The Long-Range Forecast Transient Intercomparison Project (LRFTIP): Data Specifications

Version 2, 12 August 2021

William Merryfield[1] (lead, S2D component), Mikhail Tolstykh[2,3] (lead, S2S component), Ramiro Saurral[4,5], Woo-Sung Lee[1]

1 Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, 2 Institute of Numerical Mathematics, Russian Academy of Sciences (INM RAS), 3 Hydrometcentre of Russia (HMCR), 4 Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Ciencias de la Atmósfera y los Océanos, Buenos Aires, Argentina, 5 CONICET – Universidad de Buenos Aires, Centro de Investigaciones del Mar y la Atmósfera (CIMA), Buenos Aires, Argentina

Project overview

The purpose of the Long-Range Forecast Transient Intercomparison Project (LRFTIP) is to develop an archive of hindcast climatologies and associated diagnostics that can inform investigations into the transient behavior of initialized subseasonal to decadal climate predictions, the development of model biases, and the relative merits of different initialization methods.

In support of the project, a "core" archive based on publicly available hindcast datasets including the Subseasonal to Seasonal Prediction Project (S2S), the Climate-system Historical Forecast Project (CHFP), the Coupled Model Intercomparison Project Phase 5 (CMIP5) has been developed. Additional contributions, including from forecast systems not represented in these projects, are being added as they become available. All data is publicly accessible at http://crd-data-donnees-rdc.ec.gc.ca/CCCMA/products/LRFTIP/ and is in the process of being published in Environment and Climate Change Canada's Data Catalogue with a digital object identifier (doi).

Overview of LRFTIP hindcast climatology archive

- Includes hindcast climatologies for subseasonal, seasonal, and decadal forecasting systems
- **Subseasonal** hindcast climatologies include **daily** values for the first 30-60 days
- **Seasonal** hindcast climatologies include **daily** values as above, plus **monthly** means for the entire hindcast range
- **Decadal** hindcast climatologies include **daily** and **monthly** as above, plus **yearly** means
- Selected atmosphere, ocean, land and sea ice variables are included (ocean and/or sea ice omitted if not modeled interactively)
- Data are categorized as **Priority 1** (most essential) and **Priority 2** (highly desirable) as guidance to contributors.
- Note: the priorities under **Start dates**, **Frequency**, and **Period** below supersede the "base" priorities for individual variables listed in the data tables. For example,

- for near-surface (2m) air temperature "tas" which has base priority 1, daily data for days 1-30 is Priority 1, whereas daily data for days 31-60 is Priority 2
- for near-surface specific humidity "huss" which has base priority 2, daily data for days 1-30 is Priority 2, whereas daily data for days 31-60 is also Priority 2

- For subseasonal and seasonal systems, observation-based ancillary **reference** climatologies, subject to the samegridding and temporal sampling as the hindcast climatologies, are included in order to facilitate assessments of model biases and drift.
- For decadal systems, ancillary climatologies for (i) **hindcast initial conditions** (e.g. from a supporting reanalysis, termed "**analysis**"), and (ii) **uninitialized model runs** (ideally CMIP5-style historical simulations, termed "**historical**") have been included where available because these model states represent the endpoints, beginning near (i) and evolving toward (ii), of hindcast climatological drift.

Subseasonal

Start dates:
- Near 1st day of Nov, May – *Priority 1*
- Near 1st day of Feb, Aug – *Priority 2*

Frequency:
- Daily, forecast days 1-30 – *Priority 1*
- Daily, forecast days 31-60 – *Priority 2*

Variables:
- Tables **Atmosphere 2D** & **Atmosphere 3D**, priorities as indicated

Period:
- Climatological period spanning ≥ 15 years

Seasonal

Start dates:
- Near 1st day of Nov, Feb, May, Aug – *Priority 1*

Frequency:
- Daily, forecast days 1-30 – *Priority 1*
- Daily, forecast days 31-60 – *Priority 2*
- Monthly, through longest forecast range – *Priority 1*

