nature sustainability

Article

https://doi.org/10.1038/s41893-024-01496-3

Data-driven strategies to improve nitrogen use efficiency of rice farming in South Asia

Received: 28 November 2023

Accepted: 19 November 2024

Published online: 6 January 2025

Check for updates

Sam Coggins[1], Andrew J. McDonald[2], João Vasco Silva[3], Anton Urfels[4,5], Hari Sankar Nayak[2], Sonam Rinchen Sherpa[6], Mangi Lal Jat[7], Hanuman Sahay Jat[8], Tim Krupnik[9], Virender Kumar[10], Ram. K. Malik[6], Tek B. Sapkota[11], Amaresh Kumar Nayak[12] & Peter Craufurd[13]

Increasing nitrogen use efficiency (NUE) in agricultural production mitigates climate change, limits water pollution and reduces fertilizer subsidy costs. Nevertheless, strategies for increasing NUE without jeopardizing food security are uncertain in globally important cropping systems. Here we analyse a novel dataset of more than 31,000 farmer fields spanning the Terai of Nepal, Bangladesh's floodplains and four major rice-producing regions of India. Results indicate that 55% of rice farmers overuse nitrogen fertilizer, and hence the region could save 18 kg of nitrogen per hectare without compromising rice yield. Disincentivizing this excess nitrogen application presents the most impactful pathway for increasing NUE. Addressing yield constraints unrelated to crop nutrition can also improve NUE, most promisingly through earlier transplanting and improving water management, and this secondary pathway was overlooked in the IPCC's 2022 report on climate change mitigation. Combining nitrogen input reduction with changes to agronomic management could increase rice production in South Asia by 8% while reducing environmental pollution from nitrogen fertilizer, measured as nitrogen surplus, by 36%. Even so, opportunities to improve NUE vary within South Asia, which necessitates sub-regional strategies for sustainable nitrogen management.

The Green Revolution spurred substantial increases in crop productivity and economic development in South Asia. Since the 1960s, farmers in South Asia have more than tripled their cereal productivity[1], greatly reducing poverty and food insecurity[2]. Ample use of nitrogen fertilizer has been a critical driver of increased productivity[3], with the use of inorganic nitrogen fertilizer in the region increasing from 2 to 111 kg ha^{-1} yr^{-1} over the past six decades[1].

Overuse of nitrogen fertilizer has also contributed to climate change, air pollution, water pollution and biodiversity loss[4]. Recent estimates suggest that 5% of global anthropogenic greenhouse gas emissions are linked to nitrogen fertilizer use, and that two-thirds of these emissions emanate from agricultural nitrogen surpluses at the field scale[5]. Nitrogen surplus, that is, the difference between nitrogen inputs and nitrogen outputs from cropping systems, has been proposed as a key indicator for nitrogen pollution[6–8]. Cereal production systems in South Asia have been identified as a global 'hotspot' for nitrogen surpluses as an unintended consequence of high and relatively inefficient fertilizer use[9–13]. Such large nitrogen surpluses also

[1]Australian National University, Canberra, Australian Capital Territory, Australia. [2]Soil and Crops Sciences—School of Integrative Plant Science, Cornell University, Ithaca, NY, USA. [3]Sustainable Agrifood Systems Program, International Maize and Wheat Improvement Center (CIMMYT-Zimbabwe), Harare, Zimbabwe. [4]International Rice Research Institute (IRRI), Los Baños, the Philippines. [5]Water Resources Management Group, Wageningen University & Research, Wageningen, Netherlands. [6]International Maize and Wheat Improvement Centre (CIMMYT), New Delhi, India. [7]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India. [8]ICAR—Indian Institute of Maize Research (IIMR), Ludhiana, India. [9]International Maize and Wheat Improvement Center (CIMMYT), Sustainable Agrifood Systems Program, Dhaka, Bangladesh. [10]Sustainable Impact Department, International Rice Research Institute (IRRI), Los Baños, the Philippines. [11]International Maize and Wheat Improvement center (CIMMYT), Texcoco, Mexico. [12]ICAR—National Rice Research Institute, Cuttack, India. [13]CIMMYT, NARC Research Station, Kathmandu, Nepal. e-mail: samcoggins55@gmail.com

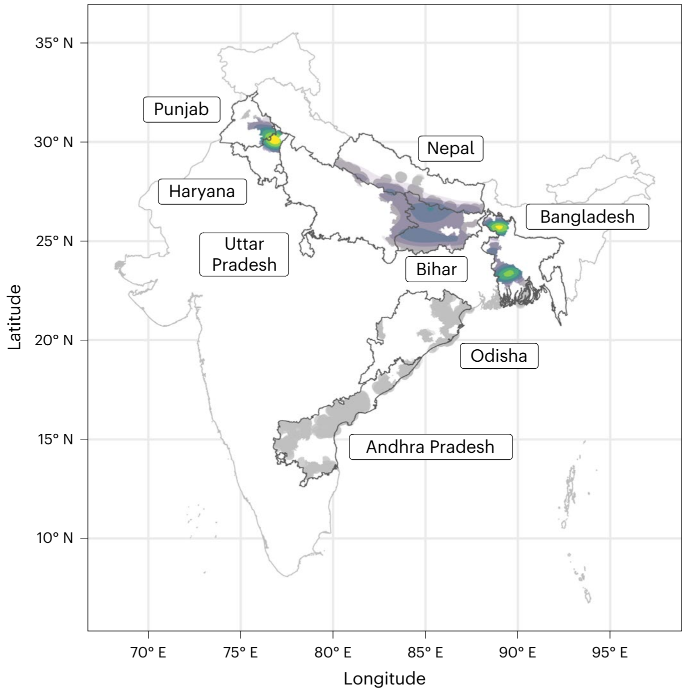

Fig. 1 | Sample location density for all rice fields used in the analysis. All surveyed rice crops are marked with a grey dot. The violet–blue–yellow colour gradient represents low–high densities of samples. Map created with geodata[67] with administrative boundaries from GADM Global Administrative Areas (Version 4.1).

create large fiscal burdens for South Asian governments that subsidize nitrogen fertilizer[14]. For example, the Indian government budgeted US$16 billion for urea subsidies in 2023/2024 alone[15].

Rice is a particularly important priority for effective nitrogen management in South Asia. Farmers in South Asia harvest about 65 million hectares of rice per year, using more land and nitrogen fertilizer than any other crop[1,16]. In response to growing domestic and international demand, India is also projected to be the largest global source of new rice production in the current decade[17]. Therefore, a central sustainable development challenge for South Asia is increasing rice production without jeopardizing fiscal sustainability and vital ecosystem services. Reconciling these competing demands hinges on improving nitrogen use efficiency (NUE), defined here as partial factor productivity of nitrogen, that is, the amount of rice produced per unit of nitrogen applied[18,19]. Increasing NUE provides a powerful lever to increase food security while reducing environmental and subsidy costs of nitrogen fertilizers[20,21]. Indeed, increasing NUE has been identified as the most powerful mechanism for reducing greenhouse gas emissions from nitrogen fertilizer[5].

Nevertheless, the overall scope and specific strategies for improving the NUE of rice farming in South Asia remain uncertain. Controlled experiments provide valuable insights into how agronomic factors influence the NUE of rice crops[22–24]. Yet controlled-condition experiments typically do not account for the diverse biophysical and socio-economic conditions under which farmers operate, rendering these experiments of limited value for ‘real-world’ decision making[25–27]. Conversely, observational studies generally rely on small-*n* surveys that miss the substantial variation within and between farming regions[28,29]. Emerging large-*n* surveys conducted over broad geographic regions offer new possibilities for data-driven approaches to NUE assessment and improvement[30].

In this study, we analysed a large and detailed dataset of 31,483 monsoon season rice fields (Fig. 1). The dataset spans six important rice-producing regions of South Asia: Bangladesh’s floodplains, the Terai of Nepal and four major rice-producing regions of India (Andhra Pradesh, Bihar and neighbouring districts in eastern Uttar Pradesh, Odisha, and Punjab and Haryana). This unique farmer field dataset was used to assess (1) the current state of rice NUE in South Asia, (2) the scope to improve rice NUE through improved agronomy and (3) the most effective agronomic levers to realize this potential across the region. Our results can be used to support more productive and sustainable rice farming systems in South Asia.

Results

Current NUE of rice crops in South Asia

We observed large variation in NUE across surveyed regions (Fig. 2a). The Indian state of Bihar and adjacent districts of eastern Uttar Pradesh featured high nitrogen inputs and low yields, with a total NUE of 32 kg kg^{-1} (Fig. 2a). Consequently, Bihar and eastern Uttar Pradesh generated the most nitrogen surplus of all regions (Fig. 2a), nitrogen surplus being the fraction of nitrogen fertilizer uncaptured by the crop and thereby susceptible to environmental loss. The average rice field in Punjab and Haryana received the most nitrogen fertilizer and thereby produced a similarly large nitrogen surplus per unit area (Fig. 2a). However, lower yields in Bihar and eastern Uttar Pradesh meant that this region still generated the most nitrogen surplus per unit area, and almost twice as much nitrogen surplus per unit of rice produced compared with all other regions (Fig. 2a). Notwithstanding this variation in NUE across regions, we also observed substantial variation in NUE within all regions (Fig. 2b).

