Qualitative modeling of the impact of the environment on early stages of salmon populations

F. Guerrin[†], **J. Dumas**[‡], **P. Davaine**[‡], **E. Beall**[‡], **O. Clément**[‡]
[†]Unité de Biométrie et Intelligence Artificielle, INRA, Toulouse, France
[‡]Station d'Hydrobiologie, INRA, St-Pée-sur-Nivelle, France
guerrin@toulouse.inra.fr

Abstract

This paper describes a qualitative model of the functioning of salmon *redds*, places where salmon reproduction occurs in the rivers. For this, we used *Qsim*, which appeared adequate for representing available qualitative knowledge of hydroecology experts.

Since the number of relevant variables was relatively large, it appeared necessary to decompose the model into two parts, corresponding to processes occurring at separate time-scales.

A qualitative clock allows us to submit the simulation of the development stage of salmon to the calculation of the moving sum of average daily temperatures, according to the clock ticks and a water temperature profile set by the user. Therefore, this introduces some way of real-time dating and duration in a purely qualitative model.

Simulating both sub-models by means of alternate QDE transitions, allows us to generate the evolutions of variables of interest, such as the mortality rates according to several factors (e.g. lack of oxygen), under various rainfall scenarios.

This work is part of a project aiming at assessing the impact of the environment on salmon populations dynamics, by the use of models of processes acting at different levels: catchment (e.g. erosion), river (e.g. sediment transport and deposition), and redds (e.g. fine particles sealing off). Only the latter level is dealt with in this paper.

Motivations and objectives

For piedmont rivers such as the Nivelle (Basque country, France), the survival rate at the egg and alevin stages has been shown critical for salmon, as it controls the population dynamics. This was evidenced by a stochastic model (called CBS) of the salmon lifecycle (Dumas *et al.*, 1996). It is assumed that the sealing of redds (spawning areas of salmon in the gravel substrate of the river bed), due to total suspended solids (TSS) transported by the flow, is responsible for a low natural survival rate, along with their destruction by floods. Filling up the empty volumes in the gravel layers, deposited particles lower the renewal of water and, consequently, decrease the amount of dissolved oxygen supplied to eggs and alevins. Also, the sealing of the upper gravel layer may prevent the fry from emerging. Understanding how human activities impact on redds is of great concern to managers of salmon populations. Therefore, in this project, we aim at modeling the known relationships between:

- the river catchment characteristics, both in terms of geographical features (climate, topography, geology, hydrology, pedology, vegetation cover, etc.) and of human activities impacting on erosion processes or TSS generation (land use, agriculture, urbanization, wastewater dispersal, etc.);
- the river hydrology and the functioning of redds according to their types and location.

For this, we need to exploit heterogenous information, ranging from numerical data to vague textual explanations and descriptions, taken either from experimental results or literature.

Involved processes

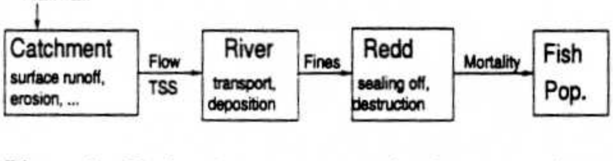


Figure 1: *Blocks of processes considered to assess the impact of the environment on a salmon population.*

Three blocks of processes (Figure 1) need to be distinguished as they relate to distinct spatial scales (catchment, river and redd) and pose different modeling problems.

Catchment level. Here, semi-quantitative (e.g. intervals) to ordinal, or even nominal, data are used to describe elementary features of the catchment (e.g. slopes, lithology, plant cover, land use). In some cases, empirical models may exist, like statistical regressions between rainfall and river flow, although such relations generally concern the whole river catchment and cannot be assessed for specific sub-parts located upstream of the redds. Hydrologic mechanistic models, describing surface runoff or erosion, are also impossible to use in practice so far as numeric parameters identification and calibration necessitate numerous field data, much too difficult to obtain from a non experimental catchment without adequate instrumentation.

River level. At this scale, one can find also mechanistic models (generally complex) or rough black-box models, to simulate TSS and sediment transportation by the river flow. Again, simulating these models requires heavy acquisition of field data, whereas it is known that, at most, they can only provide orders of magnitude of the parameters values, because of approximations made on the actual catchment hydrodynamics.

