# SCIENTIFIC DATA

**OPEN**

## Data Descriptor: A web platform for landuse, climate, demography, hydrology and beach erosion in the Black Sea catchment

Anthony Lehmann[1,2], Yaniss Guigoz[3], Nicolas Ray[1,3], Emanuele Mancosu[4], Karim C. Abbaspour[5], Elham Rouholahnejad Freund[6,7], Karin Allenbach[3], Andrea De Bono[3], Marc Fasel[1], Ana Gago-Silva[1], Roger Bär[8], Pierre Lacroix[1,3] & Gregory Giuliani[1,3]

The Black Sea catchment (BSC) is facing important demographic, climatic and landuse changes that may increase pollution, vulnerability and scarcity of water resources, as well as beach erosion through sea level rise. Limited access to reliable time-series monitoring data from environmental, statistical, and socio-economical sources is a major barrier to policy development and decision-making. To address these issues, a web-based platform was developed to enable discovery and access to key environmental information for the region. This platform covers: landuse, climate, and demographic scenarios; hydrology and related water vulnerability and scarcity; as well as beach erosion. Each data set has been obtained with state-of-the-art modelling tools from available monitoring data using appropriate validation methods. These analyses were conducted using global and regional data sets. The data sets are intended for national to regional assessments, for instance for prioritizing environmental protection projects and investments. Together they form a unique set of information, which lay out future plausible change scenarios for the BSC, both for scientific and policy purposes.

| Design Type(s) | data integration objective ● time series design ● observation design |
|---|---|
| Measurement Type(s) | environmental system process |
| Technology Type(s) | computational modeling technique |
| Factor Type(s) | DataTypes |
| Sample Characteristic(s) | Black Sea ● land use ● climate change ● demographics ● hydrological process ● sandy beach |

[1]University of Geneva, Institute for Environmental Sciences, enviroSPACE Lab., Bd. Carl-Vogt 66, Geneva CH-1211, Switzerland. [2]University of Geneva, Department F.-A. Forel of Environmental and Aquatic Sciences, Bd. Carl-Vogt 66, Geneva CH-1211, Switzerland. [3]University of Geneva, Institute for Environmental Sciences, GRID-Geneva, Bd. Carl-Vogt 66, Geneva CH-1211, Switzerland. [4]European Topic Centre-Spatial Information and Analysis, University of Malaga, edificio CAITI, Campus de Teatinos, Malaga 29071, Spain. [5]Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, Duebendorf CH-8600, Switzerland. [6]Department of Environmental Systems Science, ETH Zurich, Universitaetstrasse 16, Zurich CH-8092, Switzerland. [7]Laboratory of Hydrology and Water Management, Ghent University, Ghent B-9000, Belgium. [8]University of Bern, Centre for Development and Environment, Hallerstrasse 10, CH-Bern 3012, Switzerland. Correspondence and requests for materials should be addressed to A.L. (email: anthony.lehmann@unige.ch).

## Background & Summary

The data sets presented in this paper were essentially generated in the framework of a European research project entitled 'EnviroGRIDS: Building Capacity for a Black Sea Catchment Observation and Assessment System supporting Sustainable Development' that took place between 2009 and 2013 with the participation of more than one hundred scientists from thirty different partner institutions from fifteen countries[1].

This project addressed the gap between scientific information available in the Black Sea catchment (BSC) and the environmental policy needs of the Commission for the Protection of the Black Sea against Pollution and the International Commission for the Protection of the Danube River[2]. The main aim of the project was to build capacity on new Earth Observation techniques and data sharing through web services as a European contribution to the Global Earth Observation System of Systems (GEOSS)[3]. In order to demonstrate the benefit of system interoperability and data sharing, the project developed a pipeline of analyses that produced many new and original data sets for the Black Sea (400,000 km$^2$) and its entire hydrological catchment (2.2 mio. km$^2$).

The enviroGRIDS project is rooted in and took its name from the intended use of computer grid technology to process large amount of environmental data through complex models[4]. This data processing followed a logical suit of analyses that is illustrated in Fig. 1.

The first part of the analyses aimed at setting the scene in terms of plausible climatic, demographic and landuse change scenarios for the entire BSC until 2050. Regionalized storylines were first developed to define four scenarios (BS-Hot, BS-Alone, BS-Coop, BS-Cool) according to their position on two axes of globalisation-regionalisation on the one hand, and liberalisation-policy control on the other hand. Climatic scenarios were derived from the Delta method in order to extrapolate temperature and precipitation according to different scenarios from the Intergovernmental Panel on Climate Changes (IPCC). Demographic scenarios were developed on the basis of the extrapolation census data by political units with urban and rural models[5]. Finally, landuse change was modelled with the Metronamica[6] software with an iterative yearly stochastic process taking into account the probability of change from one landuse to another and the accessibility by roads and trains[7].

The second set of analyses is linked to hydrological models using the Soil and Water Assessment Tools (SWAT2009; ref. 8). This model was used to model and predict river flow and nutrient loads (NO$_3$) into the Black Sea.

A third independent analysis was done on the outputs from the SWAT model that were used to assess agricultural water vulnerability[9] and blue water scarcity[10].

Finally, beach erosion was modelled along the entire Black Sea coastline and across six countries by first digitizing manually every beach based on background satellite images, and then applying an ensemble of six erosion models on them[11].

This coherent set of environment geospatial data sets on the BSC constitutes a regional contribution of the enviroGRIDS project to the GEOSS. These data sets and their corresponding metadata are made freely available (1) through direct download, (2) as web services, and (3) as map composition to ease visualisation on a GeoNode platform (http://blacksea.grid.unep.ch) that is a simple implementation of a typical Spatial Data Infrastructure[4]. Full copies of these data sets and metadata have been uploaded to Dryad (Data Citation 1).
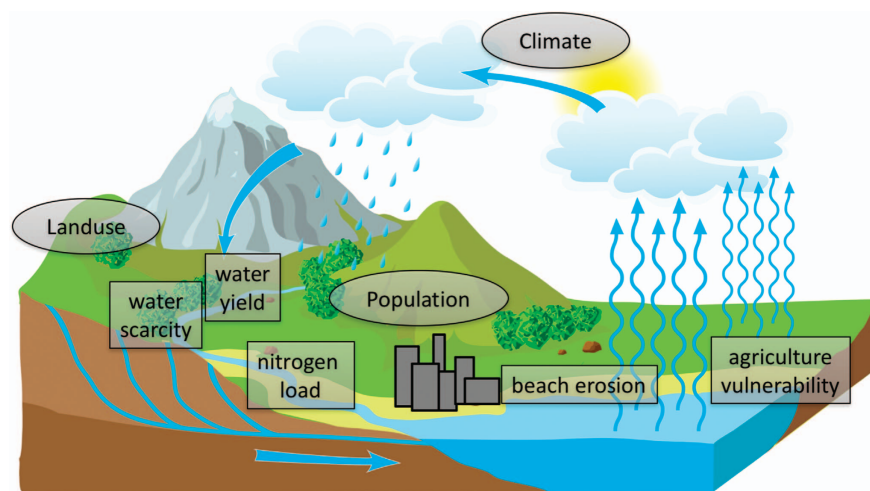


**Figure 1.** EnviroGRIDS suite of analyses: plausible landuse, climatic and demographic scenarios (circles); SWAT hydrological modelling; water scarcity and vulnerability analyses; and beach erosion (squares).

## Methods

The setting of spatially explicit **landuse change** scenarios has been defined from qualitative socio-economical enviroGRIDS storylines[7] (BS-HOT, BS-ALONE, BS-COOP and BS-COOL) based on the interpretation of the respective global scenarios from the IPCC (A1, A2, B1 and B2). Quantitative statistical downscaling techniques were used to disaggregate the outputs from global scenarios (IMAGE[12], version 2.2) at the level of administrative units (Nomenclature of Units for Territorial Statistics, level 2—NUTS2). The disaggregation of the landuse statistics has been divided into three economic groups: FSU (Former Soviet Union), OECD WEU (OECD members in Western Europe including Turkey) and REF EEU (economic reform countries in Eastern Europe). The resulting landuse indicators were spatially distributed at a 1 km resolution by an allocation model (Metronamica[6]) applying a set of factors related to the identified driving forces of change. The landuse change model[13] was calibrated by integrating spatial allocation rules translated from historical trends of landuse change (MODIS 2001 and 2008), by considering a defined stochastic perturbation, by applying suitability and constraint maps, and by respecting population trends downscaled from National statistical offices to regional (NUTS2) level[5] (see demographic section below). Physical suitability describes how suited a cell is to a specific landuse type, based on the overlay of different variables such as elevation, slope, soil quality[14], mean temperature and annual precipitation (from WorldClim[15]). The constraint maps were used to limit or stimulate particular changes or permanence in landuse by assigning a potential change level to each individual class. In detail, protected areas (Natura 2000, EEA; WDPA, UNEP/GRID-Europe), fire events and flood risk, based on the 2009 Global Assessment Report on Disaster Risk Reduction[16], were used to constrain the urban sprawl. Conversely, major and secondary roads from the Global Road Inventory Project and ESRI, as well as rail network, were included as accessibility maps to favour urban sprawl. The total potential derived by the combination of the different factors was calibrated measuring the goodness of fit between the observed and simulated landuse maps. Finally, four simulated future states were produced covering the BSC until the year 2050 for the following landuse classes: forest, grassland, cropland, built-up area, scrublands, crops/natural vegetation, and barren land.

