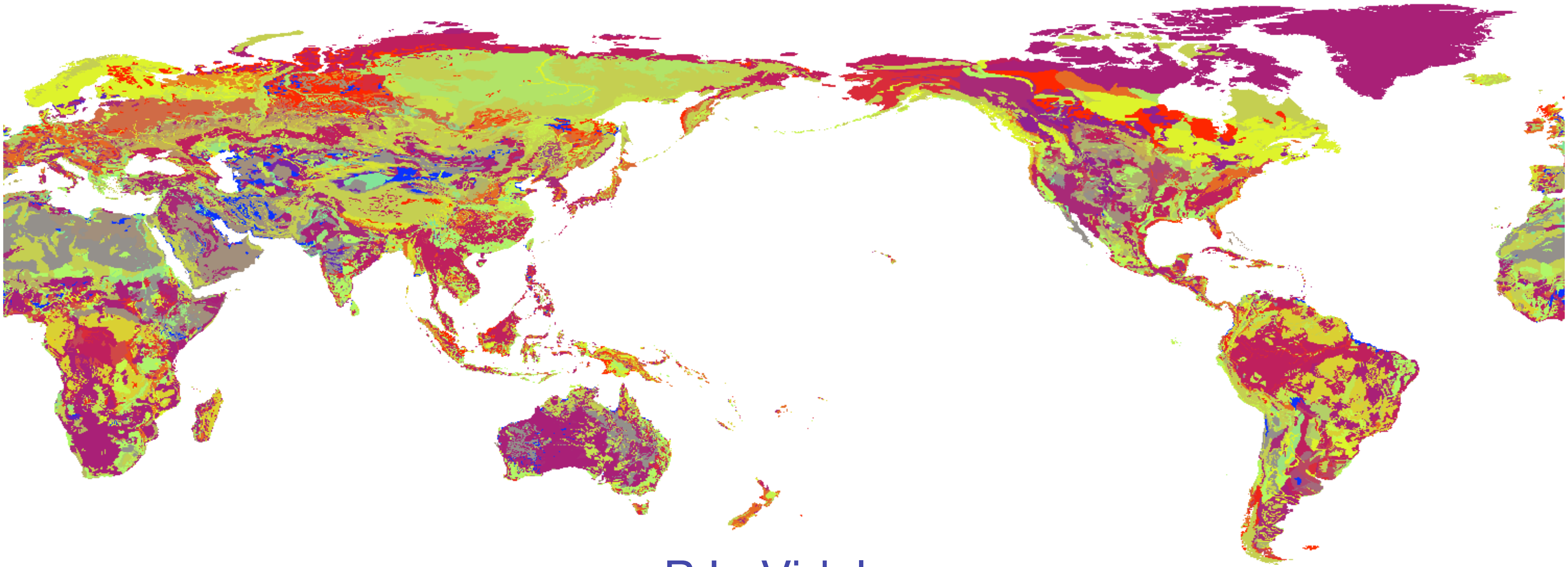
Data requirements for JULES



P.L. Vidale,

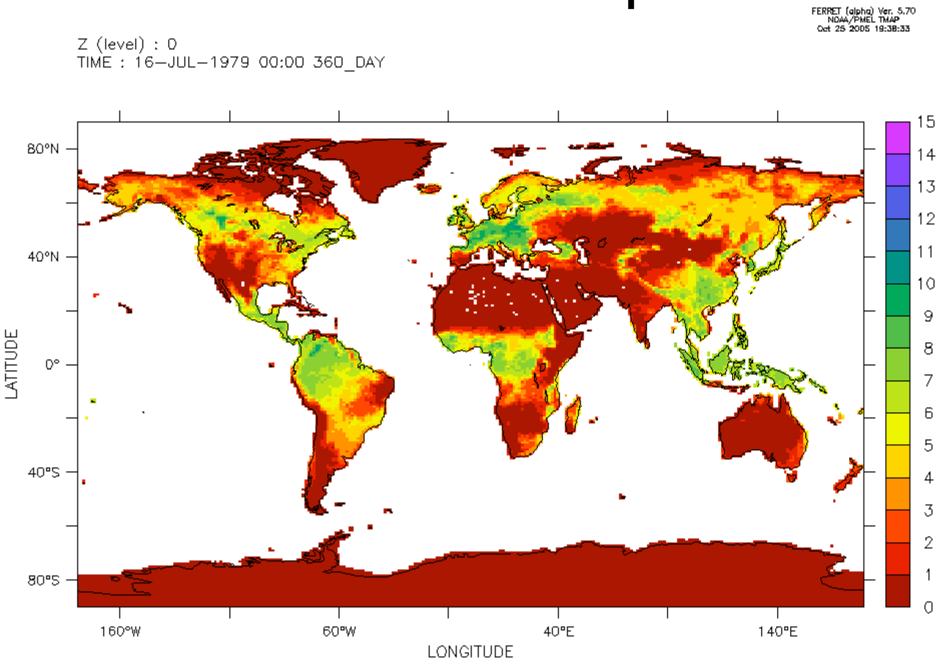
thanks to: C. Prentice, S. Quegan, R. Stöckli, A. Verhoef

