Hydrological Modeling in the Era of Big Data: Opportunities and Challenges

Yi Zheng * and Shijie Jiang

(zhengy@sustc.edu.cn)

School of Environmental Science and Engineering
Southern University of Science and Technology
1. Background

2. Three study cases

3. Concluding remarks
Global water issues

Water shortage


Vörösmarty et al., *Nature*, 2010, 467: 555-561

Water pollution

Flood risk

Water-related ecological disasters

Disappearing inland lakes (Aral Sea)

Vegetation degradation

Algae bloom

(Alfieri et al., *Earth’s Future*, 2017, 5: 171-182)
Global water issues

In Shenzhen, we have all...

Subway station flooded

Maozhou River (茅洲河)

Red tide

Old diversion project

New diversion project
Physically based modeling

Freeze & Harlan (1969), in a classical paper, provide a blueprint for physically based hydrological modeling

- Continuity between groundwater flow and unsaturated flow
- Coupling with surface water
- Role of vegetation
- Influence of meteorological phenomena
Physically based modeling

New hydro. models (water & energy balances)

Typical hydro. models (water balance only)

Climate model (energy balance)

Free Troposphere

Atmospheric boundary layer, ABL

Surface layer

Variably saturated soil zone

Saturated zone

(By courtesy of Stefan Kollet)
Physically based modeling

From the subsurface into the atmosphere!

Water and energy equations (an example)

\[
\begin{align*}
\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T &= \frac{1}{\rho c_p} \left( \frac{\partial p'}{\partial t} + \ldots \right) + \frac{Q}{c_p} \\
\frac{\partial q_k}{\partial t} + \mathbf{v} \cdot \nabla q_k &= -\frac{1}{\rho} \left( \nabla \cdot \mathbf{J}^k + \ldots \right) - \frac{1}{\rho} \mathbf{I}^k \\
\left( \bar{q}_s - \bar{q} \right) &= \frac{E}{\kappa \mu \rho} \left[ \ln \left( \frac{z-d_0}{z_0} \right) - \Psi_{sw}(\zeta) \right] \\
\frac{\partial p_s}{\partial t} &= \nabla \bar{v}_p \cdot \bar{q}_s - q_r(x) - q_s(x) \\
S_s S_w \frac{\partial p}{\partial t} + \phi \frac{\partial S_w(p)}{\partial t} &= \nabla \cdot \mathbf{q} + q_s \\
q &= K_v k_r(p) \frac{\partial (p+z)}{\partial z}
\end{align*}
\]
Model applications

- Hydropower production
- Agricultural development
- Soil and water conservation
- Flood control
- Watershed restoration
- Water quality protection
- Drinking water supply

Receive wide applications!

(Pics are from the internet)
Big Earth Data

Far space satellites
- Geosynchronous meteorological satellites

Near space satellites
- Polar orbiting meteorological satellites

Airborne sensors
- Aircraft weather radar
- Weather balloon
- Airborne lidar
- Drone

Radar-based
- X-band
- S-band
- Ground penetrating radar
- Water level radar

Ground level instruments
- Weather station
- Solar radiation
- Soil moisture
- Groundwater
**Big Earth Data**

**Multisource (variety)**
- Plain text
- Point data
- Polyline data
- Polygon data
- Raster data
- ...

**Arguably big data!**

**Large volume**

High-frequency measurements of certain variables (velocity)

