

# ON PRAGMATISM IN INDUSTRIAL MODELING

## PART II: WORKFLOWS AND ASSOCIATED DATA AND METADATA

Josip ZORIC<sup>1,\*</sup>, Alexander BUSCH<sup>2</sup>, Ernst A. MEESE<sup>1</sup>, Milad KHATIBI<sup>3</sup>, Rune W. TIME<sup>3</sup>, Stein T. JOHANSEN<sup>1,2</sup>,  
H. A. RABENJAFIMANANTSOA<sup>3</sup>

<sup>1</sup> SINTEF Materials and Chemistry, Trondheim, NORWAY

<sup>2</sup> Norwegian University of Science and Technology (NTNU), Trondheim, NORWAY

<sup>3</sup> University of Stavanger (UiS), Stavanger, NORWAY

\* Corresponding author, E-mail address: [josip.zoric@sintef.no](mailto:josip.zoric@sintef.no)

### ABSTRACT

Concepts for "pragmatism in industrial modeling", were presented and discussed at CFD2014 in Trondheim, Norway (Zoric, et al., 2014). A framework (FW) for pragmatic industrial modeling was suggested and demonstrated on industrial use cases by a process centric approach. In the present paper, these concepts will be extended through a data and information centric perspective (Analysis and use of data and metadata, their organization, syntax and semantics), exemplified by a new use case, related to drilling of oil & gas wells. The analytical problem is to produce a model which can predict the motion of a spherical particle embedded in laminar non-Newtonian flow. An overview of tasks, procedures, organization and flow of data, required metadata, and technical and quality requirements is given. In particular we clarify input and output from various modeling and experimental phases, as well as the flow of data and metadata in the modeling process. In addition we analyze the structure and organization of input data for the modeling and experimental work, such as rheometer data from the lab. The uncertainty in the characterization of the fluid will impact the accuracy of the model predictions. We next make example simulations to illustrate the steps in the solution process, and related flow of the necessary metadata and raw data. The necessary uncertainty assessments and the corresponding metadata needed to support the entire modeling process are also discussed. Quantification of uncertainty will be demonstrated using polynomial chaos to assess the effect of uncertainties in the parameters for the non-Newtonian viscosity model. Finally, we summarize the findings and discuss how the "pragmatism in industrial modeling" concept can help building more consistent industrial models, answering to customer needs for actual accuracy and delivery speed.

### 1. INTRODUCTION

#### Position and Motivation

In (Zoric, et al., 2014) we identified and discussed the process and frameworks needed for pragmatic industrial modeling. An important key element in the pragmatic modeling approach is exploitation of experimental data, organized in the optimal manner to support the prediction quality of the model. The organization of all types of data is based on "bridging" (modeling middleware, various APIs, protocols, standards etc. (Zoric, et al., 2014)) between complex scientific (aspect/phenomenon oriented) physical models, simplified industrial models, process data and experimental data. In this paper we take a data-centric perspective on pragmatic modeling: The analysis and use of data and metadata, their organization, syntax and semantics. We elaborate on creation, flow, exchange and use of modeling metadata, raw data and other analytical information. An overview of the tasks,

procedures, organization (based on (Zoric, et al., 2014)) and flow of required data, metadata, and quality requirements are given. By example simulations (A use case related to drilling of oil & gas wells – prediction of the motion of a spherical particle embedded in laminar non-Newtonian pipe flow.) we illustrate the individual steps of the process, and related flow of the necessary metadata and raw data. We demonstrate examples of the necessary uncertainty assessments and the corresponding metadata needed to support the entire modeling process.

#### Existing research body and praxis

This work complements the references of (Zoric, et al., 2014) by data-centric literature, and some newer contributions on standardization of modeling FWs.

(Slotnick, et al., 2014) stress the challenge of managing the vast amounts of data generated by current and future large-scale simulations and motivate to develop and maintain integrated simulation and software development infrastructures for integrated experimental testing and computational validation campaigns.