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

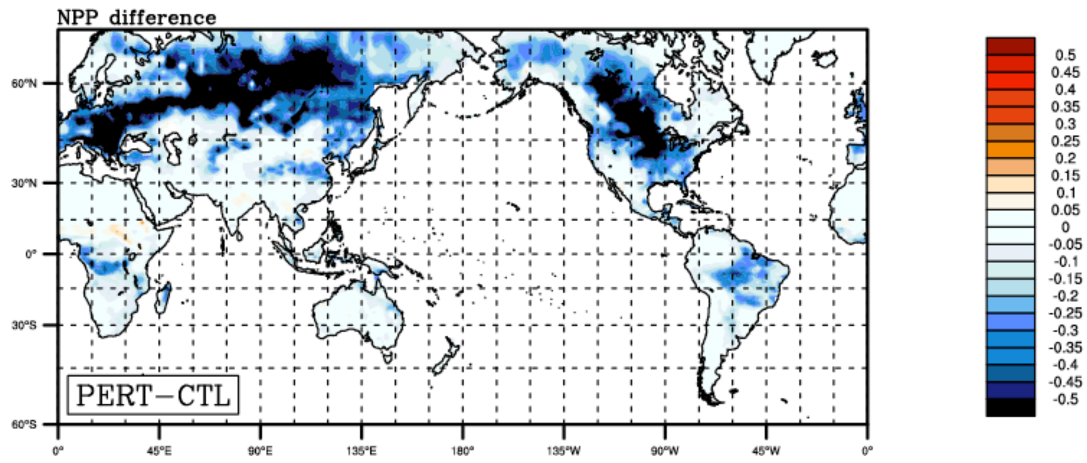
Walker INSTITUTE

JULES as a modelling system

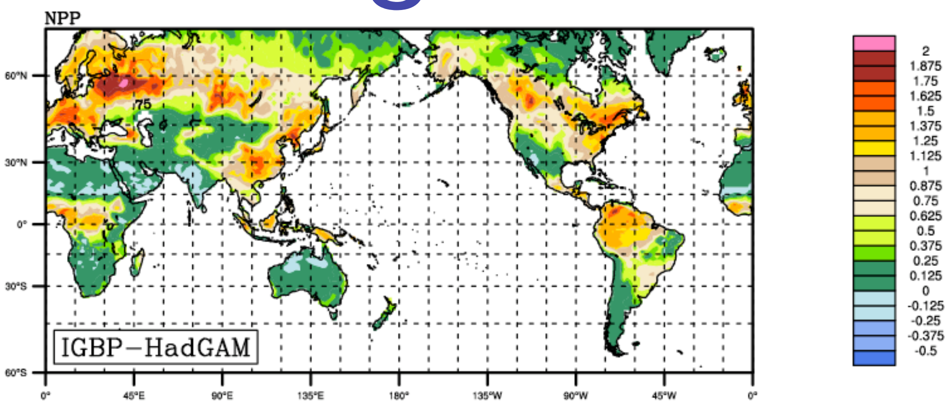
- *A model is only as good as the data that are used to force it, constrain it, verify it;*
- 2 examples:
  - <u>Left,</u> atmospheric forcing: India GPP with poor monsoonal rain;
  - <u>Right</u>, soil physical parameters: NPP change after we corrected the soil moisture parameter error.