Redds. One can notice here a total lack of valid mathematical models due to fluid mechanics complexity in such fractured porous media (existence of fissures, non-constant porosity, etc.) whose characteristics are actually unknown. However, there exists an empirical knowledge, coming either from field measurements and observations made by hydro-ecologists themselves, or from the international literature (e.g. critical reviews by Chapman, 1988, or Lisle and Lewis, 1992). Knowledge often comes from correlations made on field data, from which only qualitative features may be exploited (e.g.: *"generally there exists a decreasing relation between the amount of fine particles in the redd and the overall survival rate"*). But no precise type of such functions may be derived from the data, as in practice data vary according to time, location, hydrological regime, fish species etc. In fact, quantifying the involved phenomena could only be achieved on specific heavily instrumented and controlled experimental facilities.

Summarizing, for such large-scale complex systems one has at hand only partial, incomplete, imprecise, quantitative data (which, in the specific case of the Nivelle river, are still being acquired), and therefore cannot simulate numerically available models. Therefore, qualitative modeling and simulation were thought important to make use, on the one hand of expert informal knowledge, on the other hand of some mechanistic hydrologic models through their conversion into qualitative ones. The first aspect only is dealt with in this paper.

Qualitative model of a redd (QMR)

We focused first on representing and simulating the functioning of a redd system. Our principal aim was to predict the survival rates of fish at different egg and alevin stages, under various rainfall scenarios. For the above mentioned reasons, along with the fact that taking into account the redd dynamics as well as the continuous nature of most of relevant variables was important (e.g. redd silting in is continuous over time, at least in a *normal* operation mode), we chose *Qsim* as a modeling and simulation tool.

Knowledge representation

As a preliminary step, a *naive* model (i.e. mainly based on common sense knowledge) was first implemented with the aim to assess the adequacy of Qsim in representing knowledge and providing sound simulations. Then, after a brief review of the literature on redds entrappment, few meetings with INRA hydro-ecologists were organized. They allowed us to formalize a conceptual model, in the form of an influence graph, summarizing the relationships among relevant variables to be taken into account (Figure 2). A preliminary version of a Qsim model was then rapidly implemented, then refined in further versions.

'Hat' constraints definition. We defined two new constraints (`H+` and `H-`) in order to capture functions exhibiting 'Hat' shaped curves. `H+/H-` syntax involves four bend points. Their semantics refer to successive combination of `S` constraints: e.g., `H-` (concave downward) has the meaning of an `S+` constraint followed, after a plateau (possibly reduced to an extremum point), by an `S-` constraint. Its syntax is as follows:

`((H- X Y (x1 y1) (x2 y2) (x3 y3) (x4 y4)))`

with each `(xi yi)` pair being the landmarks coordinates of points p_i. Notice that one *must* have `y2 = y3` (plateau); in addition one *may* have `x2 = x3` (bend points $p_2 = p_3$ if the plateau is restricted to a single point — i.e. a maximum). Also, the following order should hold among landmarks:

- X-axis: `x1` < `x2` < `x3` < `x4`
- Y-axis: `y1` < `y2` = `y3` and `y4` < `y2` = `y3` (whereas any order may be set between `y1` and `y4`).

`H+` (concave upward) is derived symmetrically.

Example: sensitivity to mortality factors. Two mortality factors are taken into account to date: the lack of dissolved oxygen and the fine sediments deposition on the upper gravel layer, constituting a physical barrier capable of preventing fry from emerging. Two other factors are soon to be integrated: shocks (e.g. when the flow regime becomes too turbulent) and water temperature (optimal range specific to each stage outside of which mortality is increased). Mortality due to any specific factor is expressed as the integration over time of $L = S \times P$ (for L: intensity of lethal factor, S: sensitivity to factor, P: factor pressure) sort of generalization of the Ohm's law

