1 Introduction

1.1 ISIMIP: General concept

ISIMIP provides a framework for the collation of a consistent set of climate impact data across sectors and scales. This framework will serve as a basis for model evaluation and improvement, allowing for improved estimates of the biophysical and socio-economic impacts of climate change at different levels of global warming. It also provides a unique opportunity for considering interactions between climate change impacts across sectors through consistent scenarios.

ISIMIP is intended to be structured in successive rounds, each having its own focus topics (see section 2.1) and focus regions (see section 2.2) that inform the scenario design. The main components of the ISIMIP framework are:

- A common set of climate and other input data which will be distributed via a central database;

- A common modelling protocol to ensure consistency across sectors and scales in terms of data, format and scenario set-up;

- A central archive where the output data will be collected and made available to the scientific community.

1.2 General remark regarding adaptation

As in the ISIMIP Fast Track, simulations should not be designed to describe the effects of different adaptation measures. In contrast, it is the aim to describe the impacts of climate change on different sectors under "present-day" management assumptions. There are individual exceptions to this general rule (such as "naturalized" runs within the water sector). These exceptions are clearly specified in the sector-specific scenario set-up of the simulations (see Chapter 7). Wherever such an exception is not mentioned please choose the "present day" option regarding management.

In particular, the historical validation runs should be "as close to the real historic conditions as possible" (i.e. to the extent to which this can be achieved without major model improvement).

2 General Design of ISIMIP2a

2.1 Focus topic

"Extreme events and variability" was chosen as the focus topic of ISIMIP2a, reflecting the interest of the community as well as stakeholders in investigating and improving the representation of variability and extreme events, in particular in impact models and along the entire modelling chain. Therefore, the model evaluation and validation task specified in this protocol is explicitly designed to evaluate the models' ability to reproduce observed historical variability, responses to extreme climatic events such as heat waves, droughts, floods, heavy rains and storms, and representation of extreme impact events (ISIMIP2 A see section 2.3).

Based on these evaluation exercises, modelling teams will have the opportunity to adjust model parameters, and implement necessary model improvements. As a next step (see the ISIMIP2b protocol), future projections will be performed using these improved impact model settings. To this end it is also intended to implement an improved bias correction method, which better preserves variability and extreme events, and to apply this method to the RCP climate projections. This exercise will also make it possible to estimate the extent of impacts that are *not* captured, or possibly over- or underestimated, in already-available impact projections from the Fast Track using the original model settings.

Note that the emphasis on this focus topic does not exclude some other work not directly related to extreme events and variability. For example, in particular for regional models and in new sectors, it may be important to first calibrate and investigate performance for average condition.

2.2 Focus regions

While only global impact models participated in the ISIMIP Fast Track, ISIMIP2 will bring together both global and regional modelling groups. A common set of focus regions will allow for a comparison between global and regional (i.e. sub-global) models, as well as an integration of regional impacts from different sectors. It was developed together with participants prior to and during the side event of the Impacts World 2013 conference. Figure 1 shows the set of focus regions.

This first selection represents a compromise between maximising coverage of climatic, environmental and cultural zones, and keeping the

associated workload feasible. The set of regions may be extended to cover new regions that are highly relevant with regard to climate change impacts but are currently under researched, or to accommodate new sectors. It may also be extended to include regions of specific interest to individual sectors, such as is already indicated by the dashed boxes in Figure 1, which are particularly relevant for the regional forest and water modellers. Participants working with (regional) marine ecosystem and fishery models are encouraged to focus on coastal regions adjacent to selected river basins (e.g. Rhine – North Sea; Mississippi – Gulf of Mexico; etc.).

All participants running regional models are asked to set up and run their models for as large a part of each focus region as possible, in order to obtain maximal spatial overlap with other models.
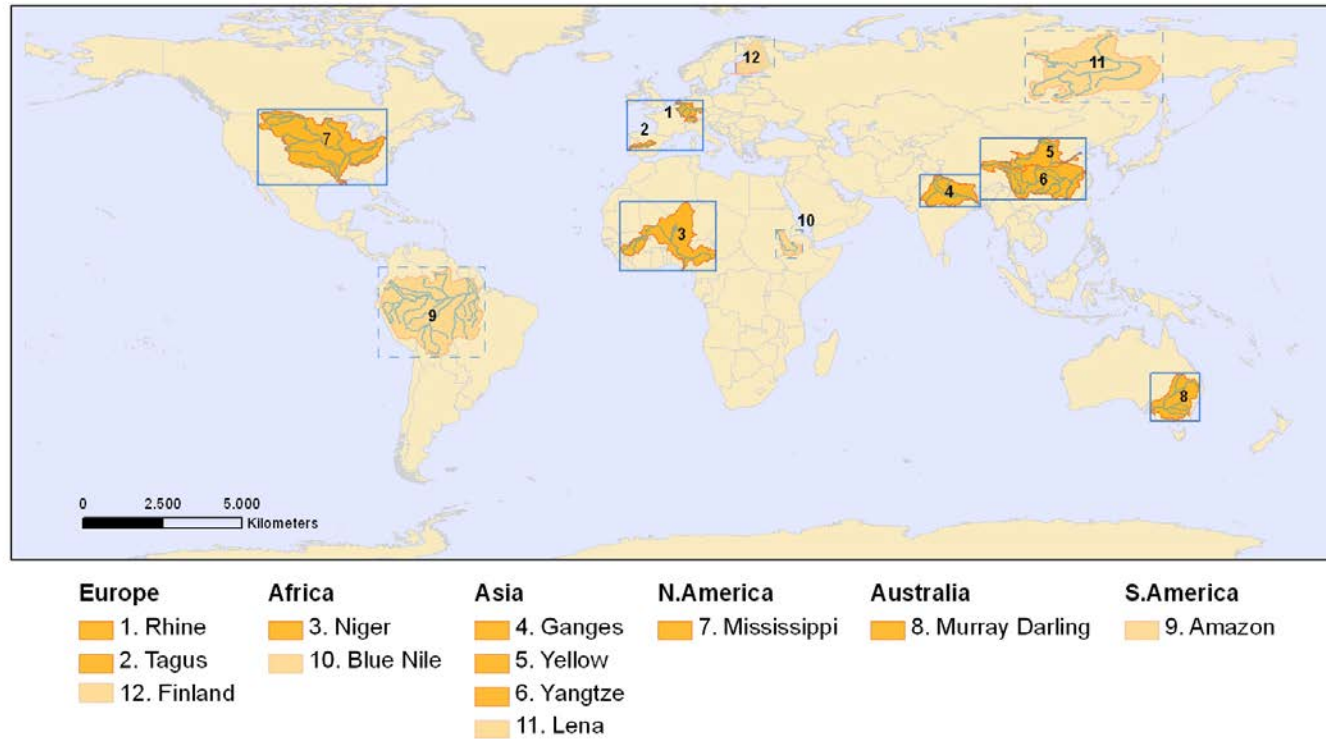
Figure 1: Focus regions of ISIMIP2. Solid boxes (centered on river basins marked in dark orange) indicate cross-sectoral focus regions. Dashed boxes and river basins/countries marked in light orange indicate possible sector-specific extensions (e.g. forests in Finland and the Amazon region, water in the Upper Amazon, Lena, and Blue Nile river basins).

