}\cdot y_{v(t)})]
\end{aligned}
$$

2. Data

This study requires two-sets of data. One set are time-series input-output tables and the corresponding water related data, including water consumption, wastewater discharge and pollutants data.

2.1 Input-Output Tables

The time-series input-output tables (IOT) used in this study contain 7 different years: 1992 [11], 1995 [12], 1997 [13], 2000 [14], 2002 [15], 2005 [16] and 2007 [17]. All the tables were in current prices. The different years had different industry classifications –33 sectors for 1992 and 1995; 40 sectors for 1997 and 2000; and 42 sectors for 2002, 2005 and 2007 – however, there was considerable overlap in the classifications. All the tables are aggregated to a uniform classification with 31 sectors (see Table 2 below). The final demand consists of 6 categories: urban households, rural household, government expenditures, fixed capitals investments, change of stock and net flows.

Table 2: Water Extended Input-Output Model Sector Name

1	Agriculture	17	Metals smelting and pressing
2	Coal mining and processing	18	Metal products
3	Crude petroleum and natural gas products	19	Machinery and equipment
4	Metal ore mining	20	Transport equipment
5	Non-ferrous mineral mining	21	Electric equipment and machinery
6	Manufacture of food products and tobacco processing	22	Electronic and telecommunication equipment
7	Textile goods	23	Instruments, meters, cultural and office machinery
8	Wearing apparel, leather, furs, down and related products	24	Maintenance and repair of machinery and equipment
9	Sawmills and furniture	25	Other manufacturing products
10	Paper and products, printing and record medium reproduction	26	Construction
11	Electricity, steam and hot water production and supply	27	Transport and warehousing, Post and telecommunication
12	Petroleum processing and coking	28	Wholesale and retail trade
13	Chemicals	29	Eating and drinking places
14	Nonmetal mineral products	30	Resident services
15	Petroleum processing and coking	31	Education, health and scientific research
16	Chemicals		

This study uses the classification categories of the State Statistical Bureau of 'non-peasants' for urban population and 'peasants' for rural population. Unfortunately, there are gross inconsistencies in the State Statistic Bureau's classification system for urban, rural, and city population, because the system mixes territorial and functional definitions. The definitions have also been changed over time and non-recorded migration from rural to urban areas further distorts the actual residency [18, 19].

2.2 Double deflation process

For the water SDA, all the IOTs are converted from current price into 2002 constant prices using the double deflation method [20-22]. This method has been widely accepted and is advocated by the United Nations [23]. The double deflation method is described as follows,

The IOT in current prices can be represented as

Z	**y**	**x**
v'		
x'		

where the $n \times n$ matrix $\mathbf{Z}$ denotes the intermediate deliveries between production sectors; the vector of $\mathbf{y}$ is the total final demand (including urban households, rural households, government, gross capital formation, changes in stocks, and net export); the vector **x** represents the total sectoral output; **v’** is a row vector of total value added. The IOT in constant prices obtained by using double deflation is

$\mathbf{Z_d} = \hat{\mathbf{d}}\,\mathbf{Z}$	$\mathbf{y_d} = \hat{\mathbf{d}}\,\mathbf{y}$	$\mathbf{x_d} = \hat{\mathbf{d}}\,\mathbf{x}$
$\mathbf{v_d}'$		
$\mathbf{x_d}'$		

The subscript *d* is used to indicate that the corresponding matrices and vectors are in constant prices after the deflation by using the double deflation method. Let p_i denote the ratio of the current price and the base year price, for product *i*. Thus, $100p_i$ is the price index. This study sets 1997 as the base year, and the price indices are adopted from official Chinese price statistics [24]. The price indices are available for four agricultural sectors, 15 industrial sectors, and eight services sectors. Since there is a total of 95 sectors in the IOT, the same index is applied to similar sectors when there was not a direct correspondence. The element d_i of the vector **d** denotes the deflator in sector *i*, which is defined as the reciprocal price ratio ($d_i = 1/p_i$). Therefore, $\mathbf{Z_d}$ is obtained by multiplying the $\hat{\mathbf{d}}$ with **Z**. One can calculate $\mathbf{y_d}$ and $\mathbf{x_d}$ in a similar way. The value added vector $\mathbf{v_d}'$ is then obtained from the balancing equation,

$$\mathbf{v_d}' = \mathbf{x_d}' - \mathbf{e}'_{(n)}\mathbf{Z_d}$$

where $\mathbf{e'(n)}$ is a row vector of ones used for summation of $\mathbf{Z_d}$.

