**Supplementary Information for:**

# Importance of spatial and depth-dependent drivers in groundwater level modeling through machine learning

Pragnaditya Malakar[1], Abhijit Mukherjee[1,2,3], Soumendra N. Bhanja[4], Dipankar Saha[5], Ranjan Kumar Ray[6], Sudeshna Sarkar[7], Anwar Zahid[8]

[1]Department of Geology and Geophysics, Indian Institute of Technology Kharagpur, West Bengal 721302, India
[2]School of Environmental Science and Engineering, Indian Institute of Technology Kharagpur, West Bengal 721302, India
[3]Applied Policy Advisory for Hydrogeoscience (APAH) Group, Indian Institute of Technology Kharagpur, West Bengal 721302, India
[4]Interdisciplinary Centre for Water Research, Indian Institute of Science, Bangalore, Karnataka 560054, India
[5]Formerly Central Ground Water Board, Ministry of Water Resources, River Development and Ganga Rejuvenation, Government of India, Faridabad, Haryana, India
[6]Central Ground Water Board (CGWB), Bhujal Bhawan, NH-IV, Faridabad, India
[7]Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, West Bengal 721302, India
[8]Bangladesh Water Development Board (BWDB), Dhaka, Bangladesh

**<u>Contents of the supplementary information:</u>**

**Tables S1 to S11**
**Figures S1 to S17**

**Table S1.** Long-term spatial mean annual precipitation between 1985 and 2015.

| Basin Name | Precipitation (mm/year) |
|---|---|
| Indus-Ganges-Brahmaputra-Meghna Basin (IGB-M) | ~1140 |
| Indus Basin (I) | ~590 |
| Ganges Basin (G) | ~1030 |
| Brahmaputra Basin (B) | ~2240 |
| Meghna Basin (M) | ~2560 |

**Table S2.** Statistics of groundwater abstraction per unit area (million $m^3/km^2$) in the year 2013.

| | Annual Groundwater abstraction (million $m^3$) | Area ($km^2$) | Groundwater abstraction per unit area (million $m^3/km^2$) |
|---|---|---|---|
| Entire IGBM basin | 194692.7 | 1151698.9 | 0.169 |
| Indus basin | 52012 | 94774 | 0.549 |
| Ganges basin | 117779.4 | 897072.7 | 0.131 |
| Brahmaputra basin | 5200 | 101017.2 | 0.052 |
| Meghna basin | 19701.2 | 58835 | 0.335 |

**Table S3.** Different data sources used in the study.

| Data type | Source | Spatial resolution | Temporal resolution |
|---|---|---|---|
| GWL data for India | Central Groundwater Board (CGWB), India | ~14000 wells | 1985 -2015 |
| GWL data for Bangladesh | Bangladesh Water Development Board (BWDB), Bangladesh | ~1300 wells | 1985 -2015 |
| Precipitation data for India | India Meteorological Department (IMD), India | 0.25˚×0.25˚ | 1985 -2015 |
| Precipitation data for Bangladesh | Climatic Research Unit (CRU TS v-4.01) | 0.5˚×0.5˚ | 1985 -2015 |
| Evapotranspiration data | Climatic Research Unit (CRU TS v-4.01) | 0.5˚×0.5˚ | 1985 -2015 |
| Temperature (max, min, mean) data for Bangladesh | Climatic Research Unit (CRU TS v-4.01) | 0.5˚×0.5˚ | 1985 -2015 |

| | | | |
|---|---|---|---|
| Temperature (max, min, mean) data for India | India Meteorological Department (IMD), India | $1^\circ \times 1^\circ$ | 1985 -2015 |
| Population data | Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2018 | ~ 5 km grids | 2000, 2005, 2010, 2015 |
| Groundwater withdrawals for India | Central Groundwater Board (CGWB), India; AQUASTAT, 2017; Minor irrigation, 2017 | District level | 2013 |
| Groundwater withdrawals for India Bangladesh | Bangladesh Water Development Board (BWDB), Bangladesh; AQUASTAT, 2017 | | 2013 |

**Table S4:** Summery of observation well number.

| | **Number of Shallow observation wells** | **Number of deep observation wells** | **Number of total observation wells** |
|---|---|---|---|
| Indus basin | 133 | 29 | 162 |
| Ganges basin | 1687 | 155 | 1842 |
| Brahmaputra basin | 164 | 3 | 167 |
| Meghna basin | 96 | 36 | 132 |
| **Entire IGBM basin** | **2080** | **223** | **2303** |

**Table S5**. Configurations for training and testing set, adapted to select the final split up of the dataset.

| Configurations | Total data | Training Set | Testing Set |
|---|---|---|---|
| Configuration - 1 | 31 years (124 seasons) | 5 Years (20 seasons) | 26 Years (104 seasons) |
| Configuration - 2 | 31 years (124 seasons) | 10 Years (40 seasons) | 21 Years (84 seasons) |
| Configuration - 3 | 31 years (124 seasons) | 15 Years (60 seasons) | 16 Years (64 seasons) |
| Configuration - 4 | 31 years (124 seasons) | 21 Years (84 seasons) | 10 Years (40 seasons) |
| Configuration - 5 | 31 years (124 seasons) | 26 Years (104 seasons) | 5 Years (20 seasons) |

**Table S6.** Basin-scale correlation coefficient (r) for the training stage

| | **Correlation coefficient (r)** | | | | | |
|---|---|---|---|---|---|---|
| | Type A | | Type B | | Type C | |
| | ANN | SVM | ANN | SVM | ANN | SVM |
| IGBM | 0.998 | 1 | 0.998 | 1 | 0.994 | 0.998 |
| IGBM SH | 0.997 | 1 | 0.997 | 1 | 0.994 | 0.997 |
| IGBM DP | 0.985 | 0.998 | 0.987 | 0.999 | 0.979 | 0.992 |
| I | 0.882 | 0.989 | 0.84 | 0.993 | 0.843 | 0.924 |
| I SH | 0.876 | 0.991 | 0.852 | 0.996 | 0.862 | 0.959 |
| I DP | 0.858 | 0.937 | 0.863 | 0.973 | 0.794 | 0.858 |
| G | 0.997 | 1 | 0.998 | 1 | 0.995 | 0.998 |
| G SH | 0.997 | 1 | 0.997 | 1 | 0.995 | 0.998 |
| G DP | 0.983 | 0.998 | 0.985 | 0.999 | 0.978 | 0.994 |
| B | 0.951 | 0.998 | 0.964 | 0.999 | 0.93 | 0.992 |
| B SH | 0.951 | 0.997 | 0.962 | 0.999 | 0.939 | 0.99 |
| B DP | 0.945 | 0.997 | 0.958 | 0.998 | 0.914 | 0.977 |
| M | 0.937 | 0.997 | 0.955 | 0.998 | 0.871 | 0.995 |
| M SH | 0.933 | 0.996 | 0.954 | 0.999 | 0.865 | 0.994 |
| M DP | 0.948 | 0.994 | 0.943 | 0.998 | 0.911 | 0.967 |

**Table S7**. Basin-scale correlation coefficient (r) for the testing stage

| | **Correlation coefficient (r)** | | | | | |
|---|---|---|---|---|---|---|
| | Type A | | Type B | | Type C | |
| | ANN | SVM | ANN | SVM | ANN | SVM |
| IGBM | 0.997 | 0.999 | 0.996 | 0.998 | 0.997 | 0.997 |
| IGBM SH | 0.997 | 0.998 | 0.996 | 0.998 | 0.998 | 0.997 |
| IGBM DP | 0.909 | 0.974 | 0.948 | 0.977 | 0.955 | 0.980 |
| I | 0.401 | 0.853 | 0.595 | 0.879 | 0.586 | 0.731 |
| I SH | 0.446 | 0.853 | 0.661 | 0.897 | 0.525 | 0.905 |
| I DP | -0.079 | 0.645 | 0.119 | 0.595 | 0.099 | -0.198 |
| G | 0.996 | 0.998 | 0.996 | 0.998 | 0.996 | 0.998 |
| G SH | 0.996 | 0.997 | 0.996 | 0.998 | 0.996 | 0.998 |
| G DP | 0.917 | 0.981 | 0.952 | 0.983 | 0.951 | 0.980 |
| B | 0.911 | 0.994 | 0.930 | 0.994 | 0.953 | 0.994 |
| B SH | 0.934 | 0.993 | 0.969 | 0.994 | 0.968 | 0.993 |
| B DP | 0.610 | 0.706 | 0.726 | 0.692 | 0.722 | 0.657 |
| M | 0.854 | 0.986 | 0.921 | 0.988 | 0.854 | 0.929 |
| M SH | 0.846 | 0.981 | 0.922 | 0.984 | 0.847 | 0.933 |
| M DP | 0.833 | 0.929 | 0.901 | 0.939 | 0.853 | 0.910 |

