Hydrological Modelling in a data rich era
Some idea about modelling

Rigon R.
Warredoc, Perugia January 30, 2019
Aristotle said a bunch of stuff that was wrong. Galileo and Newton fixed things up. Then Einstein broke everything again. Now, we’ve basically got it all worked out, except for small stuff big stuff, hot stuff, cold stuff, fast stuff, heavy stuff, dark staff, turbulence, and the concept of time.
Hydrological models, like any Physics

requires data

Physics is based on observations (actually well designed experiments)

“In general, we look for a new law by the following process. First, we guess it (audience laughter), no, don’t laugh, that’s really true. Then we compute the consequences of the guess, to see what, if this is right, if this law we guess is right, to see what it would imply and then we compare the computation results to nature, or we say compare to experiment or experience, compare it directly with observations to see if it works.

If it disagrees with experiment, it’s wrong. In that simple statement is the key to science. It doesn’t make any difference how beautiful your guess is, it doesn’t matter how smart you are who made the guess, or what his name is… If it disagrees with experiment, it’s wrong. That’s all there is to it.”

R. Feyman

https://youtu.be/OL6-x0modwY
Maybe then, the question is

Feynman was to good not to cite it, but he referred to laws, which is exactly not the same as models. For models maybe, the question is:

Where do datasets can be integrated best?

The answer, obviously, depends on models construction, as we’ll try to clarify.
Once upon a time there were simple hydrological models. Emphasis was on predicting discharges.
The water budget?

The linear model

\[ \frac{dS_g}{dt} = J^\bullet(t) - Q_g^\circ(t) \]

\[ S_g(t) = kQ_g(t) \]

The linear model
Everything is known of the linear model. In particular

\[ Q(t) = \int_{t_0}^{t} e^{-(t-t_{in})/k} J(t_{in}) dt_{in} \]

and

\[ E[ t - t_{in} ] = k \]

Travel time 

* This formula is correct but a little misleading because it move the attention from the budget to the discharges only

** Is a little more complicate than that. See Botter et al., 2011 and Rigon et al., 2016
Does it know about geology?

Does it know about vegetation?

NO!

So any geological and vegetation measure is useless

Sam Cook, What a wonderful world

....
But I do know that I love you
And I know that if you love me, too
What a wonderful world this would be
Unless \( k = k(g, v) \) vegetation and geology are controllers of \( k \)
Because measures become important, either state variables or parameters must have a hook to measures. Not necessarily measures have to be the parameters or the state variables. But “functions” of it.

\[ k = k(g, \nu) \]

Measure are useless without a model to give them significance.
Someone says, “Well, data can do it all alone.”

In fact if

\[ \frac{dS_g}{dt} = J^*(t) - Q_g^*(t) \]

the storage variation is completely known (as soon as we have measures)

If we do it we will observe that

\[ \frac{dS_g}{dt} > 0 \]

so the model is wrong because water storage does not increase indefinitely. **Models are useless without measures.**
The model is not enough. This the classical answer to the problem

\[ J_{eff} \]

\[ S_g \rightarrow Q_g \]

\[ \frac{dS_g}{dt} \approx 0 \]

We lived with it for decades but it has quite a few problems.

We have to introduce a lot of hypothesis about what \( J_{eff} \) is.
What else can we measure?

We can see the morphology of the catchments*

*It was not that easy 40 years ago. No DEMs. No LIDAR at that time
In this way we can also have a spatial distribution of rainfall.

Fig. 1. Third-order basin with Strahler’s ordering system and its trapping state.
In principle we can accommodate: spatially distributed rainfall and geomorphic characteristics (the latter through k and contributing areas). The final discharge is the convolution of discharges along path. See Rigon et al., 2016a
Eventually people became able to measure travel times with tracers.

\[ E[t - t_{in}] = k \]

And measures messed up everything.


# again not so simple, indeed

fix it up distinguishing among uniform sampling and preferential sampling of water ages.

See also:


They say the $J_{eff}$ has to do with infiltration. So various models of infiltration were used.

Obviously when you introduce a model for infiltration, you introduce new variables and new parameters.

Complication grows, you need more measures
The queen evapotranspiration

This model is better

\[ \frac{dS_g}{dt} = J^\bullet(t) - Q_g^\bullet(t) - E_T(t) - Q_{sub}(t) \]

The problem now is undetermined, even if we assume

\[ S_g(t) = kQ_g(t) \]

Can we measure

\[ E_T(t), \quad Q_{sub}(t) \]

to close the budget?
But what is evaporation function of? 

Evaporation and transpiration are from 30% to 60% of the whole hydrological budget, depending on climate.
The traditional equation for $E_T$

$$E_T = \rho C_E \frac{\epsilon}{p} \overline{u} (e^* (z_0) - e(z))$$

Dalton’s law (turbulent transport of water vapor)

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_E$</td>
<td>Dalton number</td>
<td>[-]</td>
</tr>
<tr>
<td>$e$</td>
<td>Partial pressure of water vapor</td>
<td>[F L$^{-2}$]</td>
</tr>
<tr>
<td>$e^*$</td>
<td>Partial pressure of water vapor at equilibrium</td>
<td>[F L$^{-2}$]</td>
</tr>
<tr>
<td>$E_T$</td>
<td>Evapotranspiration flux</td>
<td>[E T$^{-1}$ L$^{-2}$]</td>
</tr>
<tr>
<td>$p$</td>
<td>Air Pressure</td>
<td>[F L$^{-2}$]</td>
</tr>
<tr>
<td>$\overline{u}$</td>
<td>Horizontal direction component of mean velocity</td>
<td>[L T$^{-1}$]</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.622</td>
<td>[-]</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Dry air density</td>
<td>[M L$^{-3}$]</td>
</tr>
</tbody>
</table>
Stationary energy budget (no temporal accumulation of energy in the control volume)

\[ R_S = \lambda E_T + H + R_{ul} \]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E_T)</td>
<td>evapotranspiration flux</td>
<td>([\text{M} \ T^{-1} \ L^{-2}])</td>
</tr>
<tr>
<td>(H)</td>
<td>Sensible heat flux</td>
<td>([\text{E} \ T^{-1} \ L^{-2}])</td>
</tr>
<tr>
<td>(R_{ul})</td>
<td>Upward longwave radiation</td>
<td>([\text{E} \ T^{-1} \ L^{-2}])</td>
</tr>
<tr>
<td>(R_s)</td>
<td>Shortwave radiation</td>
<td>([\text{E} \ T^{-1} \ L^{-2}])</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>specific enthalpy of vaporization</td>
<td>([\text{E} \ M^{-1}])</td>
</tr>
</tbody>
</table>
Turbulent transport of heat

\[ H = a \rho C c_p \bar{u} (T(z_0) - T(z)) \]

\[ a = \begin{cases} 
1 & \text{if soil or water surface} \\
2 & \text{if leaf} 
\end{cases} \]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C)</td>
<td>Evaporation factor</td>
<td>[(-)]</td>
</tr>
<tr>
<td>(c_p)</td>
<td>Specific heat capacity</td>
<td>([L^2 \ M \ T^{-2} \ \Theta^{-1}])</td>
</tr>
<tr>
<td>(T)</td>
<td>Temperature</td>
<td>([\Theta])</td>
</tr>
</tbody>
</table>
Because we have four unknown and three equations, we need a further equation

\[ e^*(z_0) - e(z) = \delta e(z) + \Delta(T(z_0) - T(z)) \]

Penman assumption (a linear Taylor’s expansion of saturated tension around the air temperature gives a fourth equation)*

Usually the water budget does not enter in this estimation. One ansatz is to include in \( C_E' \) also a linear dependence on water stress

\[ C_E = \frac{S_g(t)}{S_{max}} C_E' \]

With this in mind:

* Details at http://abouthydrology.blogspot.com/2019/01/material-for-geoframe-winter-school_12.html
The 4 traditional equation in become 5

\[
\frac{dS_g}{dt} = J^*(t) - Q^*_g(t) - E_T(t) - Q_{sub}(t)
\]

\[
E_T = \rho C_E \frac{\epsilon}{p} (e^*(z_0) - e(z))
\]

\[
R_S = \lambda E_T + H + R_{ll}
\]

\[
H = \alpha \rho C_P \bar{u} (T(z_0) - T(z))
\]

\[
e^*(z_0) - e(z) = \delta e(z) + \Delta (T(z_0) - T(z))
\]

One important note is that they should be solved simultaneously. But, if we forget it we can solve the last 4 and then balance the results.

The 4 system solved

\[ 2 \quad T_\Delta = \frac{\gamma}{C_E \Delta + a C \gamma} \left( \frac{R_s - R_{ll}}{\rho \bar{u} c_p} \right) - \frac{C_E}{C_E \Delta + a C \gamma} \delta e(z) \]

\[ 4 \quad e_\Delta = \frac{\gamma \Delta}{C_E \Delta + a C \gamma} \left( \frac{R_s - R_{ll}}{\rho \bar{u} c_p} \right) - \frac{C}{C_E \Delta + a C \gamma} \delta e(z) \]

\[ 3 \quad H = \frac{a C \gamma}{C_E \Delta + a C \gamma} \left( R_s - R_{ll} \right) - \rho \bar{u} c_p \frac{a C C_E}{C_E \Delta + a C \gamma} \delta e(z) \]

\[ 1 \quad \lambda E_T = \frac{C_E \gamma}{C_E \Delta + a C \gamma} \left( R_s - R_{ll} \right) - \rho \bar{u} c_p \frac{C C_E}{C_E \Delta + a C \gamma} \delta e(z) \]

\[ \gamma := \frac{c_p \rho}{\epsilon \lambda} \] is the so-called psychrometric constant

* Some details at http://abouthydrology.blogspot.com/2019/01/material-for-geoframe-winter-school_12.html
Figure 10. Coupled energy and water budgets. These are the same budgets as shown in Figure D.1 with the addition of a new type of arc (dotted segments ending in empty squares). These arcs connect the same variables present in both budgets. In this case, the $J$'s are inputs while the $E_T$'s and $Q_G$'s are unknown variables, which must be solved simultaneously in both budgets. Because $E_T$ depends on radiation, a controller exiting from the $U_G$ place is added to reveal this further influence of the energy budget on the water budget. Other controllers of the system can be the LAI and hydraulic conductivity, which can be thought to influence flow $Q_G$.

A coupled simple model, like the one we solved the equation

This solution requires a few new measurements:

- Radiation
- Temperature of air
- horizontal wind velocity
- Vegetation cover

With these we estimate

- Leaf/Soil* temperature
- air vapor pressure
- Evapo-Transpiration
- Turbulent transport of thermal energy

We have to measure more things but we get also more state variables known.

I confess I am hiding quite a few details here.
Now, a less than **minimal** model has more than one reservoirs for HRU

The BST model.

The HBV model.

In conclusion, in **contemporary modelling**, all information is certainly absorbed to obtain a **more complete and clear description of the system**. Parameters heterogeneity require it all.

**No chance to have too much data!**

if we want to have a clear view of the system
All these models I presented assume some granularity of the processes. In a perfect world this granularity should be obtained by actions that mimics *Statistical Mechanics* in which laws and fluxes *emerge* from cancellation of degrees of freedom.

In practice we have to rely on appropriately designed *Statistics* and *Calibration*, which always depends on *Data*. Calibration of complicated models is more complicated*.

Calibration and analysis in portion is possible using the graph-structured organization of models*

*Serafin, F., (David, O & Rigon, R.) Enabling modeling frameworks with surrogate modelling capabilities and complex networks, Ph.D. Dissertation, 2019
Building such models & the issue of reusability

http://abouthydrology.blogspot.com/2019/01/pictures-from-winter-school-on-geoframe.html

R. Rigon
That’s why we use conservation laws.

They offer a big constraint on variability

The case of GLEAM v3 (model for tall vegetation)

- lateral fluxes + satellite data of soil moisture (assimilation)

Conclusions

• Models and data get along.
• Better data implies more refined models
• More refined models requires more data, specific data though.
• With data, spatial heterogeneity can be accounted for explicitly.
• Remote sensing alone is nothing.
• Budgets are what really count, because conservation laws imply constraints on dynamics.

Well, this is not the end, is just the beginning.

• Studies of fluxes of information and entropy
• Distinguishing complication from complexity
• Besides having stuff sorted out the right way, for the right reasons, which is the usual business
“While Silberstein (2006) insists that more data are needed and that models without data are not science, Anderson (2008) claims that there is already a deluge of data that can change the way science is done. Google and other similar efforts produce evidence that by analyzing huge arrays of data only, we can actually build new theories based only on correlation, ignoring causation. The new analytical tools that are going to be developed for petabyte computing in the “clouds” will require entirely different approaches to integration than the types of model integration that we were considering so far.”


Find this presentation at

Ulrici, 2000?

Other material at

http://abouthydrology.blogspot.com