Daily gridded precipitation and temperature datasets were created as well as scenarios of **climate change**. A two-step process was applied to obtain these gridded datasets. First we developed gridded precipitation and temperature (maximum and minimum values) data over the Black Sea catchment. These surfaces were created using the Integrated Nested Laplace Approximations (INLA) algorithm[17] as implemented in the R-INLA package (more information on this package is provided in Code availability section). The surfaces were obtained by interpolation of the daily observations at meteorological stations collected in the Global Historical Climatology Network (GHCN) and European Climate Assessment & Dataset (ECA & D), covering the BSC over the period 1961–1990. The INLA interpolation validation was performed by means of cross-validation, both precipitation and temperature surfaces were interpolated using R-INLA with a spatial resolution of 0.25° for both longitude and latitude. In a second step we downscale the Regional Climate Models (RCMs) outputs from the PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects). PRUDENCE was a European network program coordinate by the Danish Meteorological Institute (DMI), with the aim to create high-resolution regional simulations of present-day (1961–1990) climate and scenario simulations (2071–2100) for Europe. The scenario simulations were based on global climate models that simulate global climate response to the two regional IPCC's SRES GHG emission scenarios[18]. The two regional emission scenarios are the A2 and the B2 scenarios, and correspond in PRUDENCE to HS1 and HB1. From the available PRUDENCE outputs, we use the HIRHAM simulations (available online at http://prudence.dmi.dk/data/daily). Detailed information about how the regional simulation climate models were driven can be found on the PRUDENCE website (http://prudence.dmi.dk). The downscaling was made by using the Delta method. This method aims at perturbing observed time series with changes allowing increasing as a function of time. We derived downscaled climate projections by applying decile changes to perturb monthly rainfall and temperature probability distribution functions[19]. The validation of the Delta method was done by applying it first to the HIRHAM data set. The daily averages obtained for each scenario (HS1 and HB1) period were then subtracted from the corresponding daily averages of the present-day period (referred as HC1 in PRUDENCE). The differences are divided into deciles, and finally grouped into monthly deciles. This calculation is repeated for each climate scenario for the monthly precipitation and maximum and minimum values of temperature. The same methodology was later applied to the gridded surfaces to obtain the scenarios of climate change (2071–2100) for the BSC. Finally, we calculate the daily datasets 2025–2070 via linear interpolation between the last year from the control period (1990) and the first year of each scenario (2071).

We also integrated regional landuse and climatic change scenarios with **demographic data**. According to the enviroGRIDS storylines, the UN projection variants[20] for population were analyzed, and a methodology was proposed for the downscaling of demographic data from national to regional level (NUTS2). Results included urban and total population trends over the period 2010–2050 for 214 regions. The data on population per region was collected for all available years from 18 different sources but mainly Eurostat and National statistical offices. Data were heterogeneous and occasionally incomplete. Therefore, a three-step process of harmonization was applied: (1) formatting and translating data from National statistical offices, (2) matching the statistical and geometric (boundaries) data together, and

(3) interpolating linearly gaps for missing years. This process led to the completion of population data for the two base years (2001 and 2008) used in the landuse change analysis. Successively, missing values in time series were extrapolated linearly in order to cover the period 2000–2010. Considering that there is no uniform data on the proportion of urban population at the regional level, the urban population per region was calculated as the sum of cities with at least 10,000 inhabitants included in that region. To achieve this, a specific spatial data set containing population size of cities was generated. Except for EU 27 (ref. 21) and Russia, data for demographic projections at sub-national scale was not available. Therefore, the regional projections for total and urban population, for the period 2011–2050, were successively estimated on the assumption that future demographic developments can be derived from past population trends and hence, a continuation of observed demographic change is assumed. In the first step we analyzed how population is distributed in different regions of each country and used the trend to extrapolate until 2050. Once the trend was identified in a region over the period of interest, we transposed national United Nation Statistical Division population data for the three corresponding variants (high, medium and low fertility). Regarding the urban population, before applying the same approach, we first calculated the ratio between the urban and total national population. Subsequently, with the assumption that the urban growth rate is distributed uniformly between regions, the percentage of urban population was then calculated per region.

The Soil and Water Assessment Tool (SWAT2009; ref. 8) was used to develop the **agro-hydrological model** of the BSC. SWAT is a continuous-time and process-based hydrological model. We chose this program because it lends itself easily to climate and landuse change analyses and couples various processes to simulate natural and anthropogenic systems in watersheds including hydrology, climate, snow, nutrient, soil, sediment, crop growth, pesticide, surface depressions, and agricultural and water management. A water balance equation is solved to initiate water impoundment, which is a function of total inflow (e.g., runoff entering from the upstream subbasins, rainfall, groundwater contribution) and total outflow from the water bodies (e.g., evaporation, seepage into the subsurface). More details about the model are provided by in its theoretical manual[22].

The BSC SWAT project was built with data that are described in details in a dedicated publication[23] on observed[24,25] and gridded[26,27] precipitation and temperature, solar radiation[28], soil characteristics[29], landuse categories[30], digital elevation model[31], river networks[32], estimated nutrients point sources, and agricultural management[33]. The spatial heterogeneity of the watershed was preserved by topographically dividing the basin into multiple sub-basins (12,982) using 90-m digital elevation model. These were further subdivided into hydrologic response units (HRU) based on the unique combinations of soil, landuse, and slope characteristics (89,202 HRU). These subdivisions enabled the model to reflect differences in evapotranspiration for various crops and soils. In each HRU and at each time step, the hydrologic and vegetation-growth processes were simulated based on the curve number rainfall-runoff partitioning and the heat unit phenological development method[22]. The agro-hydrological model was run at a daily time step where a full run took 42 h on a single Windows computer (64-bit, 2.7 GHz processor, 4 cores, and 8 GB of RAM). Calibration runs where, therefore, made in a network of 50 Linux computers[34]. The Hydrological model was calibrated against observations in river discharge, nitrate concentration, and crop yields across the catchment for the period of 1970–2000. Monthly river discharge data was obtained from Global Runoff Data Centre[35], National Institute of Hydrology and Water Management (INHGA), Danube Delta National Institute for Research and Development (DDNI) in Romania, and Turkish Ministry of Forest and Water Affairs (MEF) (144 observation stations in total). Nitrate concentration in rivers was taken from International Commission for the Protection of the Danube River (ICPDR) (37 stations). Crop yield data was obtained from McGill University[33] at 5 arc-minute resolution. The calibrated model was then tested with an independent set of data on river discharge, nitrate concentration, and crop yields for the period of 2001–2006.

We used the Sequential Uncertainty Fitting program SUFI-2 (refs 36,37) to calibrate and validate the model results. The program is linked to SWAT and provides the basis for parallel processing of multi-gauge calibration and large-scale parameterization schemes[24]. It also provides a platform for sensitivity and uncertainty analysis. In SUFI-2, parameter uncertainty is described by a multivariate uniform distribution in a parameter hypercube, while the output uncertainty is quantified by the 95% prediction uncertainty (95PPU) calculated at the 2.5 and 97.5% levels of the cumulative distribution of an output variable. Latin Hypercube is used to draw independent samples from the input parameter uncertainties, which lead to the calculation of the 95PPU for a give output variable. Two statistics quantify the goodness of fit and model output uncertainty. These are the *p-factor*, which is the percentage of measured data being bracketed by the 95PPU, and *r-factor*, which is the average thickness of the 95PPU band[22,23]. The *p-factor* has the highest value of 1, while the lowest value for *r-factor* is zero. For flow, Abbaspour *et al.*[23] suggest a practical value of 0.6–0.8 for the *p-factor* and a value around 1 for the *r-factor*. In this definition, (1-*p-factor*) can be thought of as model error, or measured data not accounted for by the model.

Having the calibrated BSC model in hand, water resources components such as **blue water** (fresh water availability), **green water flow** (evapotranspiration), and **green water storage** (water stored in soil) along with their associated uncertainties were calculated for the entire BSC. These concepts give an overall picture of water resources and bring the outputs of the BSC model closer to the needs of water resources researchers and policy makers. The suggested upper and lower bounds of the 95PPU reflect primarily the prediction uncertainty (including model structure, parameters, and inputs), as well as the

temporal climate variation as a secondary effect. A comprehensive agro-hydrological database of the BSC comprising of all input data essential for building a detailed hydrological model as well as all the historic water resources modeling results was created and shared on the enviroGRIDS GeoNode platform to fill the existing gaps in water resources data in the region.

The vulnerability assessment aimed at assessing the **potential impact** of climate change on agricultural water resources, the adaptive capacity due to irrigation, and the resulting vulnerability of agricultural water resources in the BSC. The data used for the study is based on a agro-hydrological model discussed above[23]. Using SWAT[8], two strong changing forces were applied under two scenario conditions: 1) the daily temperature was increased by 3 °C, and 2) the daily precipitation was decreased by 30%. The relevant model outputs were temperature, precipitation, irrigation, potential daily evaporation, and stream flow, all provided on a daily time step and for each of the 12,892 sub-catchments of the BSC. The related uncertainties concern the predictions of the agro-ecological model and have been discussed above. The output variables were used to compute auxiliary indicators such as temperature stress, water stress, and environmental water requirement, and were subsequently processed to produce indices for potential impact, adaptive capacity, and vulnerability. The **potential impact** was measured by the changes in natural plant growth days per year. A plant growth day was considered to be a day with neither water stress nor temperature stress. Temperature stress days were defined as days when the average daily air temperature was below 5 °C or above 35 °C[38], whereas water stress days were defined as days when the actual daily evapotranspiration was less than half of the potential daily evapotranspiration[39,40]. Days when water stress would have occurred without the help of irrigation were not considered to be natural plant growth days. The **adaptive capacity** was measured by the number of potential irrigation days per year. A potential irrigation day was considered to be a day when the actual stream flow is above the environmental water requirement[41]. The latter was estimated by calculating 80% of the 10-year average stream flow for each respective day[42]. Finally, the **vulnerability** was measured by the change of the total number of plant growth days per year. Plant growth days were defined as the number of days with neither water stress nor temperature stress. Contrary to the number of natural plant growth days, the total number of plant growth days included days when plant growth was possible only with irrigation.