**Table S8.** Basin-scale Nash-Sutcliff efficiency (NSE) for the training stage

| | **Nash-Sutcliff efficiency (NSE)** | | | | | |
|---|---|---|---|---|---|---|
| | Type A | | Type B | | Type C | |
| | ANN | SVM | ANN | SVM | ANN | SVM |
| IGBM | 0.99 | 0.999 | 0.992 | 0.999 | 0.911 | 0.97 |
| IGBM SH | 0.99 | 0.999 | 0.992 | 0.999 | 0.895 | 0.966 |
| IGBM DP | 0.942 | 0.996 | 0.951 | 0.997 | 0.954 | 0.979 |
| I | 0.611 | 0.975 | 0.573 | 0.984 | 0.667 | 0.797 |
| I SH | 0.679 | 0.982 | 0.663 | 0.99 | 0.692 | 0.903 |
| I DP | 0.691 | 0.872 | 0.724 | 0.943 | -0.848 | 0.699 |
| G | 0.991 | 0.999 | 0.993 | 1 | 0.925 | 0.975 |
| G SH | 0.989 | 0.999 | 0.991 | 1 | 0.914 | 0.973 |
| G DP | 0.947 | 0.996 | 0.956 | 0.998 | 0.944 | 0.986 |
| B | 0.897 | 0.996 | 0.923 | 0.997 | 0.753 | 0.985 |
| B SH | 0.891 | 0.995 | 0.918 | 0.998 | 0.74 | 0.984 |
| B DP | 0.887 | 0.988 | 0.913 | 0.996 | 0.812 | 0.932 |
| M | 0.858 | 0.993 | 0.903 | 0.997 | 0.638 | 0.979 |
| M SH | 0.851 | 0.993 | 0.899 | 0.997 | 0.466 | 0.974 |
| M DP | 0.876 | 0.993 | 0.881 | 0.995 | 0.727 | 0.926 |

**Table S9.** Basin-scale Nash-Sutcliff efficiency (NSE) for the testing stage

| | **Nash-Sutcliff efficiency (NSE)** | | | | | |
|---|---|---|---|---|---|---|
| | Type A | | Type B | | Type C | |
| | ANN | SVM | ANN | SVM | ANN | SVM |
| IGBM | 0.980 | 0.964 | 0.985 | 0.983 | 0.906 | 0.936 |
| IGBM SH | 0.973 | 0.971 | 0.982 | 0.980 | 0.876 | 0.925 |
| IGBM DP | 0.738 | 0.942 | 0.837 | 0.949 | 0.805 | 0.902 |
| I | -0.625 | 0.547 | 0.292 | 0.718 | -0.045 | -1.280 |
| I SH | -1.421 | 0.195 | 0.307 | 0.267 | -0.204 | 0.172 |
| I DP | -0.533 | 0.186 | -0.523 | -0.096 | -2.542 | -0.860 |
| G | 0.978 | 0.964 | 0.984 | 0.980 | 0.913 | 0.957 |
| G SH | 0.975 | 0.973 | 0.982 | 0.980 | 0.903 | 0.957 |
| G DP | 0.829 | 0.960 | 0.868 | 0.965 | 0.792 | 0.917 |
| B | 0.824 | 0.976 | 0.940 | 0.986 | 0.719 | 0.986 |
| B SH | 0.841 | 0.983 | 0.924 | 0.987 | 0.706 | 0.986 |
| B DP | 0.360 | 0.427 | 0.515 | 0.440 | 0.441 | 0.410 |
| M | 0.700 | 0.961 | 0.858 | 0.962 | 0.504 | 0.948 |
| M SH | 0.686 | 0.957 | 0.843 | 0.964 | 0.286 | 0.940 |
| M DP | 0.393 | 0.714 | 0.763 | 0.753 | 0.471 | 0.602 |

**Table S10.** Normalized Root mean square error ($RMSE_n$) for the training stage

| | **Normalized Root mean square error ($RMSE_n$)** | | | | | |
|---|---|---|---|---|---|---|
| | Type A | | Type B | | Type C | |
| | ANN | SVM | ANN | SVM | ANN | SVM |
| IGBM | 0.099 | 0.026 | 0.087 | 0.025 | 0.296 | 0.172 |
| IGBM SH | 0.1 | 0.026 | 0.089 | 0.025 | 0.321 | 0.185 |
| IGBM DP | 0.238 | 0.059 | 0.221 | 0.052 | 0.212 | 0.144 |
| I | 0.62 | 0.156 | 0.65 | 0.125 | 0.573 | 0.448 |
| I SH | 0.564 | 0.132 | 0.577 | 0.098 | 0.552 | 0.31 |
| I DP | 0.552 | 0.356 | 0.523 | 0.237 | 1.351 | 0.545 |
| G | 0.093 | 0.028 | 0.083 | 0.022 | 0.272 | 0.156 |
| G SH | 0.104 | 0.027 | 0.093 | 0.022 | 0.291 | 0.165 |
| G DP | 0.229 | 0.063 | 0.208 | 0.049 | 0.235 | 0.117 |
| B | 0.319 | 0.066 | 0.275 | 0.051 | 0.494 | 0.123 |
| B SH | 0.329 | 0.071 | 0.285 | 0.048 | 0.431 | 0.127 |
| B DP | 0.334 | 0.08 | 0.293 | 0.061 | 0.507 | 0.259 |
| M | 0.375 | 0.081 | 0.31 | 0.056 | 0.598 | 0.143 |
| M SH | 0.384 | 0.083 | 0.316 | 0.057 | 0.519 | 0.161 |
| M DP | 0.35 | 0.109 | 0.343 | 0.07 | 0.726 | 0.27 |

**Table S11.** Normalized Root mean square error ($RMSE_n$) for the testing stage

| | **Normalized Root mean square error ($RMSE_n$)** | | | | | |
|---|---|---|---|---|---|---|
| | Type A | | Type B | | Type C | |
| | ANN | SVM | ANN | SVM | ANN | SVM |
| IGBM | 0.138 | 0.124 | 0.119 | 0.111 | 0.302 | 0.249 |
| IGBM SH | 0.163 | 0.137 | 0.133 | 0.120 | 0.347 | 0.270 |
| IGBM DP | 0.738 | 0.942 | 0.399 | 0.222 | 0.335 | 0.298 |
| I | 1.259 | 0.665 | 0.831 | 0.525 | 1.010 | 1.491 |
| I SH | 1.536 | 0.886 | 0.822 | 0.845 | 1.084 | 0.898 |
| I DP | 1.223 | 0.891 | 1.218 | 1.034 | 1.858 | 1.347 |
| G | 0.147 | 0.133 | 0.126 | 0.111 | 0.290 | 0.204 |
| G SH | 0.158 | 0.139 | 0.133 | 0.129 | 0.307 | 0.204 |
| G DP | 0.409 | 0.198 | 0.359 | 0.184 | 0.325 | 0.235 |
| B | 0.412 | 0.117 | 0.240 | 0.117 | 0.523 | 0.115 |
| B SH | 0.391 | 0.128 | 0.271 | 0.114 | 0.535 | 0.119 |
| B DP | 0.790 | 0.748 | 0.688 | 0.739 | 0.738 | 0.758 |
| M | 0.541 | 0.194 | 0.372 | 0.191 | 0.691 | 0.224 |
| M SH | 0.554 | 0.203 | 0.391 | 0.186 | 0.830 | 0.241 |
| M DP | 0.765 | 0.525 | 0.763 | 0.488 | 0.714 | 0.619 |