<table>
<thead>
<tr>
<th>Satellites</th>
<th>Velocity (Mbps)</th>
<th>Volumes (GB/Day)</th>
<th>Volumes (TB/Year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HJ-1B</td>
<td>60</td>
<td>57</td>
<td>20.32</td>
</tr>
<tr>
<td>HJ-1A</td>
<td>120</td>
<td>114</td>
<td>40.63</td>
</tr>
<tr>
<td>ZY-03</td>
<td>900</td>
<td>498.38</td>
<td>176.22</td>
</tr>
<tr>
<td>HJ-1C</td>
<td>320</td>
<td>187.5</td>
<td>66.83</td>
</tr>
<tr>
<td>ZY-02C</td>
<td>320.00</td>
<td>175.78</td>
<td>62.66</td>
</tr>
<tr>
<td>SPOT-4</td>
<td>50.00</td>
<td>10.99</td>
<td>3.92</td>
</tr>
<tr>
<td>LANDSAT-5</td>
<td>85.00</td>
<td>28.02</td>
<td>9.99</td>
</tr>
<tr>
<td>RADASAT-2</td>
<td>105.00</td>
<td>57.68</td>
<td>20.56</td>
</tr>
<tr>
<td>RADASAT-1</td>
<td>105.00</td>
<td>57.68</td>
<td>20.56</td>
</tr>
<tr>
<td>SPOT-5</td>
<td>100.00</td>
<td>54.93</td>
<td>19.58</td>
</tr>
<tr>
<td>ENVISAT</td>
<td>100.00</td>
<td>32.96</td>
<td>11.75</td>
</tr>
<tr>
<td>IRS-P6</td>
<td>210.00</td>
<td>46.14</td>
<td>16.45</td>
</tr>
<tr>
<td>LANDSAT8</td>
<td>440.00</td>
<td>241.70</td>
<td>86.15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3712.98</strong></td>
<td><strong>2089.06</strong></td>
<td><strong>574.6</strong></td>
</tr>
</tbody>
</table>
### NASA’s Water Cycle Missions

<table>
<thead>
<tr>
<th>Hydrological variable</th>
<th>Missions/instruments</th>
<th>Spatial resolution (km)</th>
<th>Temporal resolution (days)</th>
<th>Launch year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>GPM</td>
<td>5</td>
<td>0.125</td>
<td>2014</td>
</tr>
<tr>
<td>Snowfall</td>
<td>GPM</td>
<td>5</td>
<td>0.125</td>
<td>2014</td>
</tr>
<tr>
<td>Snowfall</td>
<td>Terra/MODIS</td>
<td>0.5</td>
<td>1</td>
<td>1999</td>
</tr>
<tr>
<td>Evaporation</td>
<td>Aqua/MODIS</td>
<td>0.5</td>
<td>1</td>
<td>2002</td>
</tr>
<tr>
<td>Evaporation</td>
<td>Suomi/VIIRS</td>
<td>0.5</td>
<td>1</td>
<td>2013</td>
</tr>
<tr>
<td>Runoff</td>
<td>SWOT</td>
<td>0.1</td>
<td>11</td>
<td>2021</td>
</tr>
<tr>
<td>Snow cover</td>
<td>Aqua/MODIS</td>
<td>0.5</td>
<td>1</td>
<td>2002</td>
</tr>
<tr>
<td>Snow cover</td>
<td>Suomi/VIIRS</td>
<td>0.5</td>
<td>1</td>
<td>2013</td>
</tr>
<tr>
<td>Surface soil moisture</td>
<td>SMOS</td>
<td>36</td>
<td>3</td>
<td>2009</td>
</tr>
<tr>
<td>Surface soil moisture</td>
<td>SMAP (radiometer)</td>
<td>36</td>
<td>3</td>
<td>2015</td>
</tr>
<tr>
<td>Surface soil moisture</td>
<td>ASCAT</td>
<td>25</td>
<td>1</td>
<td>2006</td>
</tr>
<tr>
<td>Surface soil moisture</td>
<td>GCOM-W/AMSR2</td>
<td>50</td>
<td>1</td>
<td>2012</td>
</tr>
<tr>
<td>Deep soil moisture</td>
<td>Biomass</td>
<td>0.2</td>
<td>18 days/yr</td>
<td>2021</td>
</tr>
<tr>
<td>Deep soil moisture</td>
<td>Jason-3</td>
<td>10</td>
<td>10</td>
<td>2016</td>
</tr>
<tr>
<td>Surface water elevation</td>
<td>SARAL</td>
<td>10</td>
<td>35</td>
<td>2013</td>
</tr>
<tr>
<td>Surface water elevation</td>
<td>SWOT</td>
<td>0.1</td>
<td>11</td>
<td>2021</td>
</tr>
<tr>
<td>Surface water elevation</td>
<td>ICESat-2</td>
<td>1.5</td>
<td>90</td>
<td>2018</td>
</tr>
<tr>
<td>Terrestrial water storage</td>
<td>GRACE</td>
<td>220</td>
<td>30</td>
<td>2002</td>
</tr>
<tr>
<td>Terrestrial water storage</td>
<td>Terra/MODIS</td>
<td></td>
<td></td>
<td>1999</td>
</tr>
<tr>
<td>Terrestrial water storage</td>
<td>Aqua/MODIS</td>
<td></td>
<td></td>
<td>2002</td>
</tr>
<tr>
<td>Terrestrial water storage</td>
<td>Suomi/VIIRS</td>
<td></td>
<td></td>
<td>2013</td>
</tr>
<tr>
<td>Terrestrial water storage</td>
<td>Landsat 8</td>
<td></td>
<td></td>
<td>2013</td>
</tr>
<tr>
<td>Terrestrial water storage</td>
<td>Landsat 9</td>
<td></td>
<td></td>
<td>2023</td>
</tr>
<tr>
<td>Vegetation/land cover</td>
<td>ISS/ECOSTRESS</td>
<td>0.07</td>
<td>4</td>
<td>2018</td>
</tr>
<tr>
<td>Vegetation/land cover</td>
<td>Aqua/AIRS</td>
<td>13.5</td>
<td>1</td>
<td>2002</td>
</tr>
</tbody>
</table>

Source: The future of Earth observation in hydrology, Hydrol. Earth Syst. Sci., 21, 3879-3914
Data-driven modeling

Data-driven hydrological modeling using machine learning techniques has been practiced for nearly two decades.

**Techniques**
- Artificial Neural Network (ANN)
- Support Vector Machine (SVM)
- Fuzzy Inference System
- Genetic Programming
- Regression Tree
- Gaussian Process Regression
- ...

**Applications**
- Rainfall-runoff modelling
- Sediment yield forecasting
- Evaporation forecasting
- Lake and reservoir water level prediction
- Drought forecasting
- ...

**Ad hoc applications**
Outline

1. Background
2. Three study cases
3. Concluding remarks
Some challenges

- ‘Big’ in the air, but ‘small’ at the ground.
- Direct measurements are still limited.

- Physically based models involve significant uncertainty, due to data gap and model structure.
- Data-driven models have higher accuracy in many cases, but with low transferability.

- Physically based models can offer deep insights, but are computationally expensive.
- Data-driven models are cheap to run, but can only produce case-dependent results.
Research opportunities

Key questions

What kind of roles could *opportunistic sensing* and *crowdsourcing* play in hydrological modeling?

What kind of roles could *big earth data* and *deep learning* play in hydrological modeling?

How to *merge the strengths* of both physically based and data-driven models?
A high-resolution, low-cost, ground-level precipitation monitoring network has yet to be developed.

Can existing closed-circuit television (CCTV) surveillance cameras be used for opportunistic sensing?

---

**Case 1: Camera-based Rain Gauge**

<table>
<thead>
<tr>
<th>Principle</th>
<th>Rain gauge</th>
<th>Disdrometer</th>
<th>Radar</th>
<th>Satellite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>point</td>
<td>point</td>
<td>area</td>
<td>area</td>
</tr>
<tr>
<td>Measured height</td>
<td>1 m</td>
<td>1 m</td>
<td>5 - 20 km</td>
<td>500 km</td>
</tr>
</tbody>
</table>

---
Case 1: Camera-based Rain Gauge

A new approach to measuring rainfall intensity in *real-world conditions* based on videos acquired by *ordinary* surveillance cameras.