The double-deflation method is used to compile the input-output tables in constant prices. Although the double deflation method is widely accepted, there are three drawbacks related to this approach. Firstly, by adopting this method, most sectors are assumed to produce one homogeneous product, each sector’s gross output and intermediate and final demand are deflated by this sector’s price index [25]. However, most sectors consist of more than one good, therefore to use the price index of certain goods to represent the entire sector is not always appropriate [26]. Secondly, value-added is obtained as the difference between the total input and intermediate input in each sector. Consequently, it is not accurate to use value-added to balance the input-output table after the deflation [27]. Thirdly problem arises from sectoral aggregation: a deflated IO table may be obtained in two alternative ways. The first way is to aggregate after deflation, which means that the original table is deflated first, resulting in a value added vector in constant prices, which then is aggregated. The other way, deflation after aggregation, means that the original table is first aggregated and then deflated. The two methods may produce different results unless very stringent conditions are satisfied [25, 28].

2.3 Freshwater consumption data

As stated in the main text, we obtain the *China’s Annual National Water Bulletins* [29] to provide annual consumption for surface and ground freshwater consumption for agriculture, industry and domestic consumption.

The freshwater consumption data for industrial sectors was available for 1997 [30], 2002 [31] and 2008 [32]. In order to match 7 time-point input-output tables, we scale the water consumption for each industrial sector by the following procedures:

- We obtain the total industry water consumption from the water bulletins for all available years (1992, 1995, 1997, 2000, 2005, and 2007)
- We obtain the total value added for industry for 2007 and 2008
- We back-cast the water consumption for 2007 for each industrial sector by assuming the water consumption technology remains the same between 2008 and 2007. Therefore, we could get the water consumption for all industrial sectors in 2007
- We obtain the value added for each industry sector for 1997, 2002, and 2007.
- We derive the annual water intensity growth for each industry sector during 1997 – 2007.

- We calculate the water consumption of all industrial sectors for 2000 and 2005.
- We assume the same water intensity growth rate for each industry backcasting to 1995 and 1992, so as to derive the water consumption in all industrial sectors for the two years.

The freshwater consumption data for construction and services sectors are not directly available in Chinese statistics. *China Urban Water Supply Bulletin* provides household daily water consumption. That is 171, 118 and 99 liters per person in cities, county and town level. We utilize the population statistics to estimate the water consumption for households' activities (drinking, cooking, washing etc). The difference between the 'residential water consumption' in the *National Water Bulletins* and the households' consumption can be seen as the construction and services water consumption.

We further separate water consumption between construction and services sectors. We use the GDP proportion between the two sectors to estimate the water consumption figures. By doing this, we acknowledge that the assumption of that the water intensities of construction and servicing activities are same.

2.4 Water pollution data

As stated in the main text, we obtain the wastewater and COD pollution from *China's Environmental Statistics Yearbook* [33], which provides annual wastewater discharge (measured by m^3) and COD release (measured by *tonnes*) for all in manufacturing sectors. The sectoral variation is from 13 manufacturing sectors in 1992 to 42 sectors in 2007. We acknowledge the non-point emission sources are excluded from this study as lack of data availability.

Chinese official statistics do not consist of agricultural wastewater / COD discharge. We estimate the agricultural COD discharge by using fertilizer consumption agriculture activities as a proxy. The same treatment was adopted in our previous publication by Guan and Hubacek [2].

Chinese official statistics provide wastewater discharge for residential wastewater discharge. This includes both households' wasterwater discharge as well as construction and service sectors discharge. We adopt the same mechanism as we use in water consumption (described in section 2.3) in order to separate the wastewater and COD discharge for household, construction and service sectors, respectively.

Several parameters in Equation 18 are known based on Chinese statistics, existing literatures about China's hydrology and our past published work. For example, q_0 is average runoff of Chinese surface water between 1992 – 2007, which is 2.5 trillion cubic metres [34]. q_p is amount of annual wastewater discharged, data is available from Chinese Environmental Statistics Yearbook [34]. c_0 is the COD concentration for dilution water requirement, which varies from grade I&II water (15 gram/m^3) to grade IV water (30 gram/m^3) [2]. c_p is calculated based on COD discharges and wastewater volumes statistics provided by Chinese Environmental Statistics Yearbook [34]. $c_{standard}$ is defined as 40 gram/m^3 as the minimum regulatory water quality standard [2]. Natural purification parameter k_1 is 3.64 according to Xie (1996)[3].