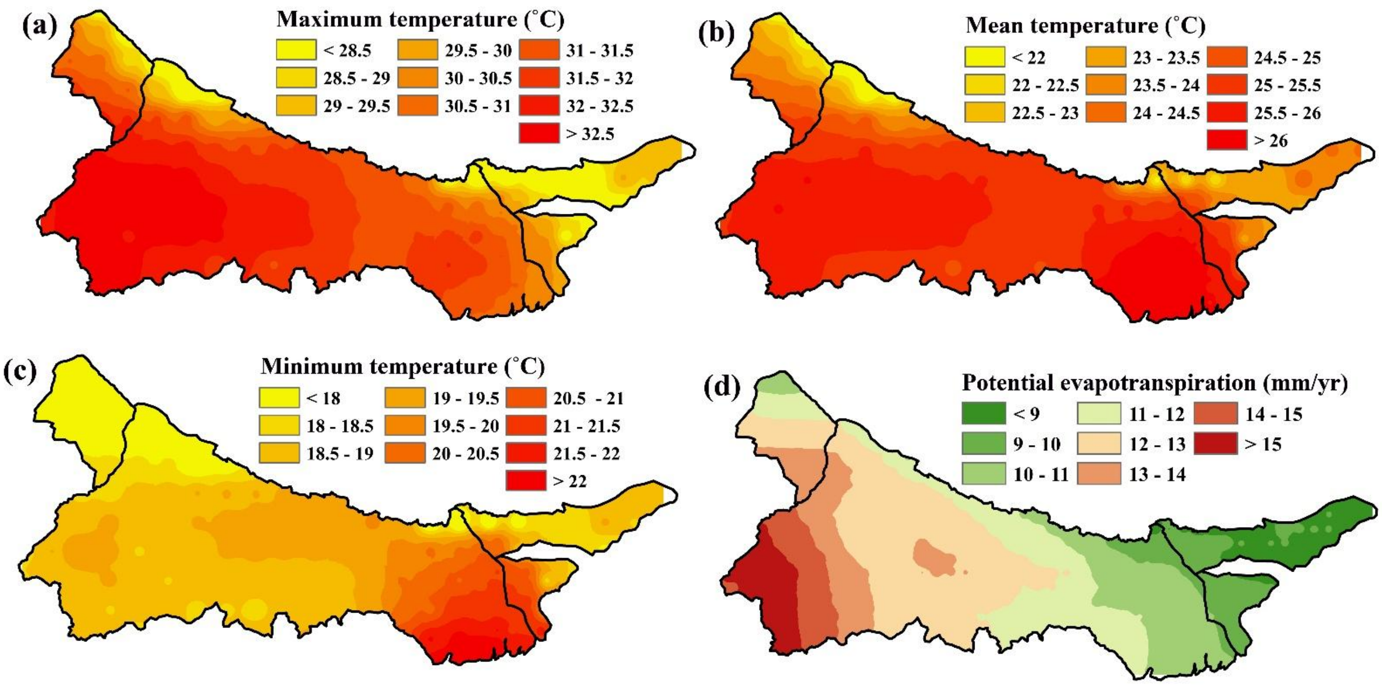



**Fig. S1.** Long-term average temperature and potential evapotranspiration map for the year 1985 to 2015.

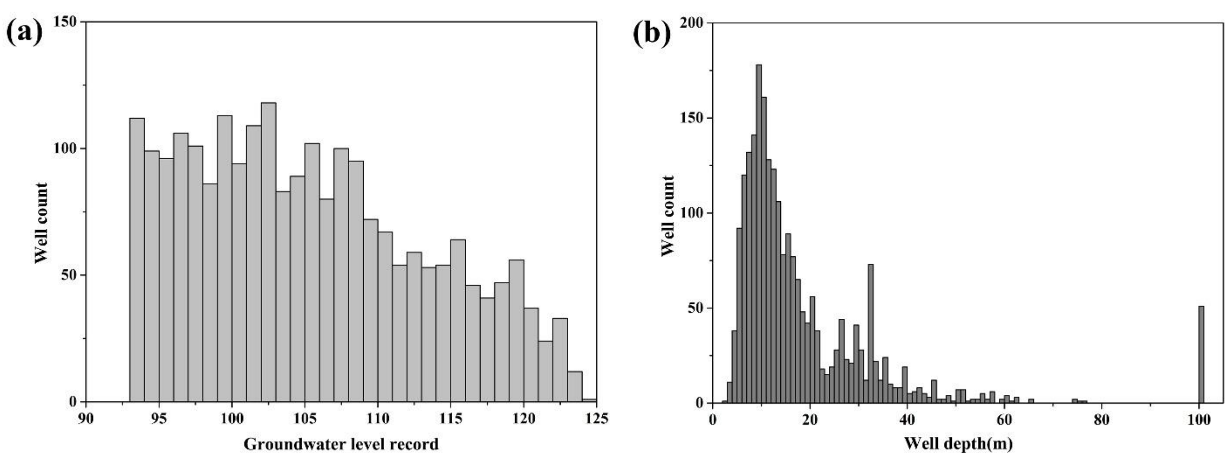
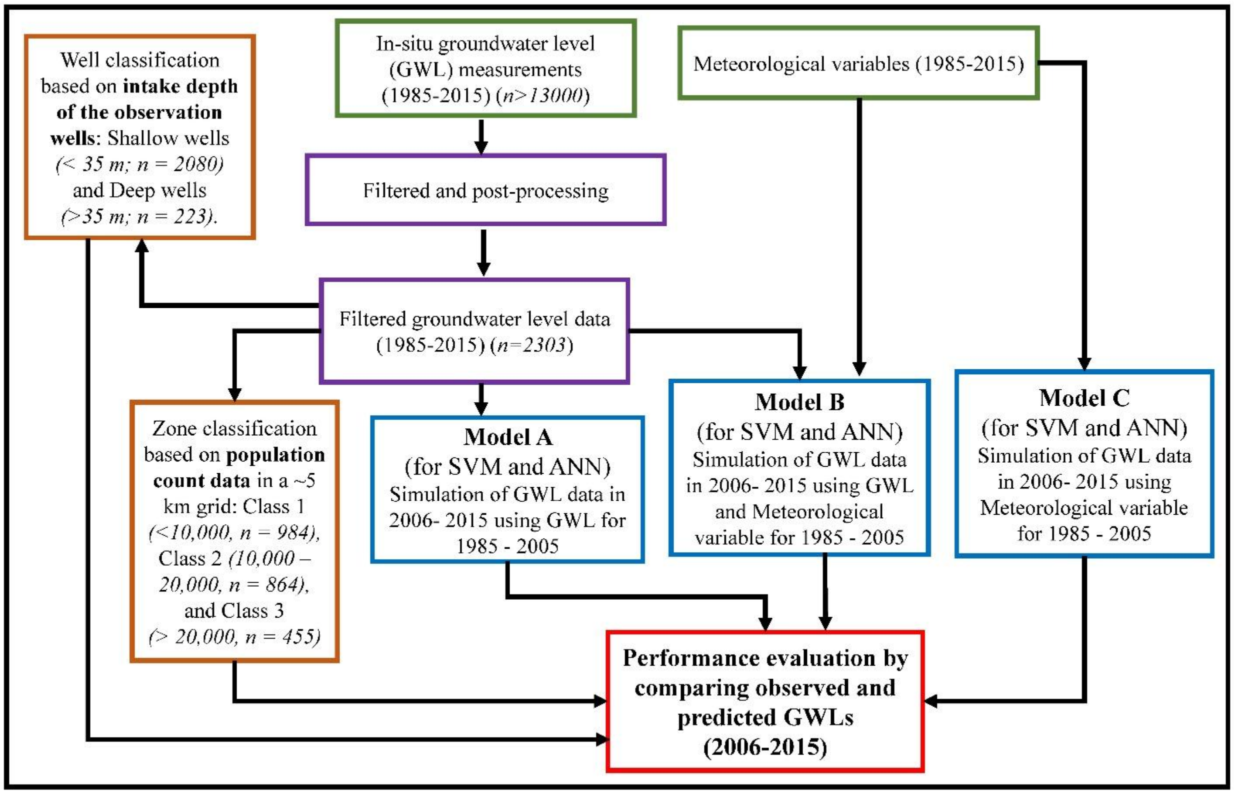



**Fig. S2.** Summary of observation well data included in the study.




95 **Fig. S3.** Flowchart showing methodology followed in this manuscript for machine learning modeling.

