Development of a New Global Dataset for Offline Terrestrial Simulations
- for Global Soil Wetness Project Phase 3 -

1,2Hyungjun Kim, 3Satoshi Watanabe, 4Eun-Chul Chang, 1,4Kei Yoshimura, 3Yukiko Hirabayashi, 2James Famiglietti, and 1Taikan Oki

1Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
2University of California Center for Hydrologic Modeling, Irvine, California, USA
3Institute of Engineering Innovation, The University of Tokyo, Tokyo, Japan
4Atmosphere and Ocean Research Institute, The University of Tokyo, Tokyo, Japan

Institute of Industrial Science, The University of Tokyo  hjkim@rainbow.iis.u-tokyo.ac.jp
• GSWP was a GEWEX project, led by COLA (P. Dirmeyer) and supported by IIS/UT involving over a dozen modeling groups on four continents.

• GSWP1 used the ISLSCP I-1 data to examine 1987-1988.

• GSWP2, a 10-year (1986-1995) global land-surface analysis, will begin next year.

• Regional studies are investigating issues of aggregation, sub-grid variability, and assimilation of remote sensing data.

Unlike the ocean, the land surface state variables (soil moisture, snow depth and coverage, soil temperature) are not routinely measured and reported. Land-surface models driven by observed meteorology give us a means to generate a surrogate ‘observed’ data set.

(Dirmeyer, et. al, 1998, 2006, BAMS)
Scope of the GSWP3

GSWP

Hydro
Water-MIP (EU-WATCH)
ISI-MIP (Phase2)

GSWP

Carbon
TRENDY
iLEAPS

ATMOS
GLASS
LAND

Coupled Hydro-Energy-Eco System Experiment
Century long timespan (EXP1: 1901-2010)
Time varying Land Use Change CO₂
“+” Shape Experiments Design

EXP1
Long-term Retrospective
1901 – 2010

EXP3
Super Ensemble
1979 – Present

EXP2
Multiple GCMs & Scenarios
2000 – 2100

Boundaries
Condition
Long-term Retrospective Experiment

Proposal: EXP1

No C-Cycle Changes
- Constant LUCC
- Constant CO₂

C-Cycle Changes Considered
- Prescribed LUCC
- Prescribed CO₂
- Dynamic LUCC
- Prescribed CO₂

MODEL GROUP 1
- No Photosynthesis
- Sim. with time-constant C-cycle boundary condition

MODEL GROUP 2
- Static Plant Physiology
- Sim. with time-varying C-cycle boundary condition

MODEL GROUP 3
- Dynamic Plant Physiology
- Sim. with unconstrained C-cycle boundary condition
We are in ‘**Data Rich Era**’, and one single flux is not enough to ensure even simple water budget closure.

**Area Representative Validations**
- Global Large Basins,
- Sub-basins,
- Large Catchments

**Pixel-based Validations**
- CEOP Reference Sites,
- FLUXNET Sites,
- Data Rich Regions (e.g., Illinois, USA)

**Multi-way Validation**
- River Discharge (In-situ or Altimetry estimated)
- GRACE Measure TWS Variations
- Tower Flux Measurements
- Soil Moisture (In-situ)
- Water Table Depth
Data Interface Development

**Global Soil Wetness Project**

**Introduction**

In nature, a huge non-linear system, components of energy, water, and carbon cycles are inter-connected through various processes. Under the climate change during the last century, the interactions between the components of hydro-energy-eco system have been altered drastically, and anthropogenic effects such as carbon dioxide emission and land use alteration have been indicated as a major cause. However, especially in large scale, our knowledge and its numerical embodiments are still lacking for representing the underlying mechanisms and understanding roles and extents of interactions, even though it is crucial to anticipate future climate and mitigating the changes. For decades, our land surface simulation systems have developed remarkably. Numbers of processes and components in the hydro-energy-eco system have become able to be simulated, the physical schemes have been enhanced and observational datasets which can be used for forcing and validating models have been expanded. *Global Soil Wetness Project Phase 2 (GSWP2)* has served the first global gridded multi-model analysis of land surface state variables and fluxes, but it has been asked to be extended because of the relatively short time span (10 years: 1986-1995). single biased forcing dataset, small number of validation sites, and so on. This document is a statement plan for the third phase of GSWP which will aim to generate a long-term comprehensive set of the art land surface simulations.