Table 1: List of focus regions (as depicted in Figure 1).

Focus region (shortname)	Zonal extent	Meridional extent	River basin(s) (shortname)
Cross-sectoral focus regions			
North America (nam)	114°0'W– 77°30'W	28°30'N–50°0'N	Mississippi (Mississippi)
Western Europe (weu)	9°30'W–12°0'E	38°30'N–52°30'N	Rhine and Tagus (rhine)
West Africa (waf)	12°0'W–16°0'E	4°0'N–24°30'N	Niger (niger)
South Asia (sas)	73°0'E–90°30'E	22°0'N–31°30'N	Ganges (ganges)
China (chi)	90°30'E–120°30'E	24°0'N–42°0'N	Yellow and Yangtze (yellow, yangtze)
Australia (aus)	138°30'E–152°30'E	38°0'S –24°30'S	Murray Darling (murrydarling)
Sector-specific extensions			
Amazon (for forest and water sectors) (ama)	80°0'W –50°0'W	20°0'S –5°30'N	Amazon (amazon)
Finland (for forest sector) (fin)	21°0'E–32°0'E	59°30'N–79°30'N	-
Blue Nile (for water sector) (blu)	32°30'E - 40°0'E	8°0'N - 16°0'N	Blue Nile (bluenile)
Lena (for water sector) (len)	103°0'E - 141°30'E	52°0'N - 72°0'N	Lena (lena)

2.3 Structure of ISIMIP2a

Depending on whether or not a given model has already participated in the fast-track phase, ISIMIP2a will comprise the following steps:

Table 2: Simulation tasks in ISIMIP2a

	Fast-track models	**New sectors/models**
ISIMIP2a	Historical runs → validation and evaluation with focus on variability and extremes	Historical runs → validation and evaluation with focus on variability and extremes (this may, particularly for regional models, include calibration and validation for average conditions as a first step)
		Run impact models driven by ISIMIP fast-track climate data using fast track protocol ("catch-up runs"; with modifications for new sectors where necessary)

Important: New sectors/models, for "catch-up runs", please consult the Fast Track simulation protocol! See www.isimip.org > Protocol > Fast Track

3 Motivation of experiment design

This chapter provides a short description of the scientific rationale behind the design of each of the experiments in ISIMIP2a. The details of the experiments are further described in the remainder of the protocol.

The overarching objective of the historical validation experiment is to gain insight into the ability of current impact models to reproduce observed features of simulated variables, with an emphasis on (but not limited to) variability and extreme events. Simulations are designed such as to match historical conditions as closely as possible, within the limitations of e.g. availability of historical forcing data, variety of model formulations, and model development resources.

Conducting such an experiment within ISIMIP can facilitate a cross-sectoral synthesis of state-of-the-art modelling skills and help identify key opportunities for model improvement efforts. Examples of research questions that could be addressed include: How good is the current generation of impact models at reproducing the observed response to a certain climatic event (e.g. the Russian heat wave in 2010) across different sectors? What are the historic extreme events that are "extreme" across different sectors and are they reproduced by the different model types? Can yield losses in certain regions only be reproduced by accounting for the limited availability of irrigation water? In addition to the conventional, standardized validation exercises one could also look at the simultaneous effects of e.g. ENSO, as a major climatic driving force, across the various sectors, and compare with observations. One could also calculate drought indices from the water model results and test the influence of those droughts e.g. on yields. Another interesting question is the spatial and temporal coherence/concurrence of impacts in different sectors.

Four different observations-based historical climate datasets will be used to force impact models, to allow for a comparison of the different historical simulations. Each data set has its own strengths and weaknesses (e.g. regarding temporal extent, quality of particular variables, previous application within the modeling community) and generally represents a plausible reconstruction of the terrestrial climate of the past ~100 years. The different historical simulations will allow a systematic quantification of the effect of the choice of forcing data on impact model results; allow comparison to previous studies using either of these datasets; and provide an extensive data base for model evaluation and impact assessment, in particular with regard to the focus topic (e.g. a certain climatic extreme event could be better reproduced in one data set than in the other).

