scientific **data**

DATA DESCRIPTOR

Check for updates

Gridded Population Structure Datasets for the Indus River Basin under Shared Socioeconomic Pathways 2020–2100

Barira Haider[1,4], Cheng Jing[2,4], Buda Su[1✉], Muhammad Asad Hussain[1], Arshad Ashraf[3], Jinlong Huang[1], Yanjun Wang[1] & Tong Jiang[2✉]

The Indus River Basin is one of the most densely populated transboundary river basins in the world and is the region with the most serious water disputes. Given the current population's rapid growth, inclusive and high-resolution datasets are urgently needed to assess how this growth will affect the distribution of resources and the sustainability of the environment. Here we present a population gridded dataset for the Indus River Basin with a resolution of 2.5 arc-minutes (~5 km). Based on the historical population distribution and the provincial (state) demographic parameters of the four countries, Afghanistan, China, India, and Pakistan, we projected the population size and structure (age, sex) changes under the Shared Socioeconomic Pathways (SSP1-5) for the period of 2020–2100 on each grid in the Indus River Basin. The dataset was well verified by comparing it with the observed population gridded dataset in 62,140 grids in the basin. The dataset can be useful sources for further research in resource management, sustainable development initiatives, and assessment of climate change impact.

Background & Summary

An essential component of South Asia's economic and social landscape is the population fluctuations along the Indus River Basin, which is shared by Afghanistan, China, India, and Pakistan and distinguished by varied socioeconomic landscape, ecological importance, and susceptibility to climate variability[1,2]. The basin is home to more than 320 million population, who maintain various environmental systems, economies, and traditions[3,4]. The Indus River Basin's water is essential for the sustaining the livelihood of the population and the cultivation of the agricultural breadbaskets of all the nations in the transboundary. In the Indus River Basin, the resided population needs more water as population is growing rapidly, and in near future the available water might become insufficient for the locals which will have negative consequences overall[5,6]. Population size and structure estimates in this basin are essential for guiding resource management decisions, policy choices, and adaptation plans to deal with the complex issues brought on by multiple environmental variables[7]. Policymakers, planners, and academics must comprehend the population growth trajectory as the region struggles with growing urbanization, environmental degradation, and changing climate impacts[8].

Global climate scenarios offer insightful analyses of future climate trends, and more focused adaptation and mitigation strategies must be developed using projections that take into account the localized particularities[9]. Through the use of Shared Socioeconomic Pathways (SSPs) concept, the Intergovernmental Panel on Climate Change (IPCC) has significantly influenced our knowledge of future socioeconomic trajectories on all levels[10]. The SSPs paradigm has been used in various research to investigate population dynamics and their consequences for global development[11]. There are studies that have carried out global and national population projections under SSP1-5[12–14]. There have also been some attempts at global population projections at the grid-scale[15,16]. However, for the Indus Basin region, which is geographically complex and in a high mountainous basin, high

[1]State Key Laboratory of Climate System Prediction and Risk Management, Nanjing University of Information Science & Technology, Nanjing, 210044, China. [2]Key Laboratory for Climate Risk and Urban-Rural Smart Governance, School of Geography, Jiangsu Second Normal University, Nanjing, 210013, China. [3]National Agricultural Research Centre, Islamabad, 44000, Pakistan. [4]These authors contributed equally: Barira Haider, Cheng Jing. ✉e-mail: subd@nuist.edu.cn; jiangtong@nuist.edu.cn

Scenarios	Countries	Fertility	Mortality	Migration
SSP1	Afghanistan China India Pakistan	Low Low Low Low	Low Low Low Low	Medium Medium Medium Medium
SSP2	Afghanistan China India Pakistan	Medium Medium Medium Medium	Medium Medium Medium Medium	Medium Medium Medium Medium
SSP3	Afghanistan China India Pakistan	High High High High	High High High High	Low Low Low Low
SSP4	Afghanistan China India Pakistan	High High Low Low	High Medium Medium High	Medium Medium Medium Medium
SSP5	Afghanistan China India Pakistan	Low Low Low Low	Low Low Low Low	High High High High

Table 1. Demographic Assumptions in Afghanistan, China, India and Pakistan under five SSPs.

resolution population projections need to be carried out in conjunction with more nuanced demographic characteristics[17]. Existing studies lack detailed population structure data on a grid such as age and gender. However, this information is important for analyzing water supply and demand, population health and disaster vulnerability at the basin scale.

Traditional high-resolution socioeconomic projections tend to have a fixed process, which involves first conducting country- or regional-scale projections, followed by downscaling of future coarse-resolution projections based on current socioeconomic distribution patterns[18]. However, such an approach tends to ignore population flows and structural changes within the region. Here we adopt a novel but intuitive grid-based projection method. We combined the gridded population of the Indus River Basin and the demographic elements of the differentiated provinces in which the basin is located to parameterize future population changes individually for each of the 62,140 grids in the basin within the SSPs framework. Population projection models are run on each grid and the final gridded population projection is obtained at 2.5 arc-min resolution for 2020–2100. The dataset contains complete demographic characteristics on each grid, including age and gender along with population size. With the provision of thorough demographic projections for the Indus River Basin, this study seeks to provide a complete set of basin population datasets to facilitate research on exposure, vulnerability, adaptation and risk under the context of climate change.

Methods

Demographic data. Two kinds of demographic data are used in this study, one is historical gridded population data and the other is provincial demographic data.

The Gridded Population of the World (GPW) collection, in the fourth version (GPWv4), models the size of population distribution at continuous global raster surface. GPW is produced by NASA's Socioeconomic Data and Applications Centre (SEDAC). Historical data of gridded population is collected at a resolution of 2.5 arc-minutes for the year 2015 and 2020[19].