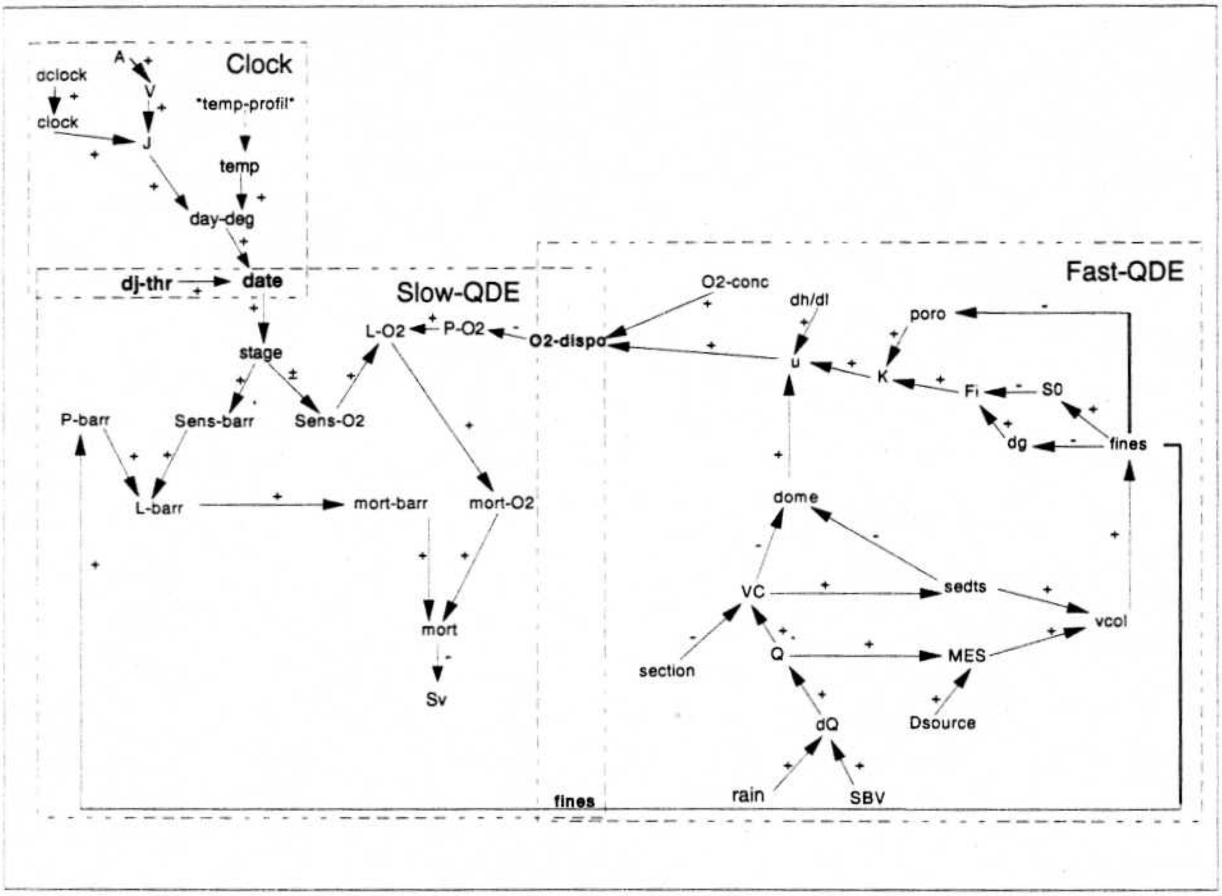


Figure 2: *Qualitative model of a redd (QMR): influence graph exhibiting the model decomposition into a slow-QDE, accounting for biological phenomena, and a fast-QDE, accounting for physical phenomena. Variables shared by both QDEs are boldface printed.*

$(I = \frac{1}{R} \times V)$.

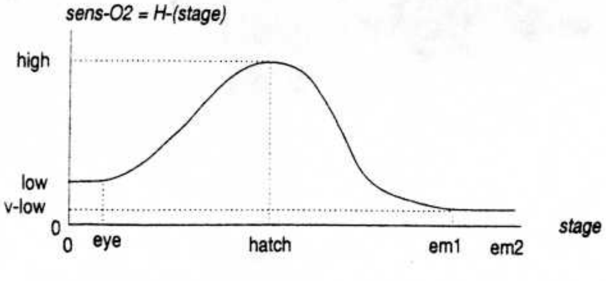


Figure 3: *Sensitivity of fish to the lack of oxygen vs. stage, computed as an H- constraint (0 on the stage axis denotes spawning).*

For example, the relationship between the sensitivity of fish to the lack of dissolved oxygen concentration as a function of their development stage, i.e. $sens{-}O_2 = f(stage)$ can be computed as an H- constraint (Figure 3):

```
((H- stage sens-O2 (eye low) (hatch high)
        (hatch high) (em1 v-low)))
```

stating that sensitivity, low at spawning, begins increasing once eyed larval stage is reached, up to a maximum at egg hatching, then decreases to the beginning of emergence and stays very low beyond.