**Purpose and Scientific Questions**

Main purpose of the GSWP3 is to generate a comprehensive set of the art land surface simulations. Uncertainties from different sources and their translated models is established. It is expected the generated background will be used as a sort of land surface reanalysis, which allows various questions to be answered through the project are:

1. How interactions between eco-hydrological processes will be?
2. What will be the water, energy, and carbon balances on the global scale? How do they vary between different ecosystems?
3. How do the state-of-the-art land surface modeling systems perform under extreme conditions?
4. How do the state-of-the-art land surface modeling systems perform under extreme conditions?

**Experiments Design**

**Project Web Page**

Provide Experiment Information

Update Progress

**Mirrored Data Center**

Distribute / Collect Data

Web-based User Interface for Data Manipulation and Analysis

http://hydro.iis.u-tokyo.ac.jp/GSWP3

http://basins.ucchm.org
## Existing Forcing Data Available Globally

<table>
<thead>
<tr>
<th></th>
<th>NCC</th>
<th>GSWP2</th>
<th>Princeton</th>
<th>ELSE</th>
<th>WATCH</th>
<th>GSWP3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference</strong></td>
<td>Ngo Duc et al., 2005</td>
<td>Dirmeyer et al., 2006</td>
<td>Sheffield et al., 2006</td>
<td>Kim et al., 2009</td>
<td>Weedon et al., 2011</td>
<td>Kim et al., in prep.</td>
</tr>
<tr>
<td><strong>Spa./Temp. Resolution</strong></td>
<td>1 deg. 6 hours</td>
<td>1 deg. 3 hours</td>
<td>1 deg. 3 hours</td>
<td>1 deg. 6 hours</td>
<td>0.5 deg. 3 or 6 hours</td>
<td>0.5 deg. 3 hours</td>
</tr>
<tr>
<td><strong>Base Reanalysis</strong></td>
<td>NCEP/NCAR 1948 - now T62 / 6hr</td>
<td>NCEP/NCAR 1948 - now T62 / 6hr</td>
<td>NCEP/NCAR 1948 - now T62 / 6hr</td>
<td>JRA25 1948 – now T106 / 6hr</td>
<td>ERA-40 1957 - 2002 TL159 / 6hr</td>
<td>20CR 1871 - 2010 2 deg. / 6hr</td>
</tr>
<tr>
<td><strong>Spa. Dis- aggregation</strong></td>
<td>Bi-linear</td>
<td>Bi-linear</td>
<td>Bi-linear, Bayesian</td>
<td>Bi-linear</td>
<td>Bi-linear</td>
<td>Dynamical Downscale</td>
</tr>
<tr>
<td><strong>Temp. Dis- aggregation</strong></td>
<td>N/A</td>
<td>Variability from Obs.</td>
<td>Variability from Obs.</td>
<td>N/A</td>
<td>Variability from Obs.</td>
<td>Dynamical Downscale</td>
</tr>
</tbody>
</table>
## Existing Forcing Data Available Globally

<table>
<thead>
<tr>
<th>Variables</th>
<th>NCC</th>
<th>GSWP2</th>
<th>Princeton</th>
<th>ELSE</th>
<th>WATCH</th>
<th>GSWP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prcp.</td>
<td>CRU (Mult.)</td>
<td>CRU (Mult., Wind Corr.)</td>
<td>CRU (Mult., Wind Corr.), TRMM(3hr)</td>
<td>GPCC, CU, CMAP, GPCP, PREC/L (Mult.)</td>
<td>CRU, GPCC (Mult., Wind Corr.)</td>
<td>GPCC (Mult., Wind Corr.), CU, GPCP-1DD (Non-parametric)</td>
</tr>
<tr>
<td>RSDN, CCOV</td>
<td>SRB (Mult.)</td>
<td>SRB (Mult.)</td>
<td>SRB (Mult.)</td>
<td>SRB (Mult.)</td>
<td>Aerosol, SRB (Mult.)</td>
<td>SRB</td>
</tr>
<tr>
<td>RLDN</td>
<td>SRB</td>
<td>SRB</td>
<td>SRB</td>
<td>SRB</td>
<td>SRB</td>
<td>SRB</td>
</tr>
</tbody>
</table>
Revising Forcing data for EXP1 \textit{(long-term retrospective)}

**Overall Strategy**

**Methods**

**Dynamical Global Downscaling**
- Spectral Nudging using GSM (Yoshimura and Kanamitsu, 2008)
- Single Ensemble Correction (Yoshimura And Kanamitsu, 2013)
- Vertically Weighted Damping (Hong and Chang, 2012)

**Two-pass Ensemble Bias Correction**
- LDMF Daily Correction (Kim et al., in prep.)
- Parametric Monthly Correction (Watanabe et al., 2012)