15 Provincial demographic data for the four regions Afghanistan, China, India, and Pakistan are collected in 2020 including population size, net migration, mortality by age, gender, and fertility rate. Afghanistan and Pakistan data is collected by their respective Demographic Health and Survey[20,21]. Data for Indian states are collected by National Family Health Survey[22]. The demographic data for China is collected by the Seventh National Population Census in 2020 from the Bureau of Statistics of China (https://www.stats.gov.cn/sj/pcsj/rkpc/7rp/zk/indexch.htm).

SSPs framework. Intergovernmental Panel on Climate Change (IPCC) has developed five major SSP to explain socioeconomic demographic technical, institutional, political, and other associated trends[23]. Table 1 compiles the parameter assumption in four regions of the Indus River Basin under 5-SSPs for fertility, mortality, and migration. Investments in health and education support the demographic transition of every nation in the sustainable development scenario depicted in SSP1. Low mortality and low fertility due to incremental advances in health and education could result a relatively small size of population. Thus, fertility is assumed to be low. Mortality will decline due to a rapid increase in educational attainment. All countries are thought to have moderate net migrations due to decreased levels of inequality. SSP2 can be considered as a middle of-the-road scenario which shows the historical trends and features. As a result, the fertility rate, mortality, and net migration assumptions are medium in all the countries. SSP3 is related to the population component which is referred to as "stalled development". There is a halted demographic change in this world: The competitiveness among countries will result in security-focused national development policies. Thus, fertility is assumed to be high in all four countries. Mortality will be higher due to low investment in education and health. Since security is prioritized and obstacles to international trade exist, it is believed that migration is minimal across all countries. Under SSP4, there will be extreme inequality both inside and between nations. Every country has a small but highly educated population

Countries	Provinces	2015	2020	2030	2040	2050	2060	2070	2080	2090	2100
Afghanistan	Kabul	0.98	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55
	Nangarhar	1.36	1.33	1.31	1.29	1.27	1.25	1.23	1.21	1.20	1.20
	Kunar	1.44	1.41	1.40	1.39	1.36	1.34	1.32	1.30	1.29	1.27
China	Xinjiang	1.56	1.40	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35
India	Jammu Kashmir	0.69	0.66	0.64	0.62	0.60	0.59	0.57	0.55	0.54	0.53
	Himachal Pardesh	1.59	1.58	1.56	1.54	1.52	1.50	1.49	1.47	1.46	1.45
	Rajistan	1.36	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.23	1.22
	Haryana	1.50	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.40
	Punjab	1.19	1.17	1.16	1.15	1.14	1.13	1.12	1.11	1.10	1.10
	Chandigarh	0.69	0.66	0.65	0.64	0.63	0.62	0.61	0.60	0.59	0.58
Pakistan	Punjab	0.97	0.94	0.93	0.92	0.91	0.90	0.89	0.87	0.86	0.85
	Sindh	1.13	1.00	0.99	0.97	0.96	0.95	0.94	0.93	0.92	0.91
	Balochistan	1.13	1.11	1.10	1.00	0.99	0.97	0.96	0.95	0.94	0.93
	KPK	1.14	1.13	1.12	1.11	1.10	0.99	0.98	0.97	0.96	0.95
	Gilgit Baltistan	1.35	1.30	1.29	1.28	1.27	1.26	1.25	1.24	1.23	1.22

Table 2. Recorded and Assumed medium TFR in provinces of Afghanistan, China, India and Pakistan.

and vast low-educated groups. Today's high fertility nations continue to have high fertility rates, whereas low fertility nations continue to have low fertility rates. It is believed that high fertility nations experience high death rates, while the other categories experience medium mortality rates. Migration is assumed to be medium for all countries due to extreme regional inequality and modest economic growth. The nation will prioritize educational investment in SSP5, the fossil-fueled development scenario, which will lower the country's rates of mortality and fertility. The vibrant capital and technology markets will promote net migration that's why assumptions for net migration are higher in all four countries[24].

Projection model. There are two components to population expansion in a model: mechanical growth and natural growth. The difference between births and deaths is natural growth; primarily, mechanical growth is a measure of the population's migration, with positive numbers indicating in-migration and negative values indicating out-migration. Equations 1, 2 present the number of persons at the age of 1 in a given year and the number of newborn in a given year respectively.

$$P_{t+1} = P_t^{'}(1 - D_{t+1})M_{t+1} \tag{1}$$

Where t refers to the calendar year, P_{t+1} is the resulting population of 1-year-olds in year $t+1$. P_t' denotes the number of individuals who were newborns (0 years old) in year t, and thus will be 1 year old in year $t+1$. D_{t+1} is the death rate of the $t+1$-year-old population in that year, and M is the migration rate of the $t+1$-year-old population in that year.

$$P_n = \sum_{\{g\}} \sum_{\{t=15\}}^{\{49\}} P_{\{t,g\}} \times R_{\{t,g\}} \times F_{\{t,g\}} \tag{2}$$

P_n is the number of new born, $P_{t,g}$ represents the population structure at age t specific to each grid, reflecting local demographic structures. R is the proportion of women in the population at the age of t in the grid, and F is the fertility rate of the population.

Setting of model parameters. *Fertility.* The recorded total fertility rate (TFR) for 2015, 2020 and assumptions for the years till 2100 used in this study for the current provinces in Afghanistan, China, India and Pakistan are given in Table 2. A medium fertility rate represents continuation of past trends and policies, can be seen as a reasonable baseline scenario. When examining the possible effects of various socioeconomic circumstances on birth rates, the medium fertility assumption provides a point of comparison. For low fertility assumptions, it will be 20% lower by 2030 and 25% lower by 2050, compared to the medium fertility rate estimate[24]. In contrast to the baseline scenario, the low fertility believes that several socioeconomic factors, including increasing female education, urbanization, and access to family planning services, will cause fertility rates to drop more quickly. When examining the possible ramifications of a quick demographic shift and how it can affect the population's age distribution and economic growth, the low fertility assumption is frequently applied. On the other hand, the high fertility assumption projects that by 2030 and 2050, it will be 25% higher, respectively, compared to the medium fertility rate estimate[24]. The socioeconomic factors such as increased financial security, societal support for families, and cultural inclinations for larger families, will be more conducive to sustaining greater fertility rates. Understanding the possible opportunities and problems that come with a young and expanding population is made easier by considering the high fertility assumption[24].

Mortality. Under the medium mortality assumption, life expectancies are projected to converge toward those found in industrialized nations, reflecting advancements in healthcare, living standards, and general quality of life. In particular, the average lifespan is projected to rise by two years every ten years. This steady rise is
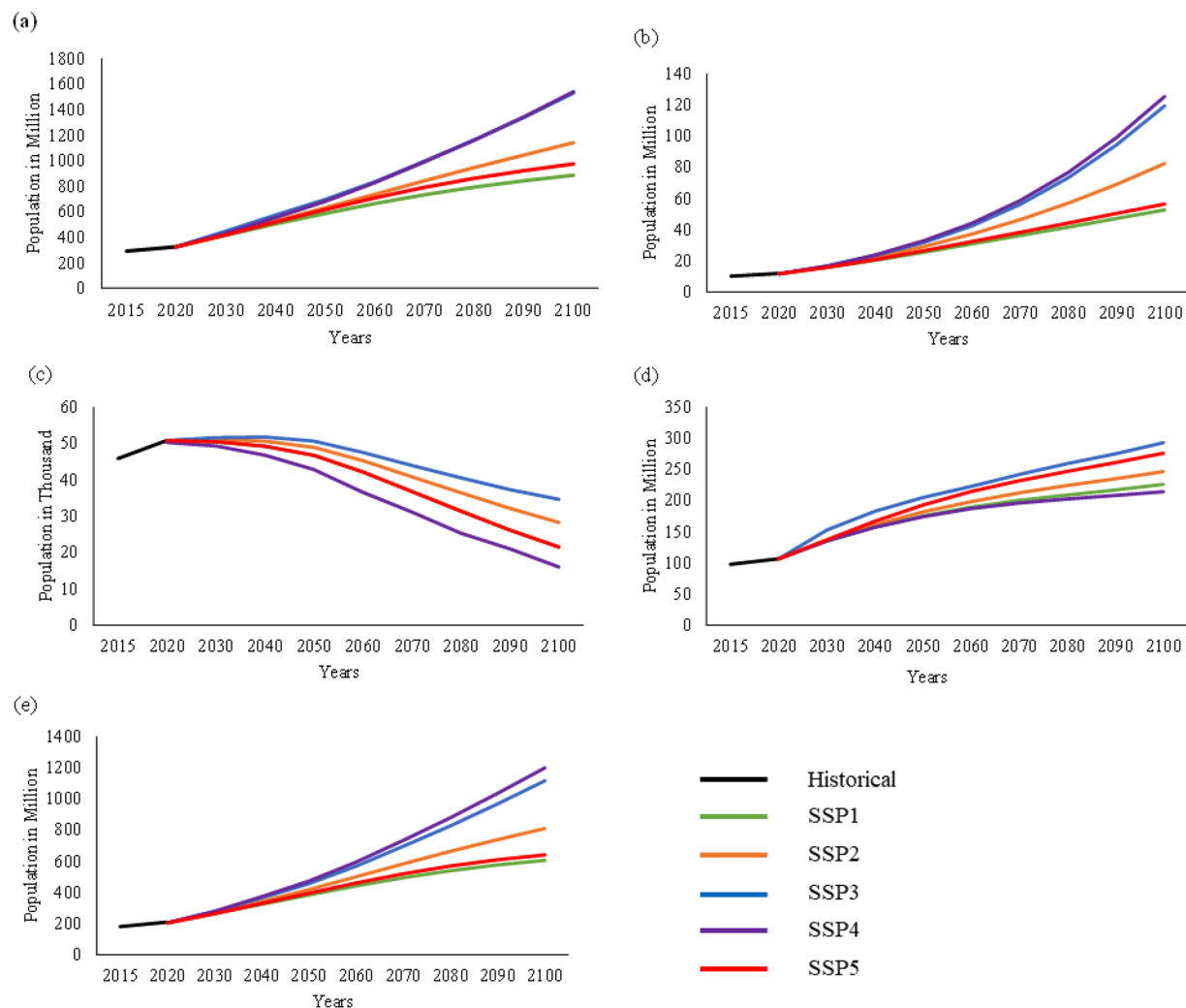
Fig. 1 Projected Population in the Indus River Basin under SSP1-5: (**a**), Indus Basin (**b**), Afghanistan (**c**), China (**d**) India, (**e**) Pakistan.

consistent with past patterns seen in many developed countries and provides a reasonable starting point for death rates in the future[12,24].

For low mortality assumptions life expectancies are projected to rise at a pace that is one year slower than the medium scenario[24]. Low mortality takes into account potential circumstances that could increase life expectancy, such as notable developments in medical technology, public health campaigns, and socioeconomic advancements. The assumption of high mortality suggests that life expectancies will rise at a rate one year lower than in the medium scenario. High mortality considers potential obstacles that could prevent increases in life expectancy, such as unstable economies, health crises, or limited access to healthcare[24].

Migration. According to the medium assumption scenario, net migration will decrease by 20% per five years until 2030, and the scale of migration will subsequently continue to decline[24]. The rates of migration between urban and rural populations will progressively decline to zero by the end of this century. According to the low and high migration assumptions, the migration rate is 50% lower and 50% higher, respectively, as compared to the medium assumption[24].

Gridded population projection. The GPW dataset from the year 2020 at 2.5 arc-minute resolution provides the basis for all future projections. Since the GPW dataset provides only the total population for each grid cell without age and sex breakdowns, we first assigned each grid cell to its respective country (Afghanistan, China, India, or Pakistan) based on geographic location. This step allowed us to derive the population structure by age and sex using country-level demographic data.

A population ratio was first computed for each grid cell using the province's total population from national demographic data to disaggregate the gridded population data. This ratio was then applied to the provincial age- and gender-specific population distributions to produce a spatially explicit population grid. Each grid cell was assigned fertility, mortality, and migration rates based on its location within the province, inheriting all
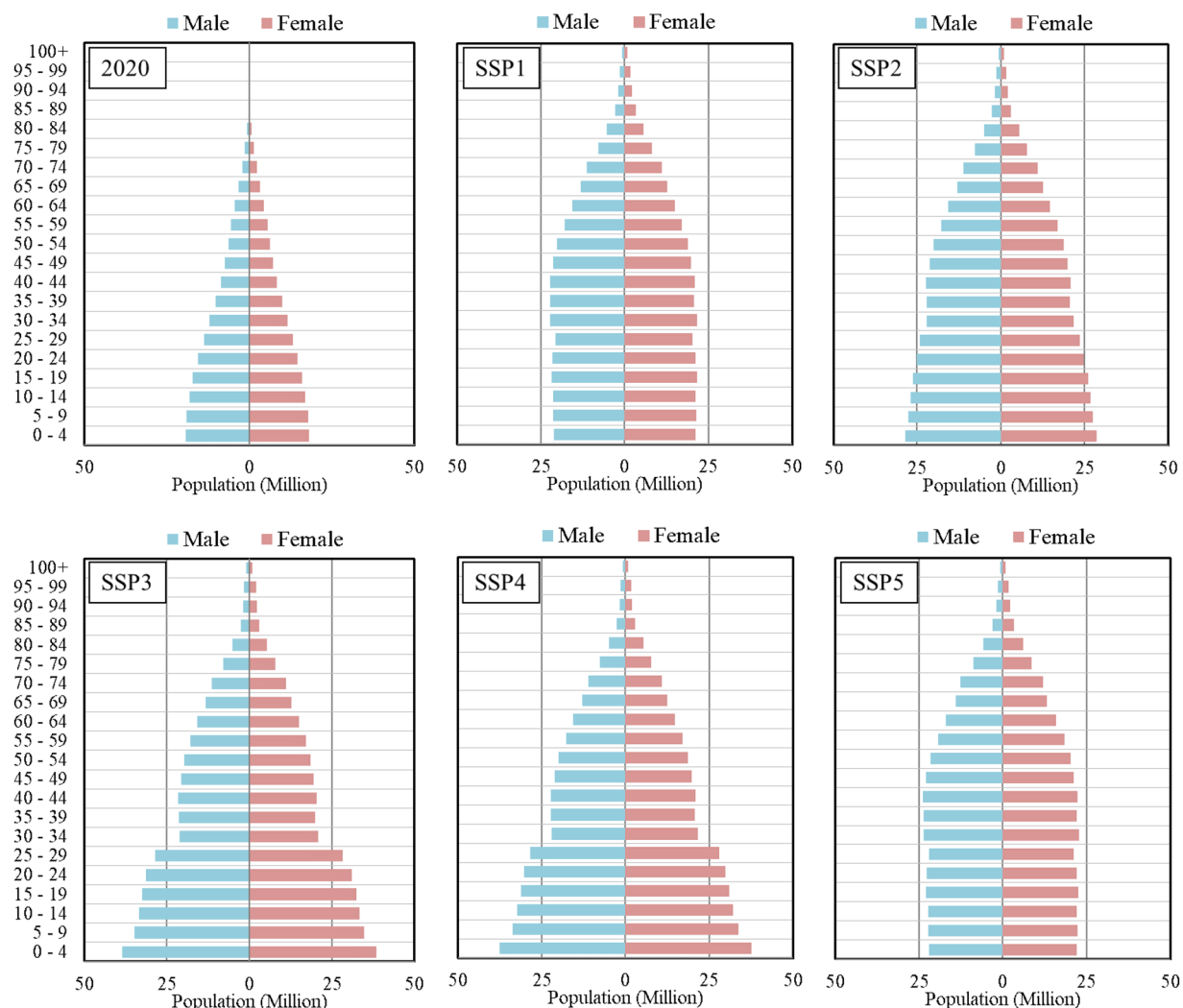
Fig. 2 Population structure in the Indus River Basin in 2020 and in 2050 under SSP1-5.

demographic characteristics from its provincial affiliation. Net migration values were downscaled to grid cells according to each grid's share of the provincial population. To maintain consistency, a standard newborn sex ratio (1: 1.05 male to female) was uniformly applied across all grid cells.

The population projection model integrated these parameters and was applied consistently across all 62,140 grid cells in the study area. This process generated age- and sex-specific population projections, which were aggregated to estimate total population under all SSP scenarios from 2020 to 2100.

Data Records

The datasets including gridded population data for different ages and genders under various SSPs for the Indus River Basin from 2020–2100 are available at https://doi.org/10.57760/sciencedb.19778[25].

The dataset includes population at 62,140 grids with 2.5′ (~5 km² around the Equator) resolution in the Indus River Basin covering 72°28′ to 79°39′ E and 29°8′ to 36°59′ N. There are 15 files in nc. format accessible. Five files are devoted to the total population of each grid for every 5 years from 2020–2100, and 10 files are of the distribution of population structure (21 ages groups) for male and female for each grid under SSP1-5. Population structure shows different genders and ages. Genders include male and female. Ages include 21 groups: "0–4", "5–9", "10–14", "15–19", "20–24", "25–29", "30–34", "35–39", "40–44", "45–49", "50–54", "55–59", "60–64", "65–69", "7074", "75–79", "80–84", "85–89", "90–94", "95–99", and "100+".

Technical Validation

Population size and structure under SSP1-5 from 2020–2100. The Indus River Basin is expected to experience a significant population increase until 2050 under all the SSPs (Fig. 1). The population was 324 million in 2020. Under "business as usual" SSP2, the population is projected to continue at the current rate, reaching 611 million people by 2050 and expected to reach about 1 billion by 2100. In SSP1 and SSP4, it is expected to increase by 75% and 86% by 2050, and by 174% and 377% by 2100, respectively. In SSP3 and SSP5, it is expected to increase by 108% and 89% by 2050, and by 373% and 201% by 2100, respectively.

Countries	Relative Error (%)							
	2016	2017	2018	2019	2020	2021	2022	2023
a) Afghanistan	0.86	1.40	2.18	1.85	1.54	1.0	0.24	2.68
b) China	0.10	0.10	0.10	0.13	0.13	0.13	0.15	0.06
c) India	0.10	0.02	0.30	0.20	0.10	0.13	0.30	0.50
d) Pakistan	1.31	2.82	3.82	4.88	5.67	4.67	3.28	2.68

Table 3. Country scale technical validation between actual and simulated population in (a) Afghanistan, (b) China, (c) India, (d) Pakistan.

The Indus River Basin has a total area of 1.12 million km². The largest share of the region, at 47%, lies in Pakistan, which is 5th most populous country in the world. Pakistan's population was 204 million in 2020. Under a medium scenario SSP2, projection indicates that the population of the Pakistani region of the Basin might reach 417 million by 2050, effectively doubling in just three decades. It is projected that Pakistan would be the third most populated country in the world by 2060, behind only India and China. According to these projections, the major population rise is seen in both Indian and Pakistani regions of the Indus River Basin. Afghanistan's population is steadily increasing across all SSPs. The Chinese portion of the Indus River Basin has a very low population.

The population structure of the Indus River Basin will undergo significant changes in the future compared to the current situation (Fig. 2). In 2020, population of the Indus River Basin mainly presented a pyramid shaped structure. Due to the high fertility rate and low level of development in healthcare, there are more new-borns and fewer elderly people. The annual new population can reach nearly 10 million, while there are only 16 million elderly people aged 65 and above, accounting for 4.9% of the total population.

Under different development scenarios in the future, there will be significant changes in population structure. SSP1 and SSP5 are scenarios with relatively higher levels of development. In these two scenarios, the fertility rate is significantly controlled, and the annual birth rate remains stable, slightly higher than the current level. At the same time, the level of medical and health development is rapidly advancing, and the life expectancy of residents is generally increasing. The elderly population is also increasing, with the distribution of elderly people aged 65 and above accounting for 14.6% and 14.8% of the total population in 2050, respectively. In both SSP3 and SSP4, the development level of the region is relatively low, and the population still presents a pyramid shaped distribution pattern. This is mainly reflected in the rapid growth of fertility rates and the increasing number of newborns every year. By 2050, the annual new population in both scenarios can reach over 20 million, while the development of medical and health care is relatively slow, with the proportion of elderly people aged 65 and above at around 11.6%. SSP2 is a relatively balanced scenario, with a relatively balanced population of all age groups, with approximately 16 million new births per year and an elderly population accounting for about 12.9% of the total population.

Errors in region-level population projection. Country scale validation was conducted to assess the reliability of demographic projection methodology using broader data availability at the national level. Specifically, for each country (Afghanistan, China), India and Pakistan), we utilized country demographic parameters fertility, mortality, and migration sourced from official government reports and United Nation (https://population.un.org/wpp). These were used to simulate population trends before 2023 starting from 2015. The resulting projected national population were validated through relative error between actual and projected. Although our core modeling focused on the Indus River Basin, this country-level validation served to verify our projection model before applying it to the basin specific context.

The relative error was used to evaluate the bias between actual and simulated population mentioned in Table 3. For Afghanistan, China, India and Pakistan's population estimation, our projection is 0.2%–2.68%, 0.02%–0.5% and 1.3%–2.68% 1.3%–5.67% greater than the actual population respectively, showing that projections are slightly overestimated, which might be caused by assuming high fertility rates.

Errors in basin-scale population projection. The Indus River Basin is project to experience significant population growth and urbanization between 2020, 2050 and 2100 (Figs. 3, 4). SSP5 and SSP3 show the highest population density, due to high fertility and urbanization, respectively. SSP4 also shows the high density due to high inequality with some urban areas being densely populated while rural areas remain sparse. SSP2 shows a moderate increase in population density and SSP1 shows the lowest density among all scenarios but still indicates a significant increase from 2020. There is a clear trend toward urbanization with denser population clusters in certain areas.