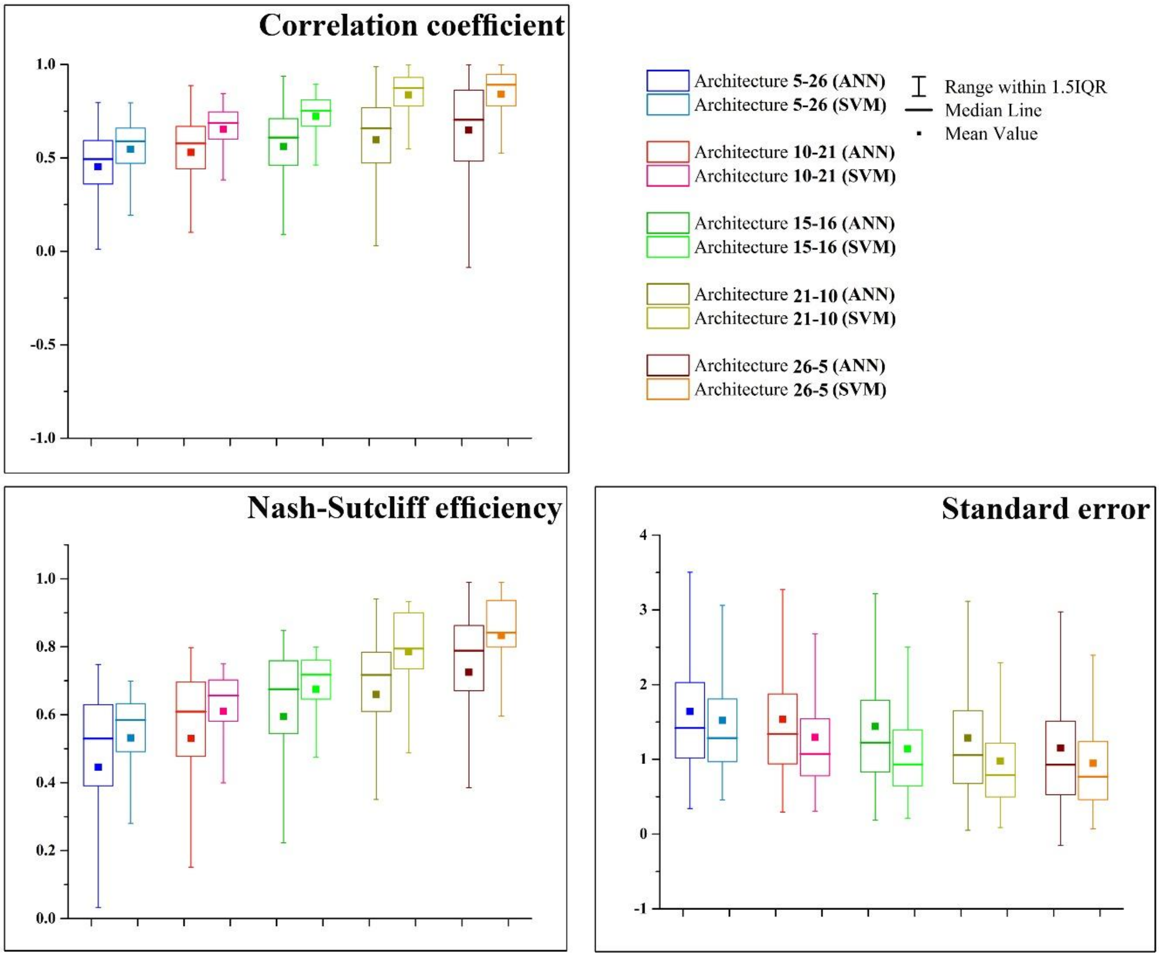



**Fig. S4.** Rational of selection of training and testing set. Configuration – 5 performance is best among other configurations. Configuration – 4 is adapted maintaining a trade-off between the model performances and highest possible test set.

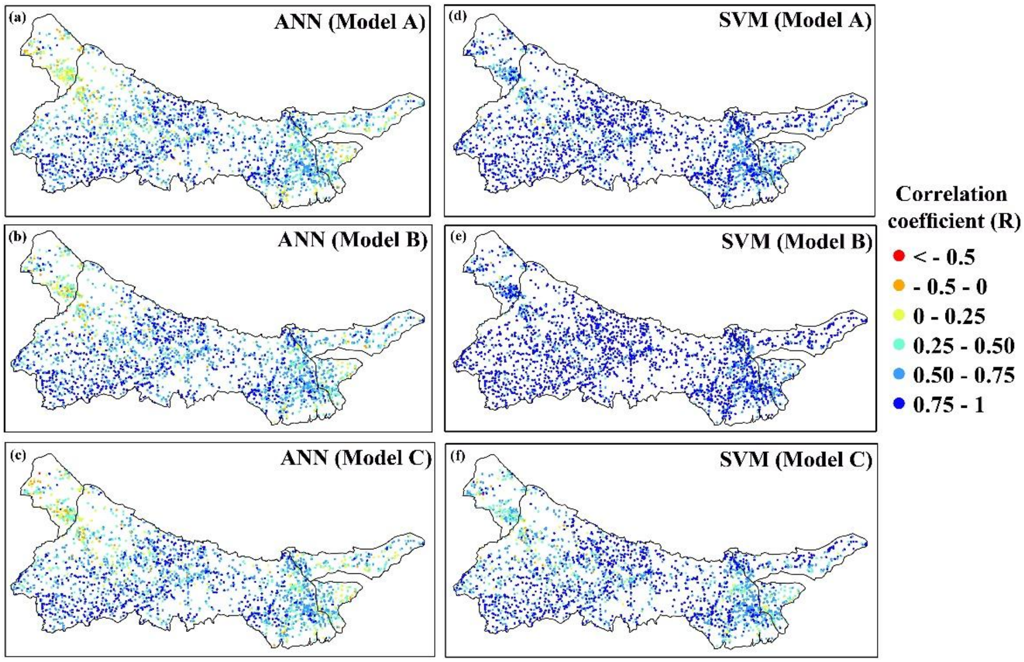



**Fig. S5.** Correlation coefficient (r) map between observed and predicted groundwater levels for the testing period.

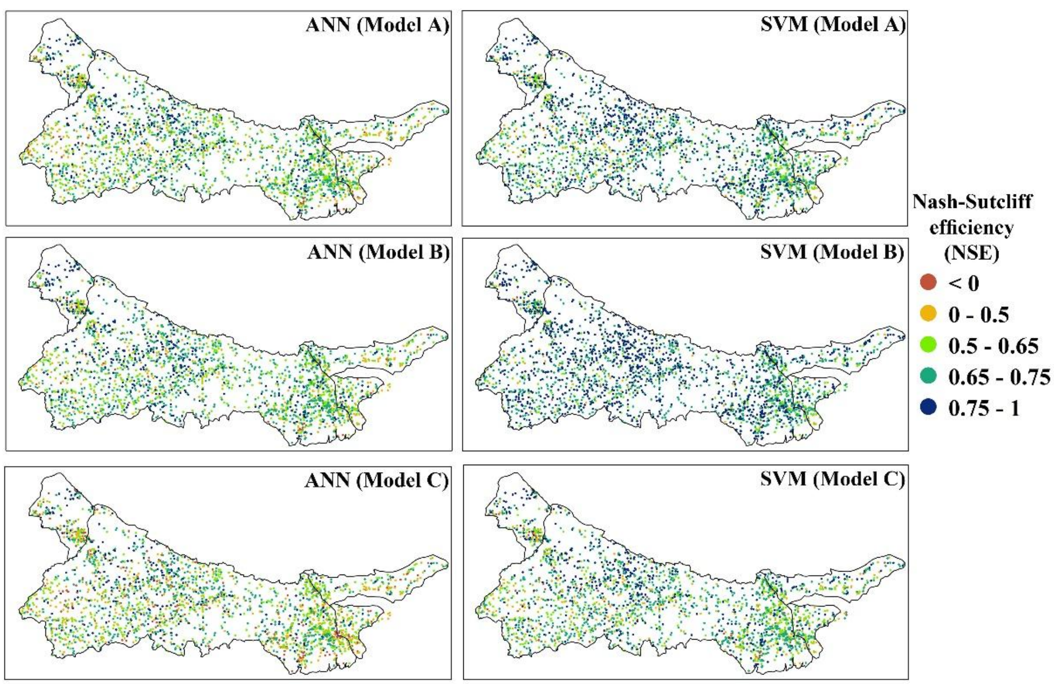



**Fig. S6**. Nash-Sutcliff efficiency (NSE) map between observed and predicted groundwater levels for the testing period.