**Part 1: Rain streaks identification**
Separate rain streaks from the rain-free background

**Part 2: Rainfall intensity estimation**
Estimate rainfall intensity from the identified rain streaks

(Jiang et al., Advancing opportunistic hydrology sensing: a novel approach to measuring rainfall with ordinary surveillance cameras, 2018, submitted to *Water Resources Research*)
Algorithm 1: Rain-streak identification

Not only based on the difference between consecutive frames, but also the visual properties of each frame.

Decomposition: $\mathcal{O} = \mathcal{B} + \mathcal{R}$

A convex optimization problem

$$\min_{U, V, W, X, R} \lambda_1 \|U\|_1 + \lambda_2 \|V\|_1 + \lambda_3 \|W\|_1 + \lambda_4 \|X\|_1$$

s.t. $U = R, V = V_y R, W = V_x (\mathcal{O} - R), X = V_t (\mathcal{O} - R)$

Alternating direction method of multipliers (ADMM) used to solve the problem
Algorithm 1: Rain-streak identification

Case 1: Camera-based Rain Gauge

Parameter sensitivity

Parameter optimization

Comprehensive tests performed

Convergence speed

Impact of wind
Case 1: Camera-based Rain Gauge

Algorithm 2: Rainfall intensity estimation
Based on geometrical optics and microphysical characteristics of raindrops.
Case 1: Camera-based Rain Gauge

- An ordinary surveillance camera is installed on SUSTech campus
- Three different scenes were shot
- Five rainfall events with varying intensities were recorded in 2018 summer
- One frame per second
- The instantaneous rainfall intensity was estimated every 30 seconds

<table>
<thead>
<tr>
<th>Event ID</th>
<th>Date &amp; Time</th>
<th>Duration (mm)</th>
<th>Number of frames</th>
<th>Exposure time (s)</th>
<th>Image size (pixel) after cropping</th>
<th>Sampling volume (m³)</th>
<th>Cumulative rainfall (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>7/26/2018 09:27:56—11:17:00 7/26/2018</td>
<td>94.1</td>
<td>6,544</td>
<td>1/250</td>
<td>200 H×960 V</td>
<td>~14.8</td>
<td>10.89</td>
</tr>
</tbody>
</table>
Case 1: Camera-based Rain Gauge
Case 1: Camera-based Rain Gauge

Comparison with previous approaches

- Lower cost, more realistic scenes, and higher estimation accuracy
- Potential of using existing CCTV networks for the opportunistic hydrology sensing

<table>
<thead>
<tr>
<th>Aspects for comparison</th>
<th>This study</th>
<th>Approaches in comparison</th>
<th>Approaches in comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of rain streaks</td>
<td>Visual and temporal properties-based</td>
<td>Temporal property-based</td>
<td>Temporal property-based</td>
</tr>
<tr>
<td>Estimation of rainfall intensity</td>
<td>The unfocused raindrops, unsatisfying size-velocity relationship, are removed</td>
<td>The blur effect of unfocused raindrops is eliminated</td>
<td>The unfocused raindrops, recognized by photometric information, are removed</td>
</tr>
<tr>
<td>Device used and its approximate cost</td>
<td>Ordinary outdoor surveillance camera (EZVIZ™ CSi, $100)</td>
<td>Digital single-lens reflex camera (Canon™ EOS 550D, $800)</td>
<td>Professional handheld camcorder (Sony™ DSR-PD198P, $2000)</td>
</tr>
<tr>
<td>Scenes considered</td>
<td>Real-world background (roads with vehicles)</td>
<td>Static background (a wall without any turbulence)</td>
<td>Dynamic background (a corner in a yard)</td>
</tr>
<tr>
<td>Data size for calibration/validation</td>
<td>Five events (video length: 388 mins)</td>
<td>Four events (video length: 104 mins)</td>
<td>Three events (video length: 9 mins)</td>
</tr>
<tr>
<td>Accuracy (MAPE)</td>
<td>21.8%</td>
<td>26.0%</td>
<td>31.8%</td>
</tr>
</tbody>
</table>
We use **computer vision (CV)** to prepare model inputs for data-driven models, which represents **a new framework to assimilate big earth data into data-driven hydrological models**.

Case 2: Flow modeling using CV

On the northern margin of the Qinghai-Tibetan Plateau (Jiang et al., *A computer vision-based approach to fusing spatiotemporal data for hydrological modeling*, *Journal of Hydrology*, 2018, 25-40)

Input data:
- Precipitation
- Temperature
- Leaf area index

Target:
- Streamflow

(Jiang et al., *A computer vision-based approach to fusing spatiotemporal data for hydrological modeling*, *Journal of Hydrology*, 2018, 25-40)
Case 2: Flow modeling using CV

Traditional strategies cannot sufficiently account for the spatial heterogeneity of data fields.

- Use data at selected points
  - Spatial information is not fully exploited

- Average over individual sub-basins
  - Spatial information is not fully exploited

- Average over entire watershed
  - Spatial information is not exploited at all

- Use data information at every spatial unit (e.g., grid cell)
  - Curse of dimensionality
Case 2: Flow modeling using CV

Key idea: Use the CV technique to transform a data field into an one-dimensional feature vector. The feature vector is then used as the model input.

Two features in this case

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Intensity</th>
<th>Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Representing pixel values</td>
<td>Representing spatial variability</td>
</tr>
<tr>
<td>Feature descriptor</td>
<td>Brightness</td>
<td>Local binary pattern (LBP)</td>
</tr>
<tr>
<td>Feature vector</td>
<td>Average brightness values of a series of sub-images</td>
<td>Histograms of LBP values of a series of sub-images</td>
</tr>
</tbody>
</table>

8-bit images of data fields

Spatial pyramid representation for image features

Feature vectors with spatial information

$\begin{bmatrix}
p_1 & p_2 & \cdots & p_n
\end{bmatrix}$ $\begin{bmatrix}
t_1 & t_2 & \cdots & t_n
\end{bmatrix}$ $\begin{bmatrix}
l_1 & l_2 & \cdots & l_n
\end{bmatrix}$
Case 2: Flow modeling using CV

The prediction well matches the observation.
Case 2: Flow modeling using CV

**Short-term forecasting**
*(compared against existing data-driven models)*

**Long-term simulation** *(compared against existing process-based models)*

<table>
<thead>
<tr>
<th>Model</th>
<th>Input data</th>
<th>Testing / Validation period</th>
<th>NSE</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Vision Model (Exp. 6)</td>
<td>Precipitation, temperature, and preceding streamflow</td>
<td>2009-2012</td>
<td>0.706</td>
<td>This study</td>
</tr>
<tr>
<td>Computer Vision Model (Exp. 7)</td>
<td>Precipitation and temperature</td>
<td>2009-2012</td>
<td>0.715</td>
<td>This study</td>
</tr>
<tr>
<td>Geomorphology-Based Hydrological Model (Distributed model)</td>
<td>Precipitation, temperature, wind speed, sunshine hours, relative humidity, DEM, land use map and soil map</td>
<td>2003-2006</td>
<td>0.70</td>
<td>Yang et al. (2015)</td>
</tr>
<tr>
<td>Distributed Hydrology Model (Distributed Model)</td>
<td>Precipitation, temperature, wind speed, sunshine hours, relative humidity, DEM, land use map, and soil map</td>
<td>2002-2009</td>
<td>0.732</td>
<td>Zhang et al. (2018)</td>
</tr>
<tr>
<td>Soil Vegetation Model (Semi-distributed Model)</td>
<td>Precipitation, temperature, wind speed, relative humidity, DEM, land use map, and soil map</td>
<td>1997-2000</td>
<td>0.678</td>
<td>Z. Li et al. (2011)</td>
</tr>
<tr>
<td>SWAT (Semi-distributed Model)</td>
<td>Precipitation, temperature, wind speed, sunshine hours, relative humidity, DEM, land use map, actual evapotranspiration data</td>
<td>2008-2013</td>
<td>0.64</td>
<td>Gao et al. (2018)</td>
</tr>
<tr>
<td>Geomorphology-Based Hydrological Model (Distributed Model)</td>
<td>Precipitation, temperature, wind speed, sunshine hours, relative humidity, DEM, land use map, actual evapotranspiration data</td>
<td>2008-2013</td>
<td>0.64</td>
<td>Gao et al. (2018)</td>
</tr>
<tr>
<td>WEP-Heihe (Distributed Model)</td>
<td>Precipitation, temperature, wind speed, sunshine hours, relative humidity, maximum evaporation, DEM, land use map, soil map, water use data, socio-economic data (GDP and population)</td>
<td>1990-2002</td>
<td>0.78</td>
<td>Jia et al. (2009)</td>
</tr>
<tr>
<td>SWAT (Semi-distributed Model)</td>
<td>Precipitation, temperature, wind speed, relative humidity, DEM, land use map, and soil map</td>
<td>1997-2000</td>
<td>0.65</td>
<td>Li et al. (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1997-2000</td>
<td>0.678</td>
<td>Z. Li et al., (2011)</td>
</tr>
</tbody>
</table>