The Pakistani part of the basin, Punjab, contributes to the high population density, especially with Lahore, the capital of Punjab and the second-largest city in Pakistan. Overall, the combination of economic opportunities, educational facilities, and well-developed infrastructure makes Punjab, specifically Lahore, the most populated area in the Indus Basin region of Pakistan with a population density of 14,372 per square kilometers. The Indian part of the basin is the most densely populated, although it covers an area less than Pakistan. Indian Punjab is a highly fertile and agriculturally productive region, contributing to its high population density. The major cities in Punjab, such as Ludhiana, Amritsar, and Jalandhar, also contribute significantly to the region's overall population. The Great Indian Desert, also referred to as the Thar Desert holds the highest population density in the world. The desert spans approximately 200,000 square kilometers, with about 85% located in India and the remaining portion in Pakistan. The total population of the Thar Desert region is around 50 people per
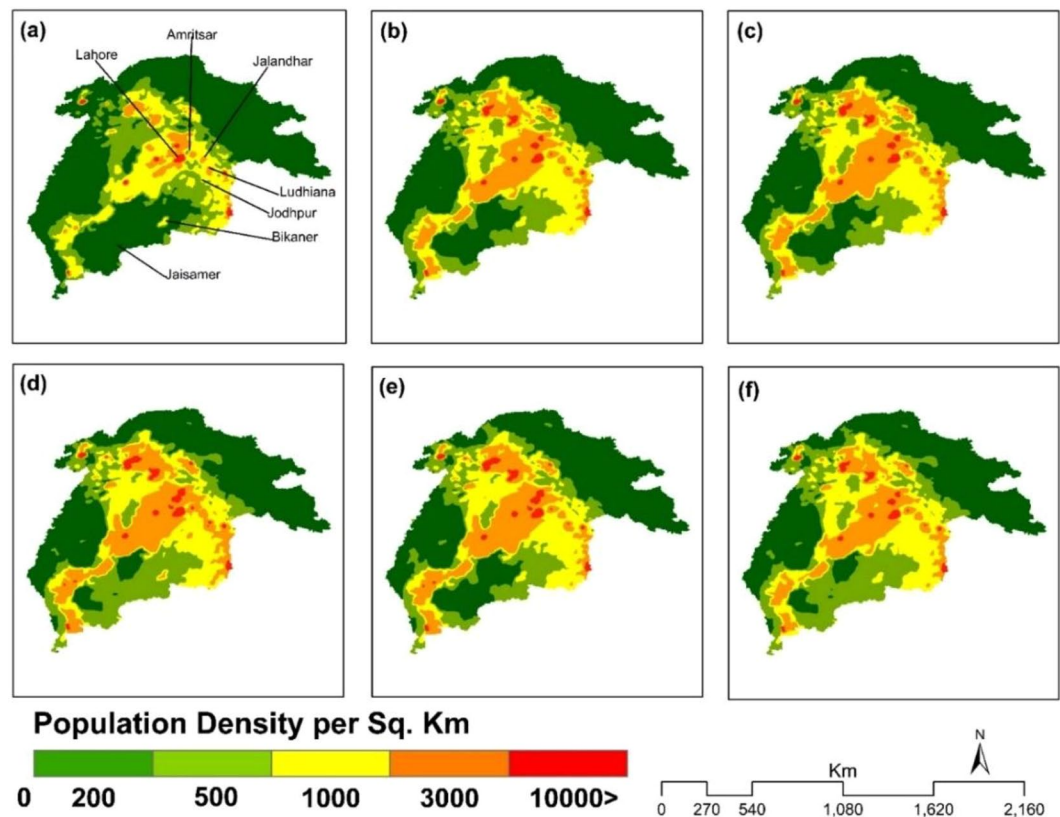
Fig. 3 Population density in the Indus River Basin in (**a**) 2020, (**b**) SSP1-2050, (**c**) SSP2-2050, (**d**) SSP3-2050, (**e**) SSP4-2050, (**f**) SSP5-2050.
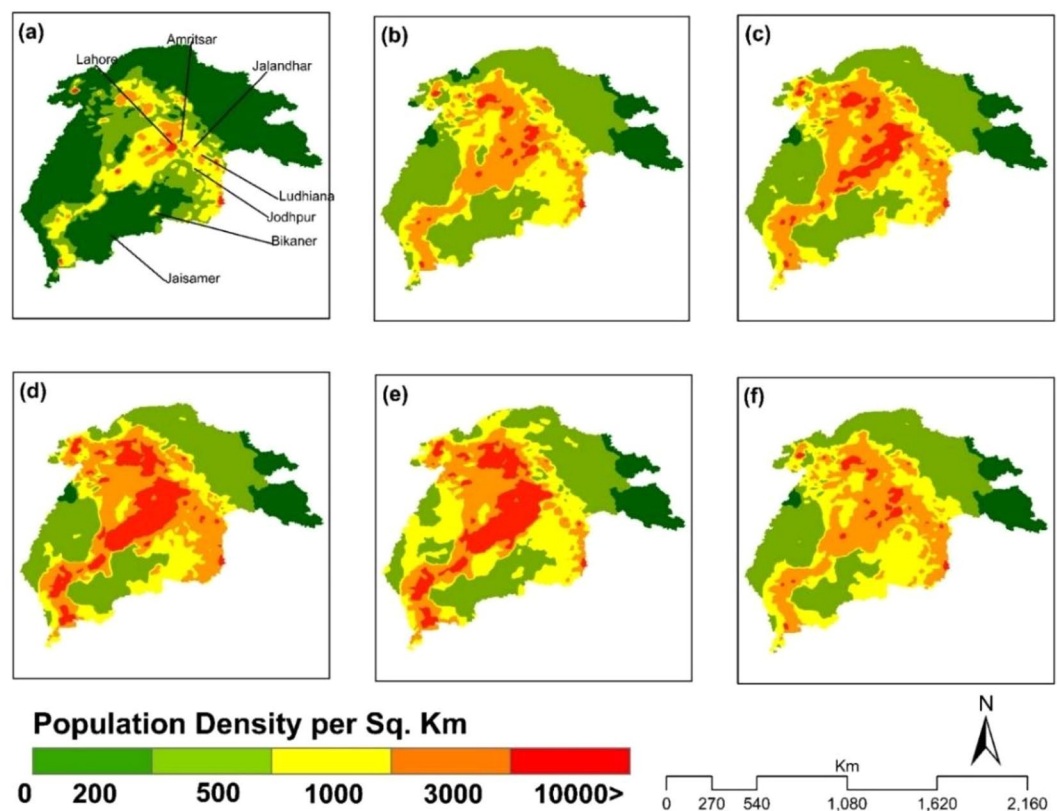
Fig. 4 Population density in the Indus River Basin in (**a**) 2020, (**b**) SSP1-2100, (**c**) SSP2-2100, (**d**) SSP3-2100, (**e**) SSP4-2100, (**f**) SSP5-2100.
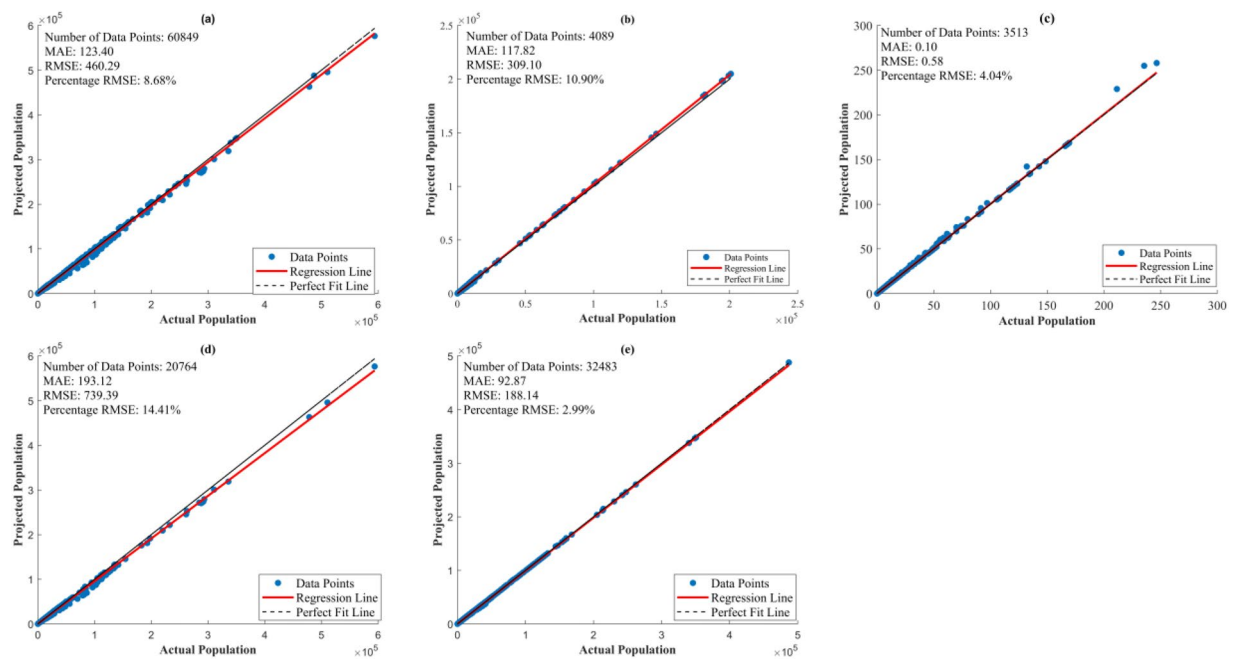
Fig. 5 Comparison of Actual and Projected values on grid level in (**a**) Indus River Basin, (**b**) Afghanistan, (**c**) China, (**d**) India, (**e**), Pakistan.

square kilometers in 2020. The Indian side, specifically in Rajasthan, hosts the majority of this population, with significant settlements in cities like Jodhpur, Bikaner, and Jaisalmer.

To validate the gridded population datasets, a separate validation step was conducted using 2015 as the starting year. we used the GPW 2015 data as the base year along with all the provincial parameters fertility, mortality and migration under SSP2 (business as usual) to simulate population projections to 2020. The SSP2 closely resembles past trends and offers a pertinent baseline for projection.

The actual and projected populations were compared per grid, and scatter plots were used to visually display the results (Fig. 5). For the fair comparison, all the zero population grids were removed from the study area. Most of the zero population grids lie in the Chinese region of the Basin, and a few are located in the Pakistani and Indian regions. The examination of the Indus River Basin indicates a noteworthy fit of the population projection model, by comparison of actual and projected population values on each grid level, as evidenced by the parameters RMSE of 460.27, MAE of 123.39, and a %RMSE of 8.68%. The moderate RMSE, despite the large number of data points, refers to the possibility of underlying regional variances or possible discrepancies in the quality of the data or the methods used to collect it throughout the basin. China exhibits remarkably low errors (RMSE of 0.10, MAE of 0.58, %RMSE of 4.04%) for the particular regions, suggesting a highly accurate model that might profit from better data quality and steady demographic patterns. India, on the other hand, has higher error metrics (RMSE of 793.33, MAE of 193.11, %RMSE of 14.41%), which may be brought on by the country's dynamic and diversified population landscape, which may make modelling more difficult. Pakistan and Afghanistan exhibit intermediate model accuracies, with errors reflecting differences in the region's population and data gathering.

Afghanistan shows moderate accuracy with slightly higher variability (RMSE of 309.10, MAE of 117.82, %RMSE of 10.90%), while Pakistan's data shows a comparatively precise projection (RMSE of 187.99, MAE of 92.82, %RMSE of 2.99%). The scatter plots collectively underscore the varying degrees of projection accuracy and the influence of regional characteristics on population modelling within the Indus River Basin. These validation results not only highlight the overall effectiveness of the modelling approach used but also suggest areas where model calibration or enhanced data gathering may further refine the projections.

Based on the grid projection results, we further validated the county results within the Indus River Basin. Given that county-level census data is only available for Pakistan in 2017, The gridded population projections for 2017 were aggregated to 25 Pakistan counties located in Indus River Basin. These projected county totals were compared with the actual county total population and computed the relative error (%) in Table 4. Most counties exhibit moderate relative errors, falling within the range of 2–4%, suggesting reasonable model accuracy overall.

Usage Notes

We offer a range of future SSPs population projections for every age group of male and female from 2020 to 2100 at a five-year interval period. Each projection result has the temporal distribution of the population at 2.5 arc-minutes (~5 km). We generate a collection of high-quality projections to gain a better understanding of demographic trends and make the population projection products publicly accessible. The values of MAE, RMSE, and %RMSE are used to validate the projected population data at each grid level in order to verify the accuracy of the population projection data. The verification findings demonstrate that our population projection

County	R.E (%)	County	R.E (%)	County	R.E (%)
Hyderabad	4.26	Malakand	2.29	Quetta	6.24
Larkana	3.21	Kohat	3.40	Sibbi	3.00
Multan	2.33	Kalat	6.99	Zhob	7.41
Peshawar	4.37	Bannu	5.77	FATA	3.99
Mardan	1.25	Islamabad	2.04	Hazara	3.88
Bahwalpur	2.49	Nasirabad	2.57	DI Khan	2.82
Rawalpindi	6.78	DG Khan	1.96	Faisalabad	6.61
Lahore	2.02	Sargodha	3.78		
Sukkar	1.90	Gujranwala	3.21		

Table 4. Technical validation between actual and simulated population in 25 counties in Indus River Basin in 2017.

has only minor errors in the majority of the areas that can truly project future population distributions and changes. However, a shorter validation period may exaggerate the performance of the model, and there may be some uncertainty in the long-term projection under SSPs. In addition, the study has some limitations. First, population projections do not account for potential future policy changes such as international agreements which could alter demographic trends. Second, economic crises including recession were not incorporated into the modeling framework which may have influence on population dynamics through fertility, mortality and migration. Third, the study did not consider any pandemic or large scale public health crisis which can alter the population structure and growth. Fourth, any natural hazard such as flood, drought, earthquake, and other climate related disaster were not integerated into the projection model which may have influence on population distribution. Lastly, There are some limitations in our simulation of population migration flows between grids. the study did not incorporate factors such as rural and urban migration and landuse cover characteristics, which could affect the spatial accuracy and overall reliability of the population projections.

Code availability

Gridded population projection datasets were created by using Matlab R2014a and ArcMap 10.8. Codes are available at https://github.com/Barira673/population_projection_R-codes/tree/main.

References

1. Smolenaars, W. J. *et al.* From narratives to numbers: Spatial downscaling and quantification of future water, food & energy security requirements in the Indus basin. *Futures* **133**, 102831 (2021).
2. Khan, A. & Idrees, M. H. The Impact of Climate Change on the Indus Basin: Challenges and Constraints. In Water Policy in Pakistan: *Issues and Options* 225–248 (Springer, 2023).
3. Ilyas, M. Pakistan, India and the Indus River Basin. *CUNY City College* https://academicworks.cuny.edu/cc_etds_theses/1102/ (2023).
4. Lytton, L., Ali, A., Garthwaite, B., Punthakey, J. F. & Saeed, B. Groundwater in Pakistan's Indus Basin (2021).
5. Fisher, F. & Cook, S. *Water, Food and Poverty in River Basins* (Routledge Press, 2012).
6. Kidwai, S. *et al* In *Coasts and Estuaries* Ch. 12. The Indus delta—Catchment, river, coast, and people (2019).
7. Yasin, H. Q., Breadsell, J. & Tahir, M. N. Climate-water governance: a systematic analysis of the water sector resilience and adaptation to combat climate change in Pakistan. *Water Policy* **23**, 1–35 (2021).
8. Helmy, H. E. *Handbook on Alternative Global Development* (Routledge Press, 2023).
9. Smolenaars, W. J. *et al.* Spatial adaptation pathways to reconcile future water and food security in the Indus River basin. *Commun Earth Environ* **4**, 410 (2023).
10. Dellink, R., Chateau, J., Lanzi, E. & Magne, B. Long-term economic growth projections in the Shared Socioeconomic Pathways. *Global Environmental Change* **42**, 200–214 (2017).
11. Jones, B. & O'Neill, B. C. Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. *Environmental Research Letters* **11**, 84003 (2016).
12. KC, S. & Lutz, W. Demographic scenarios by age, sex and education corresponding to the SSP narratives. *Popul Environ* **35**, 243–260 (2014).
13. Gao, J. & O'Neill, B. C. Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. *Nat Commun* **11**, 2302 (2020).
14. Jiang, T. *et al.* National and provincial population and economy projection databases under Shared Socioeconomic Pathways (SSP1-5) v2. *Science Data Bank* https://cstr.cn/31253.11.sciencedb.01683 (2024).
15. Wang, X., Meng, X. & Long, Y. Projecting 1 km-grid population distributions from 2020 to 2100 globally under shared socioeconomic pathways. *Sci Data* **9**, 563 (2022).
16. Arfanuzzaman, M. & Dahiya, B. Sustainable urbanization in Southeast Asia and beyond: Challenges of population growth, land use change, and environmental health. *Growth Change* **50**, 725–744 (2019).
17. Chen, Y. *et al.* Provincial and gridded population projection for China under shared socioeconomic pathways from 2010 to 2100. *Sci Data* **7**, 83 (2020).
18. Jing, C. *et al.* Gridded valueadded of primary, secondary and tertiary industries in China under Shard Socioeconomic Pathways. *Sci Data* **9**, 309 (2022).
19. Socioeconomic Data and Applications Center (SEDAC). Gridded Population of the World, Version 4 (GPWv4) https://sedac.ciesin.columbia.edu/data/collection/gpw-v4 (2020).
20. CSO, MoPH & ICF. Afghanistan Demographic and Health Survey 2015 (Central Statistics Organization, 2017).
21. National Institute of Population Studies (NIPS) & ICF. Pakistan Demographic and Health Survey 2017-18 (NIPS & ICF, 2019).
22. International Institute for Population Sciences (IIPS) & ICF. National Family Health Survey (NFHS-5), 2019–21: India: Volume I (IIPS, 2021).

23. O'Neill, B. C. *et al*. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environ. Chang.* **42**, 169–180 (2017).

24. KC, S. & Lutz, W. The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100. *Global Environmental Change* **42**, 181–192 (2017).

25. Jing, C., Haider, B., Su, B. & Jiang, T. Gridded population structure projection for the Indus River Basin under Shared Socioeconomic Pathways (SSPs). *Science Data Bank* https://doi.org/10.57760/sciencedb.19778 (2025).

Acknowledgements

Author contributions

T. Jiang and B.D. Su conceived the study. B. Haider and C. Jing contributed equally to this paper by producing and drafting the dataset. A. Ashraf and M.A. Hussain discussed the results and edited the manuscript. J.L. Huang and Y.J. Wang provided assistance and advice during the revision.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to B.S. or T.J.

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.