The highest region-level NUEs were observed in Odisha and the Terai of Nepal, at 60 and 52 kg kg^{-1}, respectively (Fig. 2a). This was expected given the lower nitrogen rates in these regions and the expected diminishing marginal returns from fertilizer. To facilitate robust ‘like with like’ comparisons of NUE across regions, we then characterized NUE as a function of nitrogen application rate. We used splines to capture both average behaviour (solid line) and the NUE frontier (dotted line), which was defined as the 75th percentile of the data distribution within each region (Fig. 3 and Supplementary Information 2). For most nitrogen application rates, the average NUE in Andhra Pradesh and Punjab and Haryana was higher than the frontier NUE in all other regions (Fig. 3). Rice crops in Punjab and Haryana achieved this NUE despite receiving the lowest phosphorus and potassium application rates (Supplementary Information 3). These substantially different NUE outcomes probably reflect regional differences in yield potential, soil nitrogen supply and propensity for nitrogen losses mediated by soil hydrology and the consistency of flooding, which influence processes such as denitrification and leaching[31].

Potential to improve NUE of rice crops in South Asia

By definition, NUE can be increased in two ways: (1) reducing nitrogen application without reducing yield (‘N-saving’ pathway) and (2) increasing yield without increasing nitrogen application (‘yield-gain’ pathway). We estimated the potential NUE gains from the N-saving pathway by supposing that all surveyed fields capped nitrogen fertilizer use at the average rate at which rice yields in each region stopped responding to increased nitrogen (see Supplementary Information 4 for the maximum productive nitrogen rate estimates for each region). We also estimated the potential NUE gains from the yield-gain pathway by supposing all surveyed fields achieved at least the average yield for their respective nitrogen application rate and region (Supplementary Information 4). Finally, we estimated the potential NUE gains of surveyed fields in situations where both pathways were possible.

For the N-saving pathway, we estimated that the average rice field could have produced the same yield with 18 kg ha^{-1} less nitrogen (Fig. 4). This potential nitrogen saving was particularly large considering the average accounts for all surveyed rice fields, not just the 55% of fields that could have reduced nitrogen application without incurring a yield loss (Supplementary Information 5).

a

Region	Monsoon rice area (million ha)	Mean N fertilizer (kg ha^{-1})	Mean rice yield (t ha^{-1})	NUE (kg kg^{-1})	N surplus per planted area (kg ha^{-1})	N surplus per rice harvested (kg t^{-1})
Andhra Pradesh	1.3	133	5.8	44	78	13
Bihar and eastern Uttar Pradesh	4.3	126	4.0	32	99	25
Odisha	3.5	67	4.1	60	40	10
Punjab and Haryana	4.6	155	6.8	44	83	12
Bangladesh's floodplains	4.8	102	4.8	47	63	13
Terai of Nepal	1.4	79	4.1	52	51	12
Aggregate	19.8	114	5.0	46	71	15

b

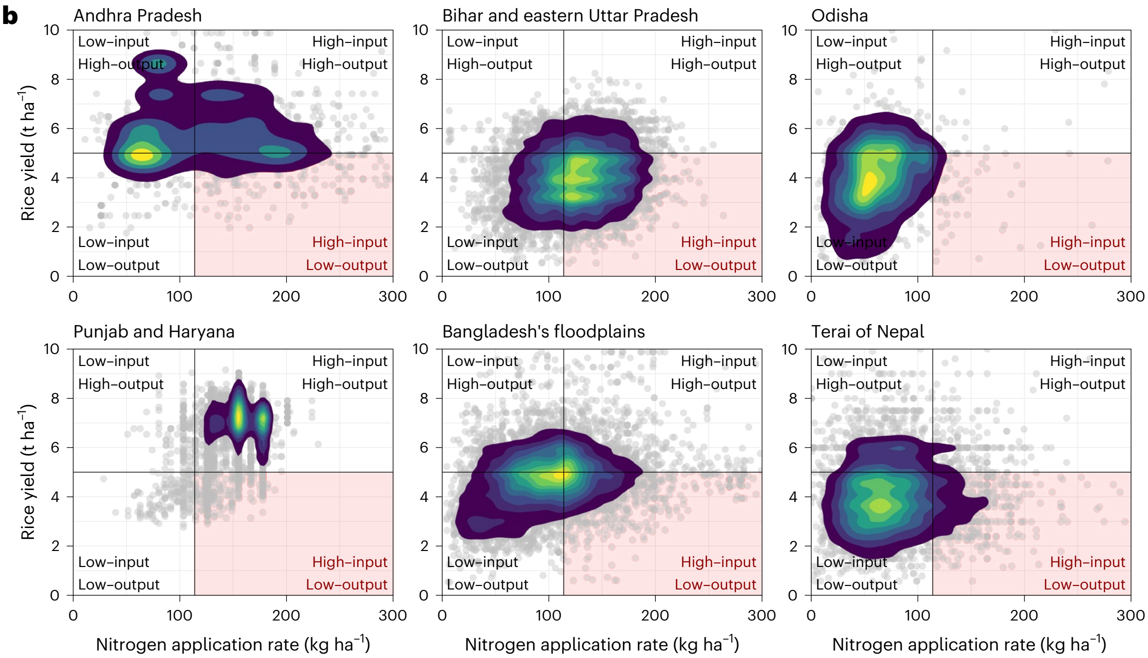

Fig. 2 | Variation of monsoon rice crops in regard to mean nitrogen application rate, mean yield, total NUE and total nitrogen surplus. a,**b**, Across (as averages) (**a**) and within (**b**) regions: Andhra Pradesh (n = 1,465), Bihar and eastern Uttar Pradesh (n = 9,579), Odisha (n = 1,204), Punjab and Haryana (n = 5,723), Bangladesh floodplains (n = 8,676) and the Terai of Nepal (n = 4,836). NUE is defined as partial factor productivity of nitrogen, that is, the amount of rice produced per unit of nitrogen applied. Nitrogen surplus was calculated per unit of planted area, as well as per unit of rice (that is, nitrogen emission intensity). Aggregate values for each metric were calculated by averaging region-level values and weighting these region-level values by monsoon rice area. Each metric in **a** is colour coded according to whether a higher value is desirable (green), undesirable (red) or without clear (un)desirability (orange), following the EU Nitrogen Expert Panel (2015) guidelines[8]. Values in **a** are shaded with a light–dark gradient, representing low–high magnitude of these metrics. Each surveyed rice field in **b** is represented by a grey dot. The blue–yellow colour gradient represents low–high densities of surveyed fields in each region. To facilitate comparisons, each panel in **b** is split into four quadrants using average rice yield (5.0 t ha^{-1}) and average nitrogen application rate (114 kg ha^{-1}) across all regions. The bottom-right quadrant in each panel is shaded red because high-input-low-output farming systems are typically undesirable for farmers, governments, food consumers and the environment[8]. See Supplementary Information 1 for the percentage of fields in each quadrant for each region, the standard error of the regional means for each variable and the statistical significance of differences between these regional means.

In aggregate, such savings in applied nitrogen would have increased NUE by 22%, reduced nitrogen surplus by 27% and reduced rice nitrogen fertilizer subsidy costs by 17% (Fig. 4). Importantly, the potential to reduce nitrogen application rates varied widely across regions (Supplementary Information 4). Only 12% of the rice fields in the Terai of Nepal received excess nitrogen fertilizer, whereas this figure ranged between 38% and 72% for the other surveyed regions (Supplementary Information 5). At present, the low cost of nitrogen fertilizers in South Asia provides limited economic incentives for farmers to pursue the N-saving pathway (Fig. 4). These incentives are particularly weak in India and Bangladesh, where nitrogen fertilizers are the most heavily subsidized (Supplementary Information 5).

Our results also indicate a substantial opportunity to increase NUE via the yield-gain pathway. On average, a yield gain of 369 kg ha^{-1} rice was possible if all rice fields reached at least the average NUE for their respective region and nitrogen rate (Fig. 4). Unlike the N-saving pathway, the scope for the yield-gain pathway was relatively consistent across all regions, ranging from an average potential gain of 322 kg ha^{-1} rice in Punjab and Haryana to 438 kg ha^{-1} rice in Andhra Pradesh. Such yield increases would increase NUE by 7%, reduce nitrogen surplus by 9% and increase rice production by 8%. These benchmarks are conservative, and positive impacts would be even greater if the 75th percentile NUE would be used as an achievable NUE target (Fig. 3).

Implementing the N-saving and yield-gain pathways in tandem across all regions would increase NUE by 32%, increase rice production by 8%, reduce nitrogen surplus by 36% and reduce rice nitrogen fertilizer subsidy costs by 17% (Fig. 4). The following section delves into how farmers and governments might realize this potential.

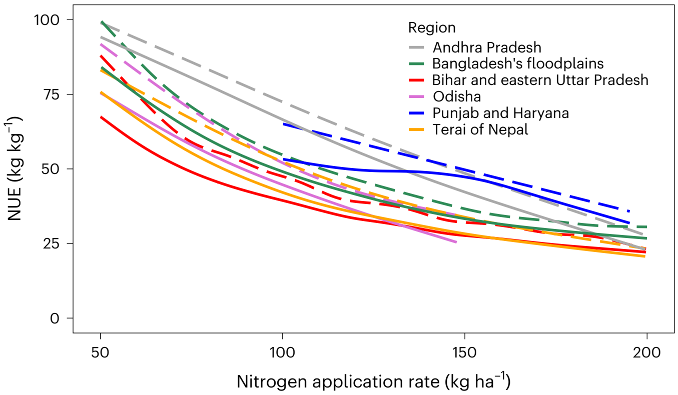

Fig. 3 | Average and 75th percentile NUE across different nitrogen application rates (kg ha^{-1}) and rice-producing environments of South Asia. NUE is defined as partial factor productivity of nitrogen, 397998809 l 81.943 -397

Agronomic mechanisms for realizing potential NUE

The agronomic mechanism for realizing the N-saving pathway entails reducing nitrogen application rates. The agronomic factors associated with the yield-gain pathway are more complex. To identify region-specific drivers of this pathway, we used Random Forest models to predict whether a given rice field achieved above- or below-average NUE for its region and nitrogen application rate. We then estimated which variables contributed most to these predictions using Shapley values, a post hoc method for ascribing variable importance based on game theory[32]. A high Shapley score for a given predictor implies its influence on the predicted NUE outcome is large (Fig. 5).