**Water scarcity indicators** aim at mapping and informing decision-makers and the public about the use of this limited renewable resource. In this section, we balanced modelled surface water availability in the BSC[23] with needs and consumptive use of key water users, i.e., municipalities, power plants, manufacturing, irrigation[43] and livestock breeding[44]. We took into account evaporation from major reservoirs and environmental flow requirements (EFR) to preserve critical ecosystem structures and functions. We evaluated EFR with the Variable Monthly Flow (VMF) method[45] allowing for multiple disturbances to occur so that socioeconomic development can be supported. The needs and consumptive use of municipalities were calculated using urban and rural population densities, rate of connection to a running water supply system, and national domestic water withdrawal intensities. Water need for cooling power plants was calculated taking into account electricity production, fuel type, and cooling system of 3,065 power plants. Those for manufacturing production were acquired from national statistics offices, FAO[46] and World Bank[47] and downscaled according to urban population density. Net irrigation water requirements were obtained from Siebert and Doell[48] and temporally allocated using the same crop coefficient method as in aus der Beek *et al.*[43]. A country specific irrigation efficiency factor was applied to obtain gross irrigation withdrawals. Withdrawals and consumptive use for livestock breeding were estimated using cattle, goat, sheep, pig and poultry densities, as well as dominant livestock production systems. In the absence of monthly data, withdrawals and consumption of all sectors except for irrigation were assumed to remain constant during the year, and to only vary from year to year. Evaporation from 325 large reservoirs was estimated using the Penman-Monteith equation assuming a daily evaporation 10% higher than reference evapotranspiration[49]. With this information, we modelled two complementary water scarcity indices. The first was about consumption ($WSI_c$), and how the reduction of water flow could threaten downstream ecosystems and users following Hoekstra's index[50]. The second ($WSI_w$) was about demand, i.e., withdrawals, and how the available water could be insufficient, even independently from upstream consumptive use according to Smakhtin[51]. The following representative scales were used for $WSI_c < = 0.2$: No scarcity; 0.2–0.3: Low; 0.3–0.4: High; and >0.4: Very high scarcity; and for $WSI_w < 0.1$: No scarcity; 0.1–0.2: Low; 0.2–0.4: Moderate; 0.4–0.8: High; and >0.8: Very high scarcity.

**Sandy beaches** are accumulation zones of unconsolidated material in the shoreline, consisting of sediments transported by littoral drift. As natural and dynamic system, beaches provide many valuable ecosystem services such as coastal protection (sediment storage and transport, wave dissipation, buffer effect to extreme events, dynamic response to sea level rise), biodiversity preservation, water filtration and purification, and recreational landscape[52]. Being squeezed between terrestrial and marine environments, this ecosystem is particularly vulnerable to coastal erosion. On the marine side, sea is rising[53] and extreme events (e.g., storm surge) could become change in frequency and intensity in some parts of the Black Sea[54], in turn modifying long or short-term coastline readjustments that are impeded in an increasingly urbanized coastal environment[55]. Coastal erosion is also exacerbated by negative riverine sediment budget supply (sediment retention in dams, river's water diversion, and decreased rainfall) coupled with sand mining and poor-design coastal protection engineering. This vulnerable ecosystem needs long-term field experiments and monitoring programs to describe its dynamics[52,56] rather than one-shot descriptive database. However, we provided a first comprehensive digital inventory of Black Sea beaches at the basin

scale obtained by photo-interpretation of freely available remote-sensed images on the web[11]. The resulting database contains 1,228 beaches (polygons) enriched with descriptive attributes (e.g., photo-based visual estimation of the sediment type, presence of coastal defences, urban development, and presence of dunes). Digitized polygons covered a total coastline length of 2,042 km with an area of 224 km$^2$. Results showed that almost half of the Black Sea beaches are associated with artificial coastal protection schemes suggesting an already considerable beach erosion problem. Cross-shore beach retreat predictions due to sea level rise (SLR) were obtained by selecting the means of the lowest and highest projections estimated by an ensemble of six 1-D analytical and numerical models (Bruun[57], Dean[58], Edelman[59], Leont'yev[60], Xbeach[61], SBEACH[62]: more details in Allenbach *et al.* 2016). All models were launched for the full range of plausible coastal environmental conditions encountered at the basin scale (i.e., using different combinations of wave conditions, sediment grain sizes and slope) for eleven scenarios of SLR (ranging from 0.1 to 2 m). Black Sea beach erosion risk was then evaluated by comparing the coastal retreat predictions for three SLR scenarios (0.5, 0.82 and 1 m) with the maximum widths of each beach. Selected scenarios corresponded to the approximate mean and high estimates of the mean SLR scenarios of the IPCC[63] for the period 2081–2100. Finally, for a rapid assessment of the beach erosion risk, percentages of beach surfaces losses were calculated for the selected three SLR scenarios and added to the attribute table of the data set.

## Code availability

The climate gridded surfaces (1961–1990 and 2071–2100) were created using R Project for Statistical Computing, freely available under the GNU General Public License, and can be downloaded at https://www.r-project.org/. Multiple R-packages were used as necessary, while the main package used was the R-INLA (http://www.r-inla.org). Implementation details are available from the corresponding author upon request.

Demographic outputs have been generated using publicly available data. Algorithms are described in detail by de Bono *et al.*[5]

The Metronamica[6] (ML v.4.2.2) software used for landuse change modelling is a proprietary package and is not freely distributed.
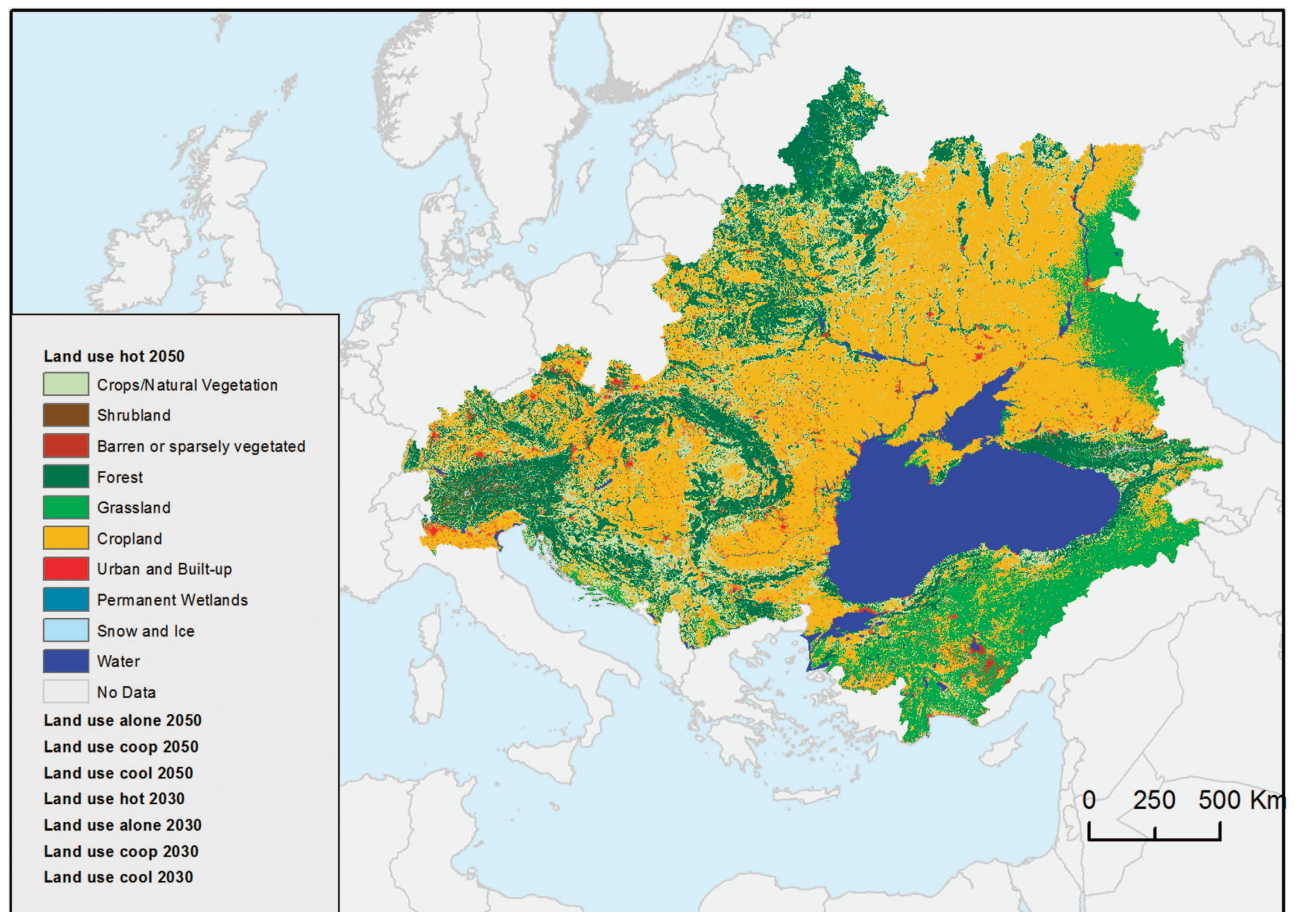


