Variational approaches intend to minimize an error function. Today solutions rely on optimization methods like gradient descent to minimize some objective function.

Data assimilation consists in using observation data (measures acquired from other scientific instruments or network of sensors) to adjust the internal parameters of the running simulation. The goal is to better control the course of the running simulation to get higher quality results as well as to limit the required number of simulation executions, thus saving CPU energy and scientist time. Data assimilation requires to analyse the outputs from the simulation, then to combine these results with the observation data to compute adjustment values for the internal simulation parameters. This process takes place periodically while the simulation is running. Often the correlation between the outputs and the internal parameters is a complex inverse problem requiring advanced algorithms. Today solutions rely on optimization methods like gradient descent to minimise some objective function.

If this workflow was not challenging enough, data assimilation can also require to handle not one single simulation execution but several simulation runs in parallel (often called an ensemble run). When working with a single simulation, the data assimilation process is called variational, while ensemble run approaches are called statistical. Variational approaches intend to minimize an error function.
function on a single simulation. Statistical approaches look at minimizing a variance. They require to run several simulations with different parameter sets enabling to provide a more complete information about the uncertainties associated with the results.

**Principales activités**

The goal of this PhD work is to investigate solutions to push data assimilation capabilities to novel scalability and performance levels, using thousands of compute cores if necessary. We will cover aspects like infrastructure (leveraging the one like Melissa and FlowVR developed at DataMove as well as other frameworks like Flink), fault tolerance, parallelization, exploration of novel data assimilation processes based on unsupervised learning algorithms for instance.

To achieve this goal we will work in close with scientists from other domains, learning specialists, as well application domains experts in particular from Hydrology or Molecular Dynamics. They will provide us the main use cases we will work with (simulation codes, observational data and data assimilation processes).

**Compétences**

We are looking for enthusiastic, motivated and creative candidates, capable of developing novel innovative approaches. The PhD candidate will have a master in computer science with expertise in one or several of the following domains: parallel, distributed computing, numerical simulation, large scale data analytics, optimization techniques. Good programming skills and a Linux expertise are required, with a taste for leading experiments on large supercomputers. The candidate will have opportunities to travel in France and abroad to present his results, in particular at some of the major international conferences of the domain. A good speaking and writing English level is sufficient in the working environment, but some bases or at least the will to learn French will greatly help for the everyday life. The candidate will integrate the DataMove team and benefit from its friendly work environment. Grenoble valley is known to offer a good quality of life in the middle of the French Alps.

**Bibliography**


**Avantages sociaux**

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours

**Rémunération**


Monthly salary after taxes: around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year. (medical insurance included).