2.5 Sector normalisation between IOT and water data

We follow standard procedures for normalization of the IOT and the water and COD discharge data,

$$\mathbf{A} = \mathbf{Z}\,\hat{\mathbf{x}}_{\mathbf{total}}^{\mathbf{-1}}$$

where **A** is the inter-industry requirements matrix which represents the technology of the Chinese economy. We need a mapping between the IO sectors and the water sectors for each year. To normalize the total water data, **T,** we must first aggregate the output to the sector classification used in the water and COD emissions data and then normalize,

$$\mathbf{F} = \mathbf{T}\left(\mathbf{P}\mathbf{x}_{\mathbf{total}}\right)^{-1}\mathbf{P}$$

where P represents the mapping between IO sectors and water sectors. We constructed P from the sector descriptions of the IOT and water data. The post-multiplication by P converts F into an IO sector row vector. This procedure assumes that all IO sectors that map to the one water sector have the same emission intensity. Since the IOT and water intensities now have the same industry classification, we can perform the calculation for each year,

$$\mathbf{f} = \mathbf{F}\left(\mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{y}$$

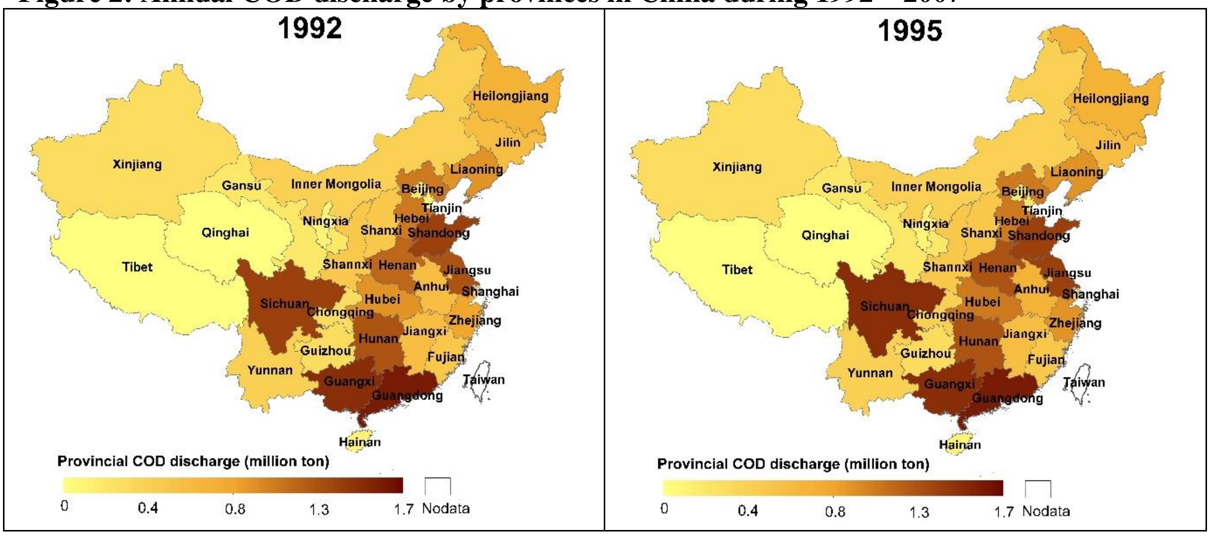
where **y** the final demand under investigation, **A** is the inter-industry requirements matrix which represents the technology of the Chinese economy, **F** is the emission intensity in each sector, and **f** is the water consumption or COD emissions required to produce the final demand.

3. Supporting results

3.1 Spatial distribution of water pollution across China

Figure 2 presents spatial distribution of annual COD discharge across China during 1992 – 2007. Manufacturing bases in coastal regions (e.g. Shandong, Jiangsu, Guangdong etc) account for 1/3 of national COD discharge while heavy industry production provinces (Hebei, Henan, Hunan etc) accounts 1/3. The remaining 1/3 is mainly contributed by western provinces that has initialised industrialisation (e.g. Sichuan and Guangxi) and engaging with labour and pollution intensive raw materials processing, for example, cement production. China's annual COD discharge has reduced 19% from 21 million tons to 17 million tons over the study period. In particular, economically most advanced regions like Beijing and Shanghai have reduced COD discharge by 62% and 32%, respectively, over the same period. Most of other provinces have reduced annual COD discharge by 15% - 25%.

