Prediction in ungauged basins

The challenge of catchment non-stationarity

Howard Wheater

CERC & Director, Global Institute for Water Security
With acknowledgements to:

- Neil McIntyre (Imperial College London, UK)
- Natasha Bulygina (Imperial College London, UK)
- Caroline Ballard (Imperial College London, UK)
- Bethanna Jackson (Victoria University Wellington, NZ)
Regionalization of stationary systems

- Major progress has been made in methods to regionalize hydrological models
- Recognition of issues of parameter identifiability has led to use of parsimoneous conceptual models
- Alternative methods have been developed based on:
  a) Relationships between model parameters and catchment characteristics
  b) Transfer of ensembles of parameter sets from ‘donor’ catchments
Data set of 293 UK catchments (daily data, > 15 years)
Comparison of locally optimised and regression model validation performance

NSE* locally optimised model

NSE* regression (all)

NSE* regression (L)

FSBM locally optimised model

FSBM regression (all)

FSBM regression (L)
Multiple parameter sets can be used from donor catchments, conditioned on:

a) **Prior likelihoods** based on calibration performance. A number of models $N$ per gauged catchment may be used

$$P_i = \frac{\left( \text{NSE}_i - \text{NSE}_{\text{min}} \right)/(1 - \text{NSE}_{\text{min}})}{\sum_{i=1}^{N} \left( \text{NSE}_i - \text{NSE}_{\text{min}} \right)/(1 - \text{NSE}_{\text{min}})}$$

for $\text{NSE}_i \geq \text{NSE}_{\text{min}}$

b) **Similarity weighting.** Consider a number $S$ of gauged catchments to be feasible ‘donor’ catchments; weight their influence by catchment similarity

$$B_j = \frac{\left( 1 - E_j / E_{\text{max}} \right)}{\sum_{j=1}^{S} \left( 1 - E_j / E_{\text{max}} \right)}$$

for $E_j \leq E_T$
Posterior likelihoods combine prior likelihoods (P) and similarity weighting (B) of parameter sets from donor sites

\[ W_{i,j} = \frac{P_{i,j} B_j}{\sum_{j=1}^{S} \sum_{i=1}^{N} P_{i,j} B_j} \]

Weighted average streamflow is thus derived

\[ \bar{Q}(t) = \sum_{i=1}^{N \times S} Q_i(t) \times W_i \]

How do we represent non-stationarity e.g. land use/land management change?

A major UK programme (FRMRC) has been examining effects of agricultural intensification on flood risk

- Analysis of national catchment scale data was unable to identify effects
- Process-based modelling has been needed to evaluate effects of field-scale management interventions
- Extension to ungauged sites has been investigated using process-based and conceptual modelling approaches
Upscaling Strategy
- The case for data-poor sites

Information about local response

Hillslope model (1d, 2d, 3d)
Parameter sets \( \theta \)

Meta model
Parameter sets \( \mu \)

Catchment scale model
Parameter sets \( \xi \)

Model structure and process mappings

Knowledge of processes and properties

Regionalised Data
i.e. HOST, Curve no.

Observations

INPUTS \( \mu \) OUTPUTS
Data-rich site
The Pontbren multi-scale experiment, Wales, UK
Physics-based Modelling Strategy

- Reproduce experimental observations at the plot and hillslope scales using detailed physically-based models
- Explore *local* effects of management strategies
- Capture detailed model response with meta-model structure at the scale of fields and hillslopes
- Develop semi-distributed catchment scale model, using meta-model for individual elements
- Investigate catchment-scale effects of land-use change
Physics-based model: Field-scale runoff for different land use types, with uncertainty bounds
Meta-model structure
Meta-model performance (woodland response)
- meta-models work!

Detailed and catchment model responses, woodland

- detailed model response
- catchment model response

Nov05  Dec05  Jan06
0
0.1
0.2
0.3
0.4
mm per timestep
Catchment modelling: Pontbren
Scenario comparisons

Flow at Starflow SF5

- 1990s landscape (mostly improved grassland, negligible hedge maintenance/tree planting)
- Present day landscape with Pontbren consortium plantings/maintenance included
- Landscape with shelter belts within all improved grassland fields
- Scenario assuming 100% woodland cover
Data sparse site
The Hodder Catchment, N.W. England

Peat models presented in today's talk

Drained Peatland Detailed Model
General Model Setup

- Boussinesq Equation
- Kinematic Wave Equation
Water Table Results

Data from a surrogate site in Upper Wharfedale, provided courtesy of Professor Joe Holden, Leeds University
Grip Blocking

Modelled as reservoirs instead of Kinematic Wave
Simulations with scenarios

• Events analysis - 79 events for the year.

• 100 parameter sets – used for Drained, Blocked and Intact simulations of a 1 year period
Difference in Peak flows – Drained minus Intact

Distribution of the mean % difference in peak flows (Drained minus intact)

- Peak flows INCREASE following drainage
- Peak flows DECREASE following drainage

Largest Events → Smallest Events

Difference in Peak flows – Drained minus Blocked

Distribution of the mean % difference in peak flows (Drained minus Blocked)

- Peak flows DECREASE following blocking
- Peak flows INCREASE following blocking
Peatland summary

- Physics-based modelling conditioned on surrogate data has been used to explore impacts of management interventions.
- Drainage of peatlands leads to an increase in the largest and smallest flows.
  - The effect of drain blocking on flooding is dependent on local conditions, increasing and decreasing flow peaks.
  - The model can be used to prioritise drain blocking activities to provide the greatest benefit in terms of peak flow reduction.
  - The model has been applied at catchment scale using the meta-modelling strategy defined above.
Bayesian conditioning of hydrological models using regionalised indices

- Model parameters are sampled from the feasible parameter space
- Regionalised indices are available as a function of soils (BFI HOST) and land management (CN)
- Parameter sets are weighted according to the consistency of model performance with the predicted indices
Bayesian parameter conditioning: data

In regionalisation
D = areal physical properties, \textit{not} direct response observations

Here, we consider
D = \{soil hydrological type (HOST), land use\}

\[ L(\vartheta \mid D) -? \]
Bayesian parameter conditioning: likelihood

Each soil type and land use are represented via *behavioural indices*:

- Base flow index \((BFI_{HOST})\)
- Curve number \((CN_{SCS})\)

\[
L(\theta | D) = L(\theta | BFI_{HOST}, CN_{SCS})
\]
Proportion of baseflow (BFI) can be estimated from soil type (HOST) using a UK regional relationship.
Behavioural indices: SCS Curve Number (CN)

*Curve Number* relates rainfall volume to direct surface runoff amount.

\[
Q = \frac{(P - 0.2S)^2}{P + 0.8S}
\]

\[
S = 25.4 \times \left( \frac{1000}{CN} - 10 \right)
\]

CN is available as a function of soil type and land management.
## Selected curve numbers

<table>
<thead>
<tr>
<th>Land use</th>
<th>Hydrological soil group</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pasture</strong>¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td></td>
<td>68</td>
<td>79</td>
<td>86</td>
<td>89</td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td>49</td>
<td>69</td>
<td>79</td>
<td>84</td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td>39</td>
<td>61</td>
<td>74</td>
<td>80</td>
</tr>
<tr>
<td><strong>Woods</strong>²</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td></td>
<td>45</td>
<td>65</td>
<td>77</td>
<td>83</td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td>36</td>
<td>60</td>
<td>73</td>
<td>79</td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td>30</td>
<td>55</td>
<td>70</td>
<td>77</td>
</tr>
</tbody>
</table>

¹ Poor: heavily grazed with no mulch.  
Fair: not heavily grazed.  
Good: lightly or only occasionally grazed.

² Poor: forest litter, small trees, and brush are destroyed by heavy grazing or regular burning.  
Fair: woods are grazed but not burned, and some forest litter covers the soil.  
Good: woods are protected from grazing, and litter and brush adequately cover the soil.
Model structure - PDM

5 parameters

1. Precipitation
2. Evapotranspiration
3. Water excess
4. Fast flow store $k_f$
5. Slow flow store $k_s$
6. Runoff

Cumulative distribution of storage capacity

$C_{max}$

$1 - \alpha$

$\alpha$

$\alpha$
Plynlimon paired catchments, Wales, UK
Parameter restrictions

*HOST class 15
Flow predictions
# Nash-Sutcliffe efficiency

<table>
<thead>
<tr>
<th>Parameter estimation</th>
<th>Severn</th>
<th>Tanllwyth</th>
<th>Hafren</th>
<th>Hore</th>
<th>Wye</th>
<th>Gwy</th>
<th>Cyff</th>
<th>Iago</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regionalisation</td>
<td>0.74</td>
<td>0.70</td>
<td>0.73</td>
<td>0.73</td>
<td>0.76</td>
<td>0.77</td>
<td>0.8</td>
<td>0.76</td>
</tr>
<tr>
<td>Calibration</td>
<td>0.78</td>
<td>0.74</td>
<td>0.75</td>
<td>0.76</td>
<td>0.85</td>
<td>0.81</td>
<td>0.88</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Land use effects can be simulated

green = Severn
blue = Wye
Conclusions from regionalisation study

- Soil type and land use are used to restrict model parameter space via regionalised BFI and CN

- BFI and CN are only partially informative for parameters

- Other regionalised behavioural indices are needed

- The proposed regionalisation:
  - significantly reduced prediction uncertainty
  - was comparable with calibrated model predictions
  - allows land use effects estimation
Modelling changing land use - conclusions

• Physics-based models can provide important insights into non-stationary responses, but uncertainty must be recognised and much work remains to be done to explore the limits of predictability

• New meta-modelling methods provide a computationally efficient way to represent local scale complexity in large scale models

• Use of hydrological indices to condition conceptual models has proved surprisingly effective

• The CN method has potential for use in conditioning models to represent land management effects – but without more research to demonstrate local (UK) validity, results are speculative