Note that one important criterion for the selection of climate forcing datasets was their temporal extent. Biomes models in particular need long spin-up, therefore it is crucial that the datasets cover a substantial extent of time before the start of the reporting period. Also, note that the GSWP3 (Global Soil Wetness Project 3) data set was only recently developed and a description paper is not yet available. This dataset is based on dynamical downscaling and thereby promises rather good quality in particular with regard to the resolution of the variability. A short description of the main characteristics of this dataset will be distributed by email and is also available upon request. Further details on the global dynamical downscaling method are given in ref (Yoshimura & Kanamitsu, 2008, 2013).

4 Common input data and settings for all sectors

This chapter describes climate forcing data and other input data that should be used by modelling groups in all sectors. See chapter 7 for additional, sector-specific input data and data for model validation. Note in particular that for some sectors the individual tasks (e.g. historical validation runs ISIMIP2a) may involve several different experiments with differences in input data and other settings. In this chapter we only describe the general rationale of the different tasks and list the common input data sets.

If you require additional input data that is neither specified in this chapter nor in chapter 7, please use your default data source. In case anything remains unclear please contact the coordination team or sectoral coordinators.

4.1 Historical validation runs

4.1.1 Atmospheric data

Please use the historical climate data listed in **Table 3** for the historical validation runs. The runs should start in 1971, or earlier if spin-up is needed (see below and section 6.1). All data will be available through the ISIMIP website, www.isimip.org. Separate historical simulations should be conducted with each of four different datasets, in the order indicated in the last column of **Table 3**. This is because each of the datasets has its own advantages and shortcomings, and thus by using several it will be possible to assess the influence of the choice of forcing data on the overall results. Moreover, this procedure serves the needs of the different participating sectors (e.g. data over ocean is needed for the fisheries sector), and facilitates consistency with other model intercomparison exercises (e.g. ISIMIP Fast Track; GSWP3). Modelling groups that cannot run all four datasets before the submission deadline should nonetheless begin in the order indicated, and inform the ISIMIP coordination team.

Table 3: Historical (atmospheric) climate data sets to be used in validation runs. All data sets contain the variables tas, pr, rhs, rlds, rsds, ps, wind, and partly also tasmin and tasmax. Note that simulations should be conducted with each of these datasets.

Dataset	Reanalysis	Years	Resolution, coverage	Bias target	Order of simulations; comments
PGMFD v.2 (Princeton) (Sheffield et al., 2006)	NCEP/NCAR Reanalysis 1	1901-2012	0.5° Land + Ocean	CRU, SRB, TRMM, GPCP, WMO validated against GSWP2	1
GSWP3 (Kim, n.d.)	20CR	1901-2010 (2011 and 2012 to be added soon)	0.5° Land + Ocean	GPCC, GPCP, CPC-Unified; CRU; SRB	2
WATCH (WFD) (Weedon et al., 2011)	ERA-40	1901-2001	0.5° Land	GPCC	3
WFDEI.GPCC (Weedon et al., 2014)	ERA-Interim	1901– 2012 (with 1901-1978 taken from WFD, WFDEI.GPCC data starting in 1979)	0.5° Land	GPCC	4 Combined forcing file (WFD 1901-1978, WFDEI.GPCC 1979-2012) will be provided by ISIMIP. **NOTE** a discontinuity in the data exists at the transition from WATCH to WFDEI, and results must be interpreted with caution.

Historical **CO2 concentrations** are also provided in the input data archive (historical_CO2_annual.txt). They are based on time series of global atmospheric CO2-concentrations from Meinshausen, Raper, & Wigley (2011) for 1765-2005 and Dlugokencky & Tan (2014) from 2006-2013.

Note that simulation results only need to be submitted for the reporting periods specified in section 6.1. The parts of the climate forcing data prior to the reporting period may be used for spin-up purposes and/or to facilitate further analyses. Simulation results for years outside the reporting period may still be submitted to the ISIMIP repository on a voluntary basis.

4.1.2 Oceanic data

See section 14.

4.1.3 Land-use/land-cover and soil data

We provide a time-varying historical land-use (LU) data set that should be used for the historical validation runs. The time series starts in 1700 and ends in 2012 (LandUse_DynamicMIRCA_1700-2012.nc) and should be applied for the spin-up as well as for the historical runs, as described above. This file includes constant LU at year-2000 level during the period 2000-2012. The HYDE3/ MIRCA data set does not contain any information about different types of natural vegetation. Modellers should follow their own default assumption regarding the partitioning of the area covered by natural vegetation between the different types; the coordination team will provide guidance if needed. Models that simulate their own natural vegetation should report that.

Table 4: Land-cover and soil data to be used in historical validation runs.