**Figure 2.** Landuse change for Scenario Black Sea Hot by 2050 *[blacksea.grid.unep.ch/maps/182]*.
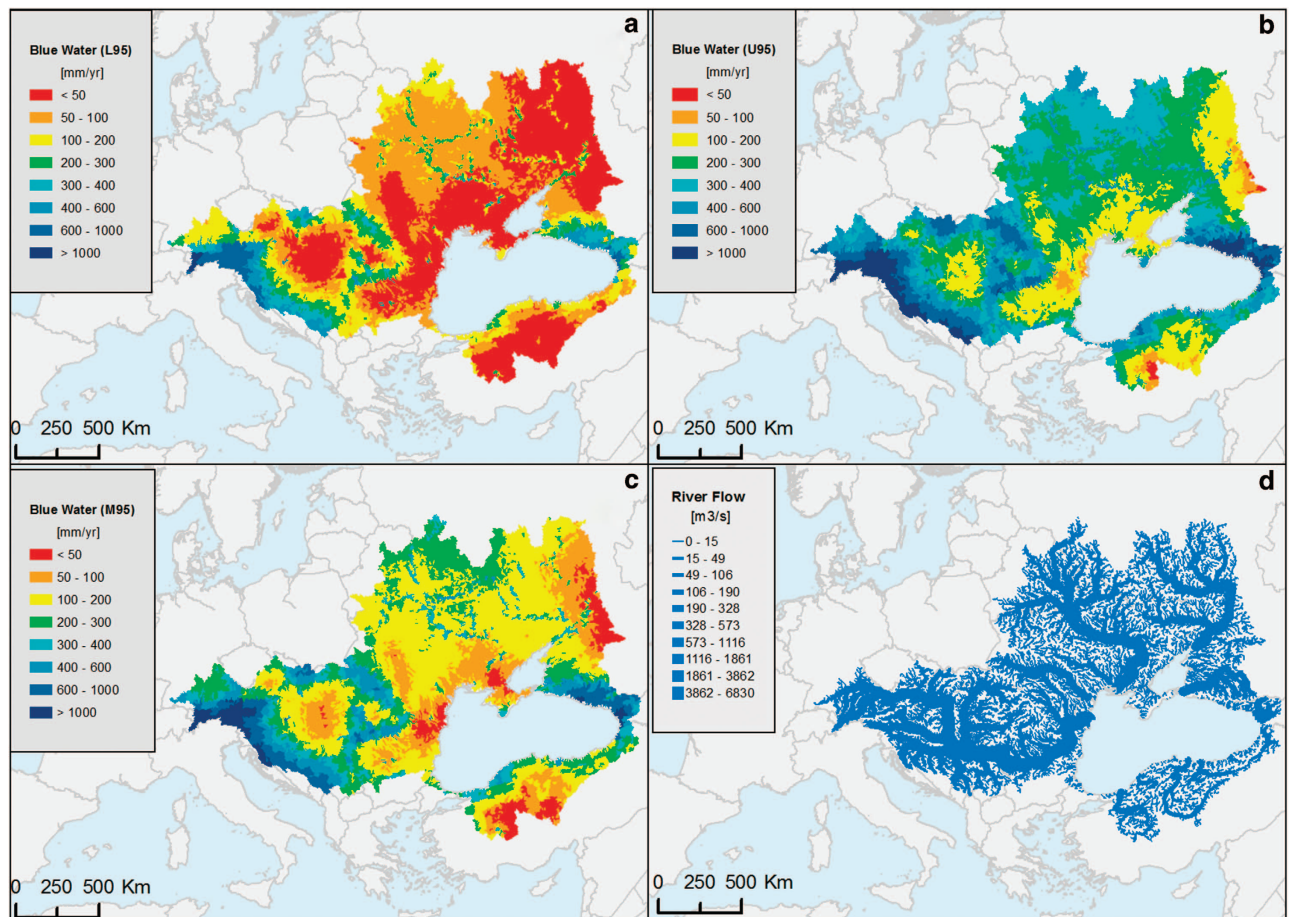
**Figure 3. Outputs of the SWAT hydrological model of the Black Sea catchment.** (**a**) Long-term annual average (1973–2006) of blue water expressed as lower (L95), (**b**) upper (U95) and (**c**) median (M95) of the 95% prediction uncertainty range. (**d**) River flow in m3/s. [blacksea.grid.unep.ch/maps/181].

SWAT (SWAT2009 rev, 528) is a public domain package used for hydrological modeling that is actively supported by the USDA Agricultural Research Service at the Grassland, Soil and Water Research Laboratory in Temple, Texas, USA. (http://swat.tamu.edu/software/swat-executables). SWAT-CUP was used for calibration of SWAT and is freely available from (www.neprashtechnology.ca).

Water vulnerability and water stress indicators were developed with scripts in R and are available from the first author of the study upon request.

The models used to predict cross-shore beach retreat induced by long and short-term sea level rise scenarios were developed with MATLAB software by the team of Prof. A.F. Velegrakis from the Department of Marine Sciences from the University of the Aegean and are freely available through the RiVAMP methodology training material[64].

The GeoNode open source platform version 2.0 has been used for sharing the enviroGRIDS data sets. This web-based application facilitates the visualization, download, sharing, and collaborative use of geospatial data through web services. It is freely available from the dedicated website (http://geonode.org/) under a GNU General Public License.

## Data Records

The entire data set is made freely available and accessible in various formats on the public domain. This includes a dedicated geoportal for data discovery, visualization and downloads, web services endpoints to ensure that data are accessible with interoperable standards (e.g., Open Geospatial Consortium (OGC) standards). Finally, a static copy of the data is made available at the data Dryad Repository (Data Citation 1).

The web-based interactive geoportal is accessible from the geoportal http://blacksea.grid.unep.ch. Users can select the Map tab to dynamically visualize and download datasets on: Landuse scenarios, Climate scenarios, Demographic scenarios, Hydrology, Water for agriculture vulnerability, Water scarcity, and Beach retreat.

Several vector formats are available such as: ESRI Shapefile, Geography Markup Language (GML) and Keyhole Markup Language (KML). Data in raster formats are available as Geotiff, ArcGrid, Gtopo30 and

| Subbasin | Sub basin ID |
|---|---|
| AreaC | Area drained by reach (km2) |
| Len2 | length of river segment (m) |
| Slo2 | slope of river segment (%) |
| Wid2 | width of river segment (m) |
| Dep2 | depth of river segment (m) |
| MinEl | minimum elevation of river segment (m) |
| MaxEl | minimum elevation of river segment (m) |
| Annual | Yearly average flow (m3/s) |
| Jan to Dec | Monthly average flow (m3/s) |
| Annual | Yearly nitrogen load (kg) |
| Jan to Dec | Monthly nitrogen load (kg) |

**Table 1. Attributes definition for rivers and subcatchments of the Black Sea catchment.**
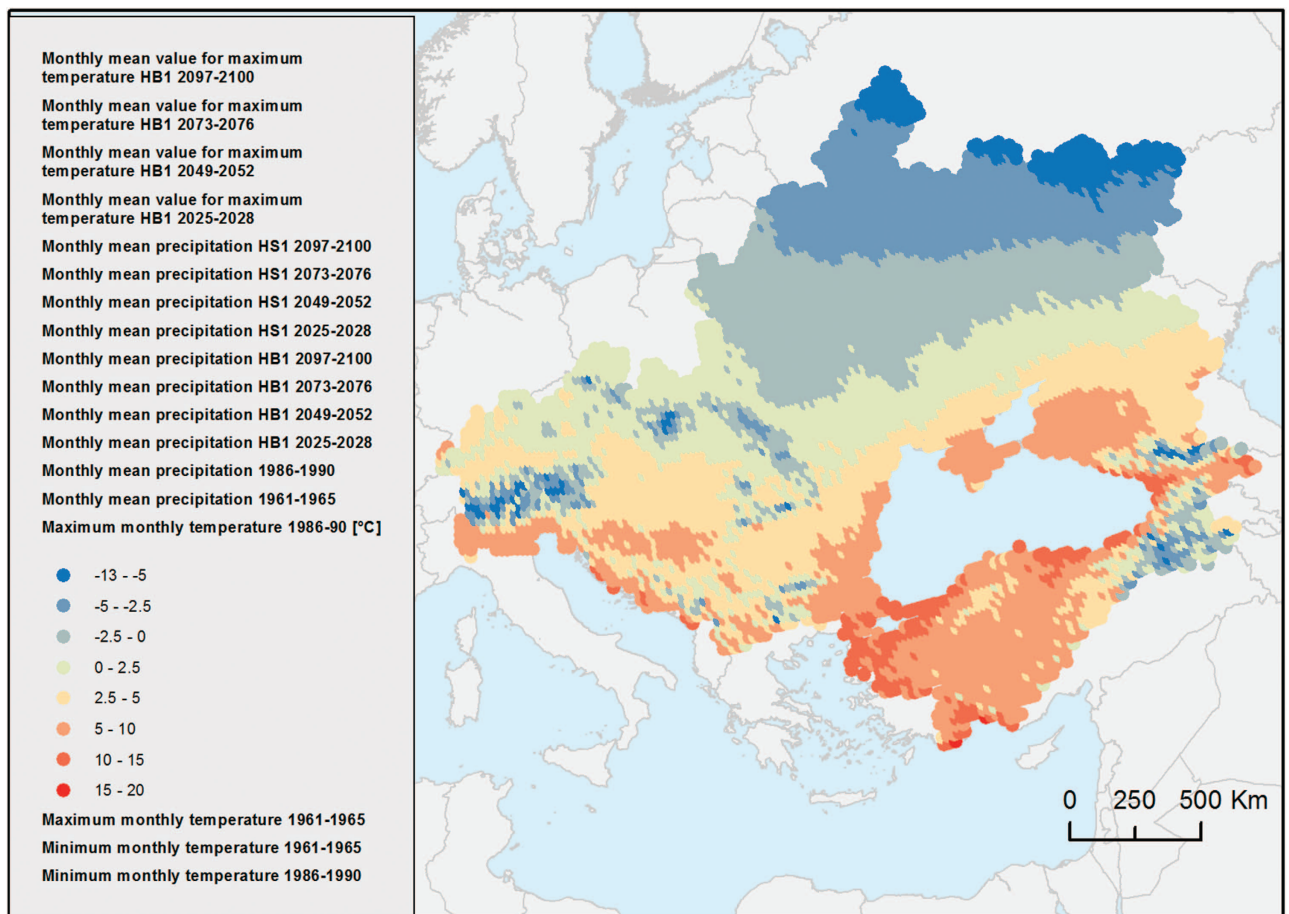


**Figure 4. Map of Maximum monthly temperature [blacksea.grid.unep.ch/maps/230].**

ImageMosaic. Metadata is available as: ATOM, Directory Interchange Format (DIF), Dublin Core, EBRIM, FGDC and ISO19115 formats.