Model decomposition

The first implemented version was an aggregate model (i.e. a single QDE) composed of more than 30 variables. It presented the major shortcoming of bringing together variables corresponding, in fact, to different time-scales:

- a *slower* scale (weeks), corresponding to the biological processes of fish development and related mortality;
- a *faster* scale (several hours or days), corresponding to the physical processes of redd sealing off under rainfall events.

The simulation was hence governed by the fastest scale, richer in event occurrences, which led the simulation to branch into numerous irrelevant distinct behaviors (several dozens according to the simulation time) from a single initial state. To solve that problem, the current version was made by partitioning this model into two connected QDEs via transitions on shared variables (principally dissolved oxygen rate and quantity of fine sediments). This decomposition was performed by hand (the causal structure was already known), not automatically as described by Clancy and Kuipers (1994). Thanks to this, a typical simulation generates only a few behaviors (3 behaviors in the example presented below) exhibiting relevant distinctions. Hence, the slower and the faster models are alternately simulated according to time-scale abstraction principles:

- the slower process views the faster one as instantaneous: continuation of the simulation is re-initiated from the latest state reached by the slow-QDE before transition;
- the faster process views the slower one as constant: all the variables of the slow-QDE are frozen (their `qdirs` are set to `std`) while the fast-QDE is simulated.

Exogenous (independent) variables are used as forcing functions, using unary constraints such as increasing, decreasing or constant.

Qualitative autonomous clock

It is known that changes in salmon development, i.e. passing from one stage to another, depend upon the moving sum of average daily water temperatures, expressed quantitatively as day-degrees. Stages are determined from spawning, taken as 0, according to the following thresholds:

- eyed larva: ≈ 250 day-degrees,
- hatching: ≈ 450 day-degrees,
- beginning of emergence: ≈ 900 day-degrees,
- end of emergence: ≈ 1100 day-degrees.

The *Timer* QDE, borrowed from Ben Kuipers' book (Kuipers, 1994 — pp. 228-229), based on QDE transitions, generates clock discontinuous ticks and their count over time. It was modified by the addition of two variables:

- a profile of average (quantitative) temperatures for each 10-days period, computed as a list of integers stored into a global variable set by the user (for sake of simplicity, values here are rounded up to the nearest 5-degrees multiple);
- the day-degrees amount, calculated as the moving sum of temperatures holding successively for each 10-days period (from the latter variable).

Two specific Lisp functions have been written to (i) pop elements up of the list of temperatures at each transition state created during the simulation, and (ii) sum them up in order to build dynamically the day-degrees variable quantity space (initialized in the define-qde clause as `(0 inf)`).

Simulating the clock motion (Figure 4) generates, in addition to the ticks sequence and their count, the

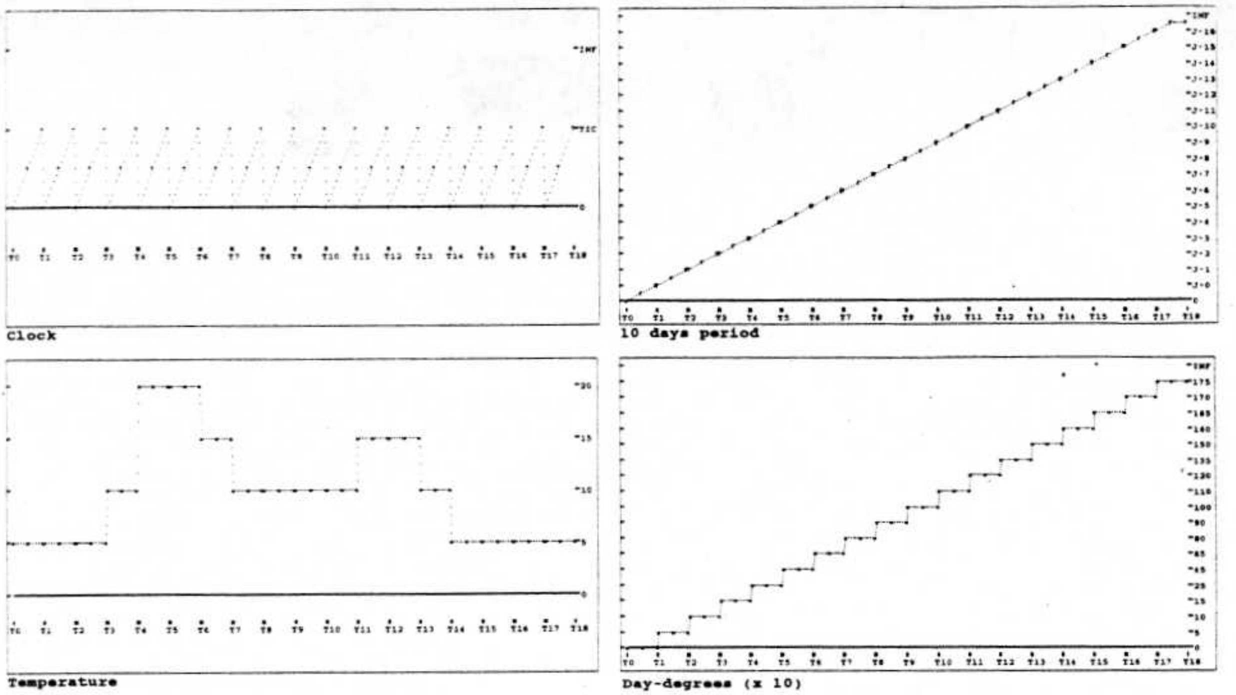


Figure 4: *An example of clock simulated behavior.*

representation of the temperature profile and the day-degrees calculation. Once performed the clock simulation, constructing an alist with both variables quantity spaces (10-days period and day-degrees) and comparing it to the critical day-degrees thresholds (see above enumeration), allow us to extract the sublist of dates at which changes in the development stage occur, allowing the slow-QDE to be simulated.

This was made as an attempt to introduce some notion of temporal dating and duration in such purely qualitative models. Suppose that each clock tick represents a 10-days period. Hence, each new created landmark of the variable accounting for their count provides a date, expressed as the number of 10-days periods elapsed since the initial T_0 state. For example (see Figure 4), at T_7, the J-6 corresponding landmark of the 10-days period variable denotes, actually, that 70 days have passed; at that time, the average temperature dropped from 15 Celsius degrees in the preceding period to 10 degrees at the current one, while the number of day-degrees, that added up to 650 before, discretely jumps to 800. Then, comparing the thresholds determining changes in the stage variable to the day-degrees accrued, allows us to generate the critical dates for stage transitions, i.e. 0, J-3, J-4, J-7, and J-9, denoting the 0th, 40th, 50th, 80th, 100th day since spawning, corresponding to the following stages: spawning, eyed larva, hatching, beginning and end of emergence, respectively (see Figures 5 and 6).

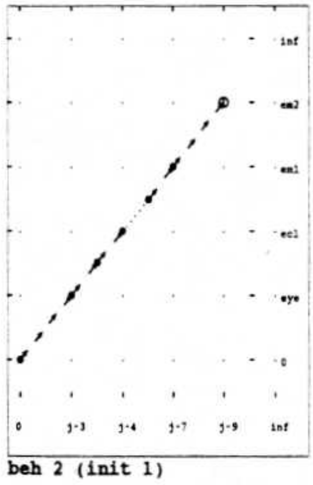
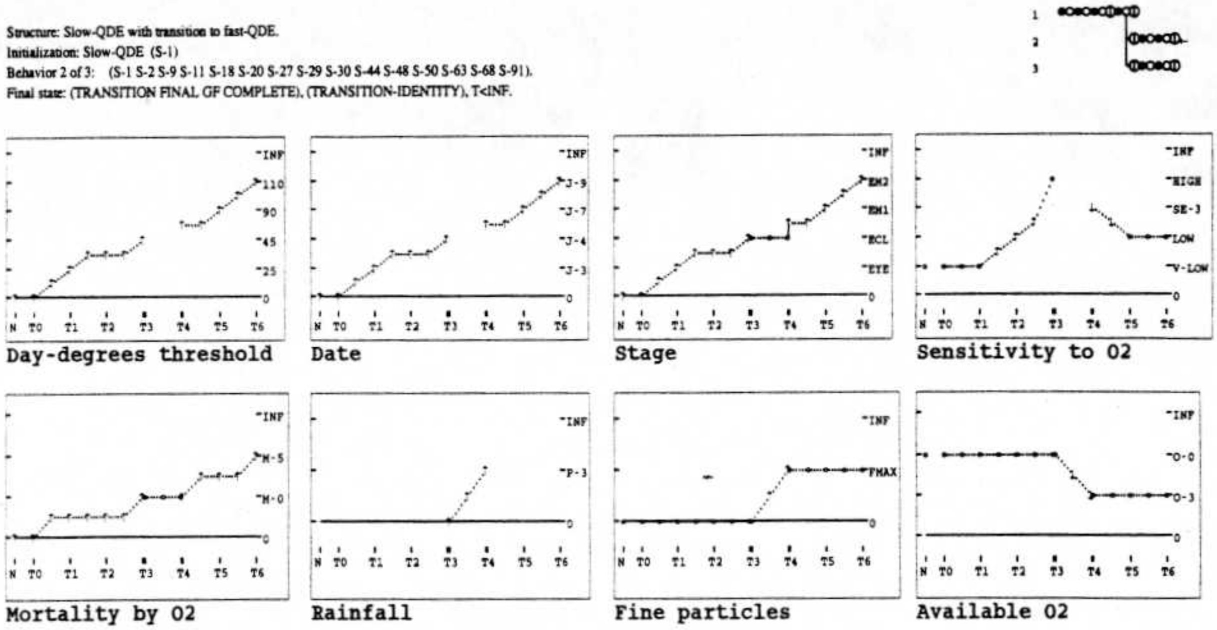


Figure 5: *An example of dating: phase-space representation of stage vs. date behaviors.*



Structure: Slow-QDE with transition to fast-QDE.
Initialization: Slow-QDE (S-1)
Behavior 2 of 3: (S-1 S-2 S-9 S-11 S-18 S-20 S-27 S-29 S-30 S-44 S-48 S-50 S-63 S-68 S-91).
Final state: (TRANSITION FINAL GF COMPLETE), (TRANSITION-IDENTITTY), T<INF.

Figure 6: *An example of QMR simulated behavior. The slow-QDE is simulated from* T_0 *to* T_3*, and* T_4 *to* T_6*; the fast-QDE is simulated from* T_3 *to* T_4*.*

An example of QMR simulation

The impact of a rainfall event is given below as an example (only some key-variables are shown on Figure 6). Typically, such a scenario is simulated as follows:

1. Specification by the user of:
    - a temperature profile to be taken as a sequence of average (quantitative) temperatures during the simulation
    - initial states of both slow and fast QDEs
    - events that will determine the transitions between both models; generally they are **qvals** reached by the shared variables during the simulation (here development stage, amount of fine sediment in the redd, and oxygen rate).
2. Clock simulation:
    - as described in the previous subsection, thus generating the critical dates values for stage changes.
3. Slow-QDE simulation:
    - completion of the initial state and generation of Qsim behaviors until a transition condition is reached;
    - from T_0 to T_3: the development stage of fish grows as a function of the amount of day-degrees added since spawning, making the sensitivities to mortality factors (and, thus, the mortalities themselves) evolve accordingly.
4. Transition slow-QDE → fast-QDE:
    - at T_3: hatching is reached (transition condition);
    - the **qvals** of shared variables in the transition state of the slow-QDE (fine sediments, oxygen rate) are passed to the initial state of the fast-QDE;
    - slow-QDE variables are frozen.
5. Fast-QDE simulation:
    - initial state completion and generation of qualitative behaviors until a transition condition is reached;
    - from T_3 to T_4: a sudden rainfall episode is triggered, which, via an increase in the river flow, along with an increase in TSS concentration and sediments carried by the flow, provoke a progressive sealing of the redd by fine particles, reducing accordingly the water flow through the gravel layers and, consequently, the oxygen available to the alevins.
6. Transition fast-QDE → slow-QDE:
    - at T_4: fine sediments are trapped to their maximum limit and available oxygen is strongly decreasing (transition condition);
    - the qualitative values of shared variables in the transition state of the fast-QDE are passed to the initial state of the slow-QDE;
    - fast-QDE variables are frozen.
7. Continuation of the slow-QDE simulation:

- the new initial state of the slow-QDE is made of its latter state frozen at the previous transition (T_3), updated with the new values taken by the shared variables;
- back to the slow-QDE simulation from T_4 to T_6 : the stage parameter continues its growth, as well as mortality, from their previous values at T_3.