Dataset Mandatory:	Description	More info	Scale	Variables included
Historical cropland patterns "HYDE3/ MIRCA"	Combination of present-day (year 2000) crop and irrigated areas from MIRCA and trends of agricultural land from HYDE	Portmann et al. (2010) for MIRCA; Hoekstra, 1998 and Appendix of Fader et al. (2010, J Hydrol) for HYDE. Contact: LPJmL team at PIK for HYDE	Global 0.5°Annual	irrigated and rainfed crop areas for the following crop and vegetation classes: TEMPERATE_CEREALS, RICE, MAIZE, TROPICAL_CEREALS, PULSES, TEMPERATE_ROOTS, OIL_CROPS_SUNFLOWER, OIL_CROPS_SOYBEAN, OIL_CROPS_GROUNDNUT, OIL_CROPS_RAPESEED, SUGARCANE, OTHERS, PASTURES

Historical area covered by natural vegetation	Derived from "HYDE3/ MIRCA" as 1-(all agricultural area). Note that forestry is counted as natural vegetation because of lack of historical forestry data.		see above	fraction of grid cell covered by natural vegetation
Optional:				
HWSD or GSWP3 (upscaled version of HWSD)	soil map	See http://hydro.iis.u-tokyo.ac.jp/~sujan/research/gswp3/soil-texture-map.html, upscaling method A. Each model has the option to use their own soil datasets if preferred.	Global 30 arc sec (HWSD) *or* 0.5° (GSWP3) fixed	soil types: Sand, Loamy Sand, Sandy Loam, Loam, Silt Loam, Silt, Sandy Clay Loam, Clay Loam, Silty Clay Loam, Sandy Clay, Silty Clay, Clay, Ice

4.1.4 Socio-economic input

The historic population data are taken from the UN World Population Prospects 2010 revision[1]. The historic GDP data are from the World Bank World Development Indicators database[2] (see **Table 5**). Both datasets are available via www.ISIMIP.org. The reporting period for model simulations making use of this information is restricted to 1960-2010 (see section 5.1).

---

[1]   Documentation accessible at http://esa.un.org/unpd/wpp/Documentation/WPP%202010%20publications.htm

[2]   Accessible at http://data.worldbank.org/data-catalog/world-development-indicators

Table 5: Socio-economic data to be used for validation runs

Variable	Name	Unit	Frequency
Population (1950 - 2009), country level on 0.5°x0.5° grid	pop	number of people	annual time steps
GDP (1960 - 2010), country level	gdp	GDP PPP (can also be written as MER)	annual time steps

4.2 Catch-up runs

4.2.1 Atmospheric Data

Within the ISIMIP Fast Track bias-corrected climate data from 5 GCMs participating in the CMIP5 were provided. Data cover the time period from 1950 to 2099, i.e. the historical period, and future projections for all RCPs (RCP 2.6 (also called RCP3-PD), RCP 4.5, RCP 6.0, RCP 8.5). The same atmospheric data should be used for the Fast-Track catch-up runs.

The data have been bias corrected (see Hempel et al., 2013, ESD). They only cover the global land area (see section 4.2.2 for oceanic variables) on a 0.5°x0.5° grid, and are available via www.isimip.org.

For the catch-up runs the **CO_2 concentration** should follow the historical time series from 1951 onwards up until 2005. In 2006, switch to the RCP CO_2 concentration data.

Table 6 Climate input variables

Variable	Name	Unit (NetCDF format)	Frequency	Bias correction method [3]
Surface air temperatures T_{avg} T_{min} T_{max} (24 hour values)	tas, tasmin, tasmax	K (K)	daily & monthly	Mean and range matched to WATCH data (1960-1999)
Precipitation (sum of snowfall and rainfall)	pr	$Kg/m^2/s$ (kg m-2 s-1)	daily & monthly	Statistical distribution matched to WATCH data
Surface radiation (short- and longwave downwelling)	rsds rlds	W/m^2 (W m-2)	daily & monthly	Statistical distribution matched to WATCH data
Near-surface wind speed (east- and north-ward)	uas vas	m/s (m s-1)	daily & monthly	Statistical distribution matched to WATCH data
Near-surface wind speed (total)	wind	m/s (m s-1)	daily & monthly	Statistical distribution matched to WATCH data
Surface air pressure	ps	Pa (pa)	daily & monthly	Statistical distribution matched to WATCH data
Near-surface relative humidity	rhs	% (%)	daily & monthly	None
CO_2 concentration	co2	ppm (ppm)	annual	None

IMPORTANT: Whenever possible, monthly data should only be used for models that run with a monthly time step. Models requiring daily input data should use the daily data provided, rather than downscaling monthly data (i.e. **do not** use built-in weather generators to downscale monthly

data to a daily time step unless multiple iterations are required for model processes). Models requiring sub-daily resolution should apply native downscaling methods to the provided daily data.

4.2.2 Oceanic data (Fast-track Equivalent)

See section 14.

4.2.3 Land-use/land-cover data

For the **catch-up runs,** use LU patterns as specified within the Fast Track protocol (e.g. pure crop runs for the crop models, pure natural vegetation runs for the biomes models, and "present day" or naturalized runs for the water models).

4.2.4 Future Socio-Economic Data

Socio-economic data (population and GDP; see **Table 7**) are provided and include all Shared Socio-Economic Pathways (SSPs 1-5, developed as part of the ongoing socio-economic scenario development process[4]). In ISIMIP2a, SSPs 1-5 should be used for the "catch-up runs" where applicable (see Fast Track protocol).

The parameter space of the Shared Socio-Economic Pathways (SSPs) is spanned by challenges for adaptation and challenges for mitigation. The former essentially means levels of development and the latter can be translated to degree of fossil energy and resource intensity. For each SSP there will be exactly one population and GDP scenario.

Note that there is no one-to-one connection of a given SSP with a given RCP. Instead the combination of SSPs with different climate policy and also adaptation assumptions will result in different forcing levels. More information on the SSPs can be found at http://www.isp.ucar.edu/socio-economic-pathways.

---

[4]   See http://www.isp.ucar.edu/socio-economic-pathways for more information.

Table 7: Socio-economic variables provided via the ISIMIP database.

Variable	Name	Unit	Frequency	Spatial scale
Population	pop	number of people	5-year time steps[5]	country level (alpha-3 country code)
GDP	gdp	PPP; can also be written as MER		

4.2.5 Future fishing pressure

See section 14.