HiGEM Jul GPP (gC/m2/day)



Range of JULES applications



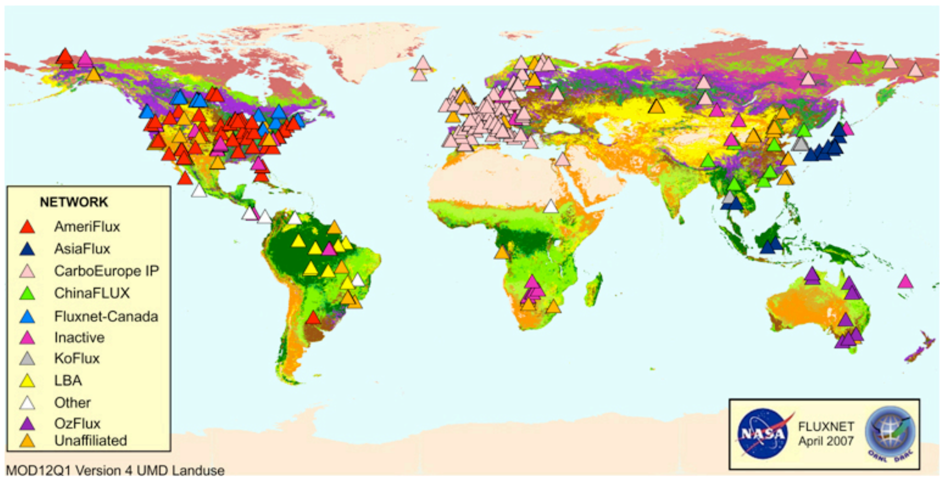
Actually, we would like to retain operational continuity across scales

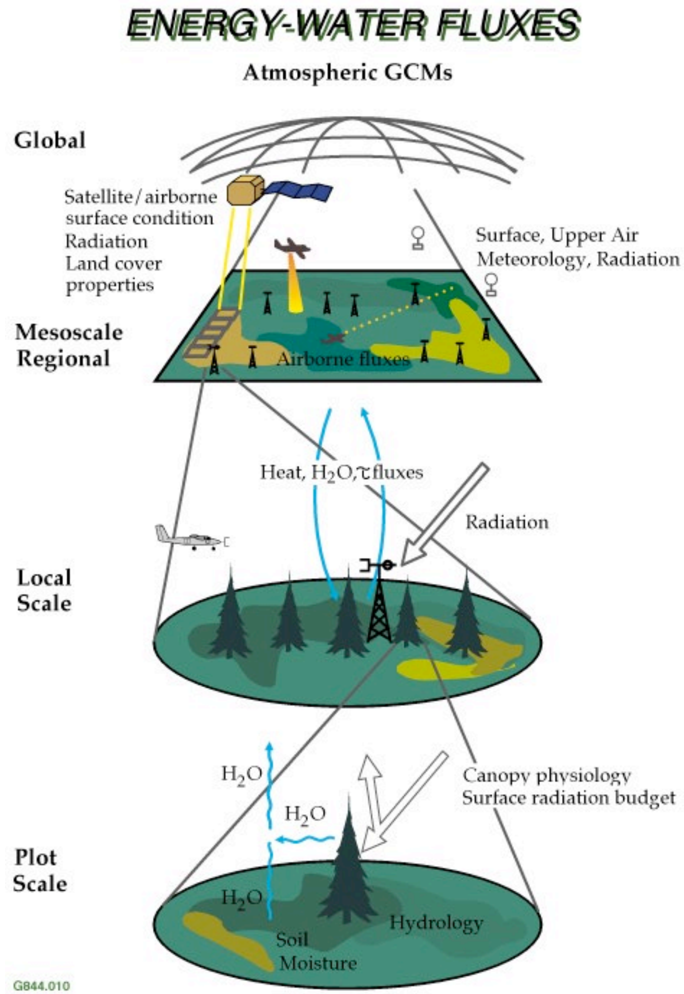
A. Coupled to a GCM, in weather forecasting and climate studies

B. In a Data Assimilation system

C. Offline, distributed: mostly for process studies and impacts, at fairly high resolution

D. Offline, site studies: 1D, long-term





LSMs require:

1. Forcing (driver) data (mostly meteo and/or climate variables, but also sub-sets of the prognostic variables, e.g. soil moisture)
2. Initial conditions (e.g. snow/ice cover, soil temperature, soil moisture)
3. Boundary conditions:
    - Static: sea/land mask, orography, albedo, land cover, glaciers, rivers
    - Semi-static: LAI, albedo
4. Parameters
    - Aerodynamic, soil physical, physiological
5. Scale-dependent choice of parameterisations/numerical methods
6. Verification data

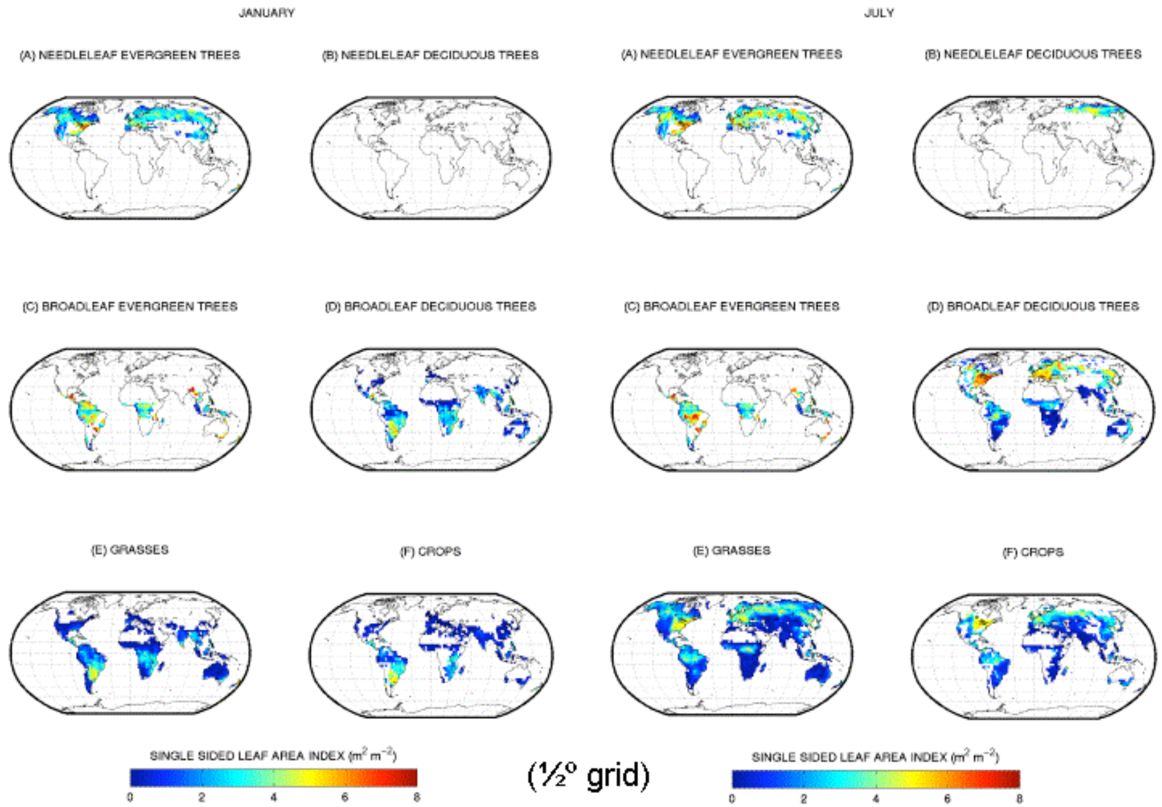
Often, these are produced ad-hoc within each application group and not shared. Some modelling groups (e.g. NCAR-CLM, GLDAS, GSWP2) have created some data sets that can be downloaded together with the models. This initially saves time and effort, but is very limiting.

Examples from the CLM3.5 site

Morphology for each plant functional type. Roughness length and displacement height are in proportion to canopy top height. Root distribution at depth z (m) is
$f(z)=1-0.5[\exp(-az)+\exp(-bz)]$.