Variation in hydrology (rainfall and irrigation) and crop timing (transplanting date and field duration) were the most important NUE predictors for the yield-gain pathway (Fig. 5). This trend was consistent across all regions. These associations were also robust across different nitrogen application rates and surveyed years, as well as for models that predicted NUE deviation from the splines in Fig. 3 as a continuous variable rather than a categorical variable (Supplementary Information 7).

Transplanting date and number of irrigations were particularly important NUE predictors in most regions (Fig. 5). More irrigation and earlier transplanting were generally associated with higher NUE for any given nitrogen application rate (Fig. 6). Responsiveness to irrigation was observed in all regions, except Andhra Pradesh, where no clear trend was observed (Fig. 6a), and in Bangladesh's floodplains, where irrigation data were not available. The trend for transplanting date was present in all regions except for the Terai of Nepal (Fig. 6b). We note that the effect of transplanting date and number of irrigations on NUE was influenced by interactions with other variables. This is indicated by the spread around the mean trend lines displayed in Fig. 6. The relationships between NUE and its other important predictors are reported in Supplementary Information 8.

Discussion

There is growing recognition that efficient nitrogen management is essential for ensuring food security while minimizing the environmental externalities of crop production systems[5,21]. South Asian cropping systems are a global hotspot for nitrogen pollution[10,11,13], and progress towards more-efficient use of nitrogen has been limited[9]. Rice is the most widely cultivated staple crop in South Asia, hence central to the sustainable development imperative of improving NUE in the region. To understand the scope and mechanisms for improving NUE in rice fields, we developed a data-driven approach that leverages a large-*n* survey of individual fields across six rice production environments spanning India, Nepal and Bangladesh. Substantial opportunities to increase NUE were identified via two complementary pathways: curbing excess nitrogen application (N-saving pathway) and addressing yield constraints unrelated to crop nutrition (yield-gain pathway).

Curbing excess nitrogen application (N-saving pathway) appears to be the most impactful mechanism for increasing NUE in South Asia (Fig. 4). Previous attempts to quantify excess fertilizer application relied on generalized assumptions regarding the relative proportion of farms that overuse nitrogen[33–35]. Here we used machine-learning models to identify region-specific nitrogen application rate thresholds beyond which negligible increases in rice yield were observed for the average rice field in each region (Supplementary Information 4). On the basis of these thresholds, substantial nitrogen-saving opportunities were identified in all surveyed regions, except the Terai of Nepal (Supplementary Information 1). In aggregate, our analysis suggests that the average rice field in South Asia could achieve the same rice yield with 18 kg ha^{-1} less applied nitrogen (Fig. 4). Importantly, not all rice fields reflect this average; clearly some farmers could reduce their nitrogen application rate much more than others (Fig. 2a,b and Supplementary Information 4). Even so, the identified region-specific nitrogen application thresholds provide actionable, evidence-based

https://doi.org/10.1038/s41893-024-01496-3

Predictor	Category	Mean	Andhra Pradesh	Bihar and eastern Uttar Pradesh	Odisha	Punjab and Haryana	Bangladesh floodplains	Terai of Nepal
Model accuracy (out-of-bag)			88%	83%	84%	85%	84%	82%
Sample size			*n* = 873	*n* = 8,707	*n* = 772	*n* = 5,160	*n* = 3,993	*n* = 2,942
Total rainfall	Rainfall	3.79	4.59	2.16	4.30		2.63	5.25
Monsoon onset date	Rainfall	3.53	2.15	4.27	2.50	6.78	3.42	2.08
Number of irrigations	Irrigation	3.32	1.26	5.23	2.39	4.06		3.64
Field duration	Crop timing	3.28	1.04	3.63	1.78	6.43		3.52
Monsoon retreat date	Rainfall	2.89	1.58	1.91	0.54	0.69	5.12	7.49
Transplanting date	Crop timing	2.87	2.63	2.24	4.28	4.73	2.22	1.14
Seed rate	Other management	2.80	2.18	1.41	5.31		2.91	2.21
Drought severity	Rainfall	2.51		2.31	2.71			
Nitrogen splitting	Nutrient management	2.41	2.33	1.39	1.07	6.05	2.44	1.16
Soil pH	Soil and landscape	2.40		1.65		4.10	2.52	1.34
Soil sandiness	Soil and landscape	2.23	0.93	2.97	2.67	2.82		1.78
Average dry-spell length	Rainfall	1.95	1.04	1.83	1.22	2.86	2.41	2.36
Nitrogen applied at planting	Nutrient management	1.91	2.44	1.32	1.23	1.71	3.58	1.19
Phosphorus fertilizer	Nutrient management	1.84	3.12	1.04	1.37	2.14	2.11	1.24
Soil carbon	Soil and landscape	1.83	0.92	1.81	1.59		2.53	2.28
Potassium fertilizer	Nutrient management	1.73	1.95	1.21	2.17	2.04	2.17	0.83
Nitrogen applied at tillering	Nutrient management	1.59	1.37	0.87	2.22		2.49	0.98
Gypsum fertilizer	Nutrient management	1.54					1.54	
Farm income dependence	Socioeconomic	1.53	1.23	2.25	1.10			
Nursery duration	Crop timing	1.49	1.41	1.24	1.93	2.23	1.64	0.49
Lodging	Other management	1.49	1.38	0.59	2.49			
Flood severity	Rainfall	1.46		0.31	2.60			
Market distance	Socioeconomic	1.33	2.19	1.16	0.75			1.22
Organic fertilizer	Nutrient management	1.22	1.43	0.62	1.41		1.18	1.47
Previous crop residue	Nutrient management	1.08	1.58	0.59				
Zinc fertilizer	Nutrient management	1.07	0.61	1.11		1.11	1.87	0.66
Transplanting method	Other management	0.84	0.55		0.14		1.82	
Soil density	Soil and landscape	0.79	1.60	0.94			0.37	0.24
Variety type	Other management	0.77	1.55	0.30	0.13		0.25	1.62
Disease severity	Other management	0.74	0.75	0.46	1.02			
Household members	Socioeconomic	0.74	0.48	0.55	0.73	1.17	0.64	0.84
Farmer social category	Socioeconomic	0.65	0.39	0.38	1.18			
Farmer education	Socioeconomic	0.60	0.37	0.77	0.67			
Landscape position	Soil and landscape	0.56	0.88	0.79	0.35			0.24
Farmer soil perception	Soil and landscape	0.50	0.86	0.15				
Weed severity	Other management	0.46	0.26	0.47	0.55	0.58	0.32	0.61
Farmer-judged soil texture	Soil and landscape	0.32	0.40	0.33	0.13			0.40
Farmer gender	Socioeconomic	0.23						0.23

Fig. 5 | Apparent drivers of rice NUE within regions. Variable importance for classification of random forest models predicting whether a given rice field exceeded the average NUE for its given nitrogen application rate and region (that is, above its region-specific solid line in Fig. 3). Each value represents the global Shapley value as a percentage, indicating the average absolute size of influence, not the direction of influence. Blank cells represent instances where a predictor is not applicable in a region's random forest model due to data limitations. The yellow–dark green colour gradient represents low–high magnitudes of these values. The predictor categories (in the 'Category' column) are colour coded for ease of interpretation. The out-of-bag prediction accuracy of each region-specific random forest model is provided in the first row. See Methods for detail and Supplementary Information 6 for variable definitions.