We agree with (Novere, et al., 2005) that most of the published quantitative models are difficult to reuse because they are either not made available or they are insufficiently characterized, e.g.: lack of a standard description format, lack of stringent reviewing and incomplete model descriptions. References in (Lombard and Yesilbas, 2006) found that almost 75% of an engineer's design work consists of seeking, organizing, modifying and translating information, often unrelated to his own personal discipline. Knowledge-based related activities consume most of a project's resources in terms of time and money (Freiberg, et al., 2012). (Lombard and Yesilbas, 2006) suggest defining a complete repository for information and knowledge management in collaborative design. When undertaking a new study, the reuse of existing models considerably reduces engineering time (Malleron, et al., 2011), and the possible lack of information in an existing model can be compensated for by coupling it to another model, thereby giving access to the missing information.

(Varnell-Sarjeant, et al., 2015) stress the importance of reuse in modeling and give an overview of reuse strategies, reusable artifacts and development constructs. (Donn, 2001) identifies commonality in the types of information that simulation and design tools produce.

Best Practice Guidelines (BPG) based approaches (Zigh and Solis, 2013; ASME, 2009) have been developed to ensure as accurate CFD results as possible given the time and budget constraints. BPG complement well the modeling platforms and technologies by offering practical guidelines, procedures, rules, criteria, requirements and other knowledge artifacts. (AIAA, 1998) was pioneering the formalization of BPG through concise definitions of the terms verification, validation, and calibration, and by recommending related work processes. Various FW approaches are suggested in the literature, each proposing its own architecture, technology,

interfacing and data structuring standards, e.g. (Lombard and Yesilbas, 2006; Freiberg, et al., 2012; Surridge, et al., 2014; Klement, et al., 2014; Mooij, et al., 2014; Gopsill, et al., 2013; Lloyd, et al., 2007; Novere, et al., 2005; Hu, et al., 2012). Unfortunately they approach differently the analysis process and related interaction with FW entities (lack of industry standards). Many FW designers focus on standardizing FWs' components and technologies, e.g.: (Bulatewicz, et al., 2014) with the Open Modeling Interface (OpenMI) data component (based on Web services and a REST technology), (Swayne, et al., 2010) present a review of available tools for code/model coupling stressing a wide range of interfaces and frameworks and challenging their interconnection and collaboration, (Hinsen, 2011) give a data and code model for reproducible research and executable papers, based on the Hierarchical Data Format 5 (HDF5) suitable for electronic publication and permitting the reproduction of computational results.

Many references focuses on data exchange in experimental and modeling FWs and processes, and from their work we can take over good ideas and modeling elements e.g.: (a) syntax and semantics of data (Gayer, et al., 2010), (b) sources of metadata and data (Tecplot, Inc., 2011), (c) organization of knowledge, information and data (Greenwald, 2004; Erdemir, et al., 2012; Frenklach, 2007), (d) requirements (Novere, et al., 2005), reusability, quality of data, validation, verification (Roache, 1998; Oberkampf, et al., 2002; Oberkampf and Barone, 2006; Babuska and Oden, 2004), uncertainty estimates (Russi, et al., 2010; Oberkampf, et al., 2002; Helton, et al., 2004; Mendenhall, et al., 2003; Robert W. Waiters and Luc Huyse, 2002; Roy and Oberkampf, 2011; Sankaraman and Mahadevan, 2015), etc.

Proprietary environments such as the (ANSYS Workbench Platform) or (modeFrontier) allow for an out-of-the box connection and orchestration of various tools, whereas open-source solutions such as (Dakota), (NLOpt), and (Porto) have to be coupled with the respective solvers/CFD software (SW) i.e. using C-based APIs and scripting. ANSYS Workbench allows Python-based scripting and uses XML for "Engineering Data files" and also integrates dedicated tools such as FLUENT/CFX (ANSYS - CFD) and MECHANICAL (ANSYS - Structural). Tools such as (Dakota) lack the sophisticated GUI but allow for much broader and cross-platform tool integration (Jareteg, 2014). As alternative to integration into a higher framework, STAR-CCM+ offers the above mentioned functionalities with the "Optimate" Add-In (Maley, 2012).

Data pre-processing standards manifest in various file formats e.g. IGES, STEP or STL for the geometry of the case. Another generic container format in use is JT, an open and standardized lightweight format for exchange of 3D data (ISO/TC 184/SC 4, 2012). For other types of pre-processing data e.g. material properties, XML seems to be the most generic file format. Data standards for post-processing are mainly based on HDF5 e.g. CGNS (CFD General Notation System), (Silo), or HDF5/XML. The CFD General Notation System (CGNS) might be considered the current CFD data standard as it has been adopted as a recommended practice by (AIAA, 2005). A coupled XDMF (for metadata) and HDF5 (for heavy data) approach has been used as an alternative in cases dealing with multiphase particle-loaded flows (Strandenes, 2013; de la Cruz, et al., 2011).