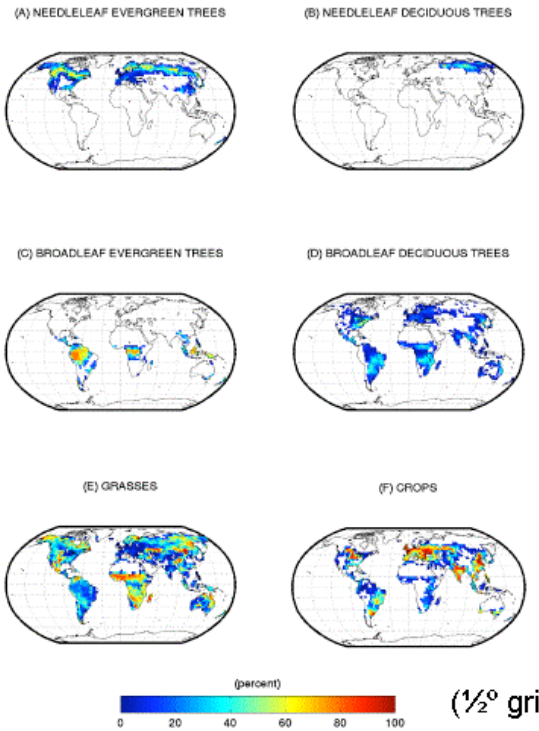
Plant Functional Type	Leaf Dimension (m)	Roughness Length	Displacement Height	Root Distribution a	Root Distribution b
NET temperate	0.04	0.055	0.67	7.0	2.0
NET boreal	0.04	0.055	0.67	7.0	2.0
NDT boreal	0.04	0.055	0.67	7.0	2.0
BET tropical	0.04	0.075	0.67	7.0	1.0
BET temperate	0.04	0.075	0.67	7.0	1.0
BDT tropical	0.04	0.055	0.67	6.0	2.0
BDT temperate	0.04	0.055	0.67	6.0	2.0
BDT boreal	0.04	0.055	0.67	6.0	2.0
BES temperate	0.04	0.120	0.68	7.0	1.5
BDS temperate	0.04	0.120	0.68	7.0	1.5
BDS boreal	0.04	0.120	0.68	7.0	1.5
C_3 grass arctic	0.04	0.120	0.68	11.0	2.0
C_3 grass	0.04	0.120	0.68	11.0	2.0
C_4 grass	0.04	0.120	0.68	11.0	2.0
Crop1	0.04	0.120	0.68	6.0	3.0
Crop2	-	-	-	-	-

Leaf Area Index



(½° grid)

Plant Functional Type Geography



(½° grid)

Photosynthetic parameters for each plant functional type. Path, photosynthetic pathway; V_{max25}, maximum carboxylation at 25°C (μmol CO_2 m^{-2} s^{-1}); a, quantum efficiency (μmol CO_2 μmol photon^{-1}); m, slope of conductance-photosynthesis relationship.

Plant Functional Type	Path	V_{max25}	a	m
NET temperate	C_3	51	0.06	6
NET boreal	C_3	43	0.06	6
NDT boreal	C_3	43	0.06	6
BET tropical	C_3	75	0.06	9
BET temperate	C_3	69	0.06	9
BDT tropical	C_3	40	0.06	9
BDT temperate	C_3	51	0.06	9
BDT boreal	C_3	51	0.06	9
BES temperate	C_3	17	0.06	9
BDS temperate	C_3	17	0.06	9
BDS boreal	C_3	33	0.06	9
C_3 grass arctic	C_3	43	0.06	9
C_3 grass	C_3	43	0.06	9
C_4 grass	C_4	24	0.04	5
Crop1	C_3	50	0.06	9
Crop2	-	-	-	-

Practical problems which we all encounter when setting up JULES

- For global simulations, we are basically using the UKMO HadCMx maps of boundary conditions and parameters (ancillary files);
- For local studies, often a particular site has one portion of the data and we "make up" the rest, not necessarily in a consistent way;
- Many parameters are <u>listed independently</u> in the namelist, but in reality they are <u>physically connected</u> and/or the relationship is scale-dependent;
- We want to alter the driver data (e.g. to study the impact of climate change), but we do not know how to do this in a consistent fashion, for instance:
  - "Delta" our climate data to simulate some aspect of climate change;
  - Create a synthetic data set for a process study, e.g. altering a precipitation or soil moisture time series;
- Very large heterogeneity of data "out there" and lack of broad expertise for selection and processing:
  - Ever tried interpolating precipitation over complex terrain ?