 https://doi.org/10.1038/s41893-024-01496-3

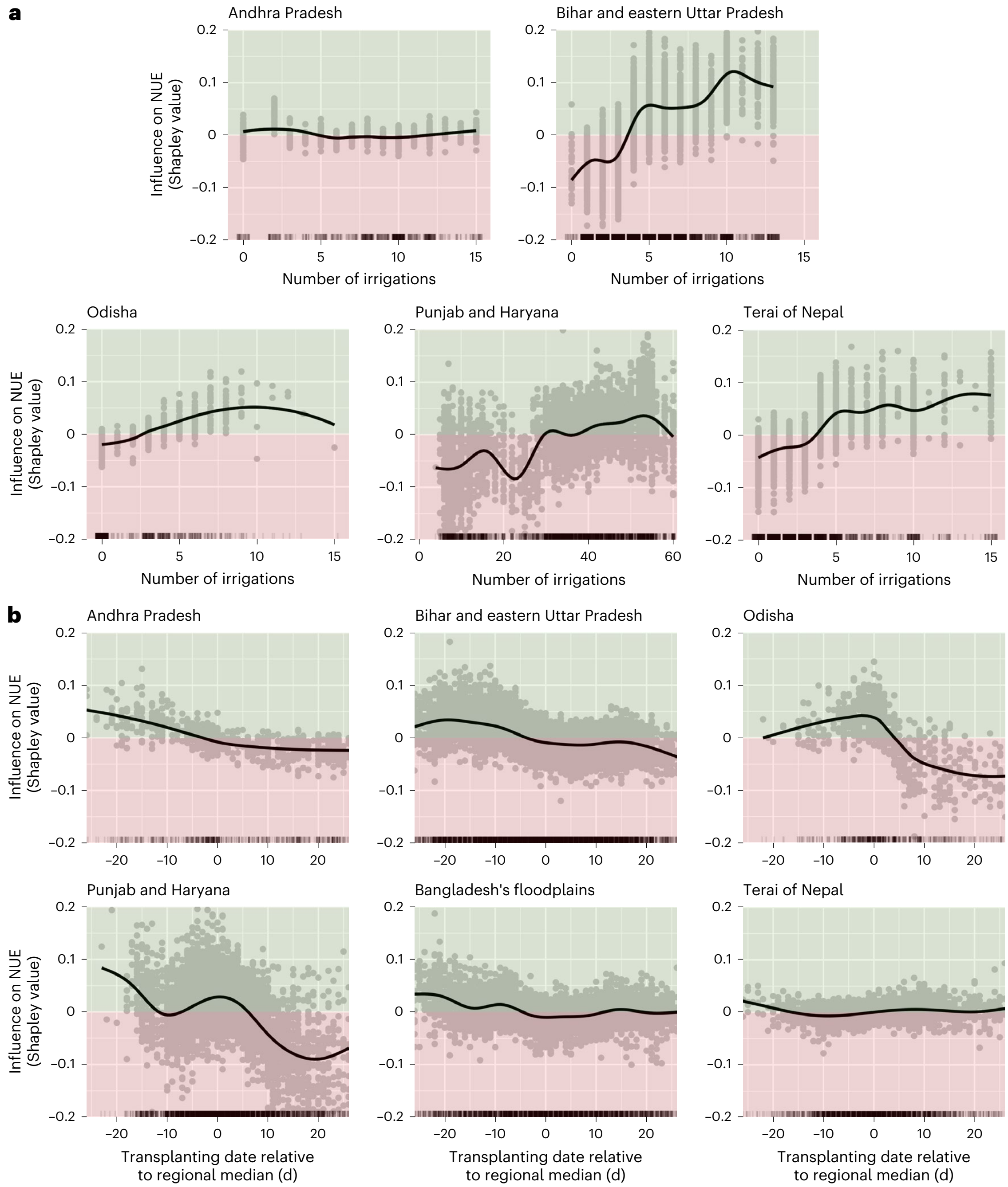

Fig. 6 | Apparent effects of irrigation and planting data on rice NUE. a,**b**, Effect of number of irrigations (**a**) and transplanting date (**b**) on rice NUE. Each grey dot represents the Shapley value for an individual field, with positive values indicating a positive impact on predicted NUE (shaded green) and negative values indicating a negative impact on predicted NUE (shaded red). See Methods for details regarding these Shapley values. The black line displays a trendline fitted to the data. Vertical x ticks represent the sample density. Irrigation data were unavailable for Bangladesh's floodplains; hence, relationships are not shown in **a**. See Supplementary Information 6 for variable definitions.

and region-level targets for policymakers to reduce nitrogen use without jeopardizing food security (Supplementary Information 4). The need for such targets was identified in the 2022 Intergovernmental Panel on Climate Change (IPCC) report on climate change mitigation[36]. Additional nitrogen fertilizer savings may be enabled by breeding rice varieties with greater NUE[37], using more-efficient nitrogen fertilizers[38] or introducing short-duration summer pulse crops in rice-based cropping systems to enhance biological N_2 fixation[39,40].

Rice farmers in South Asia currently have limited economic incentives to curb excess nitrogen application (Fig. 4). Reducing nitrogen application rates by an average 18 kg ha^{-1} nitrogen would reduce nitrogen pollution potential by 27%, measured as the nitrogen surplus uncaptured by the crop[7] (Fig. 4). However, this reduction in nitrogen application would reduce total rice production costs for farmers by only roughly 2% (Fig. 4), as governments currently subsidize the price of urea paid by farmers in Nepal, Bangladesh and particularly India (Supplementary Information 5). This draws into question the current incentives for reducing nitrogen application rates, particularly in countries with large nitrogen fertilizer subsidies[14]. The government of India has recently introduced the 'Green Credit Programme' to incentivize more environmentally sustainable practices[41]. Replacing fertilizer subsidies with commensurate levels of direct payments to farmers could provide an even stronger set of economic incentives for eliminating excess nitrogen use without reducing support to farmers[14].

Improving water management and the cropping calendar are the most powerful agronomic entry points to increase NUE without changes to current nitrogen application rates (that is, pursuing the yield-gain pathway). The top seven NUE predictors across regions included rainfall, number of irrigations, transplanting date and in-field crop duration (Fig. 5). These top predictors of NUE were also the top predictors of rice yield within each region (Supplementary Information 7), which supports the hypothesis that matching nitrogen supply (fertilizer rate) with nitrogen demand (attainable crop yield potential) is paramount to improving NUE in South Asian rice farming. This probably explains why the high-yielding regions of Andhra Pradesh and Punjab and Haryana have higher average NUE at any given nitrogen application rate (Fig. 3). In other words, NUE tends to be high when crop demand for nitrogen is high. Targeting yield gap closure to improve NUE of South Asia's rice farms could reduce nitrogen surplus by 9% and increase aggregate rice production by at least 8%, if all farmers achieve at least the average current yield for their nitrogen rate and region (Fig. 4). Further improvements in NUE would also be possible if even larger yield gains are achieved.

The 2022 IPCC report on climate change mitigation overlooked the yield-gain pathway as a means for mitigating greenhouse gas emissions[36]. Three reasons make increasing yields, without increasing nitrogen application, an attractive mitigation option. First, increasing crop yield directly reduces N_2O emissions—albeit to a lesser extent than reducing nitrogen application—because a larger proportion of reactive nitrogen is taken up by crops (Fig. 4 ref. 7,42). Second, there are often profit and food security incentives for farmers and governments to increase crop yield (Fig. 4), and these incentives will probably increase as regional and global demand for rice grows[17]. Third, increased land productivity may reduce the area dedicated to rice production in regions or seasons where other crops can be reliably cultivated, thereby reducing methane emissions[43–45]. The 2022 IPCC climate change mitigation report mentions 'nutrient management' 27 times but does not identify addressing yield constraints as an additional lever for climate change mitigation[36].

Nutrient management practices beyond nitrogen application rate did not have a strong influence on NUE in our analysis (Fig. 5). We anticipated that other dimensions of nutrient management (for example, fertilizer application timings and crop residue management) would prove to be strong NUE predictors given that research trials have consistently substantiated the role of the '4R' principles such as balanced fertility and splitting nitrogen applications for improved NUE outcomes (for example, refs. 22–24). For context, the '4R' principles of nutrient management refers to the 'right source' (fertilizer type), 'right rate' (fertilizer quantity), 'right time' (fertilizer application timing), and 'right place' (fertilizer placement). Our findings suggest that the 4Rs, beyond nitrogen application rate, become important to NUE only once yield constraints associated with water management and basic agronomy are overcome. These insights also underscore the limitations of using controlled-condition experimental trials to establish drivers of crop yield and NUE outcomes in real-world production environments.

Despite the promise of the yield-gain pathway, it must be acknowledged that there are implementation trade-offs, uncertainties and challenges associated with all agronomic practice changes that serve to improve yields. First, there may be downside risks that erode the ecosystem services associated with achieving higher NUE. For example, improving water management of rice crops commonly, but not always, involves increased irrigation[46]. This increased irrigation can increase groundwater depletion[47] and methane emissions, the most damaging greenhouse gas emitted from rice farms[36]. Second, optimal crop calendar and water management practices for increasing rice productivity vary within and across regions and seasons, posing substantial implementation challenges (Fig. 6 ref. 48). Third, it is often unclear how to align water management and cropping calendar adjustments with farmers' constraints and other priorities beyond increasing rice yields and NUE[49]. The large variation in NUE outcomes within and across the surveyed regions demonstrates the critical need for geographically targeted agricultural development interventions (Figs. 1, 2 and 4). Context-specific research with farmers and policymakers is needed to navigate the diverse factors shaping water management and crop calendar decisions and their impacts across South Asia. These factors include and are not limited to vertebrate pests, cultural differences and variable access to machinery, labour, planting material and financial capital[50,51].

In conclusion, we identify two impactful and complementary pathways for improving rice NUE: primarily by disincentivizing excess nitrogen application and secondarily by addressing yield constraints unrelated to crop nutrition. When pursuing these pathways, it is imperative to note that these opportunities are heterogeneous within and across the rice production environments of South Asia. Nuanced policies, enhanced field characterization data and effective targeting that recognizes the needs and diverse motivations of farmers are all required to manage nitrogen sustainably in the rice cropping systems of South Asia.

Methods

Dataset

All analyses relied on one database combining farmer field surveys and secondary environmental data for 31,483 farmer fields that cultivated transplanted rice in a monsoon season between 2016 and 2020. The database covers six regions across South Asia (Fig. 1): Andhra Pradesh (n = 1,465), Bihar and eastern Uttar Pradesh (n = 9,579), Odisha (n = 1,204), Punjab and Haryana (n = 5,723), Bangladesh's floodplains (n = 8,676) and the Terai of Nepal (n = 4,836). See Supplementary Information 9 for the number of samples for each year in each region. We aggregated the northwestern Indian states of Punjab and Haryana because they are commonly grouped in a distinct rice-producing region (the 'Trans-Gangetic Plains') featuring intensive and market-oriented rice–wheat cropping systems[52,53]. We also aggregated Bihar and adjacent districts of eastern Uttar Pradesh: Ballia, Chandauli, Deoria, Ghazipur, Gorakhpur, Kushinagar, Maharajganj, Mau and Siddharthnagar. Researchers and practitioners commonly aggregate Bihar and eastern Uttar Pradesh as a distinct rice production environment (the 'Middle Indo-Gangetic Plains'), featuring fertile soils, variable rainfall, high poverty rates and fragmented landholdings[53]. It should also be noted that the surveyed area of Bangladesh's floodplains encompasses mainly the Tista Meander Floodplain and Ganges River Floodplains and is not necessarily representative of Bangladesh's other floodplains[54].