---

[5]   If annual data is required, please interpolate the data linearly.

5 General Spin-up procedures for historical and catch-up runs

5.1 Spin-up procedure for validation runs

Models requiring spin-up should make use of the portion of the respective climate forcing data set prior to the reporting period (i.e. prior to 1971. See Figure 3 for an illustration of the data to use for spin-up). For extending spin-up prior to 1901, a **spin-up data set** has been prepared by generating a "constant climate" data set of 120 years as a random combination of individual years from the 1901-1930 period of the individual data sets.

For models requiring **more than 70 years** of spin-up, we have provided a script that generates up to 2000 years of spin-up data, based on random drawings from detrended climate data from 1901-1930. After running the necessary length of spin-up data, continue with the respective climate forcing data set, starting in 1901. See section 4.1.1 for more information on the historical climate forcing data sets.

Use historical CO_2 concentration, as provided in the input data archive (use at most the years 1765-1970; filename: historical_CO2_annual.txt), for spin-up. For example, when using the extended climate spin-up data set (i.e. generated using the script provided), use CO_2 concentration for the period 1780-1900. When using a longer spin-up period that extends back further than the CO_2 concentration data, please keep CO_2 concentration constant at 1780 level until reaching the year corresponding to 1780.
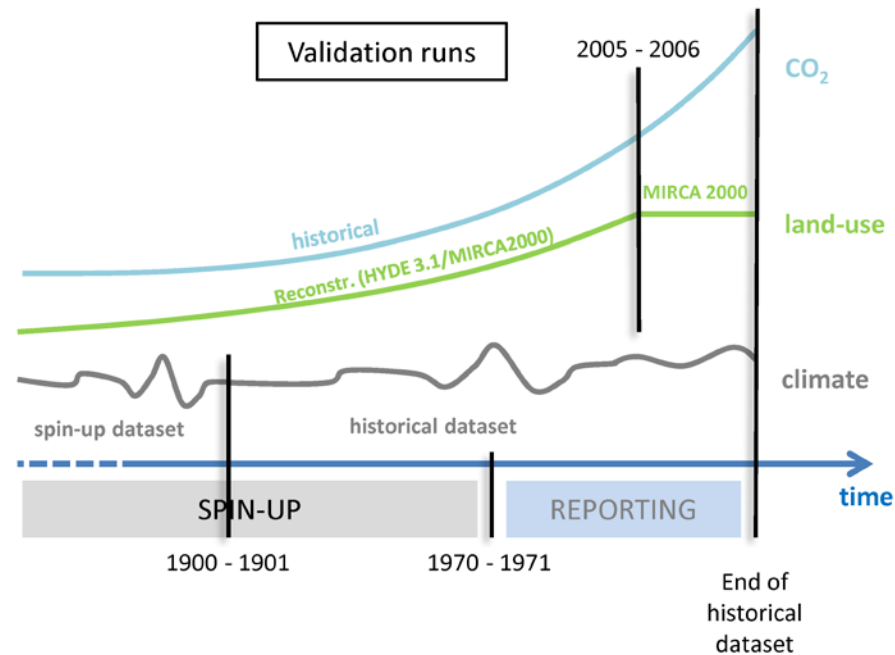
Figure 2: Input data scheme for historical validation runs. Information shown here holds for all sectors, but further details may vary across sectors – see Chapter 7.

5.2 Spin-up procedure – catch-up runs

For the catch-up runs a potentially required spin-up should be identical to the fast track spin-up (see Figure 4). With respect to the atmospheric data please apply the following procedure:

A climate **spin-up data set** has been prepared by de-trending the bias-corrected GCM data between 1951-1980, and is normalized to reflect a climate representative of the 1950 conditions. If your model requires **more than 30 years** of spin-up, the 30-year spin-up data set can be replicated and assembled back-to-back to obtain longer data sets; in this case, **make sure to reverse the order of years in alternate copies** of the 30-year

period, in order to minimize potential discontinuities in low-frequency variability. After running as many instances of the 30-year spin-up data set as required, continue using climate input data from the appropriate GCM, from 1951 to 2005 (the RCP projections start in 2006).