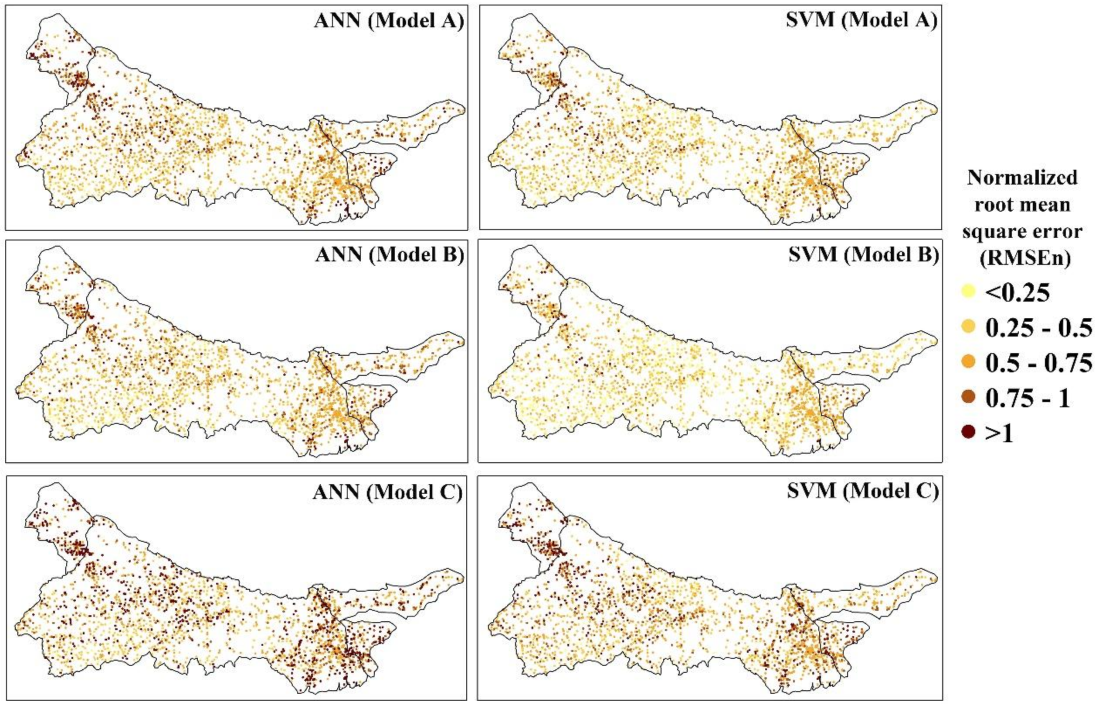



**Fig. S7**. Normalized Root mean square error ($RMSE_n$) map between observed and predicted groundwater levels for the testing period.

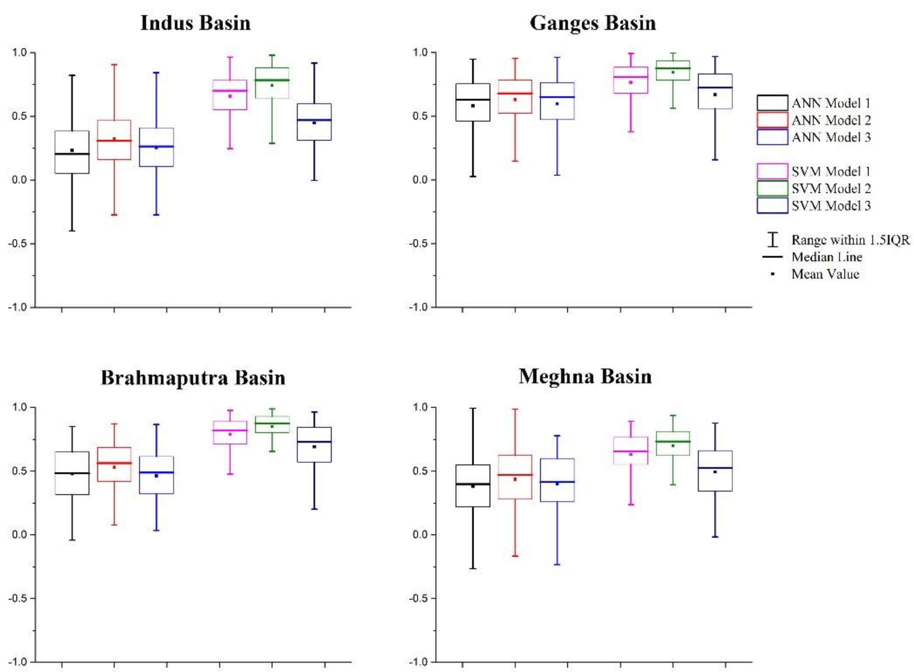



**Fig S8.** Box plots showing the range of correlation coefficient observed from individual well scale analysis for the sub-basins.

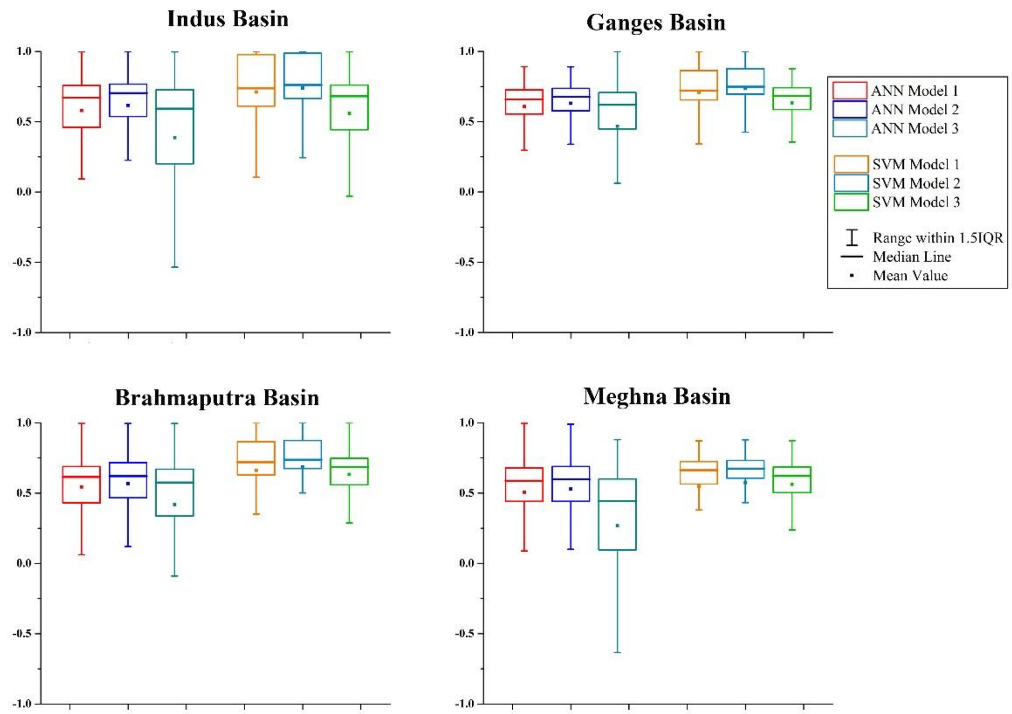



**Fig S9.** Box plots showing the range of Nash-Sutcliff efficiency observed from individual well scale analysis for the sub-basins.