In addition, several OGC-compliant web service endpoints are available to access graphical/cartographic representations of the data:

- Web Map Service (WMS): blacksea.grid.unep.ch/geoserver/wms?
- Web Map Tile Service (WMTS): blacksea.grid.unep.ch/geoserver/gwc/service/wmts?
- Vector data: Web Feature Service (WFS): blacksea.grid.unep.ch/geoserver/wfs?
- Raster data: Web Coverage Service (WCS): blacksea.grid.unep.ch/geoserver/wcs?
- Metadata: Catalog Service for the Web (CSW): blacksea.grid.unep.ch/catalogue/csw?

| Nuts2_Id | Nuts id |
|----------|---------|
| Name | Name of region |
| Y2011 | Yearly projected population for 2011 |
| Y2025 | Yearly projected population for 2025 |
| Y2050 | Yearly projected population for 2050 |

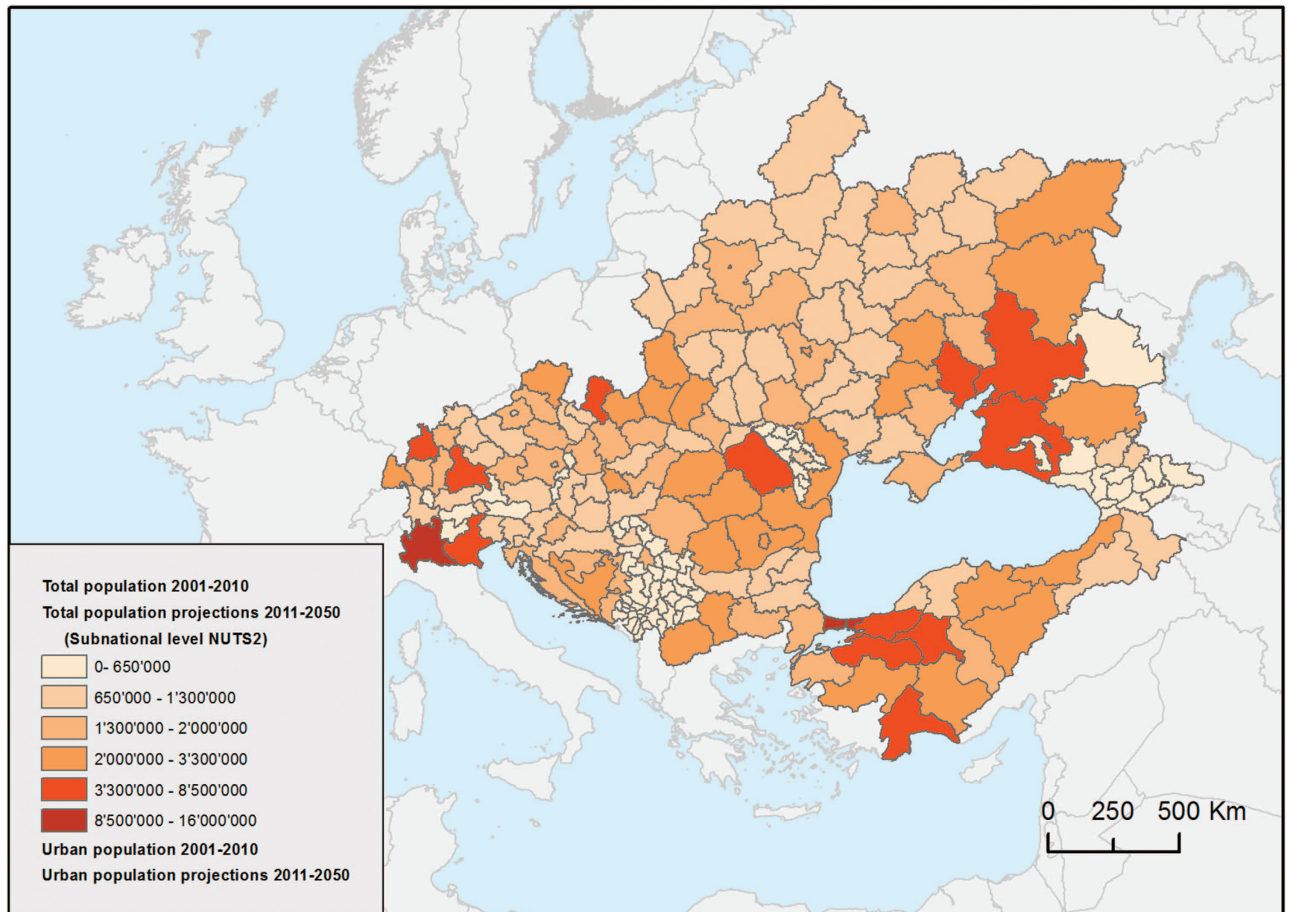Table 2. Attribute definition for population of the Black Sea sub-regional units.



Figure 5. Total human population predicted for 2050 [blacksea.grid.unep.ch/maps/228].

These interoperable web service endpoints enable user to seamlessly access and/or integrate these datasets in their desktop, web-based client, or own workflows. Moreover, the users are ensured to access the most up-to-date datasets. All the web service endpoints have also been registered into the GEOSS.

### Landuse change scenarios

The data consists of a set of nine raster maps representing the Landuse classification derived from MODIS in 2010 at a 500 m pixel resolution, and predicted with Metronamica until 2050 according to the HOT, ALONE, COOP and COOL scenarios (Fig. 2). Each map of LU is subdivided into 10 classes. Crops/natural vegetation, shrubland, barren or sparsely vegetated, forest, grassland, croplands and urban and built-up areas are changing over the years, whereas water, snow and ice, and permanent wetlands remain temporally unchanged (Landuse change scenarios, Data Citation 1).

### SWAT hydrological outputs

The long-term annual average of blue water, monthly average river flow (Fig. 3) and nitrogen loads were simulated by the Black Sea SWAT model between years 1970 to 2006. These maps are based on 12,982 subcatchments and their corresponding river reaches. The available attributes are described in Table 1. (SWAT hydrological outputs, Data Citation 1).
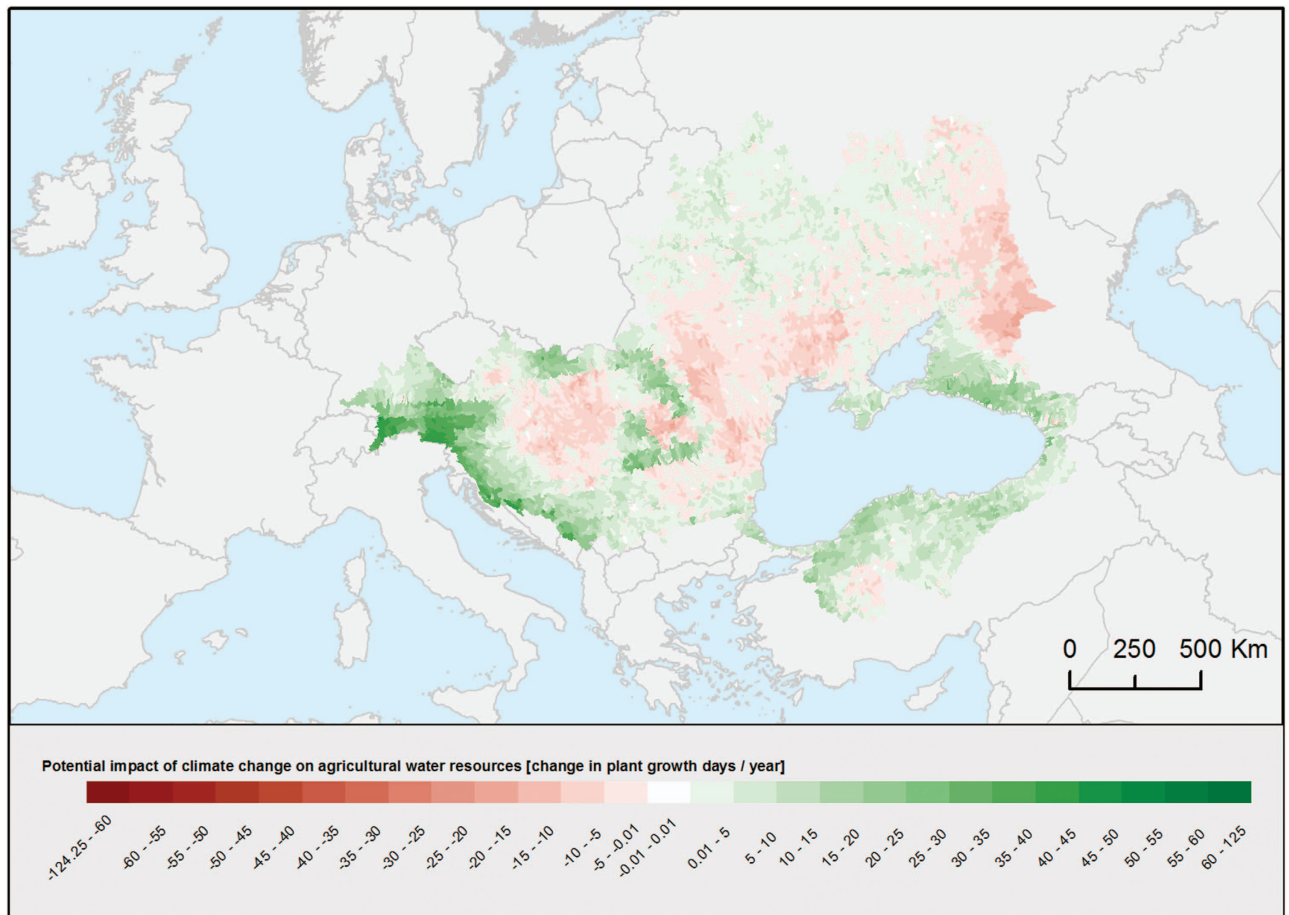
Potential impact of climate change on agricultural water resources [change in plant growth days / year]

