

4.6 Examples

We present below some numerical results obtained from recent research studies. The first two subsections present applications in river hydraulics (developed originally at INP Grenoble, then at INSA & Mathematics Institute of Toulouse by J. Monnier et al., see [15, 14, 22]). Models are based on the 2d shallow-water equations and the inverse method used is the 4D-var algorithm. Such "toy test cases" show the abilities of the approach. Also since we perform *twin experiments* (see explanations below), such test cases are the last stage of validation of the full computational software (including all the chain direct code, adjoint code and minimization process). (We refer to the next chapter to know how to validate the adjoint code only).

4.6.1 Identification of the topography in a 2d shallow-water model

We perform what we call *twin experiments*. The principle of twin experiments is as follows. First, we generate data using the direct code only. Next, we add some white noise to the 'perfect' data generated by the model. Next, we start the optimal control process from a first guess and we try to recover the set of parameters which gave the data. After the validation of the computational software (see next chapter to know how to validate an adjoint code), such numerical experiments is the necessary last step to validate both the approach and the computational code.

The first twin experiment presented concerns the identification of the topography in a small scale and academic case. The domain is 30 *m* long and 4 *m* large, and the topography is defined by:

$$\begin{aligned} z_b(x, y) &= 0.9 \exp\left(-\frac{1}{4}(x - 10)^2\right) \exp\left(-(y - 1)^2\right) \\ &+ 0.7 \exp\left(-\frac{1}{8}(x - 20)^2\right) \exp\left(-2(y - 3)^2\right) \end{aligned} \quad (4.16)$$

The inflow boundary is at $x = 0$, the outflow boundary at $x = 30$. Boundaries $y = 0$ and $y = 4$ are walls. We use a rectangular structured mesh of dimension 90×20 .

Bed roughness, defined by its Manning coefficient, is uniform ($n = 0.025$). (see the definition of the source term S_f in the shallow-water model). We impose a constant discharge $q^{in} = 8 \text{ m}^3/\text{s}$ at $x = 0$ and a constant water height $h_{out} = 1.4 \text{ m}$ at $x = 30$. We obtain a steady state solution after about 80 *s* of simulation.

Figure 4.6 shows the water height of this steady state solution and the topography.

From this steady state solution, we extract the forthcoming observations: h^{obs} and u^{obs} every 0.02 *s* during 20 *s* on each cell.

It means that in this first academic test case, we observe fully the (steady-state) flow !

The objective of this test case is to retrieve the topography. The first guess used is a flat bottom.

We run the data assimilation process with the following cost function:

$$j_1(z_b) = \frac{1}{2} \int_0^T \left(\|h(t) - h^{obs}(t)\|_{\Omega}^2 + \|\mathbf{q}(t) - \mathbf{q}^{obs}(t)\|_{\Omega}^2 \right) dt, \quad (4.17)$$

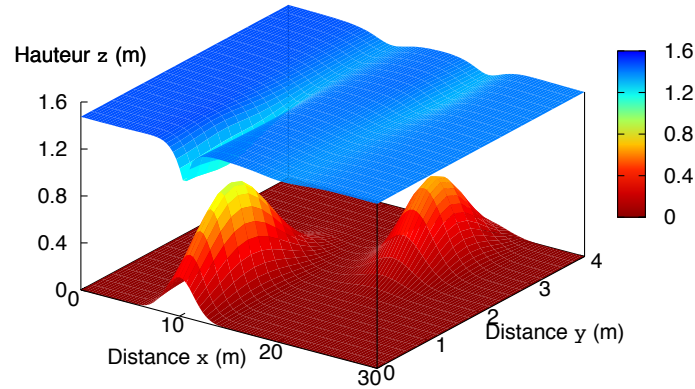


Figure 4.6: Identification of the topography. Topography and steady state elevation. From [14, 15].

Figure 4.7 shows the cost function and the norm of its gradient normalized by its initial values, vs iterates (a) and the identified topography (b). We can notice that convergence is obtained and the reference topography is well retrieved.

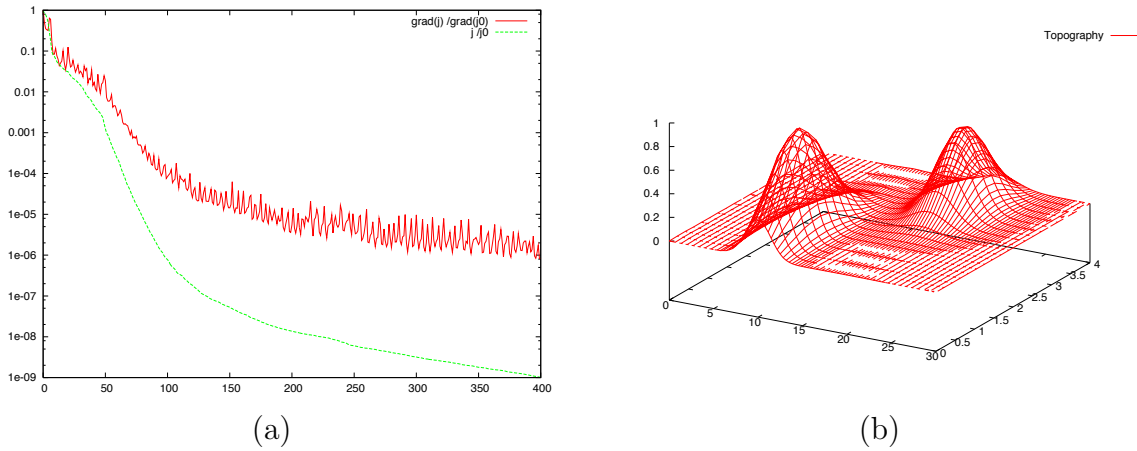


Figure 4.7: Value of the cost function and of the norm of its gradient, normalized by their initial values (a) and the identified topography (b)

This fully observed test case shows illustrate the variational data assimilation method in case of the identification of an underlying topography.

4.6.2 Identification of inflow boundary conditions in a 2d shallow-water model (floods)

We consider a toy test case which includes many features of real river flows. The computational domain contains a main channel (river) and floodplains, see figures 4.8 and 4.9).

Again, the present test case is a twin experiment. At the inflow boundary, we set the inflow discharge shown if figure 4.10 (a) simulating a flood event.

Then we perform a forward run to generate observations at points 1 and 2 shown with black stars in figure 4.9(a).

Then, we suppose that the inflow discharge is constant ($4.95 m^3 s^{-1}$), and we try to retrieve its real value by assimilating observations.

We present in Figure 4.10 the identified inflow discharge for different experiments. In Fig. 4.10(a), observations are h and $\mathbf{q}$ at each cell and each time step. In Fig. 4.10(b), observations are h at point 1 and $(h, \mathbf{q})$ at point 2, both at each time step. In Fig. 4.10(c), observations are h at point 1 only, but at each time step.

We can notice that the identified inflow discharge is good even with the observation of h at point 1 only.

In a practical point of view, such a test case show the ability of the method to identify inflow discharge in a river ...

One notice that the end of the flood event is not well identified. This is the "blind period" phenomena: for example in case (c), the inflow discharge after 270 s can not be identified because the information from the inflow boundary did not reach yet the gauge station.

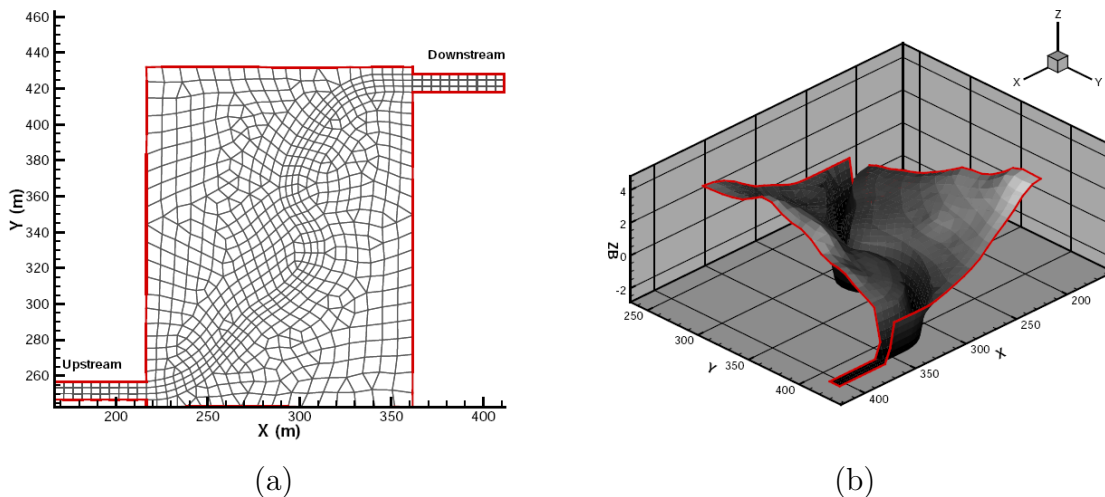


Figure 4.8: Toy test case mesh (a) and bathymetry (b)

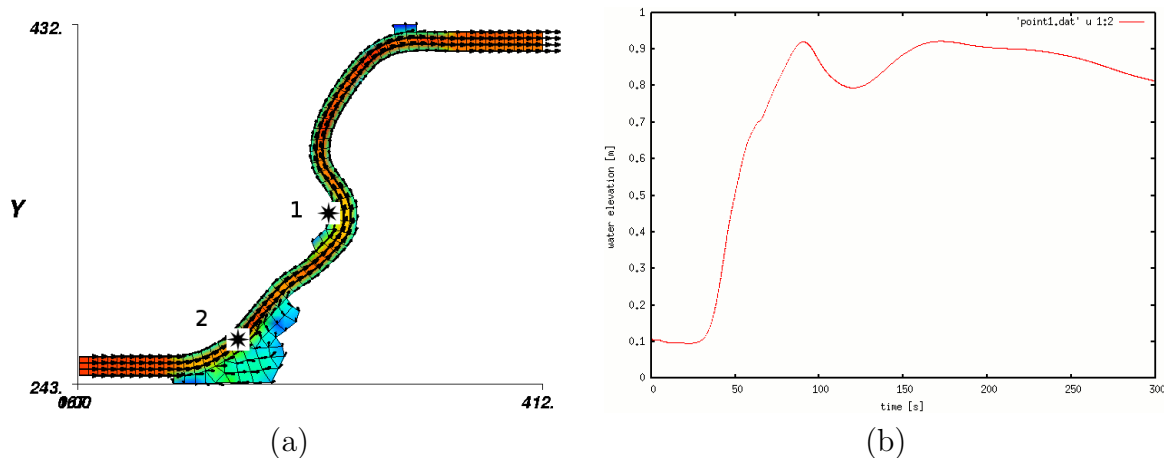


Figure 4.9: Toy test case domain with measurement points (a) and observation data available at point 1 (b)

4.6.3 Operational 4D-var in meteorology

We present below some order of magnitudes for Numerical Weather Prediction (NWP). All information and figures below come from documents written by Y. Tremolet and M. Fisher from the European Centre for Medium-Range Weather Forecasts (ECMWF) based at Reading, UK. ("ECMWF is an intergovernmental organization supported by 34 States. It provide operational medium- and extended-range forecasts and a state-of-the-art super-computing facility for scientific research. It pursues scientific and technical collaboration with satellite agencies and with the European Commission").

Forecast has improved the last twelve years in particular because of the introduction of data assimilation process into the complex dynamical atmosphere models, which allow to benefit more and more from the global various observing systems. Methods employed are 4d-var, 3d-var and ensemble methods.

Orders of magnitude presented below date from 2009 - 2010.

The unknowns of the (non-linear, fully coupled) mathematical models are 3D fields: temperature, pressure, humidity, velocities.

The dimension of the state (number of unknowns) is about 10^9 .

Time step is about 10 minutes. Discretization is about 16 kms in horizontal (triangles) and with about 90 layers in vertical (0 - 80 km meshed).

Observations are heterogeneous in nature and in space. They are in dozens of millions (see figures below).

Satellites measurements are: surface temperature, nebulosity etc. They are complex to interpret and to compare to the model outputs.

Measurements in-situ are: temperature, pressure, humidity, wind etc.

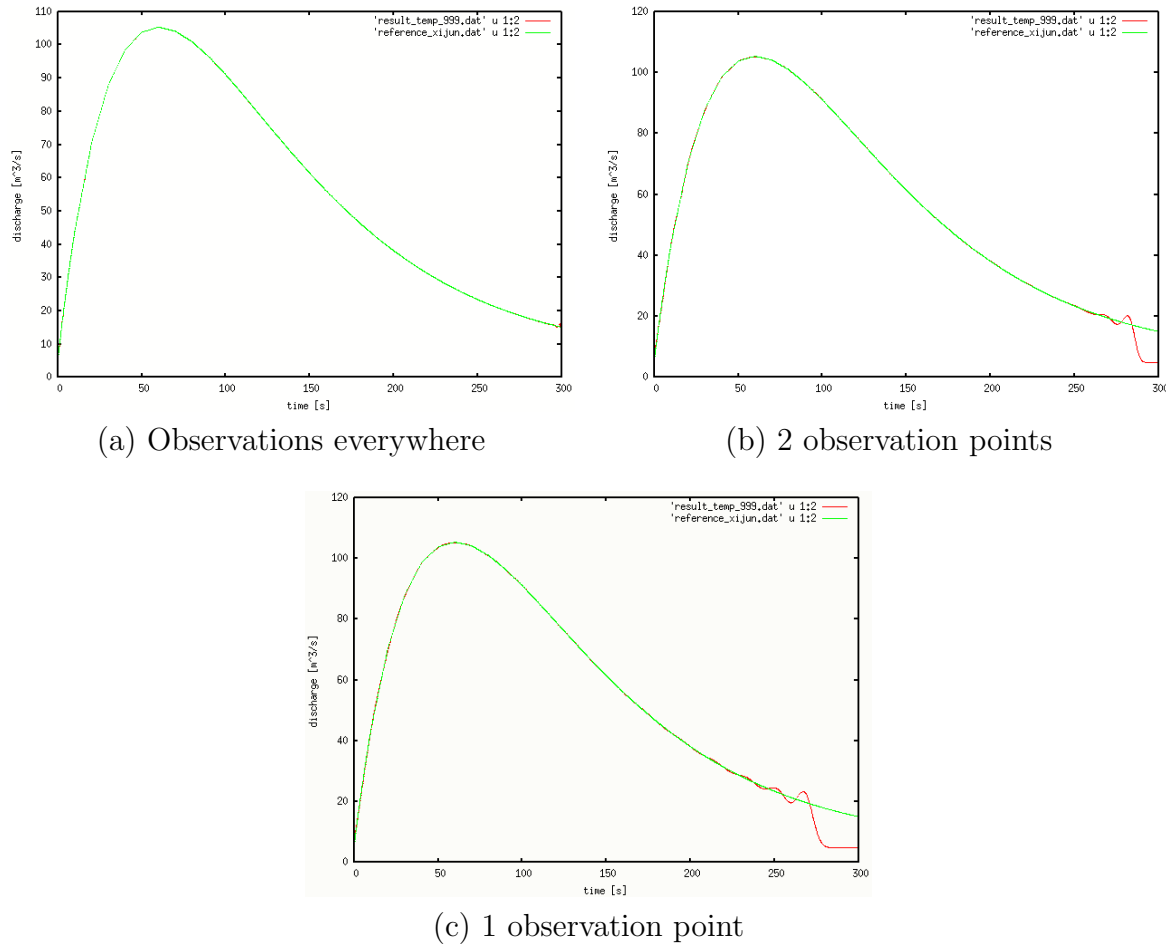


Figure 4.10: In green : reference inflow discharge. In red : identified inflow discharge. (a): Observation of $(h, \mathbf{q})$ everywhere; (b): Observation of h at point 1 and $(h, \mathbf{q})$ at point 2. (c): Observation of h at point 1 only. From [14, 15]

Orders of magnitude from the operational ECMF forecast system (in 2009). The configuration uses a 12h cycling window, with a 4D-var incremental method.

The outer loop (and forecast) resolution is 25 km. The inner loops resolutions are between 200 km and 80 km.

On average, 9 million observations are assimilated per 12h cycle. 96% of assimilated data is from satellites. On average, 4D-Var runs on 1536 CPUs in 1h10.

The incremental method with appropriate preconditioning allows the computational cost to be reduced to an acceptable level.

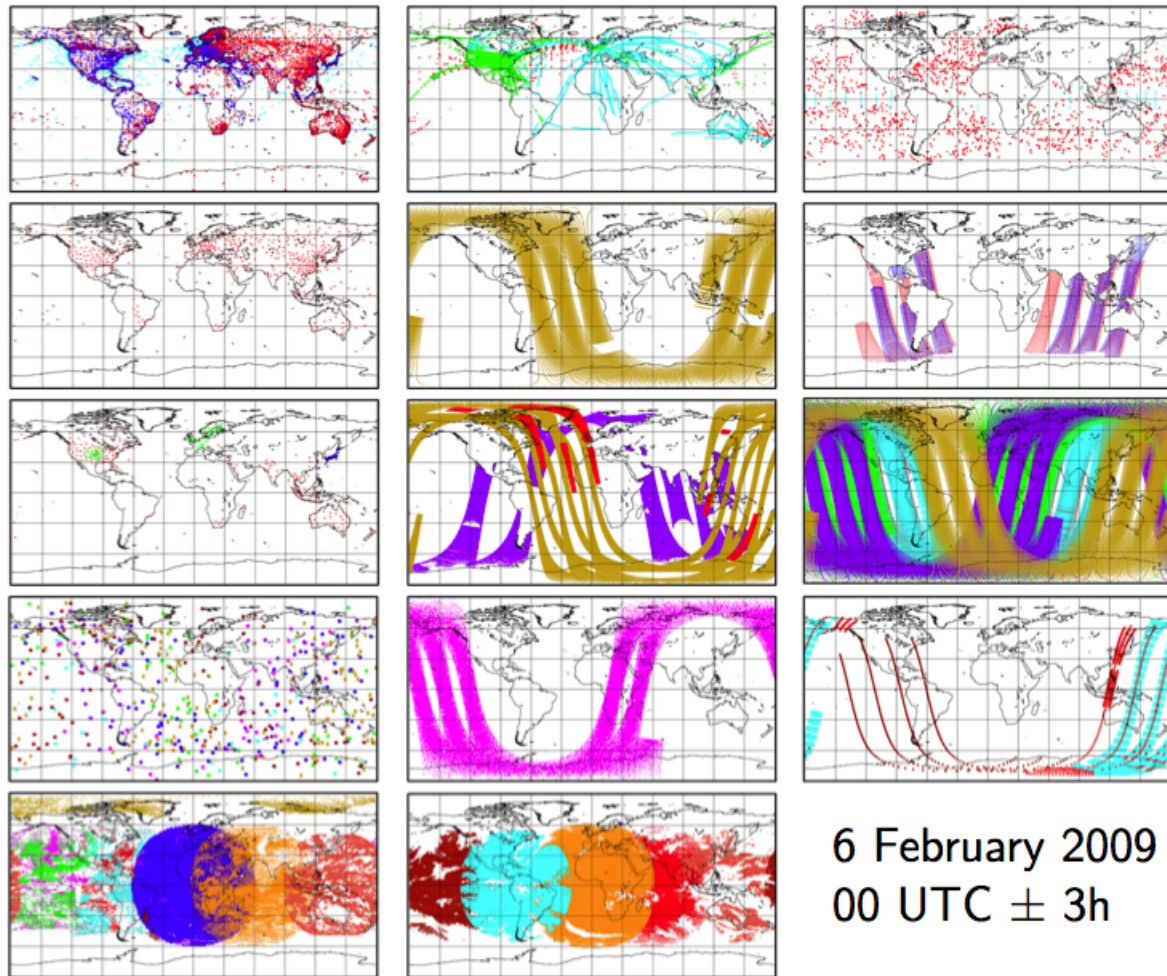


Figure 4.11: Meteorology. Some observation coverage. Figure extracted from Tremolet - Fisher, ECMWF, Colloquium 2009.

4.6.4 Some concluding remarks

Data assimilation aims to fuse in an optimal way to be defined, all information available: physics (conservation laws, mathematical models), parameters (empirical or not), initial condition (particularly in geophysics), in-situ measurements, remote-sensed observations (satellite images etc).

Any extra information, measurement, information should improve the "analysis", but only if its confidence (or accuracy) is well estimated (statistical error estimate).

The control of any uncertain parameters of all models used by engineers should (will ?) be addressed...

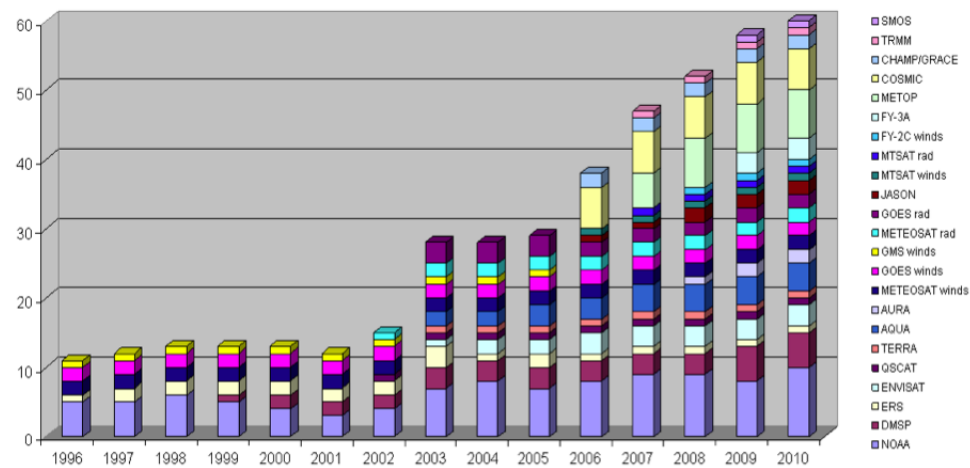


Figure 4.12: Meteorology. Observation sources. Figure extracted from Tremolet - Fisher, ECMWF, Colloquium 2009.

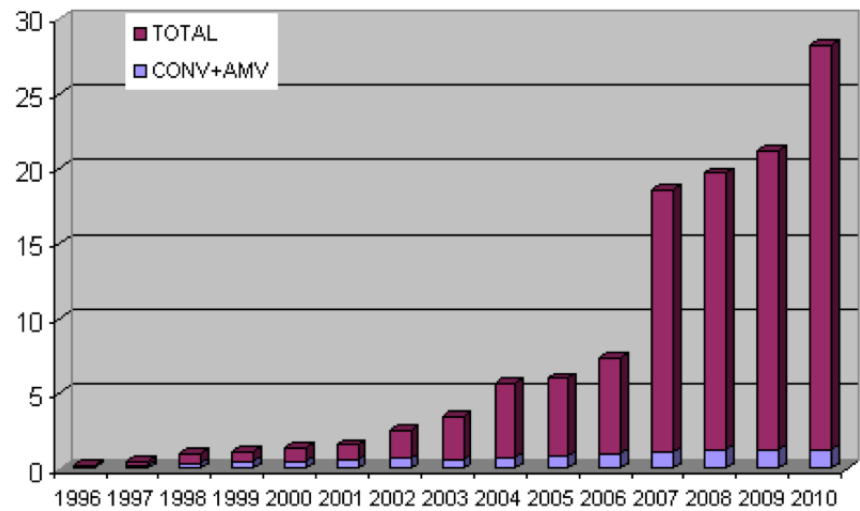


Figure 4.13: Meteorology. Observation Numbers (in millions per day). Source: ECMWF.

Data count for one 12h 4D-Var cycle
(0900-2100 UTC, 3 March 2008)

	Screened		Assimilated	
Synop:	450,000	0.3%	64,000	0.7%
Aircraft:	434,000	0.3%	215,000	2.4%
Dribu:	24,000	0.02%	7,000	0.1%
Temp:	153,000	0.1%	76,000	0.8%
Pilot:	86,000	0.1%	39,000	0.4%
AMV's:	2,535,000	1.6%	125,000	1.4%
Radiance data:	150,663,000	96.9%	8,207,000	91.0%
Scat:	835,000	0.5%	149,000	1.7%
GPS radio occult.	271,000	0.2%	137,000	1.5%
TOTAL:	155,448,000	100.0%	9,018,000	100.0%

Figure 4.14: Meteorology. Observation usage. Figure extracted from Tremolet - Fisher, ECMWF, Colloquium 2009. We are still far from using all available observations.

4.7 Exercices: optimality systems

4.7.1 Viscous Burgers' equation

The viscous Burgers' equation is the 1d simplification of the Navier-Stokes momentum equation. It is a scalar non-linear advection diffusion equation (non-linear advection term). The unknown is $u(x, t)$ the fluid velocity at point x and time t .

The control variables we consider in the present example are: the initial condition u_0 and the velocity value at one boundary extremity v .

The forward (direct) model reads as follows. Given $k = (u_0, v)$, find u which satisfies:

$$\begin{cases} \partial_t u(x, t) - \nu \partial_{xx}^2 u(x, t) + u \partial_x u(x, t) = f(x, t) & \text{in }]0, L[\times]0, T[\\ u(x, 0) = u_0(x) & \text{in }]0, L[\\ u(0, t) = v(t) ; u(L, t) = 0 & \text{in }]0, T[\end{cases} \quad (4.18)$$

We assume we have m observations points of the flow, continuous in time. Then, we seek to minimize the following cost function:

$$j(k) = \frac{1}{2} \int_0^T \sum_{i=1}^m |u(x_i) - u_i^{obs}|^2 dt$$

Exercice 4.7.1. Write the optimality system corresponding to this data assimilation problem.

4.7.2 Diffusion equation with non constant coefficients

We consider the diffusion equation (or heat equation) in an inhomogeneous media. Let u be the quantity diffused and $\lambda(x)$ be the diffusivity coefficient, non uniform. The forward model

we consider is as follows. Given λ and the flux φ , find u which satisfies:

$$\begin{cases} \partial_t u(x, t) - \partial_x(\lambda(x) \partial_x u(x, t)) = f(x, t) & \text{in } \Omega \times]0, T[\\ u(x, 0) = u_0(x) & \text{in } \Omega \\ -(\lambda(x) \partial_n u(x, t)) = \varphi & \text{in } \Gamma_1 \times]0, T[\\ u(x, t) = 0 & \text{in } \Gamma_0 \times]0, T[\end{cases} \quad (4.19)$$

with $\partial\Omega = \Gamma_0 \cup \Gamma_1$.

We assume we have measurements of the quantity u at boundary Γ_1 , continuously in time. Then, we seek to minimize the following cost function:

$$j(k) = \frac{1}{2} \int_0^T \int_{\partial\Omega} |u(x) - u^{obs}|^2 ds dt$$

Exercise 4.7.2. *Write the optimality system corresponding to this data assimilation problem.*