Farmers reported the data for most variables via primary surveys. The surveys elicited information for each farmer's largest rice plot in the most recent monsoon season. These primary farmer field surveys acquired data for rice yield, nitrogen application rate, plot size and all variables in Fig. 5 (aside from the rainfall variables and most soil and landscape variables). The GPS-recorded latitude and longitude of each surveyed field were also recorded at the time of the survey (Fig. 1). The surveys were implemented independently across all six regions but with metadata that enabled harmonization. All rice yield data were farmer-reported aside from approximately 40% of yield records from Bangladesh's floodplains and approximately 25% of yield records from Punjab and Haryana, where yields were estimated with crop cuts. Crop cut measurements were collected by government sub-assistant agricultural officers in Bangladesh and professional field technicians in Punjab and Haryana. All surveys complied with standards established by the Research Ethics Committee of the International Maize and Wheat Improvement Center, as described in policy number DDG-POL-04–2019. All survey participants gave informed consent to participate.

Surveyed rice fields were selected through a two-stage approach. In the first stage, rice-growing districts within each region were purposively selected on the basis of the districts' importance to food security in their respective states and countries. This included almost all rice-growing districts in Odisha, Bihar and neighbouring districts of eastern Uttar Pradesh. In the second stage, representative rice fields were sampled within each selected district. Villages in Andhra Pradesh, Bihar, eastern Uttar Pradesh and Odisha were selected using a 'probability proportional to population' method[55]. The only exceptions for these regions were for 25% of surveyed rice fields in Andhra Pradesh and 30% of surveyed rice fields in Odisha, where fields were only selected from villages identified to have low or medium levels of soil zinc, according to digital soil maps. The process for this second stage also varied slightly in Punjab and Haryana, where villages were selected on the basis of government-mandated travel restrictions imposed during the COVID-19 pandemic. For all surveys in Nepal, and surveys in Bangladesh up to and including the year 2018, representative rice fields were identified using satellite imagery. Specifically, LANDSAT-derived Normalized Difference Vegetative Index values were extracted to capture the variability in standing green biomass (a proxy for yield) in selected districts to identify rice fields to participate in the survey. These normally distributed Normalized Difference Vegetative Index values were then stratified into four quartiles. Rice fields to be surveyed were then randomly selected from these quantiles so that selected samples proportionally represented the bell curve. Unlike in other regions, farmers that participated in the 2019 survey in Nepal were also invited to participate in the 2020 survey. The process for selecting individual rice fields varied for surveys implemented in Bangladesh after the year 2018. In these surveys, survey enumerators selected nine representative rice fields for each sub-district survey location, three with the best quality crop, three with medium quality and three with the worst quality. In all surveyed regions, additional selections were made if farmers of selected rice fields could not be located or declined to participate.

To complement the primary survey data, we accessed publicly available secondary data. Daily rainfall data were retrieved at 180 second resolution from Climate Hazards Group InfraRed Precipitation with Station data[56]. These data were then used to develop the rainfall variables featured in Fig. 5 and defined in Supplementary Information 6. Soil sand content, pH in water, soil organic carbon content and bulk density were retrieved at 30 second spatial resolution from ref. 57. Note that the spatial resolution of the soil and rainfall data was typically larger than the size of surveyed rice fields and that the surveys were not designed to evaluate the accuracy of the soil and rainfall data at the field level. We therefore relied on the large sample size in our dataset to account for this spatial imprecision in our analyses. See Supplementary Information 6 for variable definitions and data sources.

Our data quality control criteria excluded the following fields from the analysis: rice grown outside of the monsoon season (n = 10,147; outside research scope), direct-seeded rice (n = 1,329; outside research scope), nitrogen application rate equal to 0 kg ha^{-1} (n = 2,480; probably misreported information), nitrogen application rate above 400 kg ha^{-1} (n = 298; probably misreported information), sample coordinates located outside the surveyed region (n = 360), rice yield below 0.5 t ha^{-1} (n = 577; probably an error), missing rice yield data (n = 395), rice yield above 10 t ha^{-1} (n = 541; probably misreported information), nursery duration below 0 days (n = 262; definitely an error), nursery duration above 100 days (n = 273; misreported information), field duration below 60 days (n = 457; probably an error), field duration above 300 days (n = 88; probably an error), crop duration below 90 days (n = 358; probably an error) and crop duration above 340 days (n = 95; probably an error). These exclusion criteria left 28,865 data points for analysis.

Current NUE of rice crops in South Asia

Data for nitrogen application rate and rice yields were used to calculate NUE and nitrogen surplus. NUE was defined as partial factor productivity of nitrogen, that is, the amount of rice produced per unit of nitrogen applied. Nitrogen surplus estimates were based on the nitrogen surplus assumptions in ref. 58. Specifically, total nitrogen input for every surveyed field was estimated as mineral nitrogen fertilizer application rate plus 38 kg ha^{-1} N of non-fertilizer nitrogen from manure, crop residues, deposition, seeds and irrigation water[58]. Total nitrogen output was estimated as total nitrogen in harvested grain (assuming 86% dry matter and a 1.13% nitrogen content) added to nitrogen in the rice straw (assuming a harvest index of 0.481 and nitrogen content of 0.69%)[58]. Nitrogen surplus, also commonly referred to as nitrogen balance, was then calculated as total nitrogen output subtracted from total nitrogen input.

The relationship between NUE and nitrogen application rate was modelled in each region using smoothing splines. The average NUE was modelled as a function of nitrogen application rate by fitting one smoothing spline for all rice fields in each region. This provides a conservative benchmark for NUE improvement. A more ambitious NUE benchmark was modelled with smoothing splines fitted to the 75th percentile of the NUE data. To do so, the 75th percentile NUE within each 10 kg ha^{-1} increment of nitrogen application rate in each region was identified. The relationship between NUE and nitrogen application rate for these 75th percentile rice fields was then fitted using region-specific smoothing splines, assuming each 75th percentile rice field received the midpoint nitrogen application rate of its 10 kg ha^{-1} nitrogen increment. This 75th percentile benchmark was chosen because a lower benchmark would not represent rice fields with high NUE, while a higher benchmark would not be robust to outliers and errors for regions and nitrogen rates with low sampling densities (Supplementary Information 2). All smoothing splines were fitted to the data with the 'ss' function of the 'npreg' R package. The most parsimonious number of degrees of freedom for each smoothing spline was fitted using 'ordinary cross-validation'. Degrees of freedom were subsequently set to three if the ordinary cross-validation led to overfitting[59], detected on the basis of splines not predicting NUE to monotonically decrease with increasing nitrogen application rate. This penalization was required only for the frontier splines for Odisha and the Terai of Nepal (Supplementary Information 10).

These smoothing splines were fitted only to rice fields with nitrogen application rates between 50 and 200 kg ha^{-1}. Fields with nitrogen application rates below 50 kg ha^{-1} were excluded because NUE at these low nitrogen application rates is distorted by uncaptured variation in indigenous soil nitrogen supply[60]. Fields with nitrogen application rates above 200 kg ha^{-1} were also excluded because of insufficient sample sizes at these high nitrogen application rates. Fields with nitrogen application rates below 100 kg ha^{-1} in Punjab and Haryana (**n** = 118) and above 150 kg ha^{-1} in Odisha (n = 14) were also excluded for their respective smoothing splines, given the low sample size at these nitrogen rates in these regions. See Supplementary Information 3 for the proportions of samples excluded through these added exclusion criteria.

Potential to improve NUE of rice crops in South Asia

An opportunity assessment in terms of food security, nitrogen-saving potential and input subsidy costs was conducted for different pathways of NUE improvement for each rice production region. We assessed both pathways to increase NUE: reduce nitrogen application without reducing yield (N-saving pathway) and increase yield without increasing nitrogen application (yield-gain pathway). We also assessed the scope to pursue both pathways simultaneously.

Potential NUE gains from each pathway were quantified by characterizing the average relationship between nitrogen application rate and rice yield in each region. To do so, the effect of nitrogen application rate on rice yield had to be isolated from the effects of other variables, and their interaction, on rice yield. This was achieved by fitting region-specific Random Forest models for rice yield with nitrogen application rate and all variables listed in Fig. 5 used as predictors (see Supplementary Information 6 for variable definitions). Such decision tree-based models were necessary because of their capacity to handle categorical predictor variables, nonlinear relationships between rice yield and its predictors, and interaction effects between predictors. Random Forest models, which fit multiple decision trees to bootstrapped data, were necessary because of their resilience to outliers, overfitting and low sample sizes. This was particularly important for the survey data from Odisha and Andhra Pradesh, where the sample size was smaller (Supplementary Information 9).

These region-specific Random Forest models were fitted with the 'ranger' function from the R package 'ranger'[61]. The hyperparameters for these Random Forest models were tuned using the 'train' function from the R package 'caret'[62], which identified the 'mtry' (number of variables randomly sampled as candidates at each split), as well as the minimum size of terminal nodes, that minimized the root mean squared error and maximized the R^2 of each model. Predictors were excluded from Random Forest models when data for the predictor were unavailable for more than 85% of surveyed fields, had low inference space (where more than 90% of data for the predictor shared the same value) or correlated with other predictors (correlation coefficient above 0.6). When two predictors were correlated, the less important predictor was identified using the 'Boruta' variable selection function in the 'Boruta' package[63] and then excluded. Rice fields were also excluded from the Random Forest models if the surveyed rice field had missing data for any of the included predictors. Since the resultant Random Forest models were regression models, their predictive accuracy was evaluated on the basis of out-of-bag R^2.