Variables:
- **All data tables** with priorities as indicated

Period:
- Climatological period spanning 30 years (ideally 1981-2010) is preferred, other periods spanning ≥ 15 years acceptable

Decadal

Start dates:

- At or shortly before the start of years 1961, 1966,…,2006, as per the CMIP5 Tier 1 decadal prediction experiment (Taylor et al. 2013), **OR**, at or shortly before the start of *N* consecutive years ($N \geq 15$) – *Priority 1*
- At or shortly before the start of consecutive years 1961…2010 – *Priority 2*

Frequency:

- Daily, forecast days 1-60 – *Priority 2*
- Monthly, calendar years 1-5 of forecast, plus any complete months preceding first full calendar year – *Priority 1*
- Monthly, calendar years 6-10 of forecasts – *Priority 2*
- Yearly, through longest forecast range (maximum 10 years) – *Priority 1*

Variables:

- **All data tables** with priorities as indicated

Period:

- CMIP5 Tier-1 hindcast period 1961…2006 for hindcasts initialized every 5 years ("decadal5")
- Flexible for CMIP5 or non-CMIP5 hindcasts sets initialized every year ("decadal1")

<u>File format and naming convention</u>

- NetCDF format and file naming conventions are a modified version of the original CMIP5 conventions, which are described in Taylor and Doutriaux (2011) and Taylor et al. (2012)
- In accordance with the CMIP5 convention, missing values (e.g. over land for ocean and sea ice variables) are set to 1.e20, and attributes "missing_value = 1.e+20f" and "_FillValue = 1.e+20f" are set for the corresponding model variable
- A departure from the CMIP5 convention is that the usual time variable is replaced by a leadtime variable describing time elapsed since the beginning of the forecast. This variable uses the conventions set out in Bretonnière (2014):

  time:axis = "T"
  double leadtime(time) ;
  leadtime:units = "days" ;
  leadtime:long_name = "Time elapsed since the start of the forecast" ;
  leadtime:standard_name = "forecast_period" ;
- The NetCDF4 standard is used in order to take advantage of the data compression capability of that format
- Data will be provided preferably on the original model grids on which each variable is represented. Exceptions may occur when source data is obtained from a multi-model dataset for which interpolation to a standard grid has been performed, as for the S2S, ENSEMBLES and NMME Phase 2 projects.
- Regarding file names, the general structure is the same as for CMIP5, including the CMIP5 decadal prediction experiments; however the interpretation of (i) the experiment name, (ii) the 'rip' identifier, and (iii) the time range are modified, as follows:

  (i) To emphasize that files contain hindcast climatologies as a function of lead time, the experiment names employed are

subseasonal-clim for subseasonal hindcasts
seasonal-clim for seasonal hindcasts
decadal-clim-s5 for decadal hindcasts initialized every five years
decadal-clim-s1 for decadal hindcasts initialized every year

In addition, when the “endpoint” climatologies associated with the initial conditions and uninitialized (preferably CMIP5 historical) simulations are available, these are named

analysis-clim for the hindcast initial conditions or associated (re)analysis
historical-clim for a freely running historical simulation

These climatologies should ideally be based on the same sets of initialization dates as the corresponding hindcast climatologies. For decadal hindcasts, this can correspond to **analysis-clim-s5/historical-clim-s5** for decadal hindcasts initialized every five years, or **analysis-clim-s1/historical-clim-s1** for decadal hindcasts initialized every year.

(ii) The ‘r’ component of the ‘rip’ identifier, normally indicative of realization or ensemble number, instead identifies the ensemble members that were averaged in producing the climatology. For consecutively numbered ensemble members, the first and last members are indicated separated by a hyphen. For example, when ensemble members **1** through **10** are used, the rip identifier is **r1-10i1p1** (assuming the initialization and physics identifiers are **i1** and **p1** respectively). In any unusual cases where the ensemble members that are averaged are not consecutively numbered, the ensemble members used will be indicated, separated by hyphens, in *descending* order. For example, in a case where ensemble members 1,2, 3 and 5 are averaged, the rip identifier is r5-3-2-1i1p1 (again assuming **i1** and **p1** are used).