8. Simulation halts when reaching the limits of the validity domain of the model (e.g. here `(em2 inc)` for stage).

Perspectives

Various versions of this model of a redd, have demonstrated globally the adequacy of Qsim to represent the available expertise and simulate relevant behaviors of such systems, and to handle a relatively high number of variables by means of model decomposition. However, some little difficulties arose, such as determining consistent initial states when it is dealt with a large amount of variables, specifying corresponding values in constraints among previously unspecified landmarks (e.g. in the example of dynamical construction of quantity spaces for the day-degrees and date variables).

We are now intending to take into account the two other blocks of processes (catchment and river levels; see Fig. 1) by qualitatively using hydrologic mechanistic models (to be transformed into QDEs) along with informal expertise.

Further, we are motivated by the interest of exploring extensively all the extensions built from and around Qsim, enabling semi-quantitative inferences (Q2, Nsim, SQsim,...) in order to better use available imprecise quantitative information (Kay, 1996). Particularly, connecting our redd qualitative model outputs to the salmon population dynamics quantitative model (Dumas *et al.*, 1996) is envisaged.

This poses several problems of symbolic/numeric representations interfacing, as far as hydro-ecologists knowledge incorporates also, in addition to purely qualitative ones, many pieces of numerical nature like:

- arithmetical procedures: day-degrees calculation, oxygen saturation rate (ratio measured/reference values), survival rate determined as $(1 - \frac{deads}{eggs})$,...
- average values: e.g., critical day-degrees values determining stage changes are, in fact, average values known to hold at 10°C.
- thresholds: the oxygen rate must be > 3mg/L from spawning to hatching, > 7mg/L at hatching, etc.
- numerical intervals: the range of favorable temperatures is 3–15°C from spawning to hatching, 1–17°C from hatching to the beginning of emergence...

Very often, mixed types of elements must be combined. For example, survival rates of, say, ≈ 100 % at spawning, ≈ 90 % at the eyed larva stage, ≈ 60 % at hatching, ≈ 40 % at the end of fry emergence, will be all termed *excellent* by hydro-ecologists. Similarly, for each stage, experts can define approximate percentages corresponding to labels such as *good*, *average*, *mediocre*, *bad*. Therefore, beyond the way it is dealt with that kind of problems in Q2 (that is mapping — by hand — symbolic landmarks with numerical intervals: $symbol \mapsto (number, number)$), the question is posed here how to *automatically* perform more complex mappings such as $(symbol, number) \mapsto symbol$,... and to combine many different types of such mappings within the same modeling framework. Along with other aspects of quantitative/qualitative integration, this is one of our research perspectives (Badia, Faivre, Guerrin, 1996).

References

Badia, J.; Faivre, R.; and Guerrin, F. 1996. *Modélisation de grands systèmes biologiques.* Département de Biométrie et d'Intelligence artificielle, INRA, Toulouse (F), Technical Report, 12 pp.

Chapman, D.W. 1988. Critical review of variables used to define effects of fines in redds of large Salmonids. *Transactions of the American Fisheries Society*, **117**(1):1-21.

Clancy, D., and Kuipers, B. 1994. Model decomposition and simulation. In *Proc. 8th Int. Workshop on Qualitative Reasoning (QR'94)*, Nara, Japan.

Dumas, J.; Faivre, R.; Charron, M.-H.; Badia, J.; Davaine, P.; and Prouzet P. 1996. Modélisation stochastique du cycle biologique du saumon atlantique (*Salmo salar* L.) : bases biologiques, implémentation informatique et interprétation. *$II^{ème}$ Forum Halieumétrique*, 26-28 Juin 1995, Nantes (F), Orstom éditions, p. 211-217.

Kay, H. 1996. *Refining imprecise models and their behaviors.* Doctoral dissertation, the University of Texas at Austin, 143 pp.

Kuipers, B. 1994. *Qualitative reasoning. Modeling and simulation with incomplete knowledge.* The MIT Press, Cambridge, MS, 418 pp.

Lisle, T.E., and Lewis, J. 1992. Effects of sediment transport on survival of salmonid embryos in a natural stream: a simulation approach. *Can. J. Fish. Aquat. Sci.*, **49**:2337-2344.