Given the vast and heterogenic tool landscape one could require that FW implementation standards and data management standards, interfaces and protocols get more consistently standardized, and in such a way facilitate efficient and flexible interworking (combining existing and new analyses in solution of industrial problems).

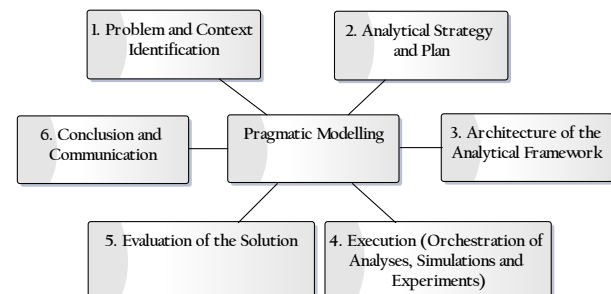
#### Structure of this work

Section 1 of this paper introduces our position and motivation for pragmatism in industrial modeling. Section 2 takes a

process view on pragmatic industrial analysis, and reflects on the interplay of practical, holistically organized and orchestrated activities as explained in (Zoric, et al., 2014). Section 3 takes a data-centric approach and focuses on creation, flow, exchange and use of modeling metadata, raw data and information in various phases of pragmatic modeling (Figure 1, (Zoric, et al., 2014)). Section 4 addresses organization and structure of modeling metadata and data. We offer a system analysis and discuss the candidate implementation technologies and standards. Section 5 discusses our experiences with a practical use-case driven exercise and suggests future steps and improvements.

## 2. PROCESS VIEW ON PRAGMATIC INDUSTRIAL ANALYSIS

Pragmatic industrial analyses require a structured approach not just with respect to models, simulations, experiments, information and data, but also regarding analytical processes, concluding with a well-structured communication of the results and their analytical and validity context. We consider these important elements being parts of the analytical framework (FW), illustrated in Figure 1.



**Figure 1: Some of the important phases, processes and results in a typical pragmatic analysis (Zoric, et al., 2014).**

(Zoric, et al., 2014) detail these modeling phases, and the results they produce. It is crucial to conclude pragmatic analysis by a communication of the analytical results (Step 6 in Figure 1). It is important to relate the analysis to its context: (a) important analytical parameters (assumptions, initial, boundary conditions, domain-specific parameters etc.), (b) information about modeling scale, (c) accuracy of the proposed solution, (d) estimates of representability, (e) predictive power, (f) computing and experimental resource consumption, etc. This information is needed both for the evaluation of existing models/analyses/experiments, and for their future reuse. One could even require more stringent standardization, in order to facilitate efficient collaboration, reuse and interworking (combining existing and new analyses in solution of various industrial problems).

## 3. DATA-CENTRIC VIEW ON INDUSTRIAL PRAGMATIC MODELING

This section takes a data-centric view on the pragmatic modeling process sketched in Figure 1. The main purpose of so-called "pragmatic modeling" is to adjust the research models to the realism/world of industrial processes, their scope, perspective and challenges. Industrial models pose requirements such as: (a) industrial scope and perspective, (b) usefulness, (c) required accuracy and predictability, (d) simplicity of use, (e) response time and speed, (f) compatibility with other (industrial) models, etc., which had to get mapped to numerous SW engineering requirements (Zoric, et al., 2014). We would like to motivate further development and standardization of the SW engineering related to pragmatic modeling frameworks, including data and metadata:

- Standardization of data exchange formats (e.g. based on XML, JSON, HDF5 and other standards).

- Standardization of modeling metadata specifying e.g. problem definition, analytical context, accuracy, sensitivity and quality of models and simulations (organized as a part of the analytical context container), syntactic and semantic description of the data entities, information and knowledge, etc. discussed below.
- Standardization of data processing, management and retrieval functionality.