(iii) Instead of the start and end time of an individual simulation, the time range indicates the start time of the earliest forecast contributing to the climatology, together with the end time of the latest contributing forecast. As an example, consider a **decadal-clim-s5** climatology for 10-year decadal hindcasts initialized every five years at the beginning of 1961, 1966,…2006. For annual means extending through the full 10 years, the first and last forecasts cover 1961-1970 and 2006-2015, and the time range would therefore be 1961-2015. For monthly data extending through the first 5 years, the first and last forecast periods are 196101-196512 and 200601-201012, so that the time range is 196101-201012. For daily values extending through the first 30 days, the first and last forecast periods are 19610101-19610130 and 20060101-20060130, so that the time range is 19610101-20060130.

Example: A monthly SST (variable name ‘tos’) climatology for decadal hindcasts initialized every five years at the beginning of 1961, 1966,…2006 from ensemble members 1-10 of the CanCM4 model using initialization method i1 is labeled as

tos_Omon_CanCM4_decadal-clim-s5_r1-10i1p1_196101-201512.nc4

List of variables

Atmosphere 2D (CMOR Tables day, Amon)

Variable name	Description	CF Standard Name	unit	realm	freq	pri-ority
clt	Total Cloud Fraction	cloud_area_fraction	%	atmos	d,m, y	1
evspsbl	Evaporatio n	water_evaporation_flux	kg m-2 s-1	atmos	d,m, y	2
hfss	Surface Upward Sensible Heat Flux	surface_upward_sensible_heat_flux	W m-2	atmos	d,m, y	1
hfls	Surface Upward Latent Heat Flux	surface_upward_latent_heat_flux	W m-2	atmos	d,m, y	1
huss	Near-Surface Specific Humidity	specific_humidity	1	atmos	d,m, y	2
pr	Precipitatio n	precipitation_flux	kg m-2 s-1	atmos	d,m, y	1
psl	Sea Level Pressure	air_pressure_at_sea_level	Pa	atmos	d,m, y	1
rlds	Surface Downwellin g Longwave Radiation	surface_downwelling_longwave_flux_in_air	W m-2	atmos	d,m, y	1*
rlus	Surface Upwelling Longwave Radiation	surface_upwelling_longwave_flux_in_air	W m-2	atmos	d,m, y	1*
rlut	TOA Outgoing Longwave Radiation	toa_outgoing_longwave_flux	W m-2	atmos	d,m, y	1**
rsds	Surface Downwellin g	surface_downwelling_shortwave_flux_in_air	W m-2	atmos	d,m, y	1*

	Shortwave Radiation					
rsdt	TOA Incident Shortwave Radiation	toa_incoming_shortwave_flux	W m-2	atmos	d,m, y	1**
rsut	TOA Outgoing Shortwave Radiation	toa_outgoing_shortwave_flux	W m-2	atmos	d,m, y	1**
rsus	Surface Upwelling Shortwave Radiation	surface_upwelling_shortwave_flux_in_air	W m-2	atmos	d,m, y	1*
tas	Near-Surface Air Temperature	air_temperature	K	atmos	d,m, y	1
tasmax	Daily Maximum Near-Surface Air Temperature	air_temperature	K	atmos	d,m, y	2
tasmin	Daily Minimum Near-Surface Air Temperature	air_temperature	K	atmos	d,m, y	2
tauu	Surface Downward Eastward Wind Stress	surface_downward_eastward_stress	Pa	atmos	d,m, y	1
tauv	Surface Downward Northward Wind Stress	surface_downward_northward_stress	Pa	atmos	d,m, y	1
ts	Surface Temperature	surface_temperature	K	atmos	d,m, y	1

uas	Eastward Near-Surface Wind	eastward_wind	m s-1	atmos	d,m,y	2
vas	Northward Near-Surface Wind	northward_wind	m s-1	atmos	d,m,y	2