Meteo/Climate example

Centre	**Res**	**Coverage**	**Type**	**Year**	**Format**
LLNL/PCMDI	~0.5-2.5°	global	IPCC Climate runs	1860-2100	NetCDF
ERA40 (or NCEP)	1°	global	Re-analysis	1959-2004	GRIB
CRU	0.5°	Global land	Analysis	20th century	NetCDF
FLUXNET	NA	local	Local observations	1992-current	ASCII

Soils physical parameters examples

Centre	**Resolution**	**Coverage**	**Type**	**Year**	**Format**
FAO	5'	global	Classes + parameters	1990	ArcInfo
IGBP	5'	global	Classes + VG parameters	?	Binary
GLDAS	5' - 0.25°	Global, multi-layer	Classes + parameters	1999	NetCDF
Wilson / Henderson Sellers	1°	global	S/S/C fractions	1995	?
ISLSCP2	1°	Global 2 layers	Classes + KH parameters	2004	ASCII
SOTER	5′	global	Orography Slope, soils	In course	?

Examples from the LDAS site

Atmospheric Forcing Datasets

DATASET NAME	DATASET TYPE	PARAMETERS	DOMAIN	SPATIAL RESOLUTION	TIME RECORD	TEMPORAL RESOLUTION
NCEP's Global Data Assimilation System (GDAS)	Model Derived	Meteorology Forcing	Global	Gaussian (~0.313 degree)	Jan. 2000 - Current	3-hourly
NASA's Goddard EOS Data Assimilation System (GEOS)	Model Derived	Meteorology Forcing	Global	1.25 x 1.00 degree	Dec. 2000 - Current	3-hourly
The European Centre for Medium Range Weather Forecasting (ECMWF)	Model Derived	Meteorology Forcing	Global	Gaussian (~0.25 degree)	Sept. 2001 - Current	3-hourly
Naval Research Laboratory Precipitation	Satellite Observed	Mean rain rate	60N - 60S	0.25 x 0.25 degree	June 2001 - Current	6-hourly
NASA/GSFC TRMM 3B42RT Realtime Huffman Precipitation	Satellite Observed	Mean rain rate	60N - 60S	0.25 x 0.25 degree	Feb. 2002 - Current	3-hourly
PERSIANN Precipitation	Satellite Observed	Mean rain rate	60N - 60S	0.25 x 0.25 degree	Jan. 2002 - Current	hourly
Disaggregated CMAP Precipitation	Merged Satellite/Gauge	Mean rain rate	Global	Gaussian (~0.313 degree)	Jan. 2001 - Current	6-hourly
Air Force Weather Agency Agricultural Meteorology modeling system (AFWA AGRMET) Radiation	Satellite Observed	Shortwave, Longwave	Global	0.25 x 0.25 degree	March 2001 - Current	hourly
NOAA/CPC CMORPH Precipitation	Satellite Observed	Mean rain rate	60N - 60S	8 km	Dec. 2002 - Current	30 minute
NASA/GSFC TRMM 3B42(V6) Precipitation	Calibrated satellite estimates	Mean rain rate	60N - 60S	0.25 x 0.25 degree	Jan. 1998 - Current	3-hourly

Land Surface Datasets

DATASET TYPE	DOMAIN	SPATIAL RESOLUTION	TIME RECORD	TEMPORAL RESOLUTION
University of Maryland Vegetation Classification	Global	1 km	Static	Static
Boston University Leaf Area Index (derived from both AVHRR and MODIS)	Global	1 km	1981-Current (AVHRR), 2000-Current (MODIS)	Monthly
Soils Database from Reynolds, Jackson, and Rawls (1999)	Global	5 minute	Not applicable	Not applicable
Elevation Database from GTOPO30	Global	30 second	Not applicable	Not applicable

Possible approach for forcing data/parameters:

Collect all data:

- For single points, using standardised sets/conventions (e.g. R. Stöckli's FLUXNET collection and in the future the "FLUXNET synthesis data set";
- Globally, at highest possible resolution (temporal, spatial), forming a shared data repository (e.g. the Earth System Atlas in IGBP)