**Figure 6.** **Potential impact of climate change on agriculture expressed as the change of annual plant growth days** *[http://blacksea.grid.unep.ch/maps/231].*

### Climate

This dataset consists of estimated values for monthly averages of maximum and minimum temperatures, as well as precipitation for the time period between 1961 and 1990, interpolated with the integrated nested Laplace approximations (Fig. 4). It also provides the estimated values of the monthly average for the HB1 and HS1 scenarios between 2071 and 2100, obtained through the Delta method. The dataset also includes the estimated values between 2025 and 2071 for both scenarios, obtained by means of a linear regression model (Climate, Data Citation 1).

### Demography

UN projection variants for population (Table 2), downscaled from national to regional level (NUTS2). Results include urban and total population trends over the period 2010–2050 for the 214 Black sea regions, consistent with BS HOT, BS ALONE, BS COOP and BS COOL scenarios (Fig. 5) (Demography, Data Citation 1).

### Water for agriculture vulnerability

This map provides an overview of the potential impact of climate change on agriculture over 12,982 subcatchments. It is expressed as change in the annual plant growth days (Fig. 6). Green indicates an increase of plant growth days due to fewer days of water stress and/or fewer days of temperature stress (positive potential impact). Red indicates a decrease in plant growth days due to more days of water stress and/or temperature stress (negative potential impact). The potential impact is very heterogeneous across the BSC. The adaptive irrigation capacity and the total agriculture vulnerability is also made available in this data set (Water for agriculture vulnerability, Data Citation 1).

### Water scarcity

This aim of this data set is to represent water stress in the BSC circa 2011, as shown in Fig. 7. Water stress is calculated according to the Water Stress Index (WSI) corresponding to the quotient of total withdrawals by surface water availability, both in $m^3$/yr. The scale of this index is as follows: $< 0.1 = $ No stress; 0.1 to 0.2 = Low; 0.2 to 0.4 = Moderate; 0.4 to 0.8 = High; $>0.8 = $ Very high stress. Several building
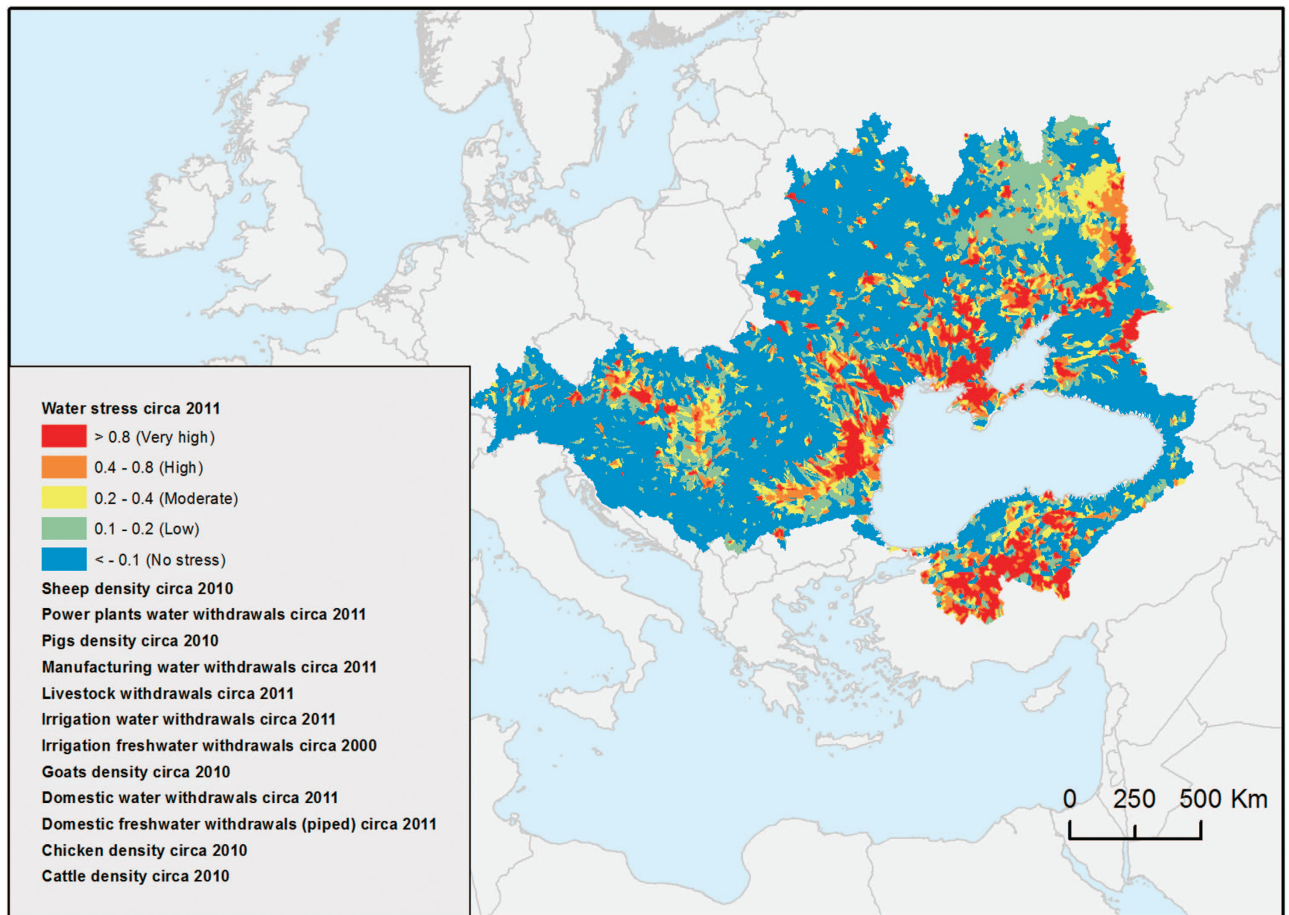
**Figure 7.** Water stress index by 2011 *[blacksea.grid.unep.ch/maps/232]*.

blocks used to assess water stress are also made available as geospatial layers in this data set (Water scarcity, Data Citation 1).

### Beaches
Maximum beach retreat (in percentage of the maximum beach width) predicted by the model ensemble for the Black Sea beaches under a 0.5 m SLR (coastal retreat estimated to 21.4 m) is shown in Fig. 8. The attributes made available for each of the 1,228 digitized beaches are described in Table 3 (Beaches, Data Citation 1).

## Technical Validation
### Climate change
The technical validation of the gridded surfaces for the maximum and minimum temperature, and precipitation, was made by means of cross-validation with other interpolation models (such as thin plate splines), for a random selection of 1 day per month for 30 years, with 10 and 50% of the weather stations used for testing. The validation of the climate scenarios obtained with the DM was made by comparing the monthly average and standard deviation results obtained for both HS1(A2) and HB1(B2), with the gridded surfaces obtained by applying the delta method to the HC1.

### Landuse
During the calibration, the model parameters were repeatedly adjusted until the goodness of fit was optimized. Validation of the calibrated model comprised statistical procedures (Kappa statistic, Klocation, and Fuzzy Kappa), visual evaluation and stakeholders' involvement in order to ensure its plausibility and accuracy.