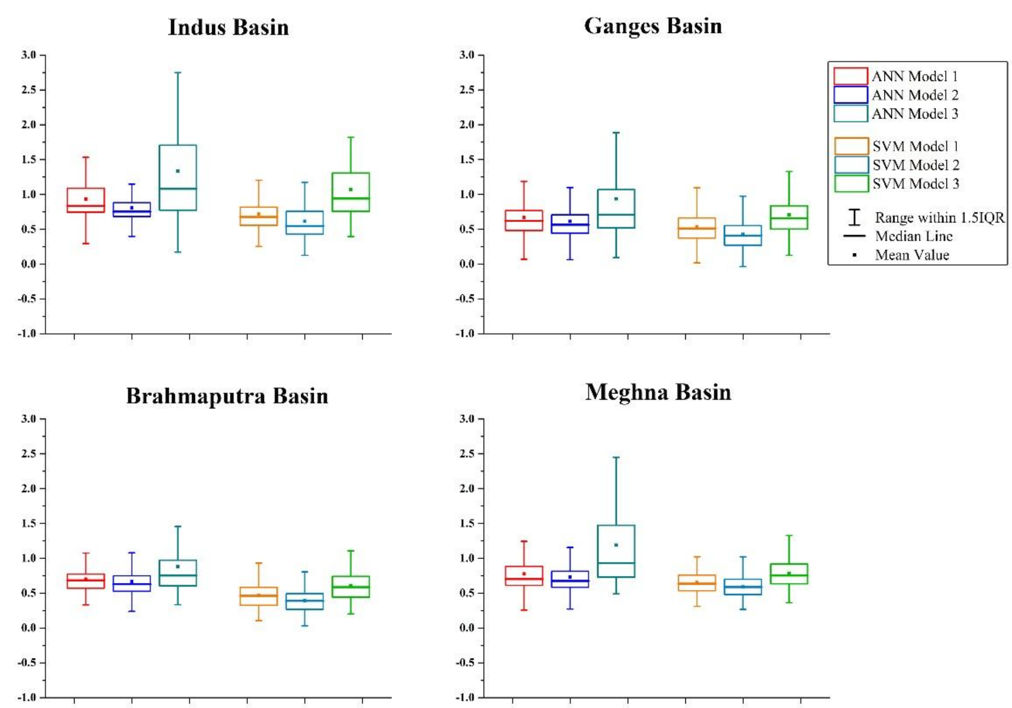
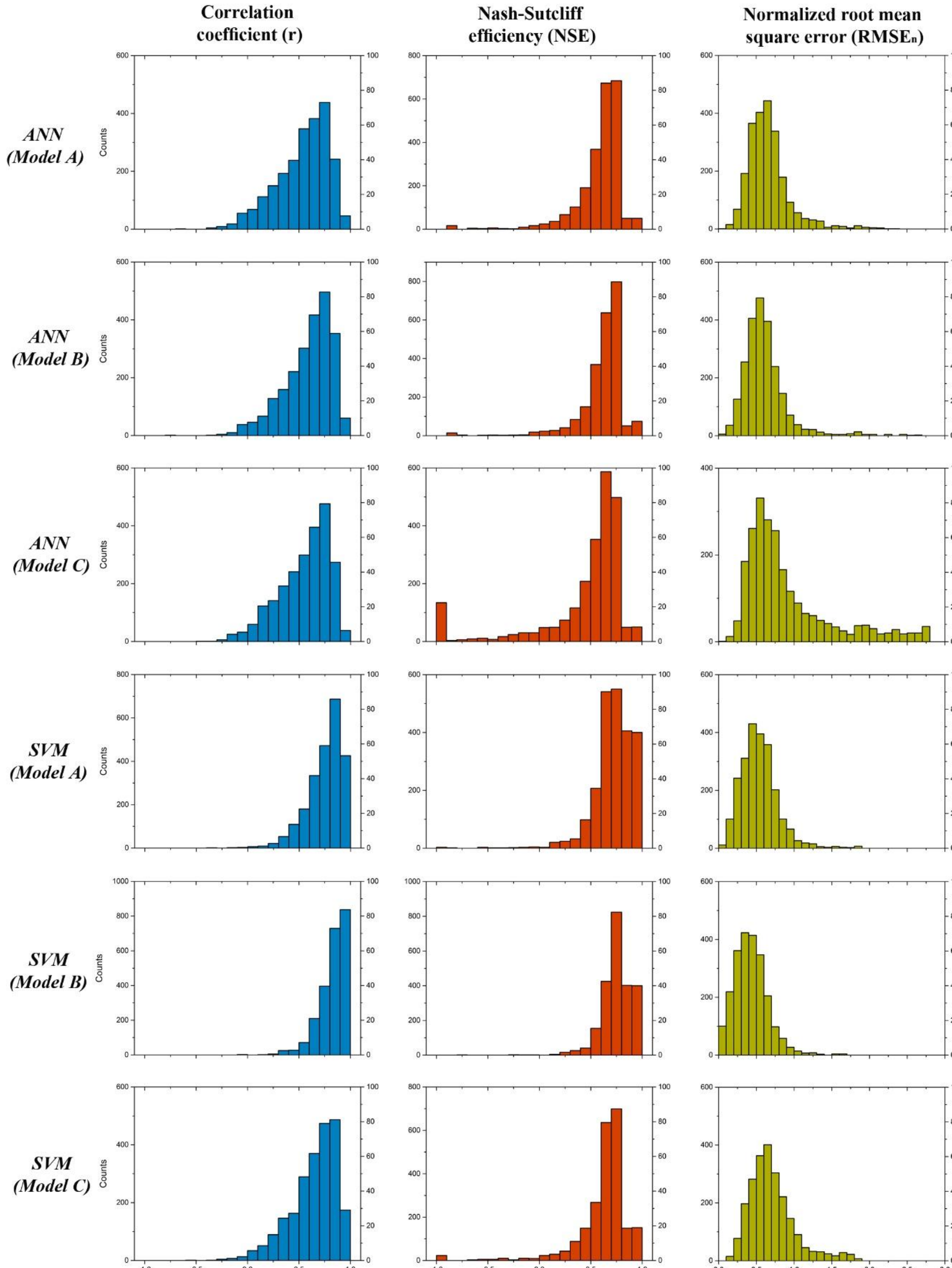



**Fig S10.** Box plots showing the range of Normalized root mean square error ($RMSE_n$) observed from individual well scale analysis for the sub-basins.

125



Correlation coefficient (r)
Nash-Sutcliff efficiency (NSE)
Normalized root mean square error (RMSEn)
ANN (Model A)
ANN (Model B)
ANN (Model C)
SVM (Model A)
SVM (Model B)
SVM (Model C)
Counts
Cumulative Percent