For the catch-up runs the **CO_2 concentration** of the spin-up should be fixed at 1950 levels for the (potentially copied) 30 year period and follow the historical time series from 1951 onwards, up until 2005. In 2006, switch to the RCP CO_2 concentration data.
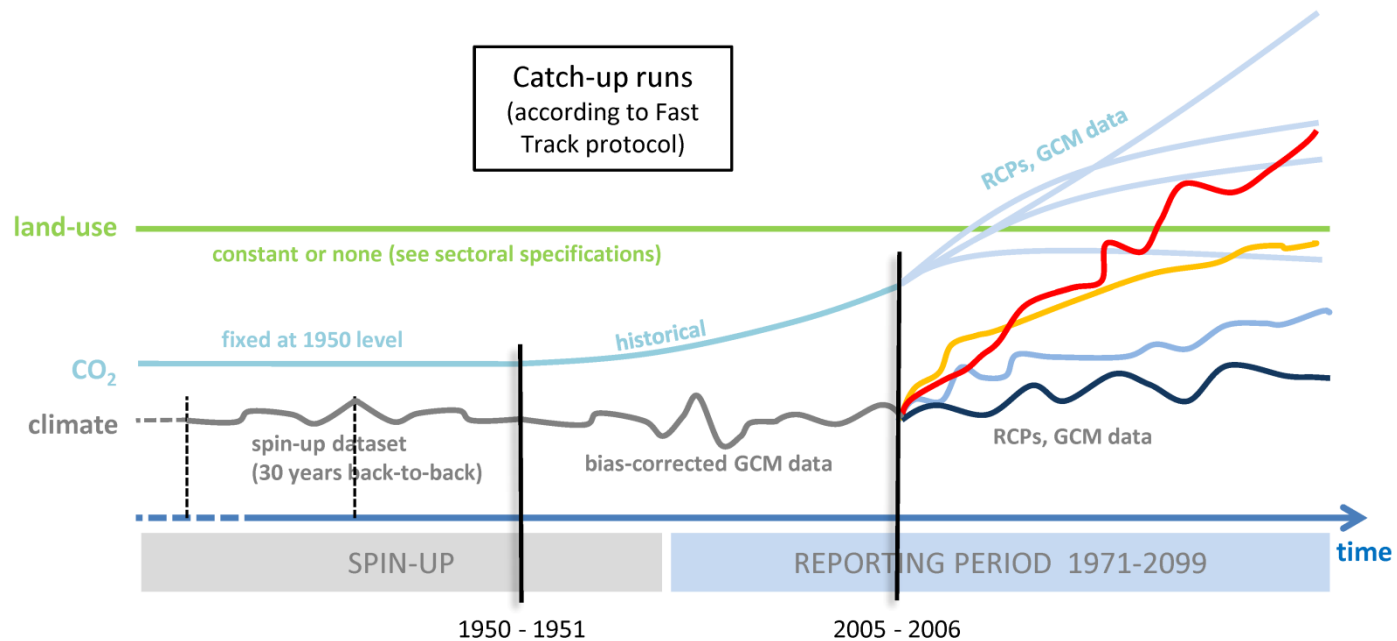

Figure 3: Input data scheme for "catch-up" runs, equivalent to the projections done in the ISIMIP Fast Track.

6 Reporting model results

6.1 Reporting and reference time periods

Table 8 Reporting and reference time periods

Item	Time period	Comments
Reporting time for historical validation runs independent of socioeconomic information about population and GDP	1971-end year of the historical climate input dataset	For transient runs. Please provide output data for the entire period. For Forest models: Please note that the reporting period starts with the first year for which stand data for model initialization is possible. Please see the introduction of the forest protocol (section 7.3.1) for further information.
Reporting time for catch-up runs	1971-2099	For transient runs. Please provide output data for the entire period.
Time slices for catch-up runs	present-day 1980-2010 (middle year 2000) (for 2006-2010 data, please use the RCP 4.5 run of each GCM)	Time slices should be provided if your model is too computationally expensive to do the transient runs, and/or if the application of transient land-use (LU) changes is not feasible. In this case, LU patterns of the corresponding middle year should be applied for the entire time slice.
	near-term 2005-2035 (middle year 2020)	
	mid-century 2035-2065 (middle year 2050)	
	end-of-century 2069-2099 (middle year 2085)	

Present-day population/GDP	2000	
Present-day land-use	2000	
Reference period	1980-2010	The reference period (and reference year) is defined only for post-processing purposes, to describe future projections relative to present-day conditions. Please fill in the years from 2005 to 2010 **by the associated** RCP (depending on the future simulations you consider). When reporting your output data, always provide absolute values.
Reference year	2000	

6.2 Convention for file names and formats

NOTE: For 'catch-up' runs, please use the (slightly different) conventions specified in the Fast Track protocol (see www.ISIMIP.org).

6.2.1 File names

One variable should be reported per file. The file names should follow this convention for the **historical validation runs**:

<modelname>_<obs>_<bias-correction><clim_scenario>_<socio-econ-scenario>_<sens-scenarios>_<variable>_<region>_<timestep>_<start-year>_<end-year>.nc4

where the parts in brackets should be replaced with the appropriate specifier, and where more than one specifier applies, the order in **Table 9** should be respected. Specifiers may be dependent on the sector. The associated sectors are given in brackets in **Table 9**. If there is not specifier for the associated sector it should be omitted from the model name. For example (cf. chapter 7):

lpjml_gswp3_ nobc_hist_pressoc_co2_qtot_global_annual_1971_2012.nc4 for the water sector and

lpjml_gswp3_ nobc_hist_firr_co2_default_yield_mai_global_annual_1971_2012.nc4 for a maize crop model run.

If there is more than one indicator specifying the sensitivity experiment (e.g. CO2, full irrigation and default model parameters for crop models simulations, please specify them in the order of the listing in **Table 9**). For example:

lpjml_gswp3_nobc_hist_pop.ssp2_gdp.ssp2_noco2_qtot_global_annual_2005_2099.nc4 for the water sector and

lpjml_gswp3_nobc_hist_noco2_firr_default_yield_mai_global_annual_2005_2099.nc4 for a maize crop model run.

Use **only lowercase** letters in the file names (necessary for some NetCDF viewers). In case of any questions, please contact the coordination team (info@isimip.org) before submitting files.

Table 9: File name specifiers for output data

Item	Possible specifiers (use lowercase letters only!)	Explanation
<modelname>	*(your model name as registered with ISIMIP)*	Name of the impacts model
<GCM/obs>	hadgem2-es, ipsl-cm5a-lr, miroc-esm-chem, gfdl-esm2m, noresm1-m (all sectors)	Name of the General Circulation Model providing the climate forcing data for the catch-up runs.
	gswp3, princeton, watch, wfdei (all sectors)	Name of the observations-based dataset providing the climate forcing data
	localclim (Forests)	For locally observed weather data from a weather station or similar which is in direct proximity of the forest stand and has been used for detailed model evaluation runs.
<bias-correction>	wfd	Indicates that the data were bias corrected to the Watch Forcing Data set – appropriate for simulations driven by GCM data.
	nobc	Indicates that no bias correction was performed on the climate data (e.g.

		observational data or ocean data)
<clim-scenario>	rcp2p6, rcp4p5, rcp6p0, rcp8p5	CO2 concentration scenario (RCP)
	hist	Historical climate information
	presclim	Present-day climate (1980-2010). For agro-economic models it means averaging biophysical information from the present day period of the biophysical model runs (Note: They differ from GCM to GCM). This run is essential to quantify the pure climate induced changes as the difference between the RCP and the presclim run.
	noclim	Demand-side runs; to compute future non-agricultural water use/ withdrawal/ consumption – only for appropriate models (e.g. WaterGAP, H08). Use population and GDP according to SSPs (as provided). For hydrological models only.
<socio-econ-scenario>		For validation experiments
	nosoc	Naturalized runs (no human impact). No irrigation. No population and GDP data prescribed. For forest models, this means running the model without any management (but nitrogen deposition should be included). If your model includes natural regeneration, please only regeneration those species previously present on the plot.
	pressoc	Present-day human impact runs: only climate varies; keep all other settings (population, GDP, land-use, technological progress etc.) constant at year 2000 values. This run will be used to quantify adaptation pressure under current socio-economic conditions. For water models that do not run separate irrigation

		sensitivity scenarios, pressoc includes present-day irrigation.
	varsoc	In addition to climate population, GDP, land-use, technological progress etc. varies over the historical period. For forest simulations this means: Manage forests according to historical management guidelines without species change. (see **Table 25**-Table 27)
	nat	A natural vegetation only run without any land-use pattern (optional for biomes models). This is a reference run to separate fluxes from natural vegetation and agriculture in runs with historic land-use. It is like the nosoc run but without land-use. If your model does not distinguish between natural and managed land, the "nat" run will be identical to the "nosoc" run.
<socio-econ-scenario>	For future runs	
	pop.<x>_gdp.<y> , where x = ssp1, ssp2, ssp3, ssp4, or ssp5 and y = ssp1, ssp2, ssp3, ssp4, or ssp5 (water, biomes) ssp<x>_pressoc (water, biomes)	Shared socio-economic pathways for future projections: population <x> and GDP <y> according to different SSPs population according to SSPx; year 2000 GDP
<sens-scenario>	A combination of the following settings:	
	co2 (only for future runs; water, biomes, crop models)	Transient CO_2 concentration (taken from the RCP and historical data) for CO_2 fertilization effects. If your model does not implement CO_2 fertilization using transient CO_2 concentrations, please use your own implementation and include that in the reporting.

	noco2 (only for future runs; water, biomes, crop models)	CO_2 concentration fixed at present-day value, i.e. run with transient historical CO_2 up to the year 2000 and keep CO_2 fixed at 369ppm thereafter. Sensitivity experiment: only applies to models that take CO_2 into account.
	co2const	Keep CO_2 concentration constant at 1971 level. For spin-up, use time-varying, historical CO_2 concentration until 1970.
	firr (for crop models)	Full irrigation, i.e. until the soil is saturated.
	noirr (for crop models)	No irrigation. Both irrigation variations are required for the agro-economic models.
	harmnon (for crop models)	full fertilizer run, i.e. ensuring no N constraints, harmonized sowing and harvesting dates
	fullharm (for crop models)	full harmonization with regard to fertilizer application rates, sowing, and harvesting dates
	default (for crop models)	"best guess" representation of historical conditions regarding fertilizer application rates, sowing and variety settings that best reproduce given harvesting dates.
\<variable\>	*(variable names listed in the sector specific output tables, see chapter 7)*	Output variable of the impact model. The identifier can also be used for information about the plant functional type (pft) in the biomes and permafrost sectors. The pft/species naming is model-specific and hence has to be reported in the impact-model database (see section 6.3 and www.isimip.org/impactmodels). In the forestry sector, the identifier might contain information about the tree species. The species names codes are

		listed in **Table 29**.
<region>	global	for global simulations
	short name of focus region as in Table 1	simulations covering a focus region
	common name of river basin as listed in Table 1 or location	for simulations covering one of several basins or a single location within a focus region
	forest site name as defined in **Table 24.**	For simulations of the regional forest models.
<timestep>	monthly, annual, decadal, daily	
<start-year>_<end-year>		For the forestry simulations, the full simulation period can be covered in one file. E.g. 1952-2012

6.2.2 Format for gridded data

Gridded data should be returned in NetCDF4 format with a compression level of at least 5. It is important that you comply precisely with the formatting specified below, in order to facilitate the analysis of your simulation results in the ISIMIP framework. Incorrect formatting can seriously delay the analysis. For questions or clarifications, please contact the ISIMIP coordination team. Further information and instructions follow in this section.

Global data are to be submitted for the ranges **-89.75** to **89.75** degrees latitude, and **-179.75** to **179.75** degrees longitude, i.e. 360 rows and 720 columns, or 259200 grid cells total. Please report the output data row-wise starting at 89.75 and -179.75, and ending at -89.75 and 179.75. The standard horizontal resolution is 0.5x0.5 degrees, corresponding to the resolution of the climate input data. Submitting data at lower resolution than 0.5x0.5 degrees is only encouraged in exceptional cases; in those cases the above numbers will change accordingly (e.g., a 1x1 degree grid would have 180 rows, from 89.5 to -89.5 degrees latitude).

Antarctica and Greenland do not have to be simulated. If you are limited by data (e.g. soil data) you can also reduce the latitude range of your simulations, however, the **minimal latitude range** to be simulated is -60 to + 67 degrees. **Important:** When *reporting* results the whole grid as specified above should be reported – pixels you did not simulate should be filled with the missing value marker (1.e+20f).