### Population
Urban and rural downscaling of regional data were kept compatible with UN statistics for the period 2001 to 2010. Projections could not be validated but were in line with UN projections for the selected regions.
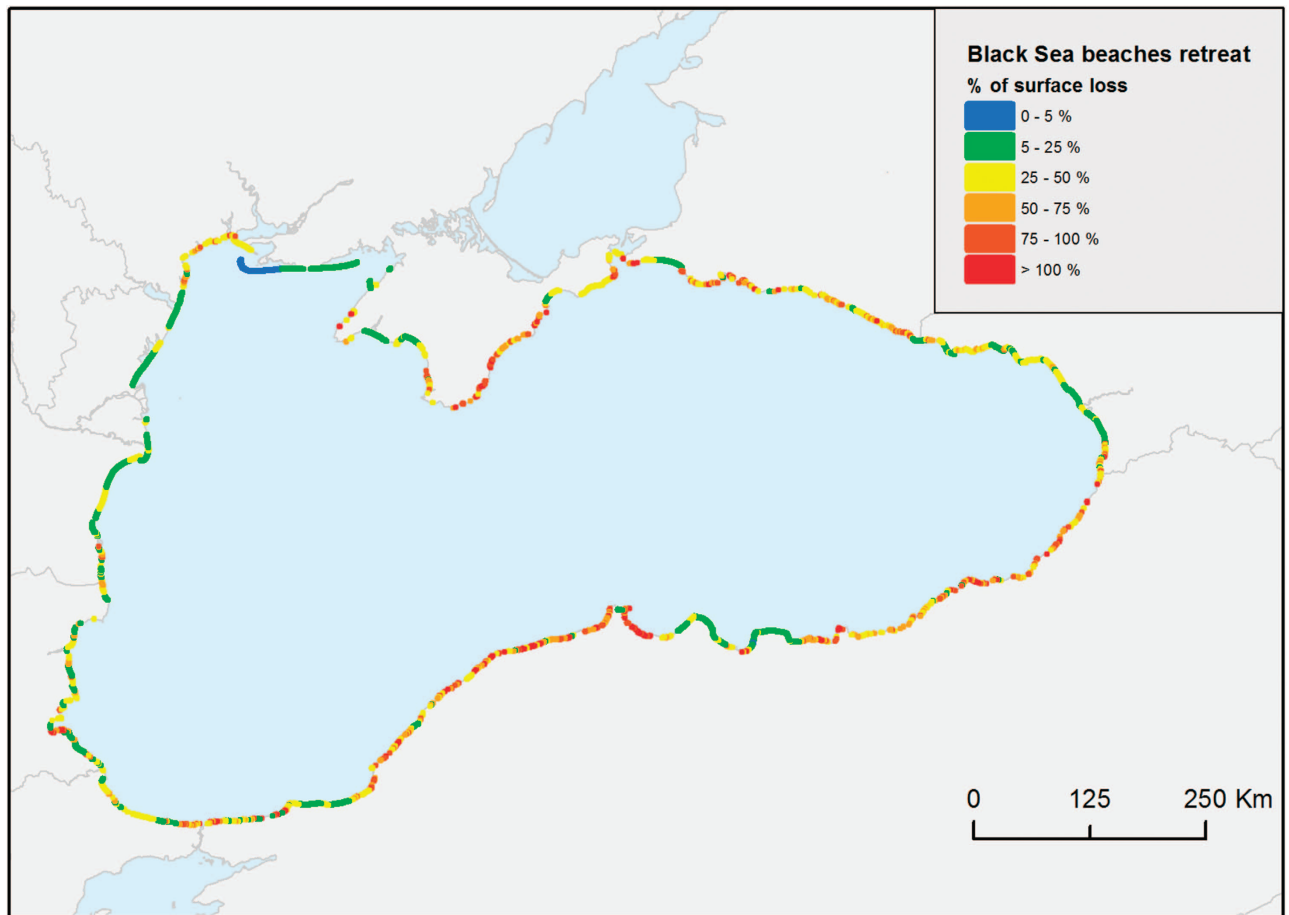
**Figure 8.** Predicted percentage of beach retreat of the Black Sea coasts [*blacksea.grid.unep.ch/maps/229*].

### Hydrology and SWAT outputs
The hydrological model was run at a daily time step (single run, 42 h). It was calibrated and validated for the period of 1970 to 2006 using observations on daily river discharges, river nitrate loads, and annual agricultural crop yields. Multi-objective, multivariable, multisite sensitivity, calibration, and uncertainty analysis were carried out using the Sequential Uncertainty Fitting program SUFI-2 (refs 36,37).

### Water for agriculture vulnerability
The main input for this analysis is the output of the SWAT hydrological model.

### Water scarcity
The main input for this analysis is the output of the SWAT hydrological model.

### Beaches
Digitalization based on photo interpretation, uncertainty analysis based on six different models of beach erosion. The uncertainty of the results can be appreciated by the difference between the max and min erosion percentages (SLR1max-SLR1min). Manual beach digitalization based on photo interpretation is subjective; accuracy was visually estimated at ±2 m. The lack of precise projection in the mosaicking images could induce shifts about 20 m between some images. Moreover, the completeness of the inventory is limited by the quality of available images (free of clouds and shadows).

## Usage Notes
A major caution should be taken concerning the resolution of the datasets. Indeed, the analysis was conducted using global/regional datasets, whose resolution are not relevant for local planning and should not be used for local decision-making. However, these datasets are intended to be used for national to regional assessments. They can be used also to compare countries or to help funding agencies to help them prioritizing environmental protection projects and investments.

In addition, some of these data sets result from models and depict plausible scenarios. Moreover, these scenarios are limited in spatial, temporal and thematic resolution due to present computing capacities.

| NAME | Code name of the beach, 3 first letters are the country followed by a clockwise dialing |
|---|---|
| ORIENT | Calculated orientation of the beach in degree |
| LENGTH | Calculated shoreline length of the beach in meters |
| MAX | Calculated maximum width of the beach in meters |
| MEAN | Calculated mean width of the beach in meters |
| AREA | Calculated area in square meters |
| DATE | Date of the satellite image used |
| SLR1min | minimum % of the beach surface loss by a coastal retreat of 4.1 meters induce by a SLR of 0.5 m |
| SLR1max | maximum % of the beach surface loss by a coastal retreat of 21.4 meters induce by a SLR of 0.5 m |
| SLR2min | minimum % of the beach surface loss by a coastal retreat of 6.9 meters induce by a SLR of 0.82 m |
| SLR2max | maximum % of the beach surface loss by a coastal retreat of 31.6 meters induce by a SLR of 0.82 m |
| SLR3min | minimum % of the beach surface loss by a coastal retreat of 8.5 meters induce by a SLR of 1 m |
| SLR3max | maximum % of the beach surface loss by a coastal retreat of 37.3 meters induce by a SLR of 1 m |

**Table 3.** **Attribute definition for beaches of the Black Sea coast.**

Black Sea beaches data sets should be used being aware of their uncertainties induced by the methodology and data availability. Moreover, the inventory does not reflect the dynamic environment of this ecosystem that is submitted to strong natural fluctuations. The available remote sensing images are not synoptic at the basin scale. They have been collected in different years and seasons (within the period October 2000—January 2011) and during different wave run-up conditions.

All layers are publicly and freely available using widely used interoperable standards and are compliant with the GEOSS Data Sharing Principles:

(1) There will be full and open exchange of data, metadata, and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation.
(2) All shared data, metadata, and products will be made available with minimum time delay and at minimum cost.
(3) All shared data, metadata, and products being free of charge or no more than cost of reproduction will be encouraged for research and education.

These principles are central to the GEOSS vision 'to realize a future wherein decisions and actions for the benefit of humankind are informed via coordinated, comprehensive and sustained Earth observations and information'.

By making these datasets available following GEOSS data sharing principles and using standardized web services, they can assist decision makers for environmental policies planning (e.g., INSPIRE Directive), environmental resource management (e.g., Water Framework Directive), support assessment processes of regional to local environmental agencies (like the International Commission for the Protection of the Danube River (ICPDR) or the Black Sea Commission (BSC)), and finally contribute to other web-based environmental platforms such as the DRDSI lead by JRC (http://drdsi.jrc.ec.europa.eu) or SCOPED-W[65].

## References

1. Lehmann, A. *et al.* Filling the gap between Earth observation and policy making in the Black Sea catchment with enviroGRIDS. *Environ Sci Policy* **46,** 1–12 (2015).
2. Myroshnychenko, V. *et al.* Environmental data gaps in Black Sea catchment countries: INSPIRE and GEOSS State of Play. *Environ Sci Policy* **46,** 13–25 (2015).
3. Giuliani, G., Ray, N. & Lehmann, A. Building Regional Capacities for GEOSS and INSPIRE: a journey in the Black Sea Catchment. *International Journal of Advanced Computer Science and Applications*, http://thesai.org/Publications/ViewPaper?Volume=3&Issue=3&Code=SpecialIssue&SerialNo=1 (2013).
4. Giuliani, G., Ray, N. & Lehmann, A. Grid-enabled Spatial Data Infrastructure for environmental sciences: Challenges and opportunities. *Future Gener Comp Sy* **27,** 292–303 (2011).
5. de Bono, A., Dao, H., Gago-Silva, A., Barbosa, A. & Mancosu, E. Proposed demographic scenario analysis and overview of driving forces and justification, model input parameters and allocation rules. EnviroGRIDS Delliverable 3.5 http://blacksea.grid.unep.ch/documents/143 (2011).
6. RIKS. *Metronamica-Model descriptions* - http://www.riks.nl/resources/documentation/Metronamica documentation.pdf (Research Institute for Knowledge Systems, 2011).
7. Mancosu, E. *et al.* Future land-use change scenarios for the Black Sea catchment. *Environ Sci Policy* **46,** 26–36 (2015).
8. Arnold, J. G. *et al.* Swat: Model Use, Calibration, and Validation. *T Asabe* **55,** 1491–1508 (2012).
9. Bär, R., Rouholahnejad, E., Rahman, K., Abbaspour, K. C. & Lehmann, A. Climate change and agricultural water resources: A vulnerability assessment of the Black Sea catchment. *Environ Sci Policy* **46,** 57–69 (2015).
10. Tallis, H., Kareiva, P., Marvier, M. & Chang, A. An ecosystem services framework to support both practical conservation and economic development. *P Natl Acad Sci USA* **105,** 9457–9464 (2008).
11. Allenbach, K. *et al.* Black Sea beaches vulnerability to sea level rise. *Environ Sci Policy* **46,** 95–109 (2015).
12. IMAGE. The IMAGE 2.2 implementation of the SRES scenarios: A comprehensive analysis of emissions, climate change and impacts in the 21st century - http://themasites.pbl.nl/models/image (Netherlands, 2001).

13. Ivanov, E. & Barbosa, A. Existing scenarios and data compilation on integrated scenarios using demographic, climatic, land cover from global and Black Sea Basin studies. EnviroGRIDS Deliverable 3.4 http://blacksea.grid.unep.ch/documents/148 (2010).

14. Fischer, G. et al. Global Agro-ecological Zones Assessment for Agriculture (FAO, 2008).

15. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978 (2005).

16. UNISDR. Global Assessment Report on Disaster Risk Reduction, http://www.preventionweb.net/english/hyogo/gar/report/index.php?id=9413&pid:34&pif:3 (2009).

17. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society 71, 319–392 (2009).

18. IPCC. Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change (2000).

19. Goyette, S. Existing data access and compilation on regional climate, historical records and prospects model. EnviroGRIDS Deliverable 3.2 http://blacksea.grid.unep.ch/documents/349 (2010).

20. UN/DESA. World Population Prospects The 2010 Revision Highlights (New York, 2011).

21. Eurostat. Regional population projections EUROPOP2008: Most EU regions face older population profile in 2030 (2010).

22. Neitsch, S. L., Arnold, J. G., Kiniry, J. R. & Williams, J. R. Soil and water assessment tool theoretical documentation version 2009 406 (Texas Water Resources Institute, 2011).

23. Rouholahnejad, E., Abbaspour, K. C., Srinivasan, R., Bacu, V. & Lehmann, A. Water resources of the Black Sea Basin at high spatial and temporal resolution. Water Resources Research 50, 5866–5885 (2014).

24. Haylock, M. R. et al. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. J Geophys Res-Atmos 113, 1–12 (2008).

25. NCDC. National Climate Data Center http://www.ncdc.noaa.gov.

26. Mitchell, T. D. & Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int. J. Climatology 25, 693–712 (2005).

27. CRU. CRU Time Series (TS) high resolution gridded datasets, http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__datent_1256223773328276 (2008).

28. Weedon, G. P. et al. Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century. J. Water and Global change (WATCH), Special Collection 12, 823–848 (2011).

29. FAO. The digital soil map of the world and derived soil properties (Rome, Italy, 1995).

30. NASA. (ed. USGS/Earth Resources Observation and Science (EROS) Center) (Sioux Falls, South Dakota, USA, 2001).

31. Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database, http://srtm.csi.cgiar.org (2008).

32. EEA. EEA Catchments and Rivers Network System (ECRINS) (Luxembourg, 2012).

33. Monfreda, C., Ramankutty, N. & Foley, J. A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. J. Global Biogeochemical Cycles 22, 1–19 (2008).

34. Gorgan, D. et al. Grid based calibration of SWAT hydrological models. Nat Hazard Earth Sys 12, 2411–2423 (2012).

35. GRDC. Long-Term Mean Monthly Discharges and Annual Characteristics of GRDC Station (Koblenz, Germany, 2011).

36. Abbaspour, K. C., Johnson, C. A. & van Genuchten, M. T. Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. J. Vadose Zone 3, 1340–1352 (2004).

37. Abbaspour, K. C. et al. Modelling hydrology and water quality in the pre-Alpine/Alpine Thur watershed using SWAT,. J. Hydrol. 333, 413–430 (2007).

38. Porter, J. R. & Semenov, M. Crop responses to climatic variation. Philos. Trans. R. Soc. London Ser. B., Biol. Sci 360, 2021–2035 (2005).

39. Dale, R. F. & Daniels, J. A. A weather-soil variable for estimating soil moisture stress and corn yield probabilities. Agron. J. 87, 1115–1121 (1995).

40. Jones, R. et al. in Definition and Scientific Justification for the Common Criteria (Technical Fiches), (eds van Orshoven J., Terres J.-M., Eliasson A.) (JRC European Commission, 2008).

41. Döll, P., Fiedler, K. & Zhang, J. Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. Hydrol. Earth Syst. Sci. 6, 4773–4812 (2009).

42. Hoekstra, A. Y., Chapagain, A. K., Aldaya, M. M. & Mekonnen, M. M. The water footprint assessment manual: setting the global standard (London, UK, 2011).

43. Tim Aus der Beek, Ellen Kynast & Martina Flörke. Modelling Current and Future Pan-European Irrigation Water Demands and Their Impact on Water Resources. Problems, Perspectives and Challenges of Agricultural Water Management (ed. Manish Kumar) doi:10.5772/29106 (InTech, 2012).

44. Fasel, M., Bréthaut, C., Rouholahnejad, E., Lacayo-Emery, M. A. & Lehmann, A. Blue water scarcity in the Black Sea catchment: identifying key actors in the water-ecosystems-energy-food nexus. Environ Sci Policy 66, 140–150 (2016).

45. Pastor, A. V., Ludwig, F., Biemans, H., Hoff, H. & Kabat, P. Accounting for environmental flow requirements in global water assessments. Hydrol Earth Syst Sc 18, 5041–5059 (2014).

46. FAO. FAOSTAT: Agri-environmental indicators - http://faostat3.fao.org/faostat-gateway (2013).

47. World Bank. World Development Indicators - http://databank.worldbank.org/data/home.aspx (2015).

48. Siebert, S. & Doell, P. The Global Crop Water Model (GCWM): Documentation and first results for irrigated crops https://www.uni-frankfurt.de/45217788/FHP_07_Siebert_and_Doell_2008.pdf (2008).

49. Hoogeveen, J., Faures, J. M., Peiser, L., Burke, J. & van de Giesen, N. GlobWat—a global water balance model to assess water use in irrigated agriculture. Hydrol Earth Syst Sc 19, 3829–3844 (2015).

50. Hoekstra, A. Y., Mekonnen, M. M., Chapagain, A. K., Mathews, R. E. & Richter, B. D. Global Monthly Water Scarcity: Blue Water Footprints versus Blue Water Availability. PLoS ONE 7 (2012).

51. Smakhtin, V., Revenga, C. & Döll, P. Taking into account environmental water requirements in global-scale water resources assessments. Comprehensive Assessment Research Report 2. 32 (Comprehensive Assessment Secretariat, Colombo, Sri Lanka, 2004).

52. Defeo, O. et al. Threats to sandy beach ecosystems: A review. Estuarine, Coastal and Shelf Science 81, 1–12 (2009).

53. Karaka, M. & Nicholls, J. Potential Implications of Accelerated Sea-Level Rise for Turkey. Journal of Coastal Research 2, 288–298 (2008).

54. Vousdoukas, M. I., Voukouvalas, E., Annunziato, A., Giardino, A. & Feyen, L. Projections of extreme storm surge levels along Europe. Climate Dynamics 47, 3171–3190 (2016).

55. Trifonova, E., Valchev, N., Andreeva, N. & Eftimova, P. Critical storm thresholds for morphological changes in the western Black Sea coastal zone. Geomorphology 143-144, 81–94 (2012).

56. Nicholls, R. J., Townend, I. H., Bradbury, A. P., Ramsbottom, D. & Day, S. A. Planning for long-term coastal change: experiences from England and Wales. Ocean Eng. 71, 13–16 (2013).

57. Bruun, P. The Bruun Rule of erosion by sea level rise: a discussion on large-scale two- and three-dimensional usages. *J. Coast. Res* **4,** 622–648 (1988).

58. Dean, R. G. Equilibrium beach profiles: characteristics and applications. *J. Coast. Res* **7,** 53–84 (1991).

59. Edelman, T. Dune erosion during storm conditions. in *13th International Conference on Coastal Engineering ASCE*, 1305–1312 (1972).

60. Leont'yev, I. O. Numerical modeling of beach erosion during storm events. *Coastal Eng* **29,** 187–200 (1996).

61. Roelvink, D. *et al. XBeach Model Description and Manual* (Deltares and Delft University of Technology, 2010).

62. Larson, M. & Kraus, N. C. *SBEACH: Numerical Model to Simulate Storm-Induced Beach Change* (U.S. Army Corps of Engineers, CERC, 1989).

63. IPCC. Climate Change 2013: The Physical Science Basis. *Summary for Policymakers* 1535 (2013).

64. Chatenoux, B., Peduzzi, P. & Velegrakis, V. *RIVAMP training on the role of Coastal and Marine ecosystems for mitigating beach erosion: the case of Negril Jamaica* (UNEP/GRID- Geneva, 2012).

65. Guigoz, Y., Lacroix, P., Rouholahnejad, E., Ray, N. & Giuliani, G. SCOPED-W: SCalable Online Platform for extracting Environmental Data and Water-related model outputs. *Transcations in GIS*, http://onlinelibrary.wiley.com/doi/10.1111/tgis.12227/abstract (2016).

## Data Citation

1. Lehmann, A. *et al. Dryad Digital Repository* http://dx.doi.org/10.5061/dryad.8t55d (2017).

## Acknowledgements

## Author Contributions

A.L.: conceptualized and coordinated the enviroGRIDS project, supervised data creation, wrote the paper, participated to quality control on data sets, and supervised the web platform. G.G.: conceptualized and supervised the SDI implementation, prepared maps, wrote the paper, participated to data creation, supervised data and metadata quality control, and reviewed the web platform. Y.G.: participated to data and metadata preparation and quality control, populated the web platform. N.R.: managed the enviroGRIDS project, coordinated grid computing workflows, and edited the paper. E.M.: landuse data creation. E.R.: hydrological data creation. K.C.A.: hydrological data creation. K.A.: beaches data creation. A.d.B.: demography data creation. A.G.S.: landuse, demography and climate data creation. R.B.: water-vulnerability data creation. M.F.: water-scarcity data creation. P.L.: participated to data creation, map preparation, and to data and metadata quality control.

## Additional Information

**Competing interests:** The authors declare no competing financial interests.

**How to cite this article:** Lehmann, A. *et al.* A web platform for landuse, climate, demography, hydrology and beach erosion in the Black Sea catchment. *Sci. Data* 4:170087 doi: 10.1038/sdata.2017.87 (2017).

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