* In cases where net longwave or shortwave surface fluxes are available but the up- and/or downwelling components are not, the net fluxes may be reported as “rls / surface_net_downward_longwave_flux” and “rss / surface_net_downward_shortwave_flux” respectively
** In cases where net longwave or shortwave TOA fluxes are available but the incoming and/or outgoing components are not, the net fluxes may be reported as “rlt / toa_net_downward_longwave_flux” and “rst / toa_net_downward_shortwave_flux” respectively

Atmosphere 3D (CMOR Tables day, Amon)

Variable name	Description	CF Standard Name	unit	realm	freq	pri-ority
hus	Specific Humidity	specific_humidity	1	atmos	d,m,y	1
ta	Air Temperature	air_temperature	K	atmos	d,m,y	1
ua	Eastward Wind	eastward_wind	m s-1	atmos	d,m,y	1
va	Northward Wind	northward_wind	m s-1	atmos	d,m,y	1

Land (CMOR Tables Lmon, LImon)

Variable name	Description	CF Standard Name	unit	realm	freq	pri-ority
mofso	Soil Frozen Water Content	soil_frozen_water_content	kg m-2	land	m,y	2
mrso	Total Soil Moisture Content	soil_moisture_content	kg m-2	land	m,y	1*
mrsov	Total Volumetric Soil	volume_fraction_of_water_in_soil	1	land	m,y	1*

	Moisture (Liquid and Solid) Content					
snw	Surface Snow Amount	surface_snow_amount	kg m-2	land	m,y	1

* If mrso is not available then mrsov may be reported. If mrso is reported, then mrsov need not be reported

Ocean 2D (CMOR Tables day, Omon)

Variable name	Description	CF Standard Name	unit	realm	freq	pri-ority
hc300*	upper 300m heat content	heat_content_to_300m_depth	K	ocean	d, m,y	2
mlotst	Ocean Mixed Layer Thickness Defined by Sigma T	ocean_mixed_layer_thickness_defined_by_sigma_t	m	ocean	m,y	1
msftbarot	Ocean Barotropic Mass Streamfunction	ocean_barotropic_mass_streamfunction	kg s-1	ocean	m,y	1
msftmyzv***	Ocean Meridional Overturning Volume Streamfunction	ocean_meridional_overturning_volume_streamfunction	m3 s-1	ocean	m,y	2
sos	Sea Surface Salinity	sea_surface_salinity	psu	ocean	m,y	1
t20d*	20 degree isotherm depth	ocean_20_degree_isotherm_depth	m	ocean	d, m,y	2
thetaoeq**	Equatorial cross section of sea water potential temperatur	equatorial_sea_water_potential_temperature	K	ocean	d,m, y	2

	e					
tos	Sea Surface Temperature	sea_surface_temperature	K	ocean	d,m,y	1
zos	Sea Surface Height Above Geoid	sea_surface_height_above_geoid	m	ocean	m,y	1

* hc300 and t20d are non-CMIP5 variables that are derivable from the Ocean 3D variable thetao
** thetaoeq is a non-CMIP5 variable that is derivable from the Ocean 3D variable thetao, and consists of a 2D cross section of thetao, averaged between 2°S and 2°N, as a function of longitude (0° to 360°) and depth (0 to 300m)
*** msftmyzv is based on the Ocean 3D resolved sea water velocity components vo and wo, and hence differs from the CMIP5 variable msftmyz, which includes parameterized eddy-induced "bolus" velocities; it is reported separately as msftmyzv_atl, msftmyzv_pac, and msftmyzv_glb for the Atlantic, Pacific+Indian, and Global Oceans respectively.

Ocean 3D (CMOR Table Omon)

Variable name	Description	CF Standard Name	unit	realm	freq	pri-ority
so	Sea Water Salinity	sea_water_salinity	psu	ocean	m,y	1
thetao	Sea Water Potential Temperature	sea_water_potential_temperature	K	ocean	m,y	1
uo	Sea Water X Velocity	sea_water_x_velocity	m s-1	ocean	m,y	1
vo	Sea Water Y Velocity	sea_water_y_velocity	m s-1	ocean	m,y	1
wo	Upward Ocean Velocity	upward_ocean_velocity	m s-1	ocean	m,y	1

Sea Ice (CMOR Table OImon)

Variable name	Description	CF Standard Name	unit	realm	freq	pri-ority
sic	Sea Ice Area	sea_ice_area_fraction	%	seaIce	m,y	1

	Fraction					
sit	Sea Ice Thickness	sea_ice_thickness	m	seaIce	m,y	1

Time-Invariant Fields (CMOR Table fx)

Variable name	Description	CF Standard Name	unit	realm	dimensionality	pri-ority
areacella	Atmosphere Grid-Cell Area	cell_area	m2	atmos	xy	1
sftlf	Land Area Fraction	land_area_fraction	%	atmos	xy	1
mrsofc	Capacity of Soil to Store Water	soil_moisture_content_at_field_capacity	kg m-2	land	xy	2
areacello	Ocean Grid-Cell Area	cell_area	m2	ocean	xy	1
basin	Region Selection Index*	region	1	ocean	xy	1
deptho	Sea Floor Depth	sea_floor_depth_below_geoid	m	ocean	xy	1
thkcello	Ocean Model Cell Thickness	cell_thickness	m	ocean	z	2

* Report on the same grid as the temperature field. flag_values=0,1,2,3,4,5,6,7,8,9,10 corresponding to flag_meanings=global_land, southern_ocean, atlantic_ocean, pacific_ocean, arctic_ocean, indian_ocean, mediterranean_sea, black_sea, hudson_bay, baltic_sea, red_sea.

Notes on naming of variables

In a few instances, the variable and CF standard names used in CMIP5 differ from names adopted in coordinated forecasting experiments such as ENSEMBLES, CHFP, and NMME. In such cases, the CMIP5 standard is adopted. These cases are summarized as follows

CMIP5 Variable name	CMIP5 CF Standard Name	non-CMIP5 Variable name	non-CMIP5 Standard Name	non-CMIP5 experiment(s)

snw	surface_snow_amount	snld	lwe_thickness_of_surface_snow_amount	CHFP
tauv	surface_downward_northward_stress	tauy	surface_downward_northward_stress	CHFP
zos	sea_surface_height_above_geoid	zoh	sea_surface_height_above_geoid	NMME

Data Archive

Hindcast and ancillary climatology datasets in accordance with the above specifications and associated diagnostics are available at http://crd-data-donnees-rdc.ec.gc.ca/CCCMA/products/LRFTIP/SEASONAL/data/CanCM4i/seasonal-clim/mon/land/ . This archive is in the in the process of being published in Environment and Climate Change Canada's Data Catalogue with a digital object identifier (doi).

Documenting reference

Saurral, R. I., W. J. Merryfield, M. A. Tolstykh, W.-S. Lee, F. J. Doblas-Reyes, J. Garcia-Serrano, F. Massonnet, G. A. Meehl and H. Teng, 2021: A dataset for intercomparing the transient behavior of dynamical model-based subseasonal to decadal climate predictions. *J. Adv. Model. Earth Syst.*, submitted.

REFERENCES

Bretonnière, P. A. , 2014: SPECS common data repository: File content and format, data structure and metadata, http://specs-fp7.eu/sites/default/files/u1/SPECS_standard_output(1).pdf

Taylor, K.E., and C. Doutriaux, 2011: CMIP5 Model Output Requirements: File Contents and Format, Data Structure and Metadata, http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5_output_metadata_requirements.pdf

Taylor, K.E., V. Balaji, S. Hankin, M. Juckes, B. Lawrence, and S. Pascoe: 2012: CMIP5 Data Reference Syntax (DRS) and Controlled Vocabularies, http://cmippcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax.pdf

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design, *Bull. Amer. Meteor. Soc.*, 93, 485–498.

<u>APPENDIX A</u> – Summary of subseasonal forecast system data holdings as of August 2021

Numbers indicate number of available variables for each model and data type

Reference (orange)   Forecasts (yellow)

Subseasonal Model	Atmosphere Daily		Ocean Daily	
	Reference	Forecasts	Reference	Forecasts
ECCC-S2S	9	**7**	1	**1**
ECMWF-S2S	11	**7**		
JMA-S2S	2	**2**	1	**1**
MeteoFrance-S2S	11	**7**	1	**1**
NCEP-S2S	11	**7**	1	**1**
UKMO-S2S	6	**4**	1	**1**

<u>APPENDIX B</u> – Summary of seasonal forecast system data holdings as of August 2021

Numbers indicate number of available variables for each model and data type

Reference Forecasts

Seasonal Model	Atmosphere Daily		Atmosphere Monthly		Ocean Monthly		Land Monthly		Sea Ice Monthly	
	Reference	Forecasts	Reference	Forecasts	Reference	Forecasts	Reference	Forecasts	Reference	Forecasts
CanCM3	18	22	22	22	12	12		4	1	2
CanCM4	18	22	22	22	12	12		4	1	2
CanCM4i	21	26	22	26	11	13	2	3	1	2
GEM-NEMO			12	12	1	1	1	1	1	1
ECMWF-S4			18	20				1		
JMAMRI-CGCM1	12	16	20	20	6	7				
JMAMRI-CGCM2			19	21	6	6		13		13
MIROC5_v1.0	12	18	16	18	6	7				
MPI-ESM-LR			20	22						
POAMA p24a/b/c			12	13						
ARPEGE			7	7						
CFS_SHFP	3	3	8	7						
CMAM	3	5	7	7				10		10
GloSea4		1	7	7						
GloSea5		1	7	7						
ENSEMBLES (CMCC-INGV, ECMWF-S3, IFM-GEOMAR, MF, DePreSys, HadGEM2	16	20	20	20				1		

APPENDIX C – **Summary of decadal forecast system data holdings as of August 2021**

Numbers indicate number of available variables for each model and data type

Analysis / Initial Conditions | Forecasts | Historical Simulations

Decadal Model	Atmosphere Daily			Atmosphere Monthly			Atmosphere Yearly			Ocean Monthly			Ocean Yearly			Land Month/yearly			SeaIce Month/Yearly		
CCSM4 (i1,i2)					24	24		24	24		8	9		8	9		3	3		2	2
MF-ENSEMBLES					20			20			11			11			1				
CFSv2(i1,i2)					26			26			7			7			1			2	
CanCM4 (i1,i2)	25	16	6	25	26	17	25	26	17	13	13	13	13	13	13	4	3	2	2	2	2
CNRM-CM5		6	6		26	26		26	26		10	10		10	10		3	3		2	2
ECMWF-ENSEMBLES					20			20			11			11			1				
GFDL-CM2p1					18	18		18	18		11	11		11	11		2	2		2	2
HadCM3		9	9		25	25		25	25		6	6		6	6		3	3		2	2
IFM-ENSEMBLES					20			20			11			11			1				
MIROC5		20	20		26	26		26	26		7			7			3	3		2	2
MRI-CGCM3		6			26			26			10			10			3			2	
UKMO-DePreSys-ENS					20			20									1				
UKMO-HadGEM2-ENS					20			20									1				
EC-EARTH					19	17		19	17		7	3		7	3					2	2
BCC-CSM1.1		6	6		26	26		26	26		7	7		7	7					2	2
CanESM5	21	21	21	26	26	26	26	26	26	11	11	11	11	11	11	3	3	3	2	2	2