Figure 2: Annual COD discharge by provinces in China during 1992 – 2007



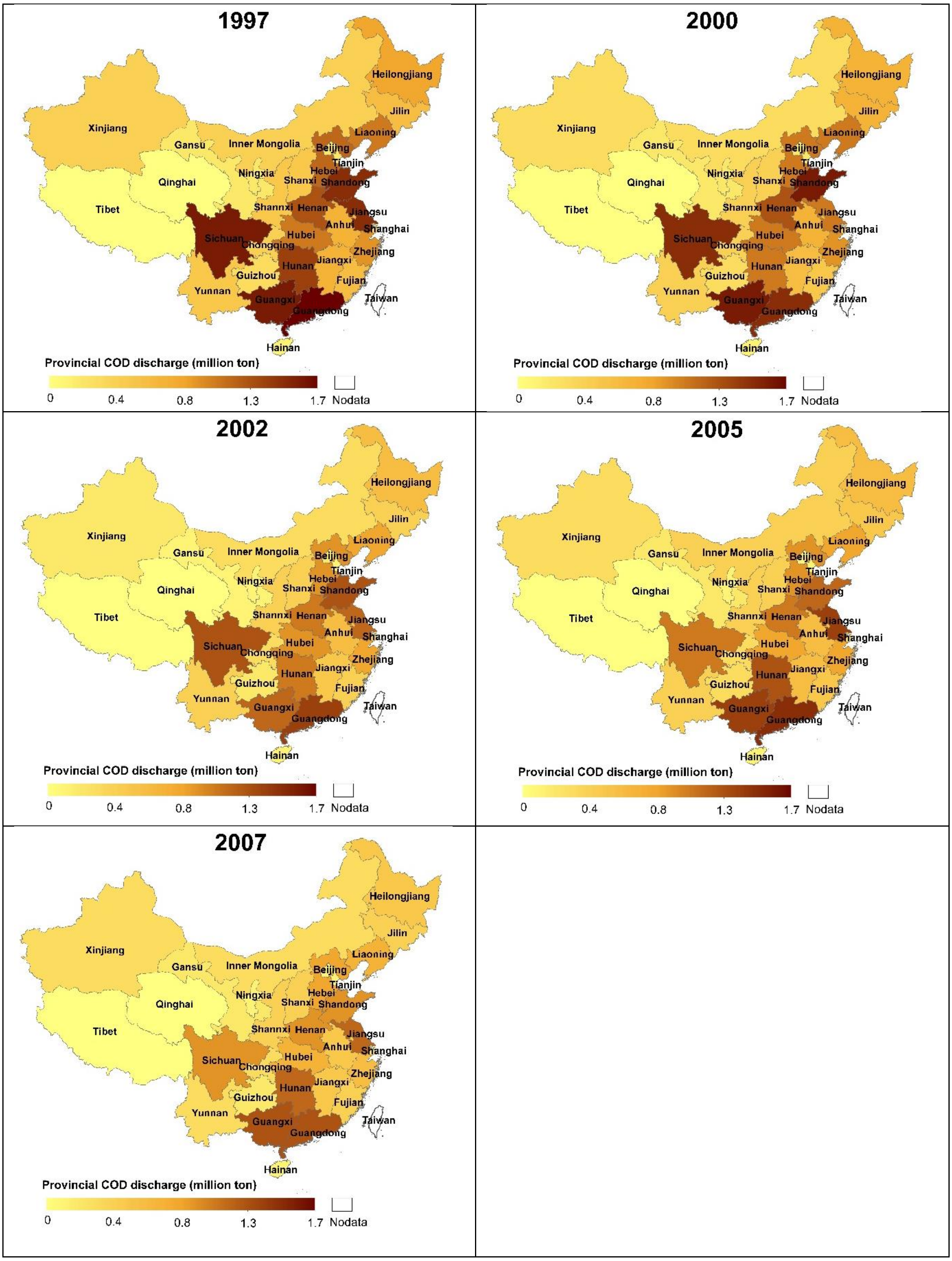
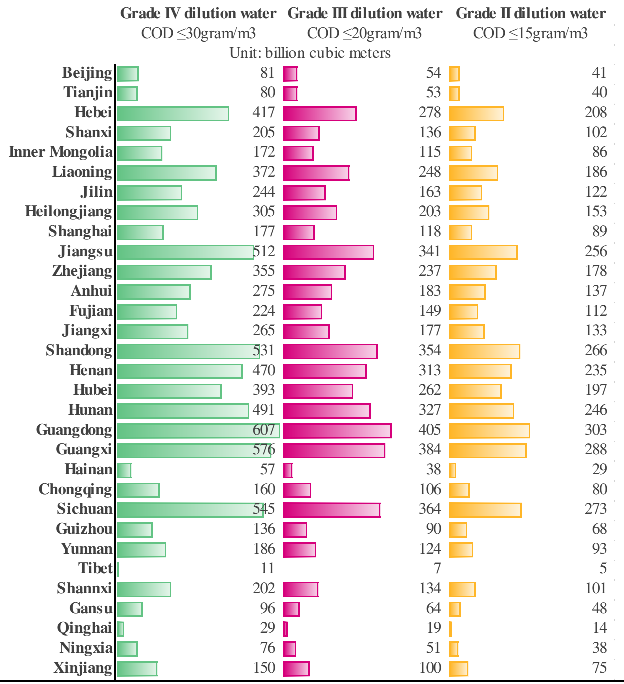


Figure 3 demonstrates the cumulative COD discharges after natural purification among 30 Chinese provinces over 1992 – 2007. The pollution distribution is similar to the annual COD discharges as shown above in Figure 2. COD level are largely accumulated in water resources abundant regions, for example, Guangdong and Jiangsu (economically advanced regions) as well as Sichuan and Guangxi (economically lack behind regions). On the other hand, our results show that cumulative COD discharge in North China (water poor regions) has further worsened the eligible water availabilities.

For example, Huaihe and Haihe region (including Hebei, Shandong, Henan, Anhui provinces) has reached 64 million tons, which is 28% of total pollution accumulations across China during 1992-2007.

Figure 3: Cumulative COD effluences during 1992 – 2007 across Chinese provinces

Figure 4: Dilution water requirements by quality types across Chinese provinces



3.2 Dilution water requirement for cumulative COD pollution

According to Equation 8, variation of initial COD concentration in dilution water (the parameter c_0) would lead to different results in dilution water requirement. Figure 4 illustrates dilution water requirements across 30 Chinese provinces by using different qualities of water resources. By using Grade IV water resources (i.e. COD concentration level is 30 gram/m^3), dilution water requirements to remedy China's cumulative COD effluence during 1992 – 2007 (i.e. 225 million tons of COD) would reach 8.4 trillion m^3 of water resources. By using Grade III water resources (i.e. COD concentration level is 20 gram/m^3), dilution water requirements would reduce to 5.6 trillion m^3. By using Grade II or I water resources (i.e. COD concentration level is 15 gram/m^3), dilution water requirements would further reduce to 4.2 trillion m^3.

Figure 4 illustrates dilution water requirements across 30 Chinese provinces by using different qualities of water resources. During our calculation, there is no distinction about water quality for dilution water requirement across different provinces. In order to dilute 16.3 million tons of cumulative COD pollution during 1992 – 2007 in Guangdong (economically advanced South Coast China, water abundant region) to grade V level, dilution water requirements is 607 billion m^3 of grade IV water or 303 billion grade II water. The cumulative COD pollution in industrialising and water abundant southwest province, Sichuan, is 14.6 million tons during same study period, which would require 545 billion m^3 of grade IV water or 273 billion m^3 of grade II water to fully remedy the polluted water resources to the minimum regulatory level. Shandong province is located in North Coast China with per capita water availability less than 300 m^3 per year. The cumulative COD pollution during 1992 – 2007 is 14.2 million tons. In order to fully dilute those pollution to the grade IV level, it would cost 531 billion m^3 grade IV water or 266 billion m^3 grade II water.

References

1. GAQSIQ; National Emblem, Code of practice for water resources bulletin. In General Administration of Quality Supervision Inspection and Quarantine; National Emblem of the People's Republic of China, Eds. Beijing, 2009.
2. Guan, D.; Hubacek, K., A new and integrated hydro-economic accounting and analytical framework for water resources: A case study for North China. *Journal of Environmental Management* **2008,** *88*, (4), 1300-1313.
3. Xie, Y., *Environment and Water Quality Model*. China Science and Technology Press: Beijing, China, 1996.
4. de Boer, P., Structural decomposition analysis and index number theory: an empirical application of the Montgomery decomposition. *Econometric Institute Report* **2006,** *EI 2006-39*.
5. Hoekstra, R.; van der Bergh, J. C. J. M., Comparing structural and index decomposition analysis. *Energy Economics* **2003,** *25*, 39-64.
6. Dietzenbacher, E.; Los, B., Structural decomposition techniques: Sense and sensitivity. *Economic Systems Research* **1998,** *10*, 307-323.
7. Rørmose, P.; Olsen, T. In *Structural Decomposition Analysis of Air Emissions in Denmark 1980-2002*, 15th International Input-Output Conference, Beijing, 2005.
8. Hoekstra, R.; van der Bergh, J. C. J. M., Structural decomposition analysis of physical flows in the economy. *Environmental and Resource Economics* **2002,** *23*, 357--378.
9. de Haan, M., A structural decomposition analysis of pollution in The Netherlands. *Economic Systems Research* **2001,** *13*, (2), 181-196.
10. Seibel, S. *Decomposition Analysis of Carbon Dioxide Emission Changes in Germany - Conceptual Framework and Empirical Results. European Commission.*; 2003.
11. State Statistical Bureau of China, *1992 Input-Output Table of China* China Statistical Publishing House: Beijing, 1996.
12. State Statistical Bureau of China, *1995 Input-Output Table of China (Extended Version)*. China Statistical Publishing House: Beijing, 1997.
13. State Statistical Bureau of China, *1997 Input-Output Table of China* China Statistical Publishing House: Beijing, 2000.
14. State Statistical Bureau of China, *2000 Input-Output Table of China (Extended version)*. China Statistical Publishing House: Beijing, 2002.
15. National Bureau of Statistics, *2002 Input-Output Table of China* China Statistics Press: Beijing, 2006.
16. State Statistical Bureau of China, *2005 Input-Output Table of China (Extended version)*. China Statistical Publishing House: Beijing, 2008.
17. State Statistical Bureau of China, *2007 Input-Output Table of China* China Statistical Publishing House: Beijing, 2010.
18. Hubacek, K.; Sun, L., A Scenario Analysis of China's Land Use Change: Incorporating Biophysical Information into Input-Output Modeling. *Structural Change and Economic Dynamics* **2001,** *12*, (4), 367-397.
19. Heilig, G. K.; Fischer, G.; Velthuizen, H. v., Can China Feed itself? An Analysis of China's Food Prospects with Special Reference to Water Resources. *International Journal of Sustainable Development and World Ecology* **2000,** *7*, 153-172.
20. de Boer, S.; Broesterhuizen, G., *The Simultaneous Compilation of Current Price and Deflated Input-Output Tables*. Oxford University Press: 1991; p 53-65.
21. Durand, R., An Alternative to Double Deflation for Measuring Real Industry Value-Added. *Review of Income and Wealth* **1994,** *40*, 303-316.
22. Folloni, G.; Miglierina, C., Hypothesis of Price Formation in Input-Output Tables. *Economic Systems Research* **1994,** *6*, 249-264.
23. United Nations, *Handbook of input-output table compilation and analysis. Studies in Methods Series F, No 74. Handbook of National Accounting*. United Nations: New York, 1999.
24. National Bureau of Statistics, *China Statistical Yearbook 2005*. China Statistics Press: Beijing, 2005.

25. Dietzenbacher, E.; Hoen, A. R., Deflation of input-output tables from the user's point of view: A heuristic approach. *Review of Income and Wealth* **1998,** *44*, 111-122.
26. Sevaldson, P., Price Changes as Causes of Variations in Input-Output Coefficients. In *Advances in Input-Output Analysis*, Polenske, K. R.; (eds.), J. V. S., Eds. Ballinger, Cambridge, MA: 1976; pp 113-33.
27. Wolff, E. N., Productivity Measurement within an Input-Output Framework. *Regional Science and Urban Economics* **1994,** *24*, 72-94.
28. Kymn, K. O., Aggregation in Input-Output Models: A Comprehensive Review, 1946-71. *Economic Systems Research* **1990,** *2*, 65-93.
29. Ministry of Water Resources, *China's Regional Water Bullets*. China Water Resource and Hydro-power Publishing House: Beijing, 1997.
30. Guan, D.; Hubacek, K., Assessment of regional trade and virtual water flows in China. *Ecological Economics* **2007,** *61*, (1), 159-170.
31. Zhao, X.; Chen, B.; Yang, Z. F., National water footprint in an input–output framework—A case study of China 2002. *Ecological Modelling* **2009,** *220*, (2), 245-253.
32. National Bureau of Statistics, *China Economic Census Yearbook 2008*. China Statistics Press: Beijing, 2010.
33. National Bureau of Statistics, *China Environmental Statistical Yearbook 1993-2010*. China Statistics Press: Beijing, 1993-2010.
34. National Environmental Protection Agency, *China Environmental Yearbook 1996-2008*. China Environmental Science Press: Beijing, 1996-2008.