Share:

1. via OpenDAP data server;
2. Data and methods documentation

Earth System Atlas Data Download Page

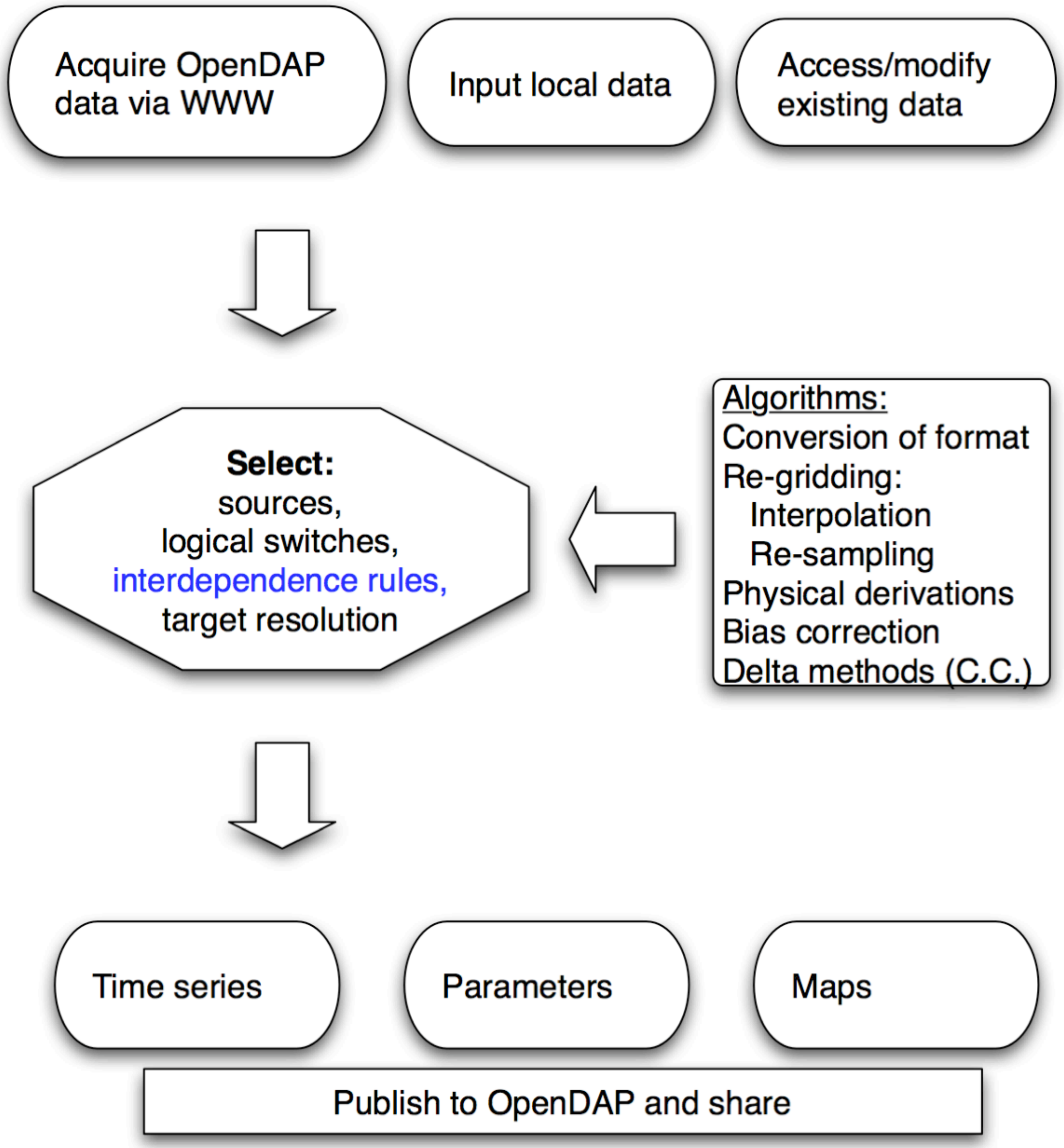
Global Data	Download File by Resolution			Comments	References
	6 min	30 min	60 min		
Countries	g_cntry6min.asc	g_cntry30min.asc	g_cntry60min.asc	Arc-Info ascii file	WSAG, UNH
River Direction		net_direction30min.asc		Arc-Info ascii file	WSAG, UNH
River Order		net_order30min.asc		Arc-Info ascii file	WSAG, UNH
Elevation	g_etopo6min.asc	g_etopo30min.asc	g_etopo60min.asc	Arc-Info ascii file	WSAG, UNH
Elevation and Bathymetry	g_smith6min.asc	g_smith30min.asc	g_smith60min.asc	Arc-Info ascii file	WSAG, UNH
Temperature, January		tempjan30min.asc	tempjan60min.asc	Arc-Info ascii file	Atlas of the Biosphere, UW
Temperature, July		tempjul30min.asc	tempjul60min.asc	Arc-Info ascii file	Atlas of the Biosphere, UW
Precipitation		precip30min.asc	precip60min.asc	Arc-Info ascii file	WSAG, UNH
Net Primary Productivity			npp60min.asc	Arc-Info ascii file	Atlas of the Biosphere, UW