**Forcing**
- 20CR (Compo et al., 2011)
  - 1901-2010 6hr / 2°x2°(91x180)
- GPCP
- Observations (Prcp: GPCC, GPCP, CPC-Unified; Tair: CRU; Rad.: SRB)
- GSWP3 (DDS T248)
  - 0.5°x0.5° 1901-2010 3hr
20th Century Reanalysis

- Using only surface pressure data historically recorded since 1871’s
- Ensemble Kalman Filter with 56 members
- T62L28 GFS with NOAH LSM
- Reanalysis skill is comparable to current Day-3 forecast skill (Whitaker et al., 2009)
- Ripple-like pattern due to spectral model interacting with the high resolution orography

Whitaker et al. (2009)
Global Dynamical Downscaling by Spectral Nudging Using GSM

\[ f(\lambda, \phi) = \sum_{m=-M}^{m=M} A_{(m, \phi)} e^{im\lambda}, \quad \text{with} \]

\[ A_{(m, \phi)} = \begin{cases} 
A_f(m, \phi) & (|m| > \frac{2\pi R_E \cos \phi}{L}) \\
\frac{1}{\alpha + 1} [A_f(m, \phi) + \alpha A_{a(m, \phi)}] & (|m| \leq \frac{2\pi R_E \cos \phi}{L})
\end{cases} \]

Successfully generate high frequency signals preserving low frequency background.

Effectively relieves ripple-like pattern (an artifact of 20CR due to high-res. topography mismatch)

Yoshimura and Kanmitsu 2008
Single Member Correction Using Ensemble Mean

\[ F_{n}^{\text{new}} = F_{n} + \langle F \rangle - \langle F_{n} \rangle \]

- Single member forecast \( F_{n} \)
- Monthly running mean of a single member \( \langle F_{n} \rangle \)
- Monthly running mean of all members \( \langle F^{\text{bar}} \rangle \)

**Data and Method**

**Yoshimura and Kanamitsu 2013**
Vertically Weighted Damping Scheme

Hong and Chang, 2013

YK08 well constrains large scale features and downscale high frequency domain, but somewhat over-ruled by original reanalysis dataset.

HC12 shows more realistic representation in Amazon avoiding pre-existing artifacts of original 20CR, but is not able to keep original large scale pattern effectively.
Vertically Weighted Damping Scheme

Data and Method

Spectral Nudging in all domains > 1000m half-width for U, V, T, P

OR

only degree 0, 1 is assimilated for T, P

(_mTP experiment)

20CR  GPCP

YK08  HC12

YK08_mTP  HC12_mTP

0.1_mTP  0.1
### Simulations Streams

<table>
<thead>
<tr>
<th>Year</th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1870</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1880</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1890</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1910</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1930</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Boundary Conditions**: ~10TB
- **Simulation Results**: ~40TB
- **CPU Time**: ~40000 Hours @ T2K Supercomputer

100% (ETA: ~7 months)
Annual Precipitation 1993 – 2008

<table>
<thead>
<tr>
<th>Products</th>
<th>20CR</th>
<th>DDS T248</th>
<th>GPCP 1DD</th>
<th>GPCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>[mm/d] (land)</td>
<td>2.6 (-)</td>
<td>2.5 (2.5)</td>
<td>1.5 (-)</td>
<td>N/A (2.0)</td>
</tr>
</tbody>
</table>
Comparison between Statistical and Dynamical Downscaling

<table>
<thead>
<tr>
<th>MEAN (2m Tair)</th>
<th>VARIANCE (2m Tair)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM T248</td>
<td>GSM T248</td>
</tr>
<tr>
<td>20CR CRU</td>
<td>20CR CRU</td>
</tr>
</tbody>
</table>

Represent high-res geographical details pretty well, and small signal loss comparing to actual 0.5 deg. Obs.

Variability in higher frequency remains depending on the model of the analysis product after bias correction.
## ‘Two-pass’ Bias Correction for Precipitation

<table>
<thead>
<tr>
<th>Name</th>
<th>Global Precipitation Climatology Centre (GPCC)</th>
<th>Timespan</th>
<th>1901.01 – 2010.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>Version 6 full data reanalysis</td>
<td>Temp. Res.</td>
<td>Month</td>
</tr>
<tr>
<td>Source</td>
<td><a href="http://gpcc.dwd.de">http://gpcc.dwd.de</a></td>
<td>Spa. Res.</td>
<td>0.5 degree</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Climate Prediction Center (CPC) Unified</th>
<th>Timespan</th>
<th>1979.01 – 2005.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>Chen et al. (2007)</td>
<td>Spa. Domain</td>
<td>Global</td>
</tr>
</tbody>
</table>