**Fig. S11.** Distribution of well counts with r, NSE RMSE$_n$ for the entire IGBM basin.

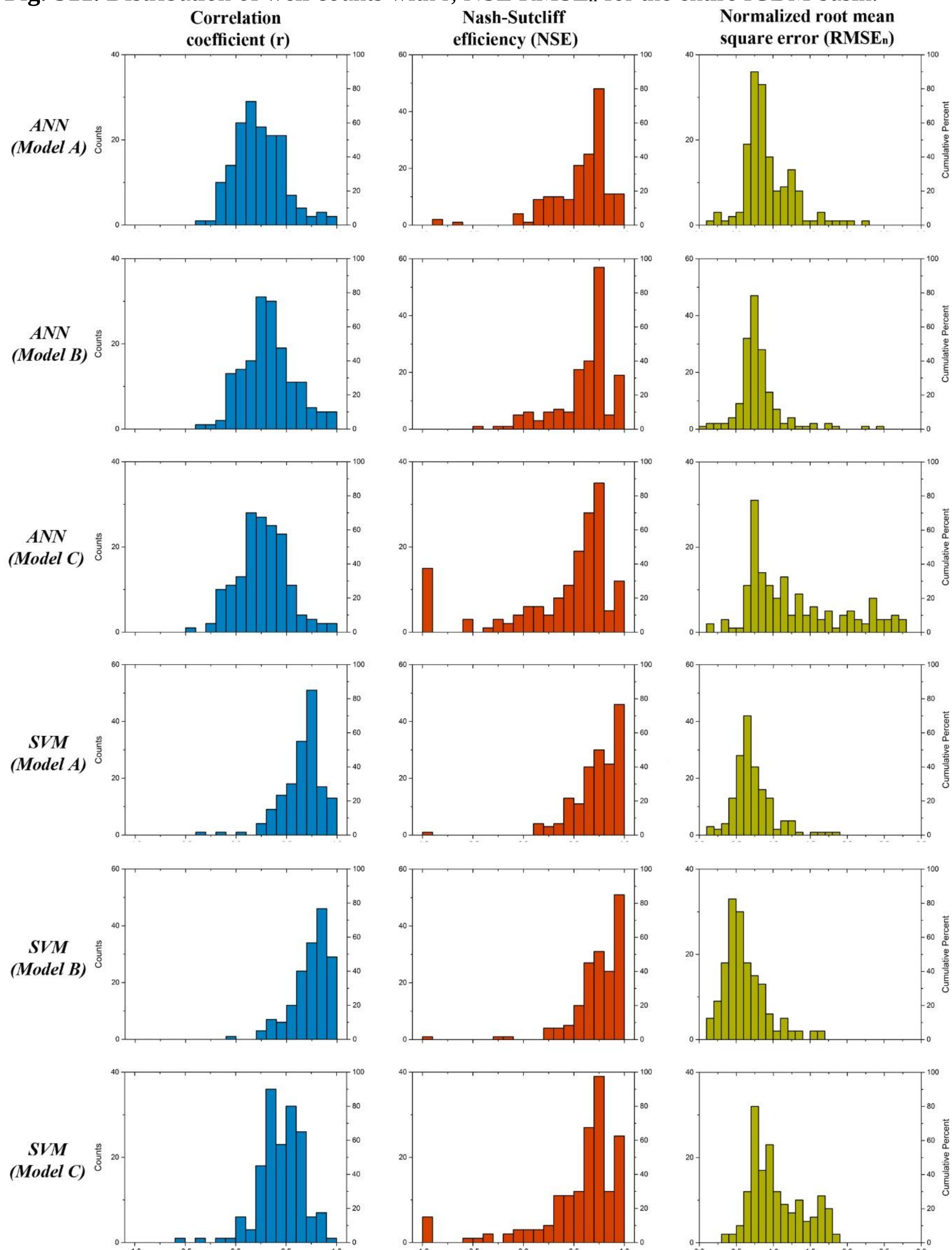



**Fig. S12.** Distribution of well counts with r, NSE RMSE$_n$ for the Indus basin.

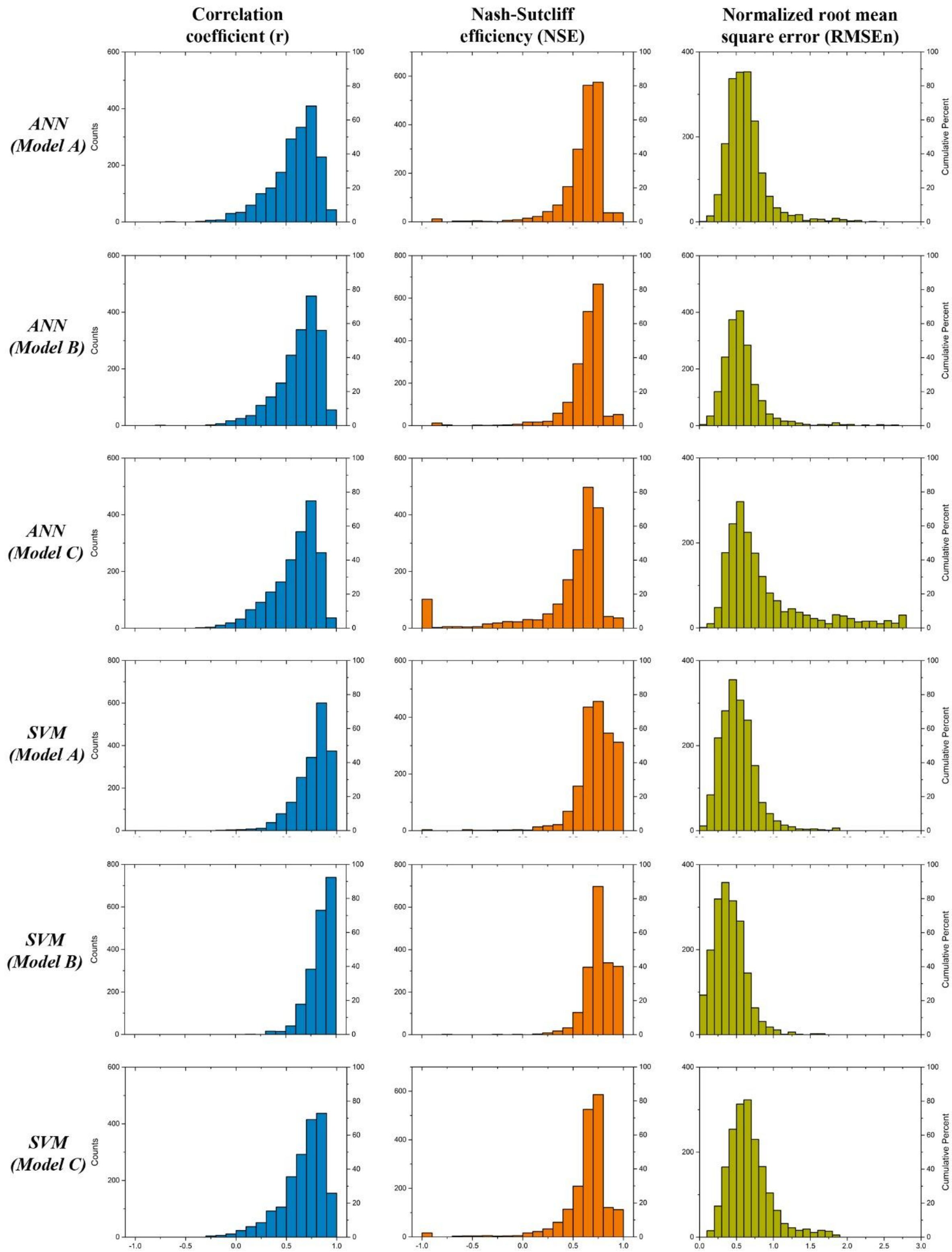



**Fig. S13.** Distribution of well counts with r, NSE $RMSE_n$ for the Ganges basin.

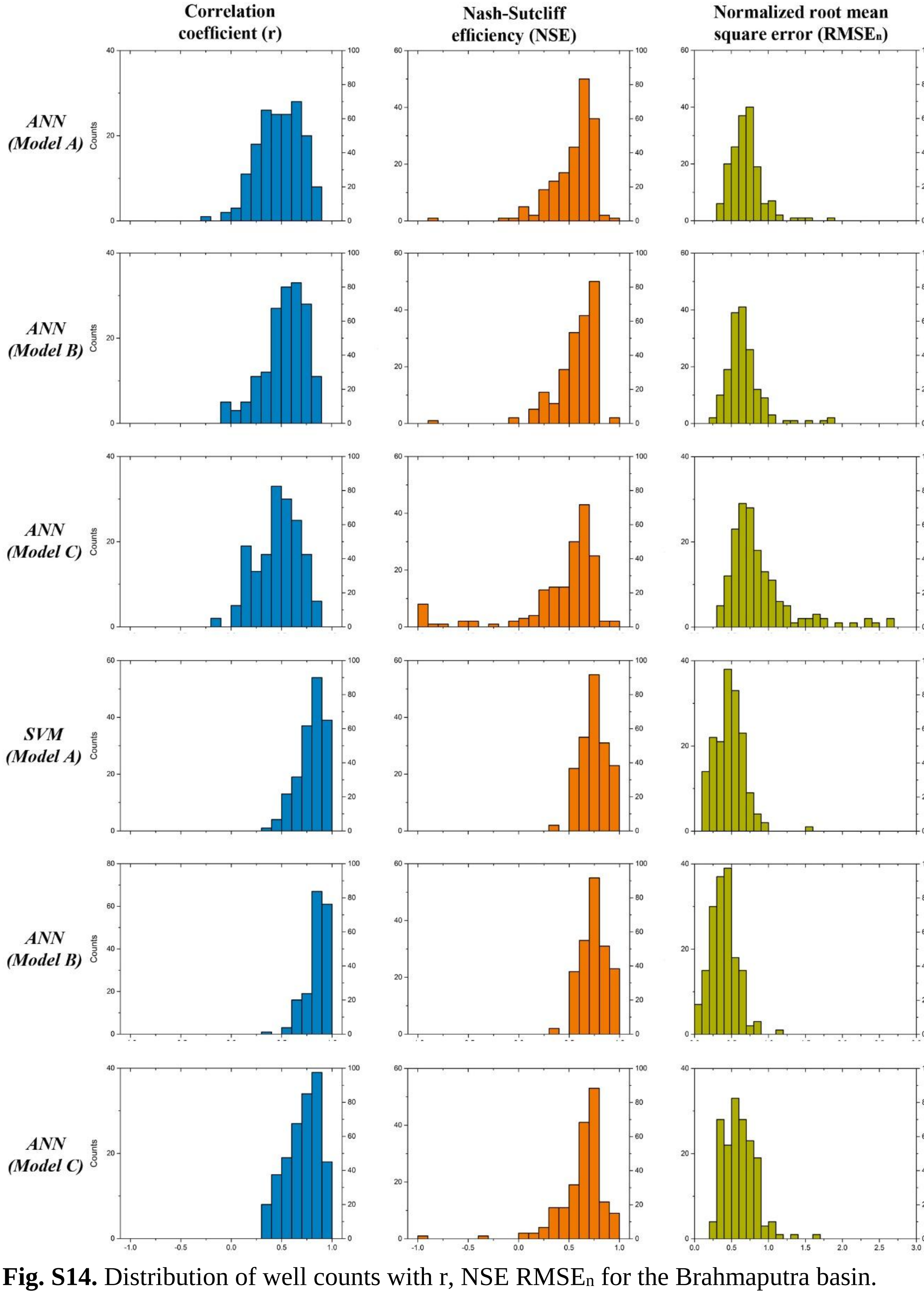



**Fig. S14.** Distribution of well counts with r, NSE $RMSE_n$ for the Brahmaputra basin.

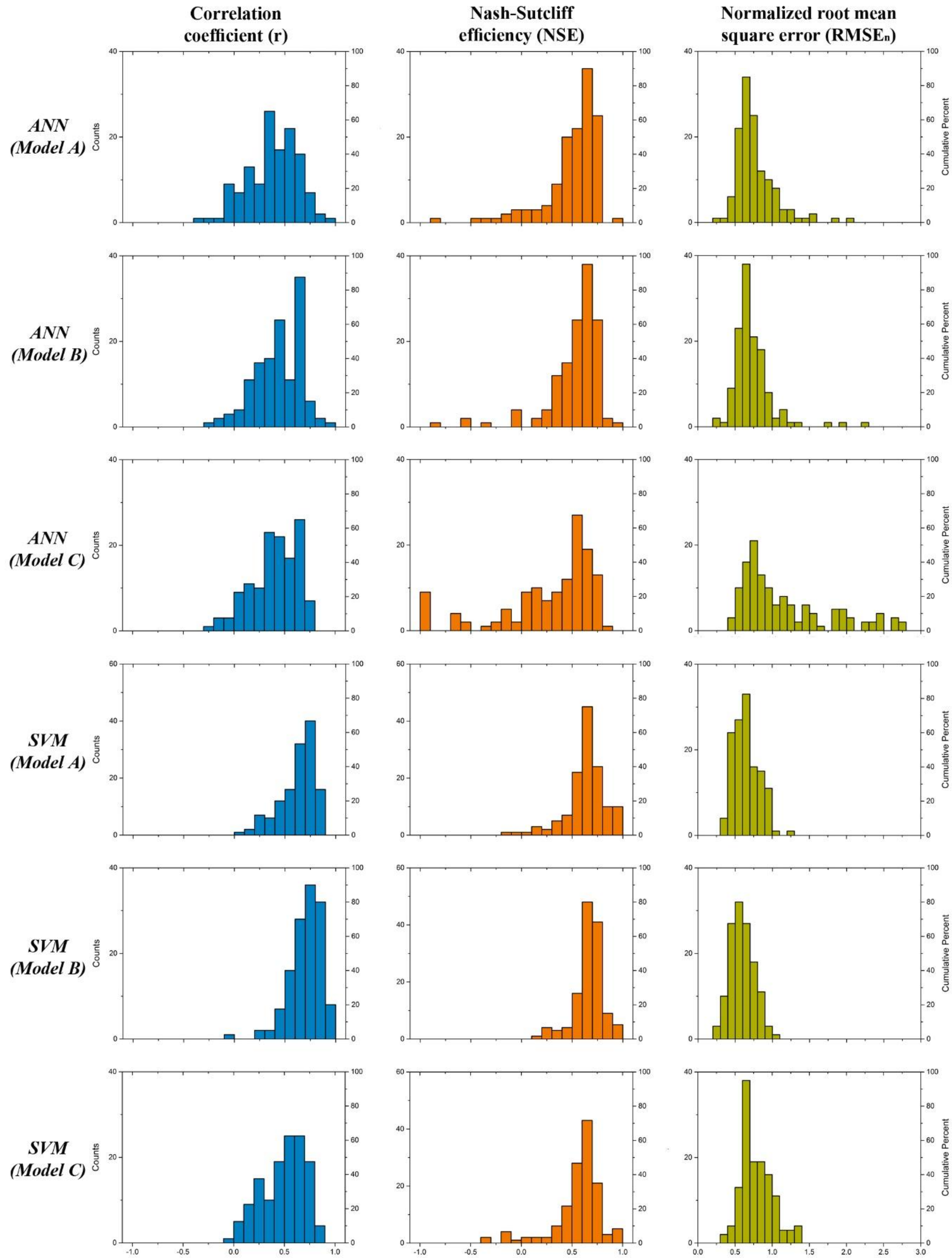



**Fig. S15.** Distribution of well counts with r, NSE $RMSE_n$ for the Meghna basin.

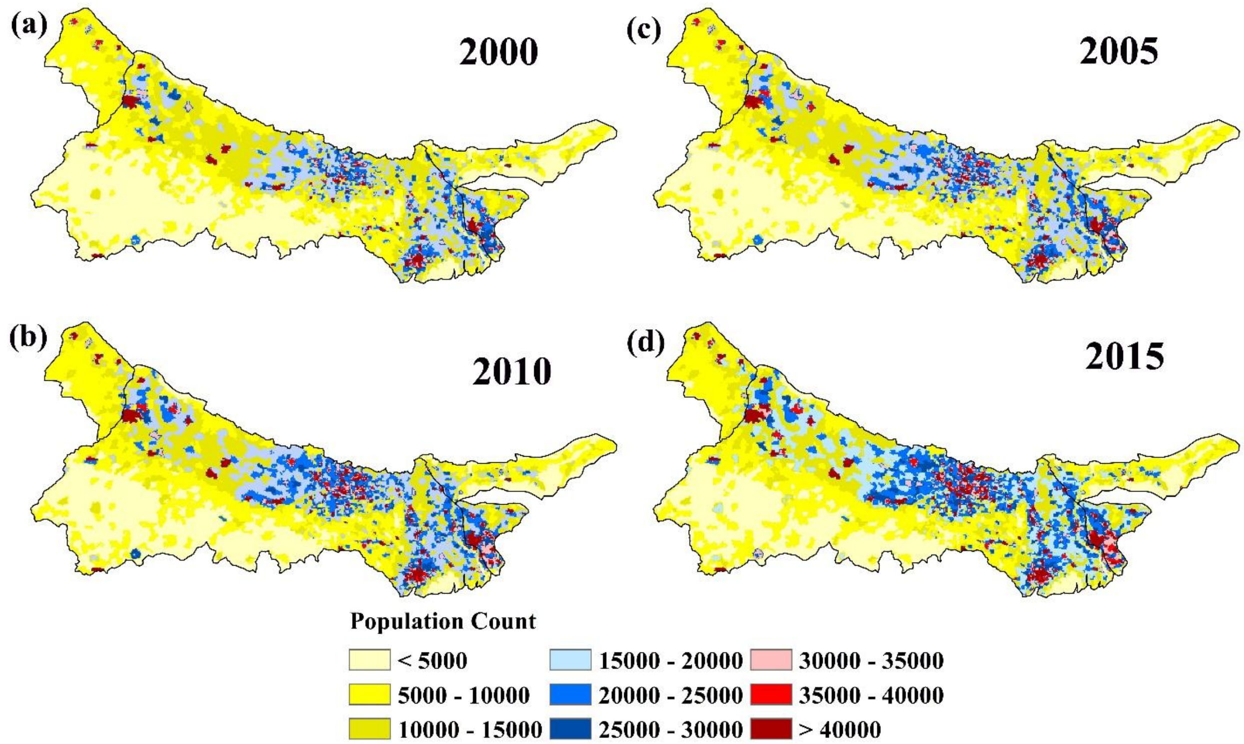



**Fig. S16.** Population count in ~5 km grids, for the year 2000, 2005, 2010, 2015.

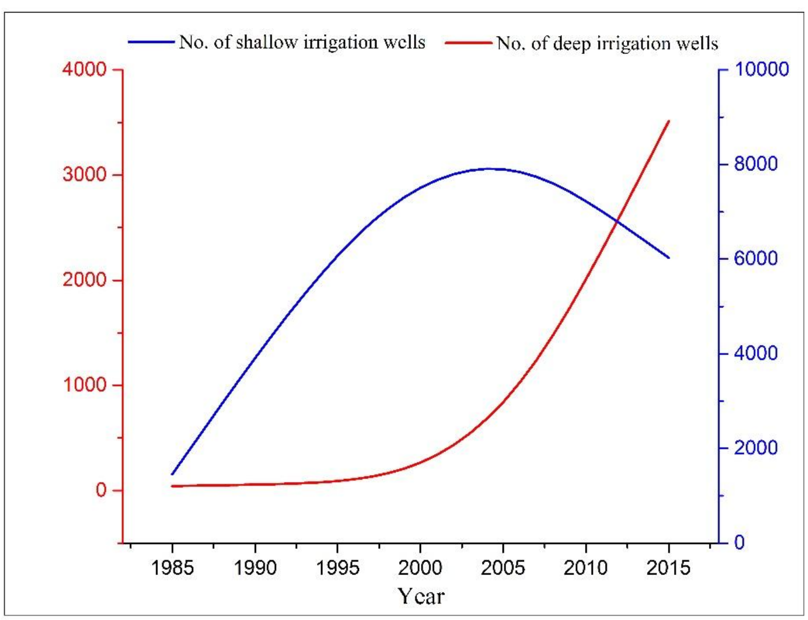



**Fig. S17.** Hodrick–Prescott (HP) of the total number (in thousands) of shallow and deep irrigational wells in the IGBM. The plot is generated using data from the Minor irrigation census reports of India and Minor irrigation survey reports of Bangladesh.

## References


Bangladesh Agricultural Development Corporation (BADC), 2014. Minor Irrigation Survey report, 2013 – 14: G.o.B.

Minor Irrigation Census (MIC), 1993, Report on census of irrigation schemes (1986-1987): G. o. I., Ministry of Water Resources.

Minor Irrigation Census (MIC), 2001, Report on census of irrigation schemes (1993-1994): G. o. I., Ministry of Water Resources.

Minor Irrigation Census (MIC), 2005, Report on 3rd census of irrigation schemes (2000-2001): G. o. I., Ministry of Water Resources.

Minor Irrigation Census (MIC), 2014, 4th census on minor irrigation schemes report (2006-2007): G. o. I., Ministry of Water Resources.

Minor Irrigation Census (MIC), 2017, 5th census on minor irrigation schemes report (2013-2014): G. o. I., Ministry of Water Resources.