Back to Atlas Homepage

Add a layer of algorithms to process the data to the desired resolution, applying:

- –Established physical relationships (e.g. canopy height, D, z0; variability of mountains and z0)
- –Checks: Canopy height-ref. height
- –Proper methods (e.g. sampling versus interpolation)

JULES data generation



Requirements:

1. Data standards, e.g. NetCDF, ArcInfo
2. Metadata, e.g. units, projection, coordinates, errors etc.
3. Portable algorithms
4. Data + algorithm verification and **maintenance*****

*** *broader scope: Colin and Shaun have reminded me how CRU has been operating as a bootleg scheme, which is absurd*

Verification data

- The typical list (we can work on this later today):
  - CRU, ERA40, NCEP, JRA25, weather/climate variables
  - FLUXNET fluxes
  - Radiation (e.g. sfc. albedo)
  - GCOS "Essential Climate Variables" ?
  - Runoff from major rivers: GRDC
  - Terrestrial Water Storage: from published studies, in the future helped by GRACE
  - Soil moisture: only available at very few locations on Earth !
  - Snow
  - Carbon (e.g. monthly CO2 concentrations, tracers, ground-based NPP, soil C, N storage)
  - Phenology and/or veg. indices (NDVI, phenological gardens etc.)
  - Land cover (every 10 years since 1860 ?)
  - Agricultural yield
  - Special data sets from field campaigns (e.g. HAPEX, FIFE, BOREAS, AMMA)
  - GSWP2 (for model intercomparison)
- I think that a major priority is to collect/use data which **resolve the diurnal cycle**, or that offer at least daily resolution;
- We should define a minimal set for JULES verification;
- The same data issues raised regarding forcing data apply to converting/regridding/deriving verification variables in order to compare to our model output. A basic example is an algorithm to compute and compare an energy and/or water balance.
- New data sets from new instruments: we need to develop observation operators (covered in other talks)

Verification tools

- Plots Atlas:
  - John Donners (Reading, UJCC) has developed a WWW-based "Plots Atlas" which automatically computes, collects and displays a wide range of model plots, allowing to:
    - instantly compare any simulation to a large number of established climatologies
    - intercompare all stored simulations
  - This tool is very useful if you are not an expert on the entire range of model science and you want to get an overview prior to focussing on your particular/favourite variables.
  - www.earthsimulator.org/PlotsAtlas
- Data assimilation (covered in other talks): need to develop a DA "wrapper", e.g. CCDAS. Crucially important for model development;
- Model intercomparison, e.g. GSWP2, C4MIP, CCMAP

Summary

- JULES should be a modelling system, comprising model and data:
  - Forcing
  - Initial conditions
  - Boundary conditions
  - Parameters
  - Data processing algorithms
  - Verification
- JULES users need a flexible, configurable data suite, with physically-based algorithms linking variables as they are in the model; it is not sufficient to have a static data repository;
- We should all contribute to the data and the data generation algorithms, so that we can exercise expert monitoring and review of what goes into the model;
- Generating and sharing data requires standards;
- I see a big danger in giving no option to the community but to supply their own data to the model, which could unintentionally produce poor science, to be published under the name “JULES”.