### Gauge Type

1. Canadian Nipher  
2. Chinese standard  
3. Hellmann like*  
4. Wild like  
5. Tretyakov  
6. Norwegian standard  
7. Japanese (average)  
8. NWS 8 inch like  
9. Unknown

[Hirabayashi et al. 2008]
Daily Precipitation Products

Experiences

CPC-Unified

1993 – 2002 (10y) for training

# of Stations_{avg} > 0.0 (6.3% coverage; 16406)
GPCC (6.8% coverage; 17612/259200)

# of Stations_{avg} > 0.5 (4.5% coverage; 11573)
GPCC (6.0% coverage; 15602/259200)
‘Two-pass’ Bias Correction for Precipitation

Second pass: Long-term monthly bias correction
Localized Dual Moment Fitting

$LDM\ (\mu, \sigma, \varepsilon)$

$\mu$: Location

$\sigma$: Scale

$\varepsilon$: Limit

$P_{day} < \varepsilon$

$P_{day} > \varepsilon$

Linear ratio bias correction
Correct mean and standard deviation
Localized Dual Moment Fitting Implementation

Median 60 70 80 N.C.

Jan. GCM#1

Jan. GCM#2

Jul. GCM#1

Jul. GCM#2
Protocol for the Analysis of Land Surface Models

Validation

Institute of Industrial Science, The University of Tokyo
Validation

**FLUXNET Marconi Conference Gap-Filled Flux and Meteorology Data**

**Variables:** carbon dioxide, water vapor, and energy exchange


**Sites:** 38 of EUROFLUX and AmeriFlux

**Sites:**
- HY: Lon: 24.29, Lat: 61.85, Elv: 181.00 (181.00)
- AB: Lon: -3.80, Lat: 56.61, Elv: 355.00 (355.00)
Thank you
Overall Strategy

Methods

Wind

- 20CR
- CRU CL1.0
- GPCC v6.0
- GPCC_Corr
- Snowf

Snowf

- CRU CL1.0
- GPCC v6.0
- GPCC_Corr
- Snowf

Prcp

- CU V1.0
- Daily Corr. (ε,μ,σ) (1883-2010)
- Rain/Snow Separation

Prcp

- Daily Corr. (ε,μ,σ) (1883-2010)
- Rain/Snow Separation

SW

- Monthly Cal. (μ) (1984-2007)
- Daily Cal. (σ) (1984-2007)
- Daily Corr. (σ) (1883-2010)

LW

- Monthly Cal. (μ) (1984-2007)
- Daily Cal. (σ) (1984-2007)
- Daily Corr. (σ) (1883-2010)

CCOV

- Monthly Cal. (μ) (1984-2007)
- Daily Cal. (σ) (1984-2007)
- Daily Corr. (σ) (1883-2010)

Press

- Monthly Cal. (μ) (1901-2010)
- Corr. (Altitude)

Qair

- Monthly Cal. (μ) (1901-2010)
- Corr. (Altitude)

Tair

- Monthly Cal. (μ) (1901-2010)
- Corr. (Altitude)
Air Temperature Correction

Sources of Bias
1. Simulation in Reanalysis & DDS
2. Reference Observation
3. Spatial Resolution
4. Elevation Mismatch

Biases to be Corrected
1. Geographical Biases [S3, S4]
2. Monthly Mean [S1]
3. Diurnal Temperature Range [S1]

<table>
<thead>
<tr>
<th>Temporal Res.</th>
<th>Spatial Res.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reanalysis</td>
<td>Hourly</td>
</tr>
<tr>
<td>Observation</td>
<td>Monthly</td>
</tr>
</tbody>
</table>

Conv. $T_{RA}$ to $T_{sl_{RA}}$

Intp. $T_{sl_{RA}}$ to $T_{sl_{RA0.5}}$

Conv. $T_{OBS}$ to $T_{sl_{OBS}}$

Diff. between $T_{sl_{RA0.5}}(m)$ and $T_{sl_{OBS}}$

Corr. $T_{sl_{RA0.5}}$ to $T_{sl_{CORR0.5}}$

Conv. $T_{sl_{CORR0.5}}$ to $T_{CORR0.5}$
MARCONI:
Griffin, Aberfeldy, Scotland
Air Temperature Correction

FLUXNET Marconi Conference Gap-Filled Flux and Meteorology Data, 1992-2000
Fluxes of carbon dioxide, water vapor, and energy exchange have been measured at 38 forest, grassland, and crop sites as part of the EUROFLUX and AmeriFlux projects. A total of 97 site-years of data were compiled, primarily between 1996 and 1998 but also for 1992-1995 and 1999-2000.