#### 4. ORGANIZATION AND STRUCTURE OF DATA / INFORMATION

Pragmatic modeling is often part of a complex analytical and/or design process. It is a team work, driven by analytical workflows (Zoric, et al., 2014). Solution architects specify a set of data/information exchange standards, protocols and interfaces (to mention just a few SW design artifacts), a number of design tools. Designers and analysts generally verify the results by use of various models, analyses, experiments and fine-tuning techniques (e.g. sensitivity analyses evaluated against physical experiments). Figure 2 illustrates the overall organization of modeling data and information on a "per-case" basis, i.e. a project divided in hierarchically organized subcases (with related data/information sets), and based on the case study described in section 5. A comprehensive specification of variables along with a detailed case description derived from the modeling customers' needs is created by the solution architect. Besides the quantities of interest this includes numerical values of all parameters, dimensions, ranges of validity, boundary conditions, fluid and particle properties, etc. This input data is being transferred to one or multiple CFD cases and experimental setups by individual modelers and experimentalists where in the latter case an additional degree of uncertainty comes into play due to natural scatter.

Simulation and experimental runs produce solution data. In the particular industry case described in section 5 the experimentally obtained raw data using PIV had been transferred by image processing techniques to the final solution data format.

Simulation and experimental activities generally share some instances of data, e.g. the above mentioned specification and case description. For validation purposes they must also share the same class of solution data as a part of the specification such as e.g. velocity profiles/fields and particle trajectories in our use case. A comparison of simulation and experimental results (with respect to the specified parameters) leads to a validated model with an associated degree of uncertainty and range of parameter validity. But, as opposed to the specified input data, the actual solution data sets are specific to the two approaches in the sense that one will end up with e.g. one particle trajectory as a result of the simulation and a second one as a result of the experiment.

Both simulations and experiments also produce their own specific sets of data: In the CFD case one usually has to deal with data related to grid properties, physical models and numerical schemes, whereas in the experimental case one has to deal with data related to e.g. rig information such as test section, inlet channel, pump and instruments such as PIV cameras and pressure transducers plus the obtained raw data.

The very same setup of data classes is then being used over and over again for multiple test runs with varying parameters accounting for different operating/design points of the system under consideration ideally following some sort of Design-of-Experiments (DoE) approach. In order to account for natural scatter one usually also performs multiple experimental test runs for the same set of input parameters.

Figure 2 illustrates the workflow and SW engineering view on modeling.