Shapley values were then calculated for the region-specific Random Forest models to isolate the relationship between nitrogen application rate and yield for each rice field. Shapley values are derived from cooperative game theory and used to describe how individual predictor variables contribute to model predictions, including through interactions with other predictor variables[32]. The magnitude of a Shapley value reflects how much the value of a predictor variable for a given observation influenced the model's prediction for that observation relative to the model's average prediction. In turn, the sign of this Shapley value reflects whether this influence was positive or negative, and the units of this Shapley value reflect the units of the outcome variable that the model predicts. In this study, we used the 'explain' function in the 'fastshap' package[64] to calculate the Shapley values of nitrogen application rate as a predictor in the region-specific Random Forest models predicting rice yield (in kg ha^{-1}). These Shapley values reflect how the rate of nitrogen applied in each surveyed field influenced its yield prediction at the margin. For example, if the Shapley value for nitrogen application rate was 100 for a given field, this reflects that the nitrogen application rate in this field increased its predicted yield by 100 kg ha^{-1}. The data were then pooled for each region, and the relationship between nitrogen application rate and corresponding Shapley values was characterized using a piece-wise linear model with the 'segmented' function from the R package 'segmented'[65]. A piece-wise linear model with a single break point was used because it reflects diminishing marginal returns to nitrogen fertilizer[66] and because it fitted the data well according to R^2 (Supplementary Information 4).

N-saving pathway

The region-specific piece-wise linear models were used to estimate how much nitrogen application rates could have been reduced for each region without reducing rice yield. Rice yields increased with increasing nitrogen application rate until the break point for each region's piece-wise linear model, after rice yield response to nitrogen was negligible (Supplementary Information 4). We defined this break point as the average maximum productive nitrogen application rate for each region. Potential nitrogen savings were then calculated by capping nitrogen application rates at the region-specific maximum. It is important to note that the maximum productive nitrogen application rate was estimated for the average rice field in each region. This means the results can be best used to estimate N-saving opportunities at the regional level, but not necessarily at the individual field level.

Yield-gain pathway

We used the same region-specific piece-wise linear models to estimate how much farmers could have increased rice yield without increasing nitrogen application rate. The piece-wise linear models delineated the average relationship between yield and nitrogen application rate in each region. We used these models to calculate the potential production gain from all rice fields achieving at least the average yield for their respective nitrogen application rate and region. If a field received more than the maximum productive nitrogen application rate for its given region, we calculated the potential yield gain from that rice field achieving at least the average yield at this threshold. We assumed fields with a nitrogen application rate below 50 kg ha^{-1} could not have increased yield given the risk of nutrient mining at these small nitrogen application rates[8]. We also tested the sensitivity of the analysis to this assumption (Supplementary Information 11).

Both pathways

We also calculated the potential NUE gains associated with pursuing both the N-saving and yield-gain pathways simultaneously. Specifically, we calculated the potential nitrogen saving and yield gain for rice fields that received more than average productive nitrogen application rate for their respective region while also achieving below-average yields for this nitrogen application rate. See Supplementary Information 12 for a visual depiction of the fields in each region we found could pursue the N-saving pathway only, the yield-gain pathway only, both pathways or neither pathway. We also estimated the implications of potential NUE improvements for rice production, on-farm profitability, nitrogen pollution and fertilizer subsidy savings using assumptions detailed in Supplementary Information 5.

Agronomic mechanisms for realizing potential NUE

We relied on slightly different Random Forest models to identify drivers of NUE in each region. NUE shares a strong association with nitrogen application rate given that nitrogen application rate partially defines NUE. To account for this association, we identified the average NUE for any given nitrogen application rate and region using the smoothing splines explained earlier and depicted in Fig. 3. We then fitted classification Random Forest models predicting whether each rice field achieved above- or below-average NUE for its given nitrogen application rate and region (Fig. 3). These NUE-predicting Random Forest models were fitted using the same predictors (excluding nitrogen application rate), tuning method and predictor exclusion criteria as for the yield-predicting Random Forest models described earlier. These NUE-predicting Random Forest models were also fitted using the same data exclusion criteria as for the splines and yield-predicting Random Forest models described earlier. Since these NUE-predicting Random Forests were binary classification models, they were evaluated on the basis of the out-of-bag prediction accuracy (%).

We then used Shapley values to identify the most important NUE predictors in the fitted Random Forest models (Fig. 5). These Shapley values estimate the marginal contribution of each predictor to the Random Forest's prediction of whether each field achieved above- or below-average NUE for its given nitrogen application rate and region. We used the 'explain' function from the R package 'fastshap' to estimate predictors' Shapley values from the fitted Random Forest models using the method proposed by ref. 64. We then applied the 'autoplot' function from the package 'ggplot2' to the Shapley values to observe how the most important NUE predictors influenced NUE (Fig. 6 and Supplementary Information 8).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The dataset presented in this paper, and used for all analyses, is available from the CIMMYT Dataverse at https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10549105.

Code availability

All analyses were performed in R (version 4.3.1). The code developed for the analysis is available at https://github.com/RiceNUE/Data-driven-strategies-to-improve-nitrogen-use-efficiency-of-rice-farming-in-South-Asia.

References

1. *FAOSTAT* (FAO, accessed 28 November 2023); http://www.fao.org/faostat/en/#data/EF
2. Pingali, P. L. Green revolution: impacts, limits, and the path ahead. *Proc. Natl Acad. Sci. USA* **109**, 12302–12308 (2012).
3. Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z. & Winiwarter, W. How a century of ammonia synthesis changed the world. *Nat. Geosci.* **1**, 636–639 (2008).
4. Erisman, J. W. et al. Consequences of human modification of the global nitrogen cycle. *Phil. Trans. R. Soc. B* **368**, 20130116 (2013).
5. Gao, Y. & Cabrera Serrenho, A. Greenhouse gas emissions from nitrogen fertilizers could be reduced by up to one-fifth of current levels by 2050 with combined interventions. *Nat. Food* **4**, 170–178 (2023).
6. Maaz, T. M. et al. Meta-analysis of yield and nitrous oxide outcomes for nitrogen management in agriculture. *Glob. Change Biol.* **27**, 2343–2360 (2021).
7. McLellan, E. L. et al. The nitrogen balancing act: tracking the environmental performance of food production. *BioScience* **68**, 194–203 (2018).
8. EU Nitrogen Expert Panel *Nitrogen Use Efficiency (NUE): An Indicator for the Utilization of Nitrogen in Agriculture and Food Systems* (Wageningen University, 2015).
9. Cassman, K. G. & Dobermann, A. Nitrogen and the future of agriculture: 20 years on *Ambio* **51**, 17–24 (2022).
10. Menegat, S., Ledo, A. & Tirado, R. Greenhouse gas emissions from global production and use of nitrogen synthetic fertilisers in agriculture. *Sci. Rep.* **12**, 14490 (2022).
11. Chang, J. et al. Reconciling regional nitrogen boundaries with global food security. *Nat. Food* **2**, 700–711 (2021).
12. Shahzad, A. N., Qureshi, M. K., Wakeel, A. & Misselbrook, T. Crop production in Pakistan and low nitrogen use efficiencies. *Nat. Sustain.* **2**, 1106–1114 (2019).
13. Zhang, X. et al. Managing nitrogen for sustainable development. *Nature* **528**, 51–59 (2015).
14. Kishore, A., Alvi, M. & Krupnik, T. J. Development of balanced nutrient management innovations in South Asia: perspectives from Bangladesh, India, Nepal, and Sri Lanka. *Glob. Food Sec.* **28**, 100464 (2021).
15. *Expenditure Budget 2023–2024* (Government of India, 2023); https://www.indiabudget.gov.in/budget2023-24/doc/eb/vol1.pdf
16. *Fertilizer Use by Crop and Country for the 2017–2018 Period* (IFA, accessed 29 September 2023); https://www.ifastat.org/consumption/fertilizeruse-by-crop
17. *OECD–FAO Agricultural Outlook 2023–2032* (OECD, 2023); https://doi.org/10.1787/08801ab7-en
18. Congreves, K. A. et al. Nitrogen use efficiency definitions of today and tomorrow. *Front. Plant Sci.* **12**, 637108 (2021).
19. Cassman, K. G., Gines, G. C., Dizon, M. A., Samson, M. I. & Alcantara, J. M. Nitrogen-use efficiency in tropical lowland rice systems: contributions from indigenous and applied nitrogen. *Field Crops Res.* **47**, 1–12 (1996).
20. Gu, B. et al. Cost-effective mitigation of nitrogen pollution from global croplands. *Nature* **613**, 77–84 (2023).
21. Cui, Z. et al. Pursuing sustainable productivity with millions of smallholder farmers. *Nature* **555**, 363–366 (2018).
22. Sharma, S., Singh, P. & Choudhary, O. P. Nitrogen and rice straw incorporation impact nitrogen use efficiency, soil nitrogen pools and enzyme activity in rice–wheat system in north-western India. *Field Crops Res.* **266**, 108131 (2021).
23. Shankar, T. et al. Productivity and nutrient balance of an intensive rice–rice cropping system are influenced by different nutrient management in the red and lateritic belt of West Bengal, India. *Plants* **10**, 1622 (2021).
24. Thind, H. S. et al. Optimal rate and schedule of nitrogen fertilizer application for enhanced yield and nitrogen use efficiency in dry-seeded rice in north-western India. *Arch. Agron. Soil Sci.* **64**, 196–207 (2018).
25. Lacoste, M. et al. On-farm experimentation to transform global agriculture. *Nat. Food* **3**, 11–18 (2022).
26. Krupnik, T. J. et al. Does size matter? A critical review of meta-analysis in agronomy. *Exp. Agric.* **55**, 200–229 (2019).
27. Whitfield, S., Challinor, A. J. & Rees, R. M. Frontiers in climate smart food systems: outlining the research space. *Front. Sustain. Food Syst.* **2**, 2 (2018).
28. Silva, J. V. et al. Revisiting yield gaps and the scope for sustainable intensification for irrigated lowland rice in Southeast Asia. *Agric. Syst.* **198**, 103383 (2022).
29. Lokers, R., Knapen, R., Janssen, S., van Randen, Y. & Jansen, J. Analysis of Big Data technologies for use in agro-environmental science. *Environ. Model. Softw.* **84**, 494–504 (2016).
30. McDonald, A. J. et al. Bigger data from landscape-scale crop assessment surveys empowers sustainability transitions. Preprint at *SSRN* https://doi.org/10.2139/ssrn.4511866 (2023).
31. Qian, H. et al. Greenhouse gas emissions and mitigation in rice agriculture *Nat. Rev. Earth Environ.* **4**, 716–732 (2023).
32. Rozemberczki, B. et al. The Shapley value in machine learning. Preprint at https://arxiv.org/abs/2202.05594 (2022).
33. Silva, J. V. et al. Agronomic analysis of nitrogen performance indicators in intensive arable cropping systems: an appraisal of big data from commercial farms. *Field Crops Res.* **269**, 108176 (2021).
34. Ladha, J. K. et al. Achieving the sustainable development goals in agriculture: the crucial role of nitrogen in cereal-based systems. *Adv. Agron.* **163**, 39–116 (2020).
35. Quemada, M. et al. Exploring nitrogen indicators of farm performance among farm types across several European case studies. *Agric. Syst.* **177**, 102689 (2020).
36. IPCC *Climate Change 2022: Mitigation of Climate Change* (eds Shukla, P. R. et al.) (Cambridge Univ. Press, 2022).
37. Hu, B., Wang, W., Chen, J., Liu, Y. & Chu, C. Genetic improvement toward nitrogen-use efficiency in rice: lessons and perspectives. *Mol. Plant* **16**, 64–74 (2023).
38. Thapa, R., Chatterjee, A., Awale, R., McGranahan, D. A. & Daigh, A. Effect of enhanced efficiency fertilizers on nitrous oxide emissions and crop yields: a meta-analysis. *Soil Sci. Soc. Am. J.* **80**, 1121–1134 (2016).
39. Snapp, S. et al. Spatially differentiated nitrogen supply is key in a global food–fertilizer price crisis. *Nat. Sustain.* **6**, 1268–1278 (2023).
40. Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J. & Garnier, J. 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *Environ. Res. Lett.* **9**, 105011 (2014).
41. *Draft Green Credit Programme Implementation Rules, 2023* (Government of India, 2023); http://www.indiaenvironmentportal.org.in/content/475298/draft-green-credit-programme-implementation-rules-2023/