More stable performance for varying prediction lengths

The CV model is much easier to build, but achieves a comparable or even better performance.
The CV model outperforms traditional data-driven models in transfer learning, especially for flow peaks.

The flexibility in fusing multisource spatiotemporal data enables a systematic analysis of the roles of different predictors.
Case 3: Surrogate-based optimization

Analyses requiring numerous model evaluations, such as optimization analysis, uncertainty analysis, etc.

Surrogate modeling (meta-modeling) approaches

Surrogate model:

1) To fully or partially replace the original complex model during the analysis
2) Computationally much cheaper than the original model
3) Usually in form of a response surface, such as support vector machine (SVM), radial basis function (RBF), etc.
Case 3: Surrogate-based optimization

GSFLOW: a complex integrated surface water-groundwater model

Zhangye Basin (张掖盆地): ~9,106 km², where intensive agricultural irrigation competes with ecosystems for precious water resources
Surrogate-based Optimization for Integrated surface water-groundwater Modeling (SOIM), which couples SVM and SCE-UA.

**Framework of SOIM**

1. **Optimization problem**
2. **Complex hydro. model (e.g., GSFLOW)**
3. **Surrogate Model (Support Vector Machine, SVM)**
4. **Heuristic search (SCE-UA)**
5. **Evaluation of the original model on the optima**
6. **Converged?**
   - **Yes**
     - Pick the best solution
   - **No**
     - Update the surrogate models
     - Add some additional samples
     - Obtain 100-200 Initial samples

**Case 3: Surrogate-based optimization**

- Obtain 100-200 Initial samples
- Add some additional samples
- e.g., 200 initial samples, converges with less than 300 samples

Case 3: Surrogate-based optimization

SVM surrogates can adequately replace the original GSFLOW model in many aspects (i.e., for many output variables).

The strengths of physically based and data-driven models are merged here!

   \[ R^2 = 0.9974 \]

b. Total storage change (2002)
   \[ R^2 = 0.9959 \]

c. Flow available at the last diversion point (March 2002)
   \[ R^2 = 0.9996 \]

d. Runoff at Zhengyixia (2002)
   \[ R^2 = 0.9823 \]

e. Total storage change (2002)
   \[ R^2 = 0.9603 \]

f. Flow available at the last diversion point (March 2002)
   \[ R^2 = 0.9978 \]
Case 3: Surrogate-based optimization

- Spatially optimize the conjunctive use of river flow and groundwater for irrigation in the 18 irrigation districts
- A number of optimization schemes were addressed, considering different hydrological conditions, management concerns and constraints.

<table>
<thead>
<tr>
<th>Year</th>
<th>Water Stress</th>
<th>$\Delta S$</th>
<th>Scenario A1 (No Storage Decline)</th>
<th>Scenario A2 (Moderate Storage Decline)</th>
<th>Scenario B1 (No Flow Decrease)</th>
<th>Scenario B2 (Normal Year)</th>
<th>Scenario B3 (Water Allocation Curve)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>35%</td>
<td>0.717</td>
<td>$\Delta S &gt; 0$</td>
<td>$\Delta S &gt; -0.15$</td>
<td>R &gt; 0.717</td>
<td>R &gt; 0.95</td>
<td>R &gt; 0.818</td>
</tr>
<tr>
<td>2001</td>
<td>47%</td>
<td>0.69</td>
<td></td>
<td></td>
<td>R &gt; 0.69</td>
<td>R &gt; 0.95</td>
<td>R &gt; 0.640</td>
</tr>
<tr>
<td>2002</td>
<td>35%</td>
<td>0.967</td>
<td></td>
<td></td>
<td>R &gt; 0.967</td>
<td>R &gt; 0.992</td>
<td>R &gt; 0.992</td>
</tr>
<tr>
<td>2003</td>
<td>27%</td>
<td>1.192</td>
<td></td>
<td></td>
<td>R &gt; 1.192</td>
<td>R &gt; 1.310</td>
<td>R &gt; 1.310</td>
</tr>
<tr>
<td>2004</td>
<td>35%</td>
<td>0.819</td>
<td></td>
<td></td>
<td>R &gt; 0.819</td>
<td>R &gt; 0.871</td>
<td>R &gt; 1.220</td>
</tr>
<tr>
<td>2005</td>
<td>32%</td>
<td>0.985</td>
<td></td>
<td></td>
<td>R &gt; 0.985</td>
<td>R &gt; 1.220</td>
<td>R &gt; 1.220</td>
</tr>
<tr>
<td>2006</td>
<td>30%</td>
<td>1.104</td>
<td></td>
<td></td>
<td>R &gt; 1.104</td>
<td>R &gt; 1.316</td>
<td>R &gt; 1.216</td>
</tr>
<tr>
<td>2007</td>
<td>25%</td>
<td>1.186</td>
<td></td>
<td></td>
<td>R &gt; 1.186</td>
<td>R &gt; 1.531</td>
<td>R &gt; 1.531</td>
</tr>
<tr>
<td>2008</td>
<td>29%</td>
<td>1.089</td>
<td></td>
<td></td>
<td>R &gt; 1.089</td>
<td>R &gt; 1.363</td>
<td>R &gt; 1.363</td>
</tr>
</tbody>
</table>


tIt is defined as the proportion of annual irrigation demand in the total water input. The total water input refers to the boundary surface water inflow from the mainstream and tributaries plus the local precipitation.

Only hundreds of GSFLOW runs are necessary, and the total computing cost is reduced from years to days (on computer clusters).
Case 3: Surrogate-based optimization

This study presents a good case in which the strengths of both physically based and data-driven models are nicely merged in a real-world application.

An example of the optimization results
1. Background
2. Three study cases
3. Concluding remarks
A. The recent progress of big data and AI open a new door for the hydrology community which has been relatively conservative in the past.

B. Hydrologists have much to learn from data scientists.

C. Opportunistic sensing can help solve the problem of “small at the ground”, and deserves further studies.

D. Computer vision can bridge the gap between big earth data and hydrological modeling.

E. Physically based models and data-driven models should be friends, not enemies, and co-evolve.

F. The existing data are still not big enough to support deep learning in classic hydrological modeling.

G. How to make use of big data from other domains?
Acknowledgments

- Funding support from National Natural Science Foundation of China (NSFC) and the Shenzhen City.
- Collaboration by Professor Vladan Babovic at National University of Singapore.

Thank you! Questions?

SUSTech Campus