Regional model teams should interpolate their output data to the same, common 0.5x0.5 degree grid as described above, and submit in the same NetCDF file format. Each file should **cover the entire globe** (even though the filename should contain the name of the region), with any grid cells outside the simulated region to be filled with missing values (1.e+20f). This will not lead to significantly larger files as long as NetCDF compression is used. **Exception:** Single (one-point) timeseries do not have to be embedded into the 0.5x0.5 degree grid, but should be reported as NetCDF 4 files with the coordinates of the point included in the header information.

Note that submitting results in this format is important in order to facilitate comparison among different models and between global and regional scale. The **ISIMIP coordination team will be glad to assist** with the preparation of these files if necessary. **In addition** to the global file, regional model teams may submit a second file containing the output data in their default format. This may be e.g. on a finer resolution than 0.5°, on a non-regular grid, etc.

Please note the corresponding file naming conventions above.

Latitude, longitude and time should be included as individual variables in each file.

6.2.2.1 NetCDF files

Table 10 Naming and format conventions for NetCDF files

Dimension	Name	Unit
X	Lon	degrees east
Y	Lat	degrees north

T	Time	<time steps> since 1901-01-01 00:00:00 (where <time steps> is replaced by days, months, etc., according to the time step the data is reported on) Note: crop models use a different time step; see section 7.4.2.
missing value	1.e+20f	

To facilitate proper formatting of your output files, shell scripts are provided on the ISIMIP website (www.isimip.org > Getting Started > ISIMIP2a Simulation Protocol) that produce "empty" NetCDF files with the correct format and header information. Please use these scripts to write your output data into NetCDF files. A detailed explanation of how to do this is given here and can also be found on the website.

The NetCDF format should adhere to the settings in the shell (bash) script "isimip2_output_<time_step>.sh". Scripts are available for the different time steps: daily, monthly, annual or decadal.

In order to create an empty netcdf file with the correct filename and the required metadata in the header, you must run the script appropriate to the time step of the output data you want to store (e.g. daily data), using the following command:

> output_<time_step>.sh x1 x2 x3 x4 x5 x6 x7 x8

The arguments are given in **Table 11**.

Table 11 Arguments for script to generate netcdf file

Arg	Details	example
x1	The first year of data to be stored in the netcdf file (e.g. '1980'). Each file should contain data for one variable only, and for a time interval as specified in sections 6.1 and 6.2.4 (e.g., monthly data for 1980-1989).	'1980'

x2	Variable name stored in the netcdf file. Please adhere to variable names given in protocol.	'pr'
x3	Long variable name.	'precipitation'
x4	Units of the variable following the CF metadata	'kg m-2 s-1'
x5	transient or time-slice run	'time-slice'
x6	Further room for comments (optional).	'includes rain and snowfall' ' ' if empty
x7	Your institution	
x8	Your email address	

Here is an example how to run the script in a bash shell:

> ./output_daily.sh '1980' 'pr' 'precipitation' 'kg m-2 s-1' '' 'includes rain and snowfall' 'PIK' 'Info@isimip.org' 'im1_hadgem-es_rcp85_ssp2_noco2'

Note: The bash shell is standard on most Linux and Mac systems. For Window systems Cygwin can be used.

Once you have created your NetCDF file, you can check the metadata by running the command "ncdump –h"; an example output is given in the Fast-Track protocol.

6.2.3 Format for non-gridded data

Data that is not defined on a grid, such as point-based data e.g. for particular gauges, or data for world regions, should nonetheless be reported in NetCDF format, e.g. as a separate file for each gauge or region, each file containing a single time series. The ISIMIP coordination team will assist with the preparation of these files where necessary. **It is important that all ISIMIP results exist in NetCDF format**, in order to be compatible with the structure and functionalities of the ESGF repository.

6.2.4 Time intervals

Please submit your output data in chunks according to the **Table 12**, depending on whether you are reporting a transient experiment or a time-slice experiment, and depending on the time step of the output variable you are reporting (the requested time step for each variable is listed in the sector-specific tables in Section 7).

Table 12 Definition of time intervals

Daily time step	Output files should cover **10 years** each (e.g. 1971-1980).
	Exceptions: The period 2001-2010 should be reported in **two separate files**, broken down in the same way as the respective input data (i.e. in the case of most GCMs, both files will cover 5 years each: 2001-2005 and 2006-2010; in the case of the HadGEM2-ES model, the files will cover 2001-2004 and 2005-2010, respectively).
	The period 2091-2099 should be reported in a file covering **9 years**
monthly, annual, or decadal time step	Output should be reported in one single file per experiment. In experiments involving future projections (catch-up runs) the historical period should be separated from the RCP period (the separation being between 2005 and 2006 except for HadGEM2-ES).

6.2.5 Submission

Data should be submitted to a dedicated file system on a central server located at DKRZ Hamburg. Access to this server will be possible via the ISIMIP website, www.ISIMIP.org. Detailed upload instructions will be circulated by email, and will also be available on the website. The ISIMIP coordination team will gladly provide assistance upon request.

6.3 Documentation of models and experiments

In addition to adhering to the common settings described in this protocol, it is essential to keep track of further details regarding the configuration of each individual model. This ensures that the simulation results can appropriately be interpreted by authors of multi-model studies, and can remain transparent and usable for a long time into the future. For this purpose, the ISIMIP coordination team provides a questionnaire that all

modelling teams are asked to answer. The questionnaire is accessible online through the ISIMIP website; for assistance please write to Info@isimip.org.