



OPEN

Nitrate transport velocity data in the global unsaturated zones

DATA DESCRIPTOR

Congyu Yang¹, Lei Wang²✉, Shengbo Chen¹✉, Yuanyin Li^{2,3}, Shuang Huang⁴, Qinghong Zeng¹ & Yanbing Chen¹

Nitrate pollution in groundwater, which is an international problem, threatens human health and the environment. It could take decades for nitrate to transport in the groundwater system. When understanding the impacts of this nitrate legacy on water quality, the nitrate transport velocity (v_N) in the unsaturated zone (USZ) is of great significance. Although some local USZ v_N data measured or simulated are available, there has been no such a dataset at the global scale. Here, we present a Global-scale unsaturated zone Nitrate transport Velocity dataset (GNV) generated from a Nitrate Time Bomb (NTB) model using global permeability and porosity and global average annual groundwater recharge data. To evaluate GNV, a baseline dataset of USZ v_N was created using locally measured data and global lithological data. The results show that 94.50% of GNV match the baseline USZ v_N dataset. This dataset will largely contribute to research advancement in the nitrate legacy in the groundwater system, provide evidence for managing nitrate water pollution, and promote international and interdisciplinary collaborations.

Background & Summary

Only 3% of the total water on the Earth is considered fresh water and approximately 30% of that is accessible as groundwater, which is vital for human development, ecosystem, the energy industry and other water-dependent activities¹. Since the 1950s, it has been realised that nitrate (NO_3^- -N), which is the most common groundwater pollutant worldwide²⁻⁴, adversely affects human health^{5,6}. Studies have shown a positive correlation between nitrate concentration in drinking water and the colorectal cancer morbidity when the drinking water quality is far below the drinking water standard (50 mg/L of nitrate as NO_3^- in the European Union⁷, or 10 mg/L of nitrogen as the maximum contaminant level regulated by the United States Environmental Protection Agency) set by policies^{8,9}. Nitrate has also been considered to be an environmental endocrine disruptor, as it has been shown to affect vertebrate reproduction and developmental processes in fishes^{10,11}. Nitrate entering wetlands, rivers or lakes can lead to eutrophication, which may lead to algae overgrowth and fish loss¹²⁻¹⁴. Moreover, nitrate has an indirect impact on the economy. Studies from the early 1990s showed that in response to groundwater pollution, many people took avoidance actions that can result in significant economic losses¹⁵. For example, in Wisconsin, USA, the direct medical cost for all adverse health consequences attributable to nitrates is estimated at between \$23 million and \$80 million per year¹⁶.

The main sources of nitrate in groundwater that cause these hazards include irrigated and rainfed agriculture and intensive animal farming¹⁷. Other sources, such as septic tanks and landfills, may leach nitrate locally¹⁸. In some urbanized areas, underground sewer leakage is also a source of nitrate in groundwater¹⁹. Nitrate pollution in shallow aquifers is mainly caused by fertilisation and the subsequent nitrate leaching^{20,21}, which is the process of nitrate migration from the upper to the lower soil with soil water. Nitrate leaves the bottom of the soil into the unsaturated zone (USZ) and finally enters the groundwater. The USZ is located below the soil and above the groundwater, which is not only an important space connecting the surface and groundwater, but also a necessary way for all kinds of pollutants to enter the groundwater. After nitrate enters the USZ from the bottom of the soils, the transformation of nitrate in the USZ mainly includes three processes, namely, adsorption²², nitrification and denitrification²³. Recent literature has indicated an increasing global concern about the effects of nitrate leaching on the environment, particularly agro-ecosystems²⁴, especially the nitrate legacy in the USZ, i.e., the nitrate time lag between the bottom of the soil layer and arrival at the water table²⁵. Some studies have termed this issue a

¹College of Geo-exploration Science and Technology, Jilin University, Changchun, China. ²British Geological Survey, Keyworth, Nottingham, NG12 5GG, United Kingdom. ³Department of Geography, Durham University, Durham, DH1 3LE, United Kingdom. ⁴MCC Smart City (Wuhan) Engineering Technology CO., Ltd, Wuhan, China. ✉e-mail: lei.wang@bgs.ac.uk; chensb@jlu.edu.cn

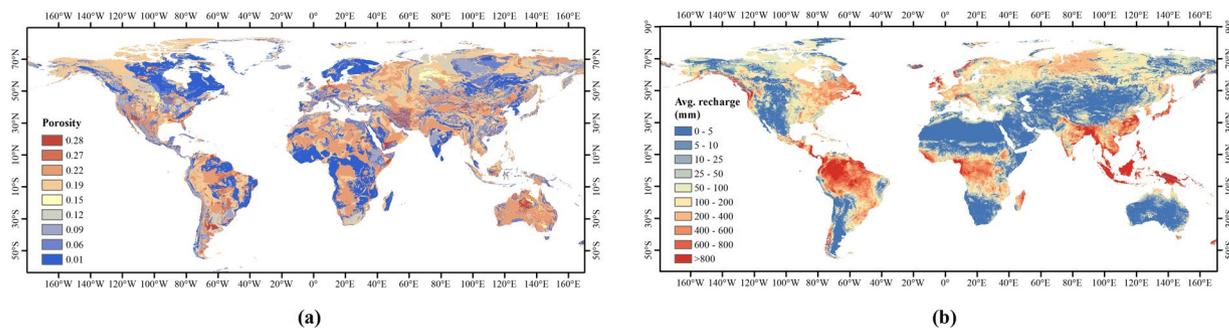


Fig. 1 Input datasets for the NTB model. **(a)** Global porosity database. Different porosity values correspond to different lithologies, including unconsolidated sediments (0.22), coarse-grain unconsolidated sediments (0.28), fine-grain unconsolidated sediments (0.15), siliciclastic sedimentary (0.19), coarse-grain siliciclastic sedimentary (0.27), fine-grain siliciclastic sedimentary (0.12), carbonate (0.06), crystalline (0.01), and volcanic (0.09). The database is available at <https://doi.org/10.5683/SP2/DLGXYO>⁶⁶. **(b)** Global groundwater average recharge from 1958 to 2015. The original dataset is available at <https://opendap.4tu.nl/thredds/catalog/data2/pcrglobwb/catalog.html>.

‘nitrate time bomb’²⁶ and indicate that countries should consider it when assessing groundwater nitrate pollution and developing pollution control policies²⁷.

To understand the nitrate legacy in the groundwater system, it is necessary to understand the nitrate transport velocity (v_N) in the USZs and hence the nitrate lag time in the USZs. In previous studies, v_N was regarded as one of the main factors affecting the nitrate concentration and distribution in the USZs of the study areas²⁸, and nitrate was also regarded as an environmental tracer to understand the transfer processes in the USZs²⁹. However, the v_N in the USZs involved in these researches is limited to specific local research areas. In terms of global-scale research, there are few studies^{28–37} on v_N estimation, and most of the research is concentrated on European aquifers^{35–37}, especially British aquifers³⁸. Although v_N maps for the UK³⁸, the Loess Plateau of China³⁹ have been generated, there is no spatial map representing the v_N distribution for the whole world. Since the USZ v_N is determined by many factors, such as rock types, permeability, porosity, and amount of groundwater recharge, it is highly regional or lithological specific⁴⁰ thus making it difficult to generate a reliable global dataset of v_N the USZs.

Based on a Nitrate Time Bomb (NTB) model³⁸, we developed a global dataset of nitrate transport velocity in the USZs (GNV) and validated it using nitrate velocity data locally observed or derived from the literature review. This first known and open-source global-scale USZ v_N dataset GNV can help scientists from other disciplines to better understand the nitrate legacy in the groundwater system at a large scale, thus contributing to developing new methods to provide sound evidence for nitrate water pollution management.

Methods

The development of the GNV consists of three steps: (1) Constructing an NTB model by preparing and inputting global datasets of rock permeability, rock porosity, and annual groundwater recharge. (2) Calibrating the NTB model based on a 22-zone baseline dataset of USZ v_N created using locally measured v_N data and global lithological data. (3) Validating the GNV dataset derived using the baseline USZ v_N dataset.

The Nitrate Time Bomb (NTB) model. The NTB model has been used to simulate the nitrate transport in the groundwater system at the national and global scales⁴¹, based on the information on nitrate leaching from the bottom of the soils, the thickness of the USZs, and the rock hydrogeological characteristics. The NTB model was used in this study to derive the GNV dataset. The NTB model was originally developed in the UK, where the transport velocity in the USZs was calculated as the ratio of average groundwater recharge to porosity⁴²:

$$V_{USZ,i} = \frac{R_{ec,i}}{P_{rock,i} \cdot R \cdot 1000}$$

where, $V_{USZ,i}$ (m/year) is the USZ v_N at the cell i ; $R_{ec,i}$ (mm/year) is the groundwater recharge in the cell i ; $P_{rock,i}$ is the porosity of the rock at the location of i ; and R is the retardation factor reflecting the influences of other factors, such as permeability, pore size, diffusion, dispersion and adsorption on the USZ v_N .

Global porosity data for constructing the NTB model. The global porosity data used in this study were derived from the GLHYMPS (Fig. 1a), which is global near-surface hydrogeology data of permeability and porosity produced by synthesising and modifying existing global databases^{43,44}. The nine classes of porosity data, which have an average polygon size of 107 km² (including Antarctica), are corresponding to nine hydrogeological categories, i.e. unconsolidated sediments, coarse-grain unconsolidated sediments, fine-grain unconsolidated sediments, siliciclastic sedimentary, coarse-grain siliciclastic sedimentary, fine-grain siliciclastic sedimentary, carbonate, crystalline, and volcanic.

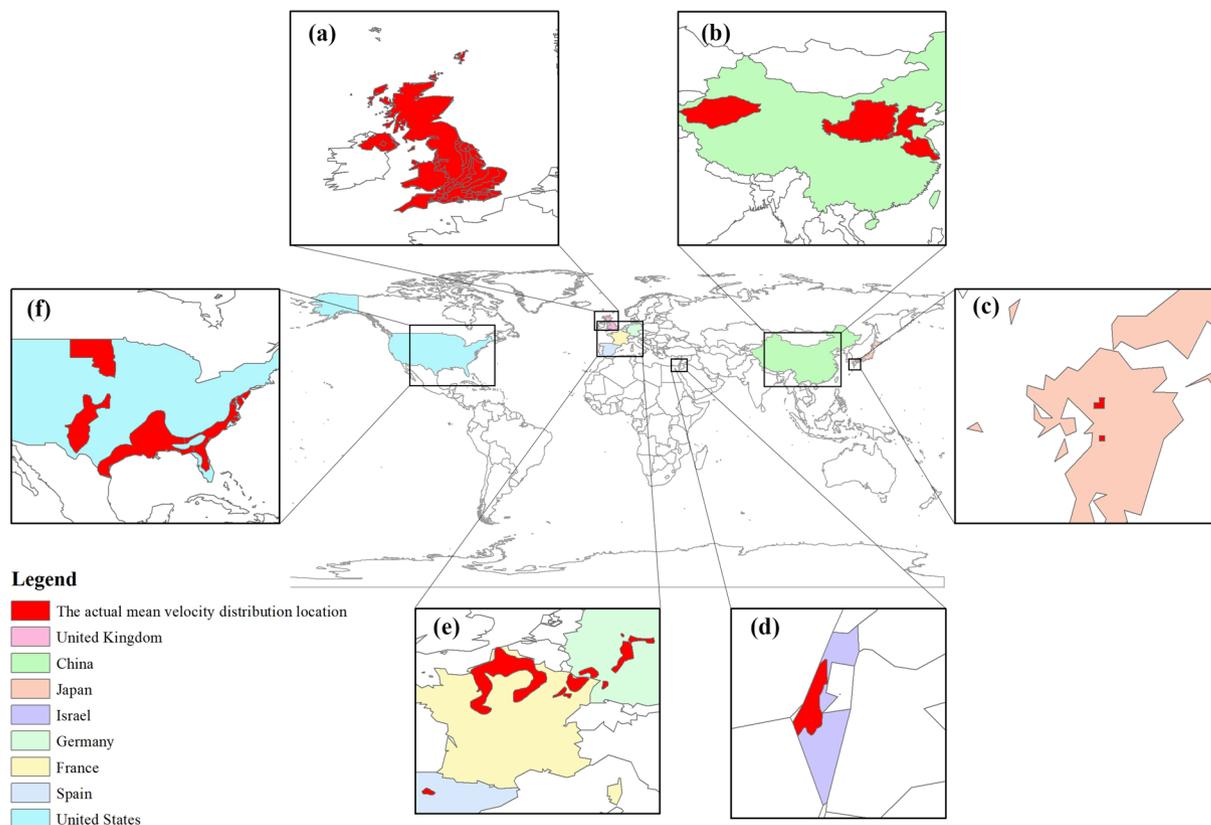


Fig. 2 The spatial distribution of the mean USZ v_N data collected in this study. The red region represents the region where the mean USZ v_N exists (the mean USZ v_N maybe one or more). (a) The UK's multiple mean USZ v_N values cover almost the whole of the UK; (b) The Tarim Basin, Loess Plateau and North China Plain in China have the same average USZ v_N ; (c) The mean USZ v_N of the metamorphic rocks in the Kumamoto region of Japan; (d) The average USZ v_N of the Loess in Israel; (e) The mean USZ v_N values of the Chalk and Triassic sandstone in Western Europe; and (f) The average USZ v_N of the US Loess.

Global annual groundwater recharge data for constructing the NTB model. The global annual groundwater recharge data used in this study were derived from a global hydrological and water resource model called PCR-GLOBWB^{45,46}, which has spatial resolutions of $0.5^\circ \times 0.5^\circ$ and $5' \times 5'$ and is available at https://github.com/UU-Hydro/PCR-GLOBWB_model^{47,48}. Similar to other large-scale hydrologic models, PCR-GLOBWB is essentially a “leaky bucket” model applied on a cell-by-cell basis⁴⁹ by considering rainfall, evaporation, canopy interception, snow accumulation and snowmelt. The monthly groundwater recharge derived from PCR-GLOBWB was used to calculate the annual average recharge from 1958 to 2015 (Fig. 1b).

Regionally measured or modelled USZ v_N data and global lithological data for generating the global-scale baseline USZ v_N . To generate a global baseline dataset of USZ v_N , the measured or modelled (but verified) v_N data of regional USZs in different countries were collected and averaged from published literature (Supplementary Table 1)^{28–38,50,51}. Figure 2 shows the distribution of the collected mean USZ v_N data from the United States, China, the UK, Western Europe, Japan and Israel. These data were then expanded to a global-scale baseline USZ v_N dataset based on the regional lithology and the global lithology data (GLiM)⁵². GLiM, which is available at the PANGAEA Database (<https://doi.org/10.1594/PANGAEA.788537>)⁵³, represents the rock types of the Earth surface with 1,235,400 polygons. The lithological classification consists of three levels: the first level contains 16 basic lithological classes, while the other two levels contain 12 and 14 subclasses respectively describing more rock details. Only 16 basic classes of the first level of GLiM were used in this paper, including: Intermediate volcanic rocks, Basic volcanic rocks, Acid plutonic rocks, Metamorphics, Unconsolidated sediments, Siliciclastic sedimentary rocks, Basic plutonic rocks, Intermediate plutonic rocks, Mixed sedimentary rocks, Water Bodies, Pyroclastics, Carbonate sedimentary rocks, Acid volcanic rocks, No Data and Evaporites. According to the GLiM, the Earth is covered by 64% sediments (a third of which are carbonates), 13% metamorphics, 7% plutonics, 6% volcanics, and 10% are covered by water or ice⁵².

Baseline datasets of USZ v_N for calibrating and validating the NTB model. The first level of GLiM classification was used in this study as a base map to interpolate the regionally measured or modelled (but verified) USZ v_N data into a global baseline dataset of USZ v_N (v_{N_base}) (Fig. 3), which is used as observed/known v_N to calibrate the NTB model. Figure 4 shows the flowchart of generating the v_{N_base} . According to the principle of the NTB model, the v_N is constrained by USZ lithology conditions, so we assumed that the same USZ lithology had

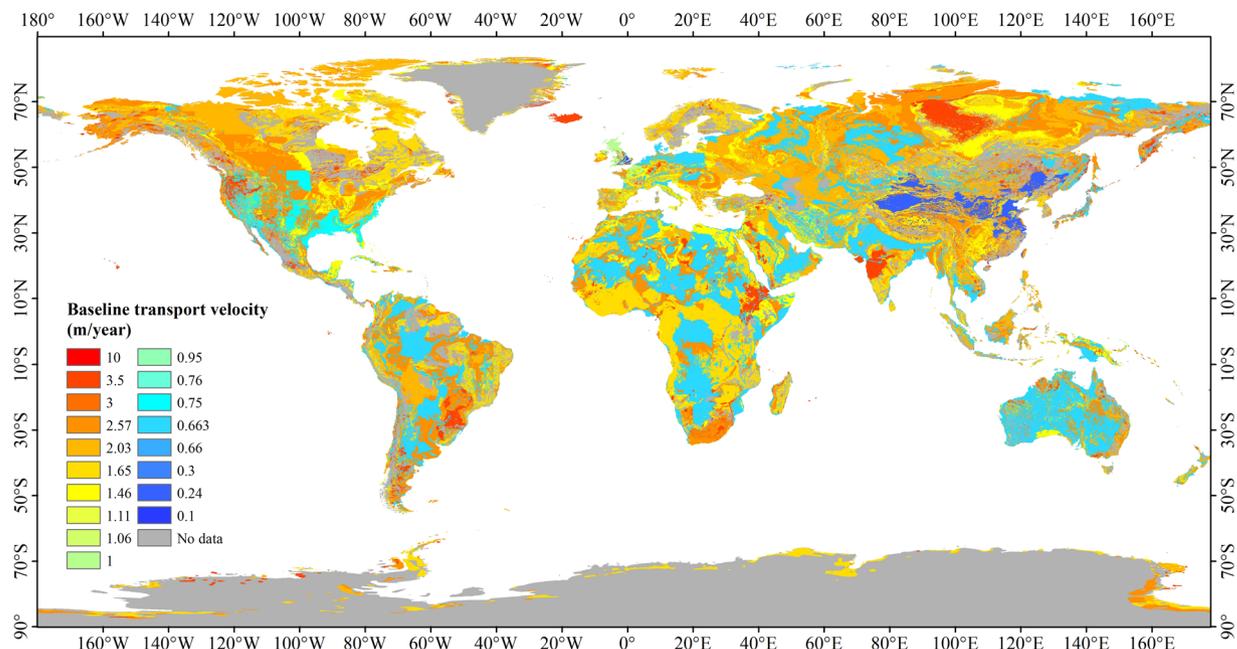


Fig. 3 The baseline USZ v_N data that contain the average nitrate transport velocities measured or modelled in the regions.

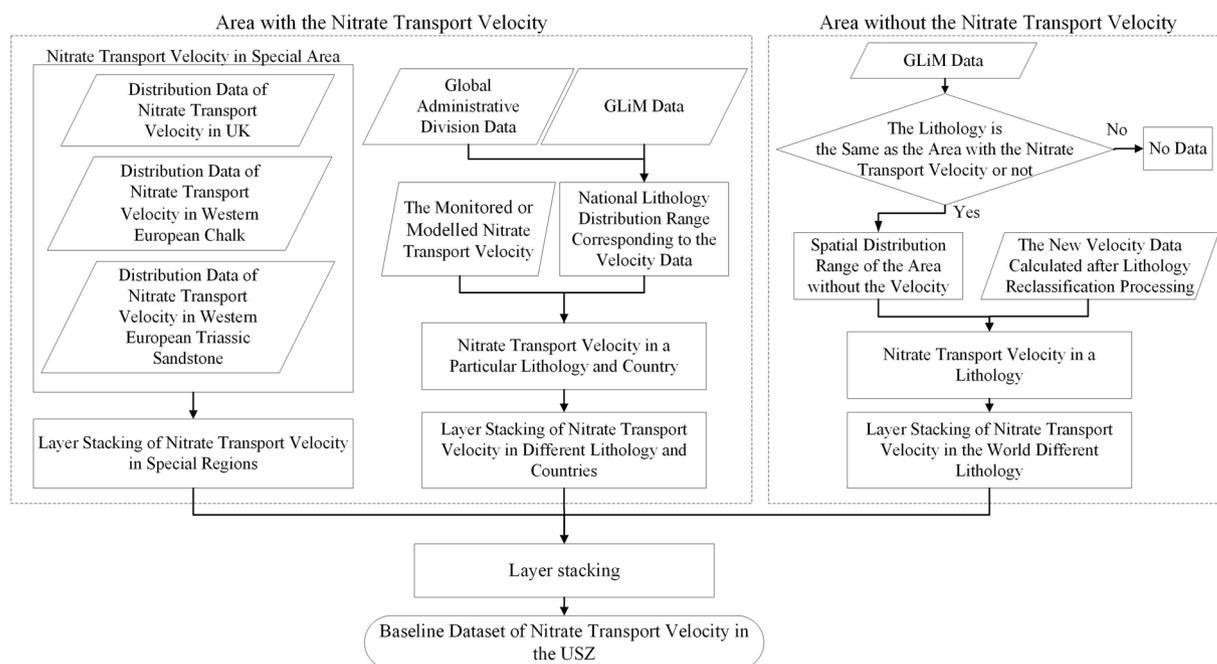


Fig. 4 The flowchart of generating a global baseline dataset of USZ v_N .

the same average USZ v_N . The collected regional monitored or modelled USZ v_N data and their corresponding USZ lithologies were collated, and the world was divided into two parts according to the existence of USZ v_N , namely, the regions with and without v_N data. For the regions with v_N data, we divided them into regions with different lithology classifications and then reclassified the lithology of these regions based on the GLiM classification (Supplementary Table 2), to calculate the mean v_N values of the reclassified lithology. Whilst, for the regions without v_N data, we derived the v_N values based on the lithology types that are the same as that in regions with v_N data. Finally, the global baseline dataset of USZ v_N was generated using the mean v_N data from all the regions.

The data processing of monitored or modelled USZ v_N collected was mainly divided into three parts: (A) the USZ v_N in the UK; (B) the USZ v_N in Chalk and Triassic sandstone of Western Europe; and (C) the USZ v_N in other regions. Their data processing are described below:

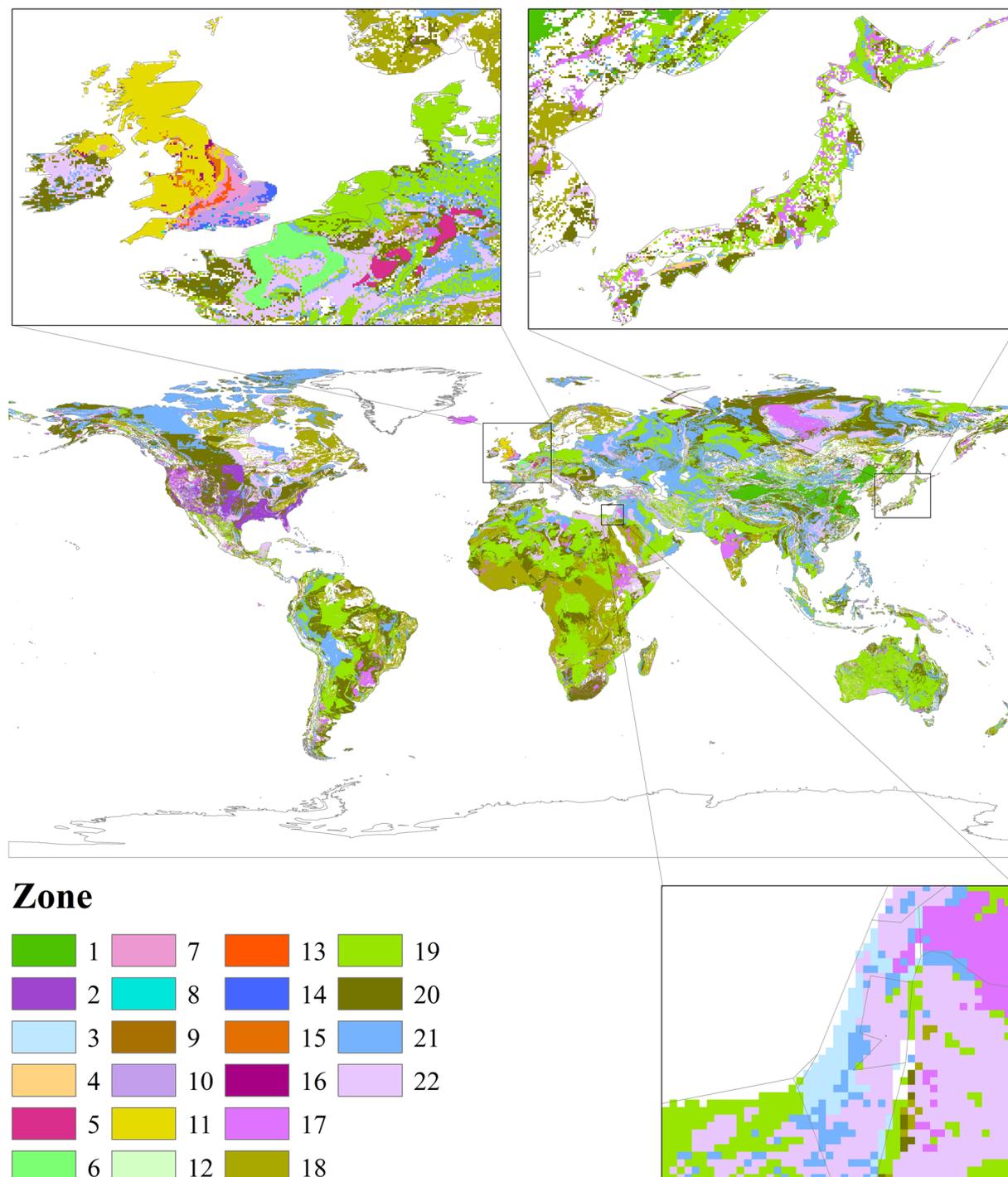


Fig. 5 The global 22 USZ zones that exclude the areas covered by water, ice and snow, or have no v_{N_base} values.

- (A) Since the UK has a complete database of USZ v_N with a detailed description of aquifers that cover almost the whole country³⁸, this UK database was used to derive the mean USZ v_N of other regions in the world based on aquifer types. Therefore, the UK aquifers were reclassified using the basic lithological classification standard of the global GLiM data (Supplementary Table 2). For example, according to the spatial distribution, Chalk, Carboniferous, Cornbrash and Great Oolite of Lincolnshire and other lithology in the UK belong to the Carbonate sedimentary rocks defined in the GLiM basic lithology. Lower Cretaceous Sands, Triassic Sandstones, Triassic and Permian and other lithology belong to the Siliciclastic sedimentary rocks of the GLiM basic lithology. The Pliocene: Corraline Crag and Quaternary Norwich and Red Craggs belong to the Mixed sedimentary rocks of the GLiM basic lithology. When more than one UK USZ lithologies were classified into one GLiM lithological type after the reclassification, the mean v_N value of these USZ lithologies was calculated and applied to calculate the mean USZ v_N of other parts of the world.

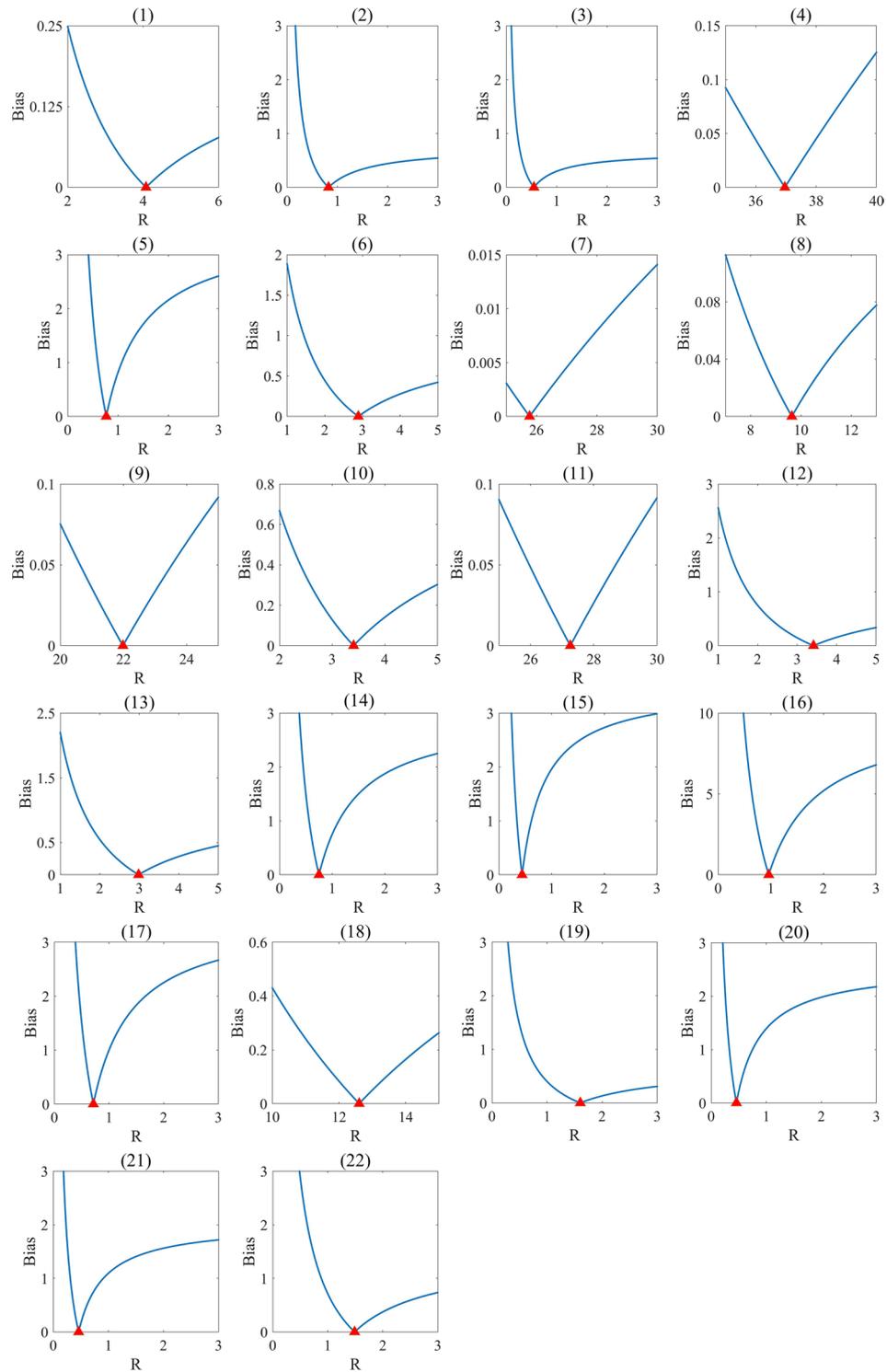


Fig. 6 Sensitivity scatter plots for 22 zones. The bias values are the absolute values of the differences between the mean simulated results and the baseline values (v_{N_base}); the blue lines consist of dots representing the bias values of 100,000 MC runs; the red triangles are the best retardation factor (R) values for 22 modelling zones (1–22).

- (B) Because of special lithological classifications in Western Europe⁵⁴ (Belgium, the former Federal Republic of Germany, Denmark, France, Ireland, Italy, Luxembourg, the Netherlands and the United Kingdom), the v_N values in USZs of the Chalk and Triassic sandstone in Western Europe were determined based on the lithological classification of Western Europe.
- (C) Regarding other countries that have USZ v_N collected from literature, the World Administrative Region data^{55,56} and the lithology classification of GLiM were used to extrapolate the USZ v_N values at the study areas in the literature to the lithological range within the boundary of the countries where the studies were undertaken (Supplementary Table 2).

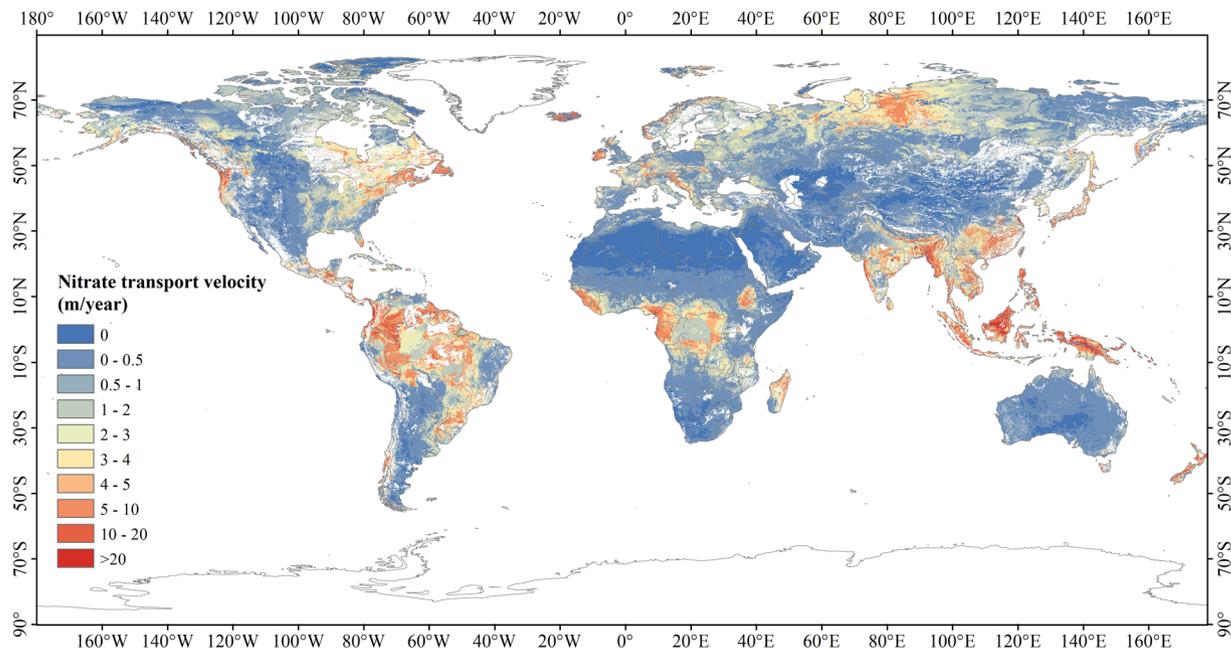


Fig. 7 Global distributed USZ v_N data (GNV) generated in this study.

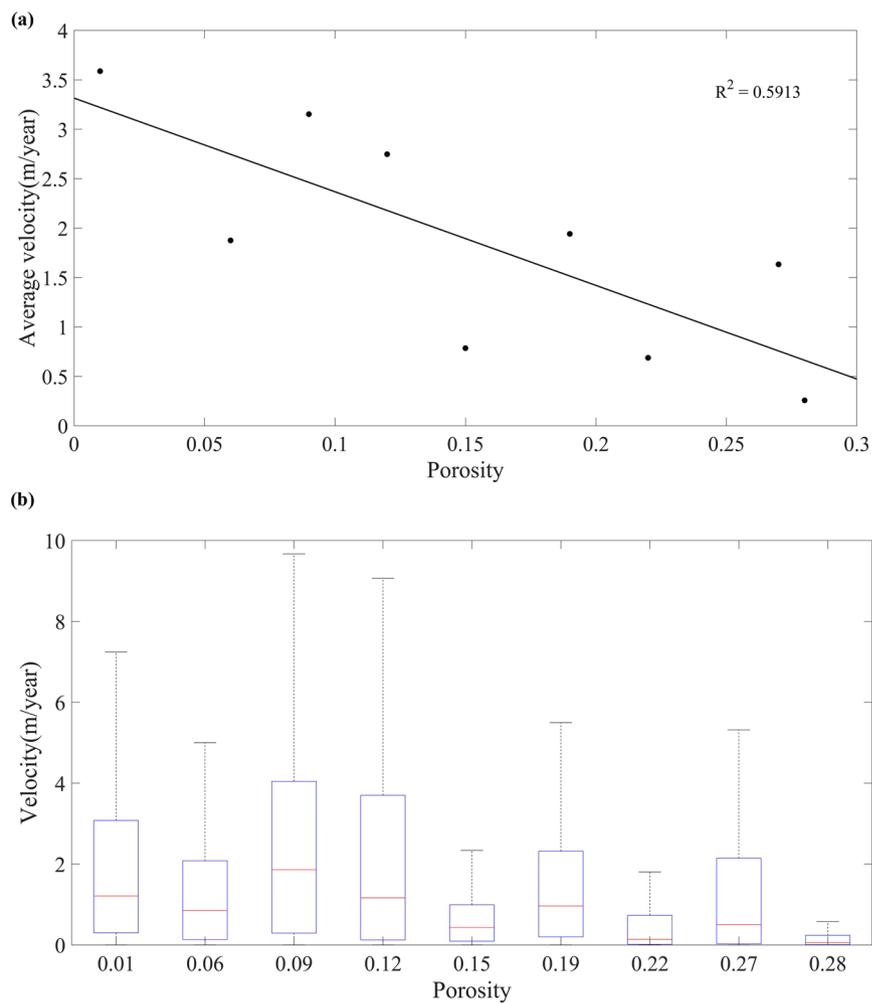


Fig. 8 Comparison of the global rock porosity data with the v_N . (a) the points in the plot represent the average v_N in different porosity zones and their corresponding porosity values; (b) the statistical distribution of v_N values at different porosity values.

Porosity	Grain size	Sediment subtype ⁵⁹	Unit proportion(%)	Clay&Silt proportion(%)
0.15	Sand+	Glacio-fluvial	7.17	41.81
		Till	24.23	
	Sand/Silt	Loess derivative	24.57	
	Sand/Clay	Fluvial-lacustrine	29.01	
	Silt	Loess	15.02	
0.22	Sand+	Alluvial terrace deposits	1.83	42.99
		Dune sands	18.22	
		Glacio-fluvial	4.93	
		Till	9.65	
	Sand/Silt	Loess derivative	7.86	
		Peat	5.36	
	Sand/Clay	Fluvial-lacustrine	19.31	
		Alluvial/Colluvial	12.23	
	Silt	Loess	10.29	
	Silt/Clay	Fluvial-eolian	2.43	
		Glacio-lacustrine	0.49	
		Salt	1.34	
Clay	Floodplain deposits	6.06		
0.28	Sand+	Dune sands	68.66	31.34
	Silt	Loess	31.34	

Table 1. When the porosity values are 0.15, 0.22 and 0.28, the corresponding type, particle size and unit proportion of unconsolidated sediments are presented. In Table 1, Unit proportion and Clay&Silt proportion represent statistics for Sediment subtype units based on data presented in Fig. 1a and the GUM database. Unit proportion is the proportion of the grid unit area of unconsolidated sediment subtype classification to the grid unit area of the corresponding porosity. Clay&Silt proportion is the proportion of the total grid unit area of clay and silt to the grid unit area of the corresponding porosity.

Zone	Max- v (m/year)	Min- v (m/year)	Avg- v (m/year)	difference
1	54.3173	0.0000	0.3354	0.0954
2	208.5967	0.0000	0.8044	0.0544
3	4.8257	0.0000	0.6954	0.0354
4	10.7158	0.0000	1.6582	0.0082
5	43.7150	0.0137	3.7747	0.2747
6	3.2351	0.0516	0.9301	-0.0699
7	0.5658	0.0226	0.1009	0.0009
8	0.6267	0.0972	0.3168	0.0168
9	3.0684	0.1431	0.9500	0.1900
10	2.4700	0.1655	0.9677	0.0177
11	10.0681	0.0000	1.0141	0.0141
12	3.4465	0.5573	1.1189	0.0589
13	2.4656	0.1783	1.1203	0.0103
14	12.8182	0.6826	3.0822	0.0822
15	14.6052	1.2340	3.5547	0.0547
16	68.8014	0.7753	10.1167	0.1167
17	327.5643	0.0000	3.7971	0.2971
18	33.7497	0.0000	1.7879	0.1379
19	166.7290	0.0000	0.8138	0.1508
20	512.0000	0.0000	2.8921	0.3221
21	588.2890	0.0000	2.3786	0.3486
22	168.9150	0.0000	1.7125	0.2525

Table 2. The maximum (Max- v), minimum (Min- v), and average (Avg- v) values of v_N and the difference between the average v_N and the v_{N_base} in 22 zones, based on data presented in Fig. 5 and the GNV dataset presented in Fig. 7.

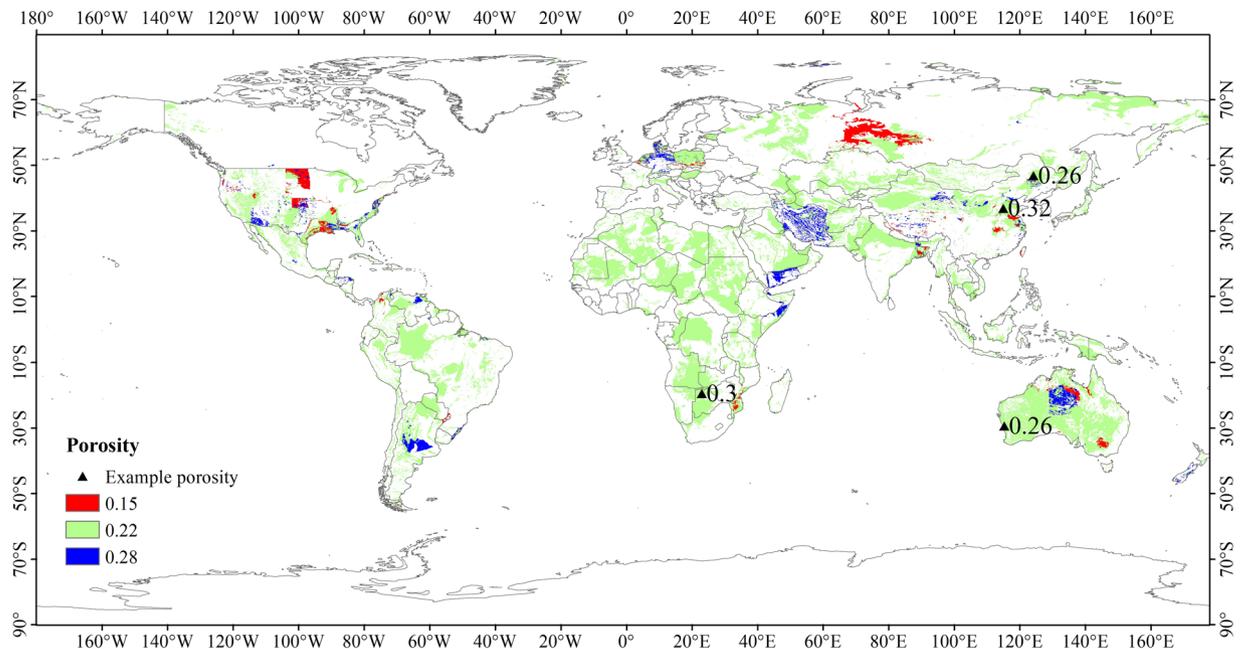


Fig. 9 The distribution of the example porosity values. The NTB model uses porosity values of 0.15, 0.22 and 0.28; and the triangle represents porosity values from literature.

The main factors affecting the value of the retardation factor R include permeability, pore size, diffusion, dispersion and adsorption, which are constrained by lithology⁴². To accurately simulate spatially distributed USZ v_N values, the R values for different lithological classifications need to be calibrated using v_{N_base} . Therefore, according to the GLiM lithology classification and the v_{N_base} of different countries, the global USZs were divided into 22 zones (excluding water, ice and snow, and no v_{N_base} value zones) (Fig. 5). The zoning method is as follows: for the whole country where there is a mean USZ v_N and the v_{N_base} in the area is the mean USZ v_N (e.g., the UK), we divide the region with the same v_{N_base} into one zone. For areas where there is a mean USZ v_N of lithology, the v_{N_base} in the area is the mean USZ v_N and the lithology boundary across several countries (such as Chalk and Triassic sandstone in Western Europe), we divide the lithology into one zone according to the boundary. For areas where there is a mean USZ v_N of lithology, the v_{N_base} in the area is obtained by using GLiM lithology to expand the space according to the subordinate relationship between the lithology and GLiM 16 lithology, and the lithology boundary exists only in one country (such as China, the United States, Japan and Israel), we divide the GLiM lithology corresponding to this lithology in this country into one zone. For example, there is a mean USZ v_N in the Loess of China, and the loess region belongs to the unconsolidated sediments of GLiM. We divide the unconsolidated sediments of China into one zone. The other areas where there is no mean USZ v_N and the v_{N_base} is obtained by interpolation are divided according to GLiM lithology. The division of 22 zones is based on the existence of the mean USZ v_N data, the calculation method of v_{N_base} and lithology. Compared with GLiM 16 lithology classifications, the 22-zone zoning method distinguishes the region where v_{N_base} is obtained by using different methods according to the mean USZ v_N in the same lithology, to better restrict the value of retardation factor of v_{N_base} directly obtained from the existence of mean USZ v_N in the region, thus increasing the accuracy of the velocity simulation results. The number and the lithology of the 22 zones are shown in Supplementary Table 3. The zone map provided regional constraints for deriving spatially distributed USZ v_N values using the NTB model.

Generating the global distributed USZ v_N data (GNV). Although some regional monitored USZ v_N data can be found from published literature, the number of these data are too limited to be directly used to derive the spatially distributed USZ v_N values in the rest of the regions of the world. However, these collected regional monitored USZ v_N data have been used to derive the baseline datasets of USZ v_N , i.e., the v_{N_base} dataset for calibrating the NTB model, which was used in this study to generate the global distributed USZ v_N data (GNV). To calibrate the NTB model using v_{N_base} in 22 zones (described in the section above), the different values of NTB retardation factors R were used and calibrated in each zone during the Monte Carlo (MC) simulation, in which, the NTB model was run 100,000 times. In each NTB run, the absolute value of the difference between the baseline datasets and the spatially distributed mean simulated values was calculated to verify the accuracy of the modelled results. The sensitivity scatter plots of the 22 zones were produced by plotting the absolute value of the difference between the baseline datasets and the spatially distributed mean simulated values of the NTB retardation factors (Fig. 6). For example, in Fig. 6(1), the number 1 corresponded to the zone1. We used the MC method to enter a random R value as R_i , ran the NTB model once, and got a mean simulated velocity (v_{N_sim}) of the zone1. The absolute value of the difference between the mean v_{N_sim} and the zone1 v_{N_base} was marked with a blue point in Fig. 6(1). The MC model had been run for 100,000 times and a total of 100,000 R_i and 100,000 scatter points had

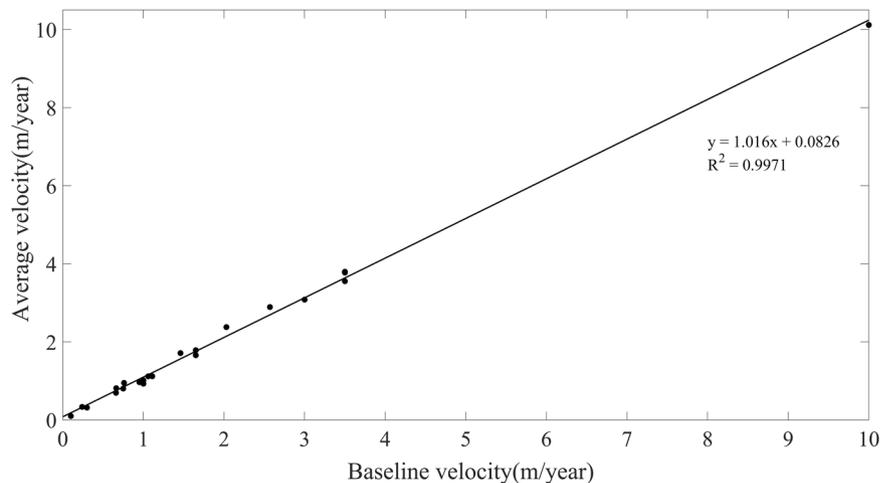


Fig. 10 The scatterplot of correlation between the average GNV and the v_{N_base} .

been obtained. Among these scatter points, the point with the value closest to 0 indicated that the mean v_{N_sim} is closest to the v_{N_base} , and we called this mean v_{N_sim} as the best mean v_{N_sim} . The R_i corresponding to this point was the best R of zone 1, marked with a red triangle in Fig. 6(1). The values of v_{N_base} , the simulated velocity (v_{N_sim}) and retardation factor in 22 regions are shown in Supplementary Table 3. After the R values of 22 zones were determined, the NTB model was run again, and the GNV dataset was obtained. The GNV dataset generated using the NTB model is shown in Fig. 7.

Data Records

The GNV dataset and its quality details are made available to the public free of charge in GeoTIFF format through an unrestricted public repository (Figshare⁵⁷). The data is provided in a $5' \times 5'$ spatial resolution with the velocity unit of m/year. The GNV dataset represents the global distribution of USZ nitrate transport velocities, which are mainly affected by rock types, rock hydrogeological characteristics, long-term groundwater recharge, etc. Data quality information, which will be discussed in the following section, is the precision estimation of nitrate transport data based on the v_{N_base} values in different zones. Upon the availability of new regionally measured or modelled USZ v_N data, the repository will be updated with a newer version of the nitrate transport data graph.

Technical Validation

Since the global rock porosity dataset is one of the input parameters when estimating USZ v_N in the NTB model, the correlation analysis of USZ v_N distribution and porosity data are performed. In order to eliminate the influence of zero groundwater recharge on this analysis, the zero USZ v_N results calculated by zero average annual groundwater recharge (e.g. Sahara Desert, Arabian Desert, Iranian Desert, Turkish Desert, Taklimakan Desert, Gobi Desert, Australian Desert, Namib Desert and Karari Desert) were not considered. Figure 8a shows that the mean value of v_N is inversely proportional to the rock porosity on the whole; and this is consistent with the basic formula of the NTB model⁴². However, when the porosity values are 0.15, 0.22 and 0.28, the value ranges of v_N are smaller than that of other porosity values (Fig. 8b). To explain this phenomenon, we checked the lithology classification of GLiM and found that these three porosity values belong to the same lithology category, namely unconsolidated sediment⁵⁸. We compared the spatial distribution of these three kinds of porosity with the spatial distribution⁵⁹ of unconsolidated sediment subtypes in the Global unconsolidated sediment Map Database (GUM)⁵⁸. Through statistical comparative analysis, it was found that under the condition of excluding undifferentiated sediments, the unit area of clay and silt (assuming that different particles in the mixture were mixed in the same volume) corresponding to these three kinds of porosity accounts for more than 30% of the corresponding porosity area (Table 1). This shows that there was a large amount of clay and silt in the unconsolidated sediments with porosity values of 0.15, 0.22 and 0.28. The reference values of porosity of clay and silt are 0.4~0.7 and 0.8 respectively⁶⁰, which are much higher than the 0.15, 0.22 and 0.28 used to calculate the v_N . In order to verify the accuracy of this conclusion, we obtained the example porosity values from literature (Fig. 9)^{61–64}. Figure 9 shows that it is possible that the actual porosity values can be higher than that used in the NTB model. Based on the above analysis, the actual porosity of the unconsolidated sediments may be higher than 0.15, 0.22 and 0.28 used in the NTB model, thus leading to overestimating v_N . However, the v_N calculation uncertainties, which were introduced by porosity errors, can be reduced due to the existence of the retardation factor in the NTB model. When calibrating the NTB model using the baseline v_N , the retardation factor can be adjusted to make the v_N modelled closer to the real v_N .

To verify the accuracy of the GNV dataset derived in this study, we compared the simulation results with the baseline velocity in 22 zones. Firstly, the average value of simulation results in each region was compared with that of the v_{N_base} . Table 2 shows that the maximum error between the average value of v_N and v_{N_base} is 0.4252 in zone 21, which has the main lithology of the mixed sedimentary rocks. The scatter plot of correlation between the average v_N and v_{N_base} shows a strong positive correlation (Fig. 10, $R^2 = 0.9956$). To further evaluate the

Region	Standard deviation	Confidence interval(m/year)	Outlier proportion (%)
1	1.2045	(0.0000,1.4445)	5.23
2	3.6148	(0.0000,4.3648)	2.00
3	0.8672	(0.0000,1.5272)	8.25
4	1.5315	(0.1185,3.1815)	15.00
5	7.3985	(0.0000,10.8985)	4.82
6	0.5110	(0.4890,1.5110)	40.92
7	0.0651	(0.0349,0.1651)	17.57
8	0.1614	(0.1386,0.4614)	32.14
9	1.4143	(0.0000,2.1473)	25.00
10	0.3603	(0.5897,1.3103)	24.25
11	1.6345	(0.0000,2.6345)	16.22
12	0.5917	(0.4683,1.6517)	13.33
13	0.4810	(0.6290,1.5910)	25.93
14	2.5857	(0.4143,5.5857)	18.09
15	2.5848	(0.9152,6.0848)	11.83
16	9.1387	(0.8613,19.1387)	10.47
17	8.5980	(0.0000,12.0980)	4.98
18	2.0383	(0.0000,3.6883)	15.02
19	2.9207	(0.0000,3.5837)	3.26
20	9.2670	(0.0000, 11.8370)	2.96
21	8.0919	(0.0000, 10.1219)	3.35
22	3.6457	(0.0000, 5.1057)	5.75
Total			5.50

Table 3. The standard deviation of the v_N , confidence intervals, and outlier proportions for 22 zones. Standard deviation represents the standard deviation for the zones based on data presented in Fig. 5 and the GNV based on data presented in Fig. 7. Since there is no negative nitrate transport velocity, when the confidence interval endpoint appears negative, the endpoint value is replaced by 0 value.

accuracy of the GNV data, the $v_{N_base} \pm$ the standard deviation of v_N in each zone were taken as the confidence interval, and the v_N values outside the confidence interval were taken as the outliers, and then the cell proportion of outliers in each region was calculated (Table 3). The results show that zone 6, which has the lithology of Chalk in Western Europe, has the largest proportion of outliers (40.92%). Besides, the outlier proportions in zone 8, 9, 10 and 13 are also relatively large, accounting for 32.14%, 25%, 24.25% and 25.93%, respectively. However, outliers in these regions only occupy a small proportion globally. Therefore, the overall percentage of outliers is 5.50%, indicating that the accuracy of GNV is 94.50%. Figure 11 shows the outlier proportion of each zone. It can be found that the outliers in Western Europe and southern Britain have relatively large proportions. This is because the total areas of these regions are comparatively small, and a single grid cell takes up a large proportion of the region, resulting in a relatively high proportion of outliers.

Usage Notes

In this paper, the global-scale USZ v_N dataset named GNV was generated using the NTB model based on the global porosity data and global groundwater average recharge datasets from 1958 to 2015. This GNV dataset was derived by constraining the NTB model and has been carefully analysed and verified using the measured values in various regions of the world from published literature.

Generally, the information on nitrate transport velocity in the USZs is valuable when better understanding the legacy of nitrate in the groundwater system and investigating and forecasting its impacts on nitrate in groundwater on the environment, human health, ecological quality, plant and animal growth. In detail, the GNV dataset can be used by different numerical models, such as groundwater and USZ pollution transport models and surface water models, in conjunction with other datasets. For example, the GNV dataset can be combined with the USZ thickness data to calculate the lag-time in the USZs (the time for nitrate to travel from the bottom of the soils to the water table). Similarly, this GNV dataset could be used to estimate the time when the peak value of nitrate leaching reaches the water table, thus informing policymakers to be prepared for the possible increase or decline of nitrate concentrations in groundwater in the future.

This global study can help funders, policymakers and practitioners of a country better understand the feasible time scale for expecting the benefits of nitrate mitigation measures, thus guiding setting regional priorities of groundwater nitrate management plans at the country scale. However, further localised work needs to be undertaken to get detailed information when handling local groundwater nitrate pollution issues.

This calibrated GNV dataset is available in GeoTIFF and ASC formats, thus making it easy to be imported into ESRI ArcMap and any other geospatial software.

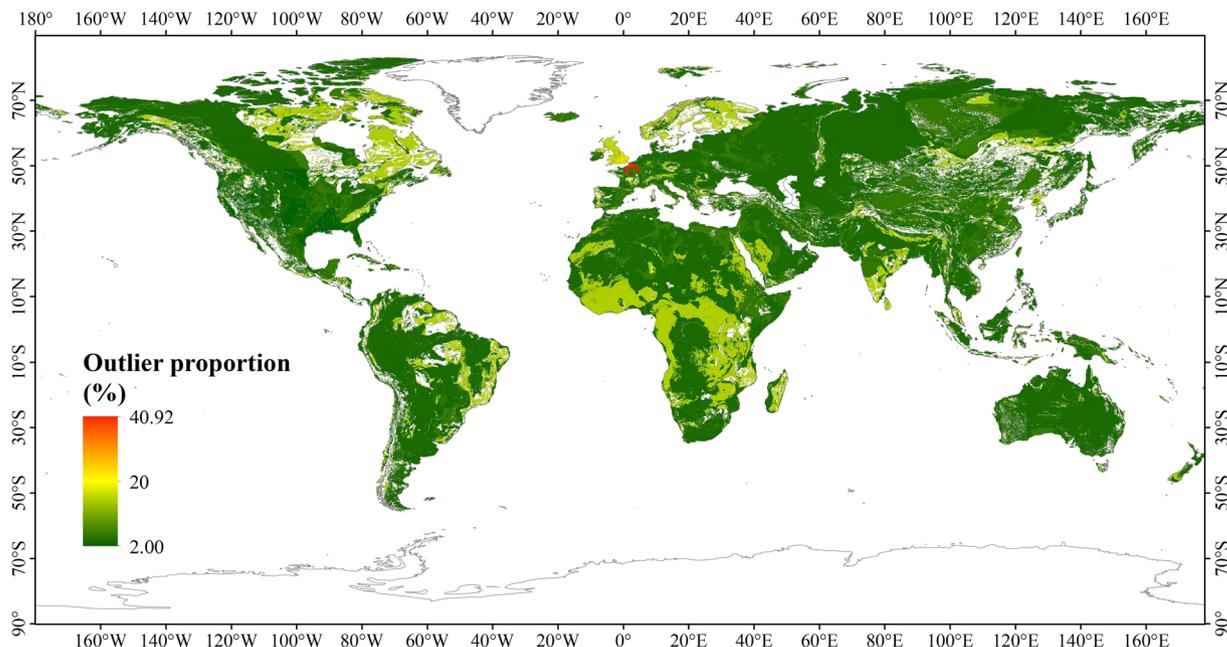


Fig. 11 Spatial distribution of the outlier proportions (shown in Table 3).

The limitation of this study is that deriving the GNV dataset relies on global annual groundwater recharge and global porosity data, thereby possibly passing the uncertainties in these two datasets to this GNV dataset. Besides, 8.30% of the global area in GNV have no values of nitrate velocity in the USZs due to the lack of measured v_N data for the rock types in these areas. According to the classification of 16 basic lithological types of GLiM, the lithological classes, which have no measured USZ v_N , includes Intermediate volcanic rocks, Acid plutonic rocks, Basic plutonic rocks, Intermediate plutonic rocks, Pyroclastics, Acid volcanic rocks and Evaporites. However, these data can be updated once the measured USZ v_N for the rocks in these areas become available.

Code availability

The NTB model code involved in generating GNV was developed using VC++, and the code is available in Figshare⁶⁵. Usage methods and important parts of the code have been commented.

Received: 25 August 2021; Accepted: 21 July 2022;

Published online: 11 October 2022

References

- Shiklomanov, I. A. World fresh water resources. In: Gleick, P. H. (Ed.), *Water in Crisis*. New York (1993).
- Bailey, R. T. *et al.* Spatially distributed influence of agro-environmental factors governing nitrate fate and transport in an irrigated stream-aquifer system. *Hydrology and Earth System Sciences*. **19**(12), 4859–4876 (2015).
- Gilmore, T. E. *et al.* Quantifying the fate of agricultural nitrogen in an unconfined aquifer: Stream-based observations at three measurement scales. *Water Resources Research*. **52**(3), 1961–1983 (2016).
- Nakagawa, K. *et al.* Spatial trends of nitrate pollution and groundwater chemistry in Shimabara, Nagasaki, Japan. *Environmental Earth Sciences*. **75**(3), 1–17 (2016).
- Comley, H. H. Cyanosis in infants caused by nitrates in well water. *Journal of the American Medical Association*. **129**, 112–116 (1945).
- Su, X., Wang, H. & Zhang, Y. Health risk assessment of nitrate contamination in groundwater: A case study of an agricultural area in Northeast China. *Water Resources Management*. **27**(8), 3025–3034 (2013).
- EU. Directive 98/83/EC on the quality of water intended for human consumption. *CELEX-EUR Off J*. **330**, 32–54 (1998).
- Espejo-Herrera, N. *et al.* Colorectal cancer risk and nitrate exposure through drinking water and diet. *Int. Cancer*. **139**, 334–346 (2016).
- Schullehner, J., Hansen, B., Thygesen, M., Pedersen, C. B. & Sigsgaard, T. Nitrate in drinking water and colorectal cancer risk: a nationwide population-based cohort study. *Int. Cancer*. **143**, 73–79 (2018).
- Hannas, B. R., Das, P. C., Li, H. & LeBlanc, G. A. Intracellular conversion of environmental nitrate and nitrite to nitric oxide with resulting developmental toxicity to the crustacean *Daphnia magna*. *PLoS One*. **5**(8), e12453 (2010).
- Poulsen, R., Cedergreen, N., Hayes, T. & Hansen, M. Nitrate: an environmental endocrine disruptor? A review of evidence and research needs. *Environ. Sci. Technol.* **52**, 3869–3887 (2018).
- Peña-Haro, S., Pulido-Velazquez, M. & Sahuquillo, A. A hydro-economic modeling framework for optimal management of groundwater nitrate pollution from agriculture. *Journal of Hydrology* **373**(1–2), 193–203 (2009).
- Cameron, K. C., Di, H. J. & Moir, J. L. Nitrogen losses from the soil/plant system: a review. *Annals of Applied Biology* **162**, 2 (2013).
- Hansen, B. *et al.* Groundwater nitrate response to sustainable nitrogen management. *Scientific Reports* **7**(1), 8566 (2017).
- Charles, W. Abdalla. Measuring economic losses from ground water contamination: an investigation of household avoidance costs I. *JAWRA Journal of the American Water Resources Association* **26**(3), 451–463 (1990).
- Mathewson, P. D., Evans, S., Byrnes, T., Joos, A. & Naidenko, O. V. Health and economic impact of nitrate pollution in drinking water: a wisconsin case study. *Environmental Monitoring and Assessment*, **192**(11) (2020).
- EEA (European Environment Agency). Nutrients in European Ecosystems. *Environ. Assess. Rep.* 4, Copenhagen, Denmark (1999).

18. Meisinger, J. J., Delgado, J. A., Alva, A. K. Nitrogen leaching management. *Encyclopedia of Soils in the Environment* 2. 1122–1124 (2006).
19. Lasagna, M. & De Luca, D. A. Evaluation of sources and fate of nitrates in the western Po plain groundwater (Italy) using nitrogen and boron isotopes. *Environmental Science & Pollution Research* **26**(3), 2089–2104 (2019).
20. Suthar, S. *et al.* Nitrate contamination in groundwater of some rural areas of Rajasthan, India. *J. Hazard. Mater.* **171**, 189–199 (2009).
21. Lapworth, D. J., Krishan, G., MacDonald, A. M. & Rao, M. S. Groundwater quality in the alluvial aquifer system of northwest India: new evidence of the extent of anthropogenic and geogenic contamination. *Sci. Total Environ.* **599–600**, 1433–1444 (2017).
22. Ryan, M. C., Graham, G. R. & Rudolph, D. L. Contrasting Nitrate Adsorption in Andisols of Two Coffee Plantations in Costa Rica. *Journal of Environmental Quality* **30**(5), 1848–1852 (2001).
23. Desimone, L. A. & Howes, B. L. Nitrogen transport and transformations in a shallow aquifer receiving wastewater discharge: a mass balance approach. *Water Resources Research* **34**(2), 271–285 (1998).
24. Padilla, F. M., Gallardo, M. & Manzano-Agugliaro, F. Global trends in nitrate leaching research in the 1960–2017 period. *Sci Total Environ.* **643**, 400–413 (2018).
25. Gardner, S. G., Levison, J., Parker, B. L. & Martin, R. C. Martin. Groundwater nitrate in three distinct hydrogeologic and land-use settings in southwestern Ontario, Canada. *Hydrogeology Journal.* **4**, 1–18 (2020).
26. Wang, L., Butcher, A. S., Stuart, M. E., Goody, D. C. & Bloomfield, J. P. The nitrate time bomb: a numerical way to investigate nitrate storage and lag time in the unsaturated zone. *Environmental Geochemistry & Health* **35**(5), 667–681 (2013).
27. Vero, S. E. *et al.* Review: The environmental status and implications of the nitrate time lag in Europe and North America. *Hydrogeology Journal.* **26**(1), 7–22 (2018).
28. Huang, T., Pang, Z. & Yuan, L. Nitrate in groundwater and the unsaturated zone in (semi)arid northern China: baseline and factors controlling its transport and fate. *Environmental Earth sciences* **70**(1), 145–156 (2013).
29. Chen, N. *et al.* Water, nitrate and atrazine transfer through the unsaturated zone of the Chalk aquifer in northern France. *The Science Of The Total Environment.* **652**, 927–938 (2019).
30. Baran, N., Richert, J. & Mouvet, C. Field data and modelling of water and nitrate movement through deep unsaturated loess. *Journal of Hydrology* **345**(1–2), 27–37 (2007).
31. Gvirtzman, H. & Magaritz, M. Investigation of Water Movement in the Unsaturated Zone Under an Irrigated Area Using Environmental Tritium. *Water Resources Research.* **22**(5), 635–642 (1986).
32. Alberts, E. E. & Spomer, R. G. NO₃-N movement in deep loess soils. *Transactions of the American Society of Agricultural Engineers.* **28**(5), 2030 (1985).
33. Bobier, M. W., Frank, K. D. & Spalding, R. F. Nitrate-N movement in a fine-textured vadose zone. *Journal of Soil and Water Conservation.* **48**, 350–354 (1993).
34. Okumura, A., Hosono, T., Boateng, D. & Shimada, J. Evaluations of the downward velocity of soil water movement in the unsaturated zone in a groundwater recharge area using $\delta^{18}O$ tracer: the kumamoto region, southern japan. *Geologia Croatica.* **71**(2), 65–82 (2018).
35. Brouyère, S., Dassargues, A. & Hallet, V. Migration of contaminants through the unsaturated zone overlying the hesbaye chalky aquifer in belgium: a field investigation. *Journal of Contaminant Hydrology.* **72**(1–4), 135–164 (2004).
36. Geyh, M. A. Carbon-14 concentration of lime in soils and aspects of the carbon-14 dating of groundwater. *Isotope Hydrology Symposium.* (1970)
37. Butcher, A. *et al.* Investigation of rising nitrate concentrations in groundwater in the Eden Valley, Cumbria. 2, unsaturated zone studies. *Yonsei Medical Journal.* **57**(5), 1282–1285 (2008).
38. Wang, L. *et al.* Prediction of the arrival of peak nitrate concentrations at the water table at the regional scale in Great Britain. *Hydrol. Process.* **26**, 226–239 (2012).
39. Turkeltaub, T., Jia, X., Zhu, Y., Shao, M. A. & Binley, A. Recharge and nitrate transport through the deep vadose zone of the loess plateau: a regional-scale model investigation. *Water Resources Research* **54**(7), 4332–4346 (2018).
40. Moore, S. L. Quantifying nitrate transport rates in the unsaturated zone below agricultural fields. *Hydrogeology I Tracers & Other Field Techniques* (2005)
41. Wang, L. *et al.* The changing trend in nitrate concentrations in major aquifers due to historical nitrate loading from agricultural land across england and wales from 1925 to 2150. *Science of The Total Environment.* **542**(Pt.A), 694–705 (2016).
42. Psc, R., Hornsby, A. G. & Jesup, R. E. Indices for ranking the potential for pesticide contamination of groundwater. *Proceedings Soil & Crop Science Society of Florida.* **44**, 1–8 (1985).
43. Gleeson, T., Moosdorf, N., Hartmann, J. & Beek, L. P. H. V. A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity[*J*]. *Geophysical Research Letters.* **41**(11), 3891–3898 (2014).
44. Gleeson, T. *et al.* Mapping permeability over the surface of the Earth. *Geophysical Research Letters.* **38**(2), 93–104 (2011).
45. Beek, L. P. H. V. and Bierkens, M. F. P. The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification. *Tech. rep., Department of Physical Geography, Utrecht University* (2009).
46. Bierkens, M. F. P. & Beek, L. P. H. V. Seasonal Predictability of European Discharge: NAO and Hydrological Response Time. *Journal of Hydrometeorology.* **10**, 953–968 (2009).
47. Sutanudjaja, E. H. *et al.* PCR-GLOBWB_model: PCR-GLOBWB version v2.1.0_beta_1, *Zenodo.* (2017).
48. Edwin H. *et al.* PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. *Geoscientific Model Development* **11**(6), 2429–2453 (2018).
49. Beek, L. P. H. V., Wada, Y. & Bierkens, M. F. P. Global monthly water stress: 1. water balance and water availability. *Water Resources Research.* **47**(7) (2011).
50. Bogardi, I., Kuzelka, R. D. & Ennenga, W. G. Nitrate Contamination: Exposure, Consequence, and Control. *Springer-Verlag.* 333–347(1991).
51. Allen, D. J. *et al.* The physical properties of major aquifers in England and Wales. *British Geological Survey Tech. Rep. WD/97/34 and Environment Agency R&D Pub.* **8** (1997).
52. Hartmann, J., Moosdorf, N. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry Geophysics Geosystems.* **13**(12) (2012).
53. Hartmann, J. & Moosdorf, N. Global Lithological Map Database v1.0 (gridded to 0.5° spatial resolution), PANGAEA, <https://doi.org/10.1594/PANGAEA.788537> (2012).
54. Wendland, F. *et al.* European aquifer typology: a practical framework for an overview of major groundwater composition at european scale. *Environmental Geology.* **55**(1), 77–85 (2008).
55. *DIVA-GIS* <https://www.diva-gis.org/gdata>.
56. *Geographical Information Monitoring Cloud Platform* <http://www.dsac.cn/dataproduct/index/2019>.
57. Yang, C. *et al.* Nitrate transport velocity data in the global unsaturated zone. *figshare* <https://doi.org/10.6084/m9.figshare.14782959.v2> (2021).
58. Freeze, R. Allan, and J. A. Cherry. *GroundWater. Groundwater.* (1979).
59. Bekele, E. B., Salama, R. B. & Commander, D. P. Hydrogeology and hydrochemistry of the Parmelia aquifer, northern Perth Basin, Western Australia. *Australian Journal of Earth Sciences: An International Geoscience Journal of the Geological Society of Australia* **53**(6), 891–904 (2006).

60. Börker, J., Hartmann, J., Amann, T. & Romero-Mujalli, G. Terrestrial sediments of the earth: development of a global unconsolidated sediments map database (gum). *Geochemistry, Geophysics, Geosystems*. **19**(4) (2018).
61. Huscroft, J. *et al.* Compiling and Mapping Global Permeability of the Unconsolidated and Consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). *Geophysical Research Letters*. **45**(5) (2018).
62. Milzow, C., Kgotlhang, L., Bauer-Gottwein, P., Meier, P. & Kinzelbach, W. Regional review: the hydrology of the okavango delta, botswana—processes, data and modelling. *Hydrogeology Journal* **17**(6), 1297–1328 (2009).
63. Qiao, Y., Miao, S., Lu, X. & Wang, T. Obstacle-factor Analysis of Soil Profile in Corn Field Located in Aeolian Sand Soil Zone of Northeast China. *Journal of Maize Sciences* **28**(3), 8 (2020).
64. Sun, Z., Han, J., Liu, Z. & Lü, Y. Effect of Planting Patterns on Soil Micro Structure in Typical Farmland of Huabei Plain. *Journal of agricultural machinery* **48**(5), 8 (2017).
65. Wang, L., Li, Y., Yang, C. & Chen, S. Nitrate Time Bomb - the version of global nitrate velocity in the unsaturated zones, *figshare*, <https://doi.org/10.6084/m9.figshare.19264700> (2021).
66. Gleeson, T. GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, *Borealis*, <https://doi.org/10.5683/SP2/DLGXYO> (2018).

Acknowledgements

The work was funded by the Research on Large-area and Multi-target Classification Method for Massive Multi-source Remote Sensing Based on Graph-spectrum Feature Model (2020YFA0714103); and this work was also supported by the British Geological Survey via NERC national capability.

Author contributions

C.Y.Y., L.W. and S.B.C. conceived and supervised the study, designed the analyses. C.Y.Y. wrote the paper. L.W. provided code for the NTB model and edited the manuscript. C.Y.Y. and Y.Y.L. managed the collection and processing of datasets for the baseline database. S.H. and Q.H.Z. helped with the GNV data verification. Y.B.C. helped mapping.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-022-01621-x>.

Correspondence and requests for materials should be addressed to L.W. or S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© British Geological Survey, UKRI 2022