Information related to reproducibility, code availability and datasets for manuscript:

Common irrigation drivers of freshwater salinisation in river basins worldwide

Josefin Thorslund[1,2*], Marc F.P. Bierkens [2,3], Gualbert Oude Essink[2,3], Edwin H. Sutanudjaja[2], Michelle T.H. van Vliet[2]

[1] Department of Physical Geography and the Bolin Centre for Climate Research, Stockholm University, Sweden
[2] Department of Physical Geography, Utrecht University, The Netherlands
[3] Unit Subsurface and Groundwater Systems, Deltares, The Netherlands

[*] Corresponding author, e-mail: josefin.thorslund@natgeo.su.se

Computer code and software environment:

All computer code for this study was generated and run in R, using the 3.6.0 version. R is released under the GNU General Public License (GPL), which is open source and the latest version (85 megabytes) can be downloaded and installed from; https://cran.r-project.org, with instructions for both Windows (https://cran.r-project.org/doc/FAQ/R-FAQ.html#How-can-R-be-installed-_0028Windows_0029), Mac (https://cran.r-project.org/doc/FAQ/R-FAQ.html#How-can-R-be-installed-_0028Mac_0029) and Unix-like (https://cran.r-project.org/doc/FAQ/R-FAQ.html#How-can-R-be-installed-_0028Unix_002dlike_0029) options.

Once the R software is installed, only minimal changes (such as file paths and selection of sub-basin and variable values) are necessary to reproduce the analyses of this work. The script 'Thorslund-et-al-example-code.R' includes functions for installing and loading of required packages and code used to perform most of the analyses (produce necessary data for) and to reproduce associated figures of the main manuscript and of the Supplementary Materials (Fig. 3a, Fig. 3b, Supplementary Figure 3a-b, Fig. 4, Supplementary Figure 4, Fig. 5a, Supplementary Figure 2, Supplementary Figure 5 and Supplementary Figure 6).

The script ‘'Thorslund-et-al-RF.R' includes functions for installing and loading of required packages and code used within the Random Forest (RF) modelling approach and example code for running both the CPI and RF analyses on the associated RF dataset ‘'Thorslund.et.al.RF.data.csv'.

Sub-basin shapefiles:

All 401 derived river sub-basins used in the analyses are included in the zipped folder; ‘Sub_basin_shapefiles’, divided by the regional basin to which they belong (i.e. 7 shapefiles; 1 for each regional basin). The sub-basins were projected to the WGS84 reference coordinate system.

Variable included in the processed datasets of this study:

All variables included in the uploaded datasets belonging to this paper are described in Table 1-2 below.

Table 1. Variable names and descriptions, including reported units, of the dataset "Thorslund.et.al.salinity.data.csv".

Variable Name	*Description*	*Unit*
Station_ID	Unique EC sampling station ID	-
OBJECTID	Unique sub-basin shapefile ID	-
Area	Sub-basin area	m^2
Lat	Latitudinal coordinate of sample location	Decimal Degrees
Lon	Longitudinal coordinate of sample location	Decimal Degrees
Country	Geographic location	-
Monthly_lenght	Number of monthly EC data	number
Regional_b	Regional river basin	name
EC_long_term_ann_av	Long term annual average electrical conductivity (EC) value	$\mu S\ cm^{-1}$
Date	Date of observed EC value	yyyy-mm-dd
EC	Observed EC value	$\mu S\ cm^{-1}$

Table 2. Variable names, description and units of the dataset "Thorslund.et.al.drivers.data.csv" used in the analyses of driver contributions and the Random Forest approach. All values are estimated per sub-basin (and given for corresponding Station_ID of each sub-basin). More information on each variable name, data source and resolution are given in Table 1 of the main manuscript.

Variable Name	*Description*	*Unit*
Station_ID	Unique EC sampling station ID of each sub-basin	-
MK_trend	Classified MK trend based on tau and sl values	no trend/inc/dec
EC_long_term_ann_av	Long term annual average electrical conductivity (EC) value	$\mu S\ cm^{-1}$
T	Temperature	°C
P	Precipitation	$m\ month^{-1}$
PET	Potential evapotranspiration	$m\ month^{-1}$
AET	Actual evapotranspiration	$m\ month^{-1}$
PET_P	Evaporative ratio	-
AET_P	Evaporative ratio	-
Q	Discharge	$m^3\ sec^{-1}$
Irr_ww	Irrigation water withdrawals	$m\ month^{-1}$
Non_irr_ ww	Non-irrigation water withdrawals	$m\ month^{-1}$
Irr_rf	Irrigation return flows	$m\ month^{-1}$
Non_irr_rf	Non-irrigation return flows	$m\ month^{-1}$
tot_dam_storage_Mm3	Total dam capacity	Mm^3
n_dams	Number of dams	number
Dam_storage_norm	Ratio of dam capacity to sub-basin area	-
Dam_area_ratio	Ratio of dam area to sub-basin area	-
Tot_cropland	Total cropland area	m^2
Irr_area_ratio	Ratio irrigated area to sub-basin area	-
N_ appl	Nitrogen application	Tons
P_ appl	Phosphorous application	Tons
EC_top_soil	Soil salinity of top layer	$dS\ m^{-1}$
EC_sub_soil	Soil salinity of sub layer	$dS\ m^{-1}$
EC_soil_average	Average soil salinity	$dS\ m^{-1}$
Elevation	Elevation	m.a.s.l.
Distance_coast_km	Actual distance from coast	km

relative_dist_coast	Relative distance from coast	-
Tau_*variable_name*	Mann-Kendall Kendall rank correlation coefficient of each variable	-
sl_*variable_name*	Mann-Kendall two-sided p-value of each variable	-
Sen_slope_*variable_name*	Sen slope value of each variable	-
P_value_*variable_name*	P value of Sen slope analysis of each variable	-