42. van Groenigen, J. W., Velthof, G. L., Oenema, O., Van Groenigen, K. J. & Van Kessel, C. Towards an agronomic assessment of N_2O emissions: a case study for arable crops. *Eur. J. Soil Sci.* **61**, 903–913 (2010).
43. Bajželj, B. et al. Importance of food-demand management for climate mitigation. *Nat. Clim. Change* **4**, 924–929 (2014).
44. Kyle, P., Müller, C., Calvin, K. & Thomson, A. Meeting the radiative forcing targets of the representative concentration pathways in a world with agricultural climate impacts. *Earths Future* **2**, 83–98 (2014).
45. Calvin, K. et al. Implications of simultaneously mitigating and adapting to climate change: initial experiments using GCAM. *Climatic Change* **117**, 545–560 (2013).
46. Devkota, K. P., Devkota, M., Paudel, G. P. & McDonald, A. J. Coupling landscape-scale diagnostics surveys, on-farm experiments, and simulation to identify entry points for sustainably closing rice yield gaps in Nepal. *Agric. Syst.* **192**, 103182 (2021).
47. Bhattarai, N. et al. The impact of groundwater depletion on agricultural production in India. *Environ. Res. Lett.* **16**, 085003 (2021).
48. Silva, J. V. et al. Big data, small explanatory and predictive power: lessons from random forest modeling of on-farm yield variability and implications for data-driven agronomy. *Field Crops Res.* **302**, 109063 (2023).
49. Urfels, A. et al. Farm size limits agriculture's poverty reduction potential in Eastern India even with irrigation-led intensification. *Agric. Syst.* **207**, 103618 (2023).
50. McDonald, A. J. et al. Time management governs climate resilience and productivity in the coupled rice–wheat cropping systems of eastern India. *Nat. Food* **3**, 542–551 (2022).
51. Urfels, A. et al. Social-ecological analysis of timely rice planting in Eastern India. *Agron. Sustain. Dev.* **41**, 14 (2021).
52. Nayak, H. S. et al. Rice yield gaps and nitrogen-use efficiency in the northwestern Indo-Gangetic Plains of India: evidence based insights from heterogeneous farmers' practices. *Field Crops Res.* **275**, 108328 (2022).
53. *Production and Technological Gaps in Middle Indo-Gangetic Plains* (ICAR, 2016); https://www.researchgate.net/publication/308918630_Production_and_Technological_Gaps_in_Middle_Indo-Gangetic_Plains
54. *Yearbook of Agricultural Statistics of Bangladesh 2022* (Bangladesh Bureau of Statistics, 2023); http://www.bbs.gov.bd/site/page/3e838eb6-30a2-4709-be85-40484b0c16c6/Yearbook-of-Agricultural-Statistics
55. Ajay, A. et al. Large survey dataset of rice production practices applied by farmers on their largest farm plot during 2018 in India. *Data Brief.* **45**, 108625 (2022).
56. Funk, C. et al. The Climate Hazards Infrared Precipitation with Stations—a new environmental record for monitoring extremes. *Sci. Data* **2**, 150066 (2015).
57. Poggio, L. et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* **7**, 217–240 (2021).
58. Ladha, J. K. et al. Global nitrogen budgets in cereals: a 50-year assessment for maize, rice and wheat production systems. *Sci. Rep.* **6**, 19355 (2016).
59. Perperoglou, A., Sauerbrei, W., Abrahamowicz, M. & Schmid, M. A review of spline function procedures in R. *BMC Med. Res. Method.* **19**, 46 (2019).
60. Dobermann, A. R. *Nitrogen Use Efficiency—State of the Art* (Agronomy–Faculty Publications, 2005).
61. Wright, M. N. & Ziegler, A. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Softw.* **77**, 1–17 (2017).
62. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
63. Kursa, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).
64. Štrumbelj, E. & Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* **41**, 647–665 (2014).
65. Muggeo, V. M. Segmented: an R package to fit regression models with broken-line relationships. *R News* **8**, 20–25 (2008).
66. Ferreira, I. E., Zocchi, S. S. & Baron, D. Reconciling the Mitscherlich's law of diminishing returns with Liebig's law of the minimum. Some results on crop modeling. *Math. Biosci.* **293**, 29–37 (2017).
67. Hijmans, R. J., Barbosa, M., Ghosh, A. & Mandel, A. Package 'geodata'. *CRAN* https://doi.org/10.32614/cran.package.geodata (2024).

Acknowledgements

Analytical and written efforts from S.C., A.J.M., J.V.S., A.U., H.S.N., S.R.S., M.L.J., T.K., V.K., R.K.M., T.B.S. and P.C. were funded by Excellence in Agronomy (EIA; https://www.cgiar.org/initiative/excellence-in-agronomy/), the CGIAR Regional Integrated Initiative Transforming Agrifood Systems in South Asia (TAFSSA; https://www.cgiar.org/initiative/20-transforming-agrifood-systems-insouth-asia-tafssa/), the Australian Government Research Training Program Scholarship (S.C.), the Westpac Future Leaders Scholarship (S.C.) and CGIAR Fund Donors (https://www.cgiar.org/funders/). Field research was conducted under the USAID (USAID grant number BFS-G-11-00002), and Bill and Melinda Gates Foundation (BMGF grant numbers OPP1052535 and OPP1133205) supported Cereal Systems Initiative for South Asia (CSISA; http://csisa.org/) and CGIAR Research Program on Climate Change, Agriculture, and Food Security (https://www.cgiar.org/research/program-platform/climatechange-agriculture-and-food-security/). ChatGPT was used to refine the wording of <15 sentences included in the text. It was also used to simplify ~5% of the R code used for data analyses. The authors critically reviewed all suggestions provided by ChatGPT. This study is built on the generous participation in surveys from thousands of smallholder farmers. It also reflects the diligent efforts of field scientists from the Indian Council of Agricultural Research (ICAR) in India, Krishi Vigyan Kendra (KVK) system in India, the Department of Agriculture in Nepal and the Department of Agricultural Extension in Bangladesh. This work was also made possible by the careful data curation efforts of S. Karki (CIMMYT), M.K. Hossain (CIMMYT), P. Peramaiyan (IRRI), A. Ajay (CIMMYT), G.P. Paudel (CIMMYT), and A. Samaddar (IRRI), as well as considered advice from H. Rajan (CIMMYT). Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of any of the acknowledged organizations.

Author contributions

S.C. and A.J.M. conceived and designed the overall project. A.J.M., A.U., H.S.N., M.L.J., H.S.J., T.K., R.K.M., T.B.S, A.K.N. and P.C. played key roles in acquiring the data. S.C. harmonized and analysed the acquired data and interpreted the analysis, all with supervision from A.J.M., J.V.S., A.U., H.S.N., S.R.S., R.K.M., V.K. and P.C. S.C. drafted the paper. All authors edited and commented on the paper, in particular A.J.M., J.V.S. and A.U.

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/s41893-024-01496-3.

Correspondence and requests for materials should be addressed to Sam Coggins.

https://doi.org/10.1038/s41893-024-01496-3

Peer review information *Nature Sustainability* thanks Xu Zhao, Tai McCellan Maaz and Ana Meijide for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Sam Coggins
Last updated by author(s): Nov 16, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NA
Data analysis	All analyses were performed in R (version 4.3.1). The code developed for for the analysis is available at: https://github.com/RiceNUE/Data-driven-strategies-to-improve-nitrogen-use-efficiency-of-rice-farming-in-South-Asia

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The dataset presented in this manuscript, and used for all analyses, is available from the CIMMYT Dataverse at: https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10549105

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Both male and female farmers participated in the analyzed surveys. The gender of surveyed farmers was collected in some but not all surveys. In regions with sufficient data availability, we analyzed how strongly the gender of farmers predicted the nitrogen use efficiency of their surveyed rice crops (Figure 5).
Population characteristics	Surveyed farmers represented a wide range of social groups. After applying the data exclusion criteria explained in the 'Methods', the surveyed farmers had the following demographic characteristics: FORMAL EDUCATION: the median surveyed farmer had completed primary education only. However, formal education levels varied between no formal education, up to PhD completion. SOCIAL CATEGORY: 10% General, 21% Other Backward Classes, 2% Scheduled Tribes, 5% Scheduled Castes, 61% no data or not applicable. HOUSEHOLD MEMBERS: surveyed farmers had a median numbers of household members of 6. GENDER: 46% no data, 50% men, 4% women. In regions with sufficient data availability, we tested how strongly the above social factors predicted nitrogen use efficiency of rice farmers (Figure 5). Despite the best efforts of survey designers and implementors, few other population characteristics were identified in the surveys.
Recruitment	Surveyed rice fields were selected through a two-stage approach. In the first stage, rice-growing districts within each region were purposively selected based on the districts' importance to food security in their respective states and countries. This included almost all rice-growing districts in Odisha, Bihar and neighboring districts of eastern Uttar Pradesh. In the second stage, representative rice fields were sampled within each selected district. Villages in Andhra Pradesh, Bihar, eastern Uttar Pradesh, and Odisha were selected using a 'probability proportional to population' method [55]. The process for this second stage varied slightly in Punjab & Haryana, where villages were selected based on government-mandated travel restrictions imposed during the COVID-19 pandemic. For all surveys in Nepal, and surveys in Bangladesh up to and including the year 2018, representative rice fields were identified using satellite imagery. Specifically, LANDSAT-derived (NDVI Normalized Difference Vegetative Index) values were extracted to capture the variability in standing green biomass (a proxy for yield) in selected districts, in order to identify rice fields to participate in the survey. These normally distributed NDVI values were then stratified into four quartiles. Rice fields to be surveyed were then randomly selected from these quartiles so that selected samples proportionally represented the bell curve. Unlike in other regions, farmers that participated in the 2019 survey in Nepal were also invited to participate in the 2020 survey. The process for selecting individual rice fields varied for surveys implemented in Bangladesh after the year 2018. In these surveys, survey enumerators selected nine representative rice fields for each sub-district survey location, three with the best quality crop, three with medium quality, and three with the worst quality. In all surveyed regions, additional selections were made if farmers of selected rice fields could not be located or declined to participate. Despite the best efforts of survey designers and implementors to mitigate recruitment biases, two types of recruitment biases influenced results. First, biases introduced by the subtle variations in recruitment protocols across regions (described above). Second, biases inherent in each protocol. For example, men generally participated in the surveys more than women due to cultural norms, including gendered norms regarding who is socially categorized as a 'farmer'.
Ethics oversight	All surveys complied with standards established by the Research Ethics Committee of the International Maize and Wheat Improvement Center (CIMMYT), as described in policy number DDG-POL-04–2019. All survey participants gave informed consent to participate.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	An observational analysis of nitrogen use efficiency of rice farms across South Asia.
Research sample	All analyses relied on one database combining farmer field surveys and secondary environmental data for 28,865 farmer fields that cultivated transplanted rice in a monsoon season between 2016 and 2020. The database covers six regions across South Asia (Figure 1): Andhra Pradesh (n = 1,469), Bihar & eastern Uttar Pradesh (n = 9,338), Odisha (n = 1,211), Punjab & Haryana (n = 5,723), Bangladesh's floodplains (n = 5,953), and the Terai of Nepal (n = 4,944).

RATIONALE: Controlled experiments provide valuable insights into how agronomic factors influence the NUE of rice crops. Yet, controlled-condition experiments typically do not account for the diverse biophysical and socioeconomic conditions under which farmers operate, rendering these experiments of limited value for 'real world' decision making. Conversely, observational studies generally rely on small-n surveys that miss the significant variation within and between farming regions. We therefore relied on large-n surveys conducted over broad geographic regions for data-driven approaches to NUE assessment and improvement.

Sampling strategy

See above response to the 'Recruitment' field.

Data collection

Farmers reported the data for most variables via primary surveys. The surveys elicited information for each farmer's largest rice plot in the most recent monsoon season. These primary farmer field surveys acquired data for rice yield, nitrogen application rate, plot size, and all variables in Figure 5 (aside from the rainfall variables and most soil and landscape variables). The GPS-recorded latitude and longitude of each surveyed field were also recorded at the time of the survey (Figure 1). The surveys were implemented independently across all six regions but with metadata that enabled harmonization. All rice yield data was farmer-reported aside from approximately 40% of yield records from Bangladesh's floodplains and approximately 25% of yield records from Punjab & Haryana, where yields were estimated with crop cuts. Crop cut measurements were collected by government sub-assistant agricultural officers in Bangladesh, and professional field technicians in Punjab & Haryana. All surveys complied with standards established by the Research Ethics Committee of the International Maize and Wheat Improvement Center (CIMMYT), as described in policy number DDG-POL-04–2019. All survey participants gave informed consent to participate.

To complement the primary survey data, we accessed publicly available secondary data. Daily rainfall data were retrieved at 180-second resolution from Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS; [56]). This data was then used to develop the rainfall variables featured in Figure 5 and defined in SI#6. Soil sand content, pH in water, soil organic carbon content, and bulk density were retrieved at 30-second spatial resolution from [57]. It should be noted that the spatial resolution of the soil and rainfall data was typically larger than the size of surveyed rice fields, and that the surveys were not designed to evaluate the accuracy of the soil and rainfall data at the field-level. We therefore relied on the large sample size in our dataset to account for this spatial imprecision in our analyses, See SI#6 for variable definitions and data sources.

Timing and spatial scale

TIMING: The surveys elicited information for each farmer's largest rice plot in the most recent monsoon season between 2016 and 2020. The robustness of results across seasons was tested in Figure S6 in Appendix 7.
SPATIAL SCALE: See above response to the 'Research sample' field. See Figure 1 for a map of this spatial scale.

Data exclusions

Our data quality control criteria excluded the following fields from the analysis: rice grown outside of the monsoon season (n = 10,147), direct-seeded rice (n = 1,329), nitrogen application rate equal to 0 kg ha-1 (n = 2,480; likely misreported information), nitrogen application rate above 400 kg ha-1 (n = 298; likely misreported information), sample coordinates located outside the surveyed region (n = 360), rice yield below 0.5 t ha-1 (n = 577; likely an error), missing rice yield data (n = 395), rice yield above 10 t ha-1 (n = 541; likely misreported information), nursery duration below 0 days (n = 262; definitely an error), nursery duration above 100 days (n = 273; misreported information), field duration below 60 days (n = 457; likely an error), field duration above 300 days (n = 88; likely an error), crop duration below 90 days (n = 358; likely an error), and crop duration above 340 days (n = 95; likely an error). This exclusion criteria left 28,865 data points for analysis.

Reproducibility

All analyses were performed in R (version 4.3.1). The code developed for for the analysis is available at: https://github.com/RiceNUE/Data-driven-strategies-to-improve-nitrogen-use-efficiency-of-rice-farming-in-South-Asia

Randomization

NA - this was an observational analysis, not an experimental analysis.

Blinding

NA - this was an observational analysis, not an experimental analysis.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions

Thousands of surveys were completed across multiple years in a variety of conditions. All surveys were implemented in rice-growing regions of India, Bangladesh and Nepal (Figure 1). These regions typically feature high temperatures, high humidity and monsoon seasons with high rainfall.

Location

See above response to the 'Research sample' field. See Figure 1 for a map of these locations.

Access & import/export

All surveys were implemented in formal partnerships with relevant national government organizations (see Acknowledgments for detail). Material samples were not exported.

Disturbance

All survey participants gave informed consent to participate. No material samples were taken, aside from a minority of surveys that involved crop cut measurements. In these cases, crop cut samples and measurements were taken in situ by trained field technicians, from minimal subsets of farmers' fields, and only with consent from participating farmers.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging