Model-data fusion: using observations to understand and reduce uncertainty in hydrological models

A.I.J.M. Van Dijk

CSIRO Water for a Healthy Country National Research Flagship, CSIRO Land and Water, P.O. Box 1666, Clunies Ross Street, Canberra, ACT 2601, Water Information Research and Development Alliance
Email: albert.vandijk@csiro.au

Abstract: Without corroborating observations, hydrological models should not be trusted. Hydrometric observations are the basis for any statistical or empirical hydrological model, and are widely used for ‘calibration’ – tweaking model parameter values to reduce the disagreement between model and observations. However there are other, less travelled paths to understand, quantify, and reduce hydrological model uncertainty using observations. A wealth of new observation types has become available from in situ sensor networks, airborne data collection and satellite remote sensing. With conventional hydrometric observation networks in decline, using these new observations is not merely of scientific interest but a bare necessity. This paper discusses opportunities and challenges to expand the application of model-data fusion techniques – using observations to make models behave better.

Assessing the outcome of future boundary conditions (predictions or hypothetical scenarios) on water resources remains the most common hydrological model application. New observations can be used to improve model inputs (e.g. spatial parameter maps) or be considered in calibration to derive a more robust parameter set. Confronting the model with new observations helps to assess model uncertainty and build our confidence in it, though not usually without introducing some tough questions. (Is past performance actually relevant to uncertainty in future predictions? What about unknown or unquantifiable uncertainty components; are partial uncertainty estimates better than none? How does one determine whether a model is too poor to suit its intended purpose; what is the alternative? What is the uncertainty in the observations?)

When looking into the past rather than the future, some of these questions disappear, while new ones emerge. Estimating past (and current) conditions is becoming increasingly important: as the basis of monitoring systems (of drought, flood, water resource generation, availability and use); to initialize forecasting systems; and to inform water resource accounting and situation analysis. The Australian Water Resources Assessment system (AWRA) is such an application, developed to support the Bureau of Meteorology’s water information services. Having a predictive model is no longer an objective in this case. Rather, models take on a subservient role to reconcile and interpolate the available observations. In this case, observations are still used before the model is run (in model configuration and calibration), but now can also be applied during the model run (via data assimilation) and/or afterwards (e.g. via bias correction). Available techniques vary from the mathematically near-trivial to the very complicated, but assumptions about error in both model and observations are always critical. The generation of large volumes of relevant observations and derived products (e.g. from time series remote sensing) grows opportunities for model-data fusion, but equally grows the conceptual, mathematical and computational challenges. Several examples are discussed.

Keywords: water resources, model-data fusion, data assimilation, hydrological model, remote sensing
1. ALL MODELS ARE WRONG, WHITHER WATER SCIENCE?

Without corroborating observations, hydrological models should not be trusted. In fact, strictly speaking, we can never trust our models. Oreskes et al. (1994) pointed out that we can certainly never claim to have ‘validated’ our models, because we already know that they must be wrong: no matter how detailed our model, no matter how many algorithms we string together, they will always remain a highly simplistic representation of real world complexity. Oreskes’ point was not new and may sound trite or semantic. However, they put it forward not as a philosophical fancy but as a practical warning: telling others we have ‘validated’ our model discourages them from appreciating the necessary crudeness and uncertainty of our models and their limited predictive power.

One could counter that ‘all models are wrong, but some are useful’ (Box, 1979). How do we decide which models are the useful ones? We would need to know ‘how wrong’ a model is, have a particular model purpose in mind, and then be able to determine at what point wrong become useless. The complexity of these deceptively simple questions can be mindboggling (e.g. what about ‘things we don’t know we don’t know’?). An optimistic conclusion can sometimes be argued, but rarely proven. Perhaps as a consequence, few modelling studies conclude that the model applied is useless. At best there may be a post-hoc justification that involves a standard disclaimer of non-fatal caveats and perhaps for good measure an inconclusive parameter sensitivity analysis, almost as a fig leaf to hide scientific embarrassment.

The article “Stationarity is dead, whither water management?” (Milly et al., 2008) and its reception by part of the hydrological community may illustrate a tendency towards uncritical modelling. Milly and co-authors claim that the idea that precipitation is stationary (i.e., it varies stochastically but without long-term changes) is a ‘foundational concept’ taught to generations of water engineers. Anthropogenic climate change has assassinated our foundational concept, our models are all useless, and water management is doomed. If their argument came as a surprise (it certainly promoted ‘(non-)stationarity’ as a hydrological research topic), then we must have collectively forgotten to test our model assumptions against observations. Surely, climate has always been non-stationary, over decades, centuries, and beyond – with or without anthropogenic climate change. Milly and co-authors argue that we should increase the sophistication of our modelling tools and embed them within ‘earth system simulators’. It would certainly create even more sophisticated ways for our models to be wrong, though not necessarily more useful: global climate models can reproduce some observed processes and phenomena quite impressively, but precipitation is not one of them. Instead, the historic rainfall record may be more instructive on how ‘non-stationary’ climate has been or is becoming. Increasingly, palaeoclimate predictors are identified, understood and interpreted to extend the frustratingly short gauge records (e.g. Gallant and Gergis, 2011). We may still want to consider climate model predictions, but perhaps should not do so without testing whether the models have any explanatory power.

The above was intended as an example only. Other examples of common but demonstrably poor modelling practices are not hard to find. I certainly don’t pretend to be without fault, but blame low community standards rather than poor education in foundational concepts. Perhaps we have become complacent in ignoring empiricism in our modelling: by itself, increasing model complexity will not reduce uncertainty; only observations can help to understand and reduce uncertainty (e.g. Andréassian et al., 2010). Sooner (if our modelling is of relevance) or later (if it is inconsequential), we may be held to account over our low standards and much of our modelling exposed as pseudo-science.

Against this dark background, I will now attempt to paint a bright future. In the remainder of this paper, I will propose how observations can be used at each step of the modelling process. I have used the term ‘model-data fusion’ in the title of this paper to capture the idea that bringing models and observations into closer agreement makes it possible to both better understand and reduce uncertainty.

2. THE SIX HABITS OF HIGHLY EFFECTIVE MODEL-DATA FUSION

With gratuitous reference to the title of a popular self-help book, I describe the modelling process in six steps (Figure 1). Together they form a cycle, to emphasise that each step generates insights about uncertainty and opportunities for improvement that together inform future modelling efforts. It is a Catch-22 of modelling that observations used to bring models closer to reality cannot be used once again to quantify this improvement. This idea is shown by the outer ring: agreement with observations increases with each step (the broadening blue part), but further possibilities for independent evaluation diminish (the narrowing white). In the sections that follow I will discuss each step, focusing on the role of underutilized or new observations.
Figure 1. Six steps of model-data fusion. The use of observations is shown in the outer ring.

3. MODEL FORMULATION: NEW OBSERVATIONS AT THE SCALE OF INTEREST

The formulation of predictive equations is obviously a critical step in any hydrological model prediction. Their development, encoding and combination into hydrological models are a major undertaking. There already is a range of hydrological model code and software, which potentially saves much effort. However, inevitably each model will reflect a specific set of assumptions about dominant hydrological processes, application type, time and space scales, boundary conditions, available observations, and so on. Where these provide an ill fit to the application at hand, and if there are observations available that can lead to a better model formulation, then opportunities for improved model formulation should not be dismissed easily.

Hydrological processes are characterised by only few physical principles; the laws of conservation and classical mechanics. However, our limited knowledge and information on system properties and boundary conditions (e.g. subsurface hydraulic properties, precipitation) means that these principles cannot be applied directly. As a consequence, even our most ‘physically-based’ hydrological models are essentially empirical, or phenomenological; derived from observation rather than theory. One of the main constraints on ‘physically-based’ modelling is the frequent lack of scalability of processes and properties at a feasible observation scale (e.g. a field plot or drill core) to the modelling scale (e.g. a catchment or aquifer). Despite this, the essentially mechanistic fallacy, that the most realistic model predictions are logically obtained by the most complex and detailed process descriptions, still appears widespread.

This is not to suggest that physical processes should be ignored, but their outcome needs to be observed, interpreted and described at the appropriate scale. Fortunately new observation technology increasingly provides opportunities to derive empirical process relationships directly at the scale of modelling. Examples of such technologies include rainfall radar, cosmic ray soil moisture sensors, and evapotranspiration (ET) measurement with eddy covariance equipment or scintillometers. Similarly, airborne laser ranging provides unprecedented opportunities to characterise surface topography and roughness and increasingly powerful geophysical techniques (e.g. magnetic resonance imaging) allow three-dimensional hydraulic characterisation of the subsurface. Satellite remote sensing provides additional opportunities, with often global coverage, ever increasing detail in time and space, and in some cases with an accuracy that is commensurate to field techniques (e.g. Van Dijk and Renzullo, 2011). There have been rapid advances in precipitation estimation...
from multiple satellites, surface elevation has been mapped globally at 30 metres, and total water storage
(combining surface, soil and groundwater storage) is now estimated every 10 days from satellite gravity
measurements. Several alternative remote sensing methods exist to measure open water and moisture at the
surface, vegetation properties and ET, and their estimates are increasingly merged to increase and
characterise measurement accuracy.

These new large-scale observations make it possible to derive and test new process equations directly at the
scale of interest, particularly when combined with hydrometric measurements already made at that scale (e.g.
catchment streamflow). This approach is sometimes referred to as ‘top-down’ modelling (Klemes, 1983) and
need not be any more or less ‘physically-based’ than ‘bottom-up’ approaches. The latter, assembling process
equations derived at smaller scales, tend to be impossible to reliably parameterise at larger scales, where they
require calibration of ‘effective’ parameter values and fall victim to equifinality problems. We should re-
evaluate our hydrological models and equations as soon as new, more appropriate observations become
available, and not hesitate to replace complexity with simplicity if the observations suggest an opportunity to
do so.

4. MODEL CONFIGURATION: SPATIAL MODELS REQUIRE SPATIAL INPUT DATA

Spatial models require spatial input data: as input ‘maps’ to configure model structure, parameterise surface
and subsurface properties (i.e. static boundary conditions), and as model forcing (i.e., dynamic boundary
conditions). Common static data requirements include topography, soil properties and land cover. New
spatial data sources are usually adopted fairly rapidly in hydrological modelling where they can substitute
existing data without too much additional development. Maps from field surveys and airborne photography
are increasingly being replaced or enhanced with remotely sensing, e.g. of surface elevation data and land
cover. It has proven more challenging to spatially estimate soil properties using remote sensing, but progress
is being made. For example, Warren et al. (in prep.) use remotely sensed ET dynamics to estimate plant
available water content across Australia.

Precipitation gauges and climate stations are likely to remain a foundation of hydrological modelling, but
where these data were previously interpolated using necessarily simple assumptions (e.g. Thiessen polygons,
kriging) new observations from rainfall radar and satellite remote sensing provide an opportunity to improve
spatial estimation, particularly in regions with few gauges. For example, Renzullo et al. (2011) blended
gauge observations with satellite-based rainfall products to improve precipitation estimation across Australia
(Figure 2). A caveat is that satellite instruments and derived products change over time, and going back in
time satellite products become fewer and less accurate. This caveat applies to all observations, but satellite
records are shorter than many conventional records. For example, satellite rainfall products are available
from 1998 onwards, whereas Australia already counted ca. 3,000 precipitation gauges a century before
(compared to ca. 6,500 in more recent years).

Figure 2. (a) current distribution of rainfall gauges in Australia; (b) areas where remote sensing can (white)
and cannot (black) improve daily precipitation estimates; (c) blended gauge-satellite rainfall grid for
February 2, 2011 (cyclone Yasi) (figures provided by L.J. Renzullo, CSIRO).
5. MODEL CALIBRATION: AVOIDING THE EQUIFINALITY TRAP USING MULTIPLE CONSTRAINTS

Where parameters cannot be estimated from theory or input data, they may need to be iteratively estimated to achieve a model with satisfactory predictive power – in hydrological vernacular referred to as ‘model calibration’ (in the statistical rather than instrument-related meaning of the word). Model calibration techniques are arguably more commonly applied, more developed, and more widely accepted in hydrology than in many other fields of environmental modelling. It may reflect a history of models that were sufficiently simple (i.e. had sufficiently few free parameters) to allow optimisation, and sufficiently poorly constrained (e.g. because of the mentioned scaling problems) to require it. By comparison, parameter optimisation is uncommon in weather prediction, for example: presumably mainly due to the complexity of posing the optimisation problem (the term ‘calibration’ is still used, but usually refers to the adjustment of model outputs, e.g. through bias correction).

As hydrological models have grown more complicated, the use of observation has not kept pace, resulting in the well published problem of under-determination or equifinality (Beven, 2004). Increasingly, new observations are being used to address this problem. For example, Zhang et al. (this volume) demonstrate how remotely-sensed surface soil moisture and leaf area index can be used along with streamflow data to derive a more robust model calibration. An important challenge in doing so is that most optimisation methods require a single cost (or objective) function. This requires that an appropriate metric of agreement is chosen for each of the observation sets, and that these metrics are subsequently combined into a single cost function. In principle, this can be avoided if the measurement errors in each of the observation types are understood and quantified in time and space. In practice, such information is not widely available, making it very challenging to define agreement metrics and relative weightings for unlike quantities in a statistically robust way. Another approach to address the equifinality problem is to make better use of prior knowledge to limit the tendency of the calibration process to throw up improbable parameter values or value combinations. For example, Lerat et al. (this volume) use Bayesian methods to incorporate prior knowledge into the calibration of a river model. Their analysis also challenges the (often implicitly made) assumption that model calibration necessarily improves predictive power.

6. MODEL SELECTION: THE NEED FOR BENCHMARK TESTING

Model selection here refers to the comparison of predictions from alternative models to the same set of independent observations in order to select the one(s) with the most desirable performance. (One may want to maintain an ensemble of plausible models.) Such a model selection process supports evidence-based decision making and serves as a benchmark to guide further model improvement. Alternative models may not differ only in structure (equations, code), but also differ in choices made in configuration (e.g. input data sources, model resolution or discretisation) and calibration.

Model selection through comparison against observations is uncommon, and comparison of alternative model structures less common again. Confronted with practicalities, one is tempted to simply select a model that (1) can answer the question (e.g., it can be applied in a way to represent the scenario(s) and produce the outputs of interest); (2) has a certain reputation (e.g. it has been documented, widely published, used previously for a similar purpose, etc); (3) is technically convenient (e.g. runs on a desirable computer operating system, has an intuitive interface, requires minimal additional training); and (4) is preferable on non-technical grounds (e.g. the individual, employer or client may have developed it, invested in a license, or have another vested interest in its use). Without dismissing the weight of such practical considerations, no pretences should be made about their scientific merit.

A formal comparison of models against observations is more desirable, but strewn with pitfalls. Some of these were discovered during development of the Australian Water Resources Assessment (AWRA) system. The AWRA system is being developed to support the Bureau of Meteorology’s water information services, in particular the production of annual water accounts and regular water resource assessments. It integrates on-ground and satellite observations with hydrological models and may be interpreted as a continental, 5-km resolution water resources monitoring system. As part of system development, selection and benchmark tests were performed to compare existing methods for the estimation of precipitation, ET, streamflow and groundwater recharge. These tests have increased transparency in model selection and provide an ongoing benchmark environment, but several challenges and ambiguities needed to be confronted. Firstly, the limited availability of observations meant that they were not always independent from the estimation methods; in some cases a subset of observations was used in model development or calibration. Secondly, model performance was the consequence of all three prior steps (cf. Figure 1): model development, configuration and calibration. An example for streamflow is shown in Figure 3a.
The five estimation methods that performed best were all configured with parameters obtained from calibration for a neighbouring catchment, whereas the three poorer performing models used a single, largely uncalibrated parameter set for all catchments. One might be tempted to conclude that what distinguished the five models was a better structure or a better calibration approach (the input data were comparable). However, calibration of one of the poorer performing models (still of a single uniform set of parameters, rather than using a nearest-neighbour approach) brought its performance in line with the other models (Figure 3b). Moreover, the graphs do not show that the nearest-neighbour calibration approach becomes comparatively less successful as the distance to the nearest catchment increases (Viney et al., in press). This demonstrates (1) that any model comparison can only lead to a statement on the compound effect of model structure, input data and calibration, unless one or more factors are purposefully controlled for; and (2) that performance metrics need to be interpreted with considerable caution.

7. DATA ASSIMILATION: TOWARDS HYDROLOGICAL REANALYSIS

Data assimilation techniques ‘nudge’ model estimates closer to observations, and can be used to improve forecasts, ‘now-casts’ (i.e., monitoring) or ‘hindcast’ (estimating past conditions). Data assimilation does not pre-empt the need for careful model development, configuration, calibration and selection, but provides a powerful additional step to improve estimates of past and current conditions, and thereby potentially forecasts of future conditions. Available techniques vary from the mathematically near-trivial (e.g. direct replacement of modelled states by observed values) to the very complicated (e.g. variational schemes). Assumptions about error in both model and observations are always critical. The generation of large volumes of relevant observations and derived products (e.g. from time series remote sensing) grows opportunities for model-data fusion, but equally grows the conceptual, mathematical and computational challenges.

Data assimilation is central to numeric weather prediction systems, which ingest an increasingly large range of global in situ and satellite observations, resulting in rapid advances in forecast skill. A well-developed application in streamflow forecasting is the adjustment of initial model state at the time of forecast by assimilating the latest streamflow observations. Output bias correction – correcting systematic deviations between model predictions and observations – may also be seen as a form of data assimilation. Now-casting applications of data assimilation are increasingly used in monitoring systems for drought, floods, water resource generation, availability and use (see review in Van Dijk and Renzullo, 2011). In this case, the initial conditions (or the ‘now-cast’) themselves are of most interest. The use of data assimilation to estimating past conditions is known as reanalysis in atmospheric modelling terminology. Here the initial state itself and the (re-)forecast up to the next assimilation step are of the most interest – they provide the best comprehensive estimate of atmospheric conditions at the time. Reanalysis data provide a great resource for climate research, and hydrological reanalysis are likely to be equally useful for water resources research. The mentioned...
Model-data fusion in hydrological modelling

AWRA system can be considered such a hydrological reanalysis system. Having a purely predictive model is no longer an objective in this case, rather, the models take on a subservient role to reconcile and interpolate the available observations. Data assimilation is not likely to have an important role in scenario modelling studies, although in some cases initialisation and output bias correction may be desirable.

8. MODEL UNCERTAINTY ASSESSMENT: MULTIPLE LINES OF EVIDENCE

All previously discussed steps provide information about the different sources of model uncertainty and ways to reduce it: model development provides information about structural uncertainty; model configuration about uncertainty from model resolution and boundary conditions; model calibration about parameter uncertainty; and model selection and data assimilation about uncertainty from all three sources combined. The inevitable consequence of a more intensive use of observations in each of these steps is that it prevents their use as an independent data source for model uncertainty assessment as a final modelling step, although careful data withholding experiments (e.g. leave-one-out or jack-knifing procedures) can minimise this problem.

Conversely, an incidental advantage of the currently relatively poor use of new observations is that it does leave them available for uncertainty assessment. Such an application was pursued in a recent project that aimed to assess the possible consequences of climate change and changes in land use and water management across the Murray-Darling Basin (the so-called Murray-Darling Basin Sustainable Yields project) (Van Dijk et al., 2008). For pragmatic reasons, the project needed to rely on the use of existing river models. To obtain some insight into model uncertainty, separate river water balance estimates were derived by combining satellite-based estimates of irrigation water use, wetland water use and flooding with streamflow and diversion records. The different data types were integrated by what may be considered a fairly unsophisticated form of reanalysis. While far from perfect and not fully independent (e.g. the hydrometric data used in the river water balance has partially also been used to calibrate river models), the exercise provided valuable insight into some important sources of model error and uncertainty.

For scenario applications, evaluation against observations can only provide partial information on uncertainty. Additional uncertainty comes from future boundary conditions (e.g. the scenarios and methods used to translate them into model input) and the possibility of wholly unanticipated responses (e.g. Refsgaard et al., 2006). In the mentioned project, which considered conditions over a 30-year time horizon, natural or ‘background’ climate variability was overall considered a greater source of uncertainty than any of the scenarios or hydrological model uncertainties. A quantitative integration of all uncertainties was not possible, but the fairly comprehensive attempt at uncertainty assessment made it possible to identify which projected outcomes suffered most from which source of uncertainty, with implications for water management planning, monitoring and research. For example, the unattributed losses in several of the lower river reaches were a major source of uncertainty for some projected outcomes (see Van Dijk et al., 2008, for details).

9. CONCLUSIONS

Few of the ideas brought forward in the previous sections are new, if any. My objective was to emphasise our scientific responsibility to use observations rigorously and exhaustively in our modelling. Even though practical water management questions frequently require us to apply our science to hypothetical (and hence untestable) future scenarios and in observation-poor environments, this does not absolve us of making the best use of observations we reasonably can. Model calibration is fairly entrenched in hydrological modelling, but there are several other, less travelled and sometimes more promising paths opportunities for model-data fusion, through development, configuration, calibration, selection, assimilation and evaluation. I described several recent or new applications and observations from in situ sensor networks and airborne and satellite remote sensing. With conventional hydrometric observation networks in decline, using these new observations is not merely of scientific interest but a bare necessity. Together, they provide unprecedented possibilities to understand and reduce uncertainty in hydrological modelling.

ACKNOWLEDGMENTS

There was insufficient space here to do full justice to the many publications that have previously stated the various points reiterated in this paper. I am indebted to the work of, among others, Beven, Bredehoeft, Grayson, Klemes, Oreskes, and Refsgaard. I also thank Luigi Renzullo, Damian Barrett, Neil Viney and Julien Lerat for inspiring discussions and suggestions. Development of the AWRA system is funded through the water information research and development alliance between CSIRO’s Water for a Healthy Country Flagship and the Bureau of Meteorology.
REFERENCES