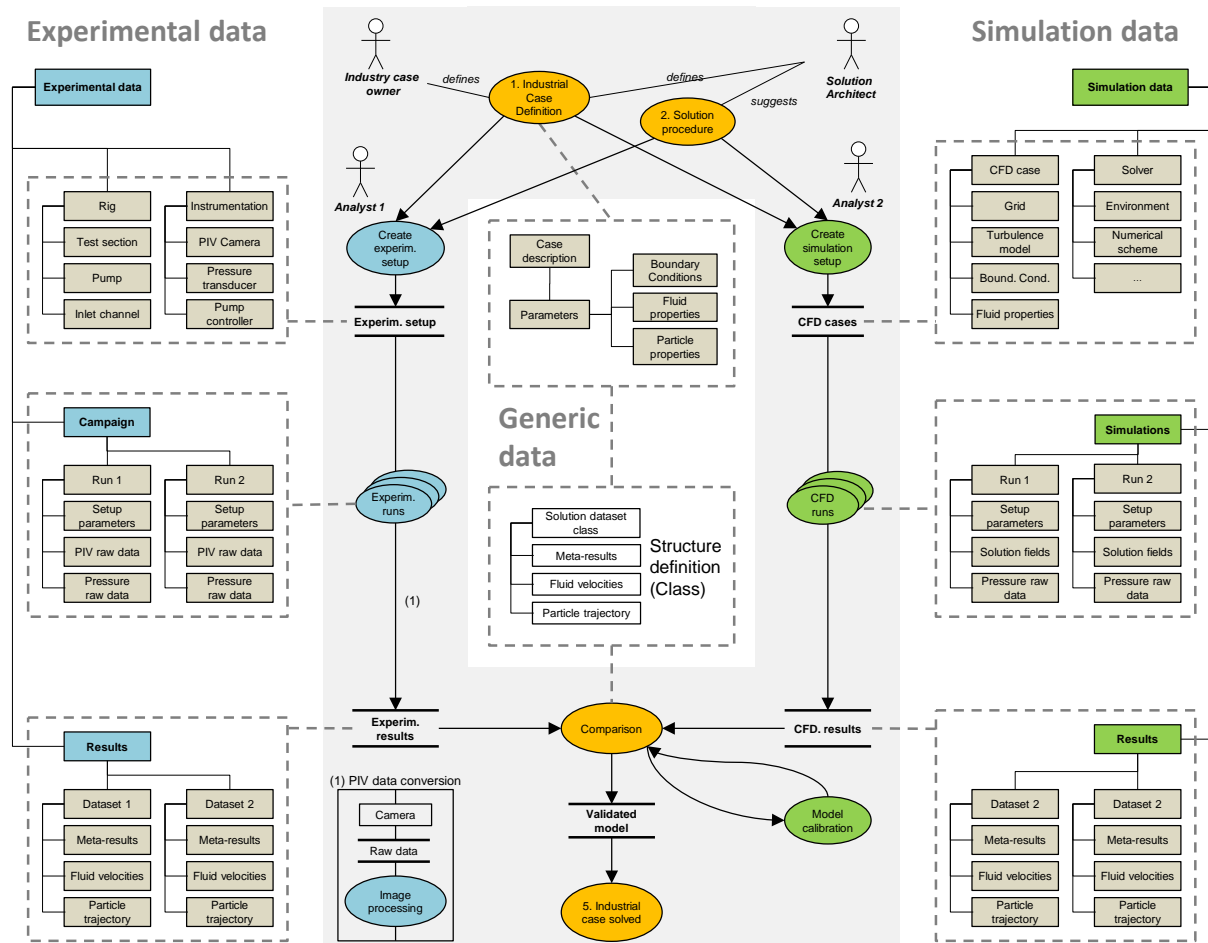


Figure 2: Organization of data flow and structure on the case study (section 5), see also Figure 1 and (Zoric, et al., 2014).



We will detail it by dedicated examples in Section 5. We use a data-centric view to discuss the requirements and SW Engineering issues related to model interaction, data/information exchange, interfacing, standardization and other important elements for design of pragmatic models. Our analysis (Section 5) is case-based. We follow an industry case and its modeling, experimental and analysis parts, and follow creation, usage and flow of data and information. Our process is structured as a data-focused interview of participants during the six analytical phases (Figure 1), by means of templates for description of use-cases, metadata lists (attribute-value pairs), and descriptions of other data types (hyperlinks, raw-data containers, figures, documents etc.). Some attribute-value-pairs will be exemplarily indicated by [colored text] with reference to Figure 2 in the following description of our case.

## 5. CASE STUDY - MOTION OF SPHERICAL PARTICLES IN NON-NEWTONIAN LAMINAR CHANNEL FLOW

Adequate cuttings transport in drilling wellbores [Application = Cuttings transport in drilling wellbores] is a major challenge. This case study is limited to a sub-problem that might become an element in a larger and more complex model describing the physics of cuttings transport in a wellbore. The task of this sub-problem is to predict particle trajectories in a horizontal channel flow according to Figure 3.

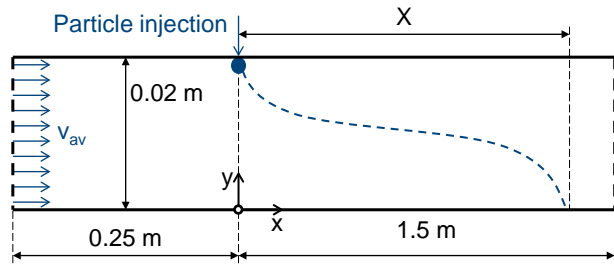


Figure 3: Particle settling domain.

The fluid is either Newtonian (Water) or non-Newtonian (Poly-Anionic Cellulose) [Fluid = H<sub>2</sub>O; PAC], [Density = kg/m<sup>3</sup>; 1000]. A spherical particle with [Diameter = mm; 1; 2; 3] and [Density = kg/m<sup>3</sup>; 2560] is released at the centerline of the top wall. Besides predicting the particle trajectory there is a particular interest in the position X at which the particle impinges on the lower horizontal wall. One major requirement is that the axial particle displacement, from the release position to the impingement position, should be predicted by a relative accuracy of better than 5 % [Overall accuracy = %; 5] with the working model being available in [Delivery time = Month; 1].

Both experimental and CFD activities will be part of the solution approach as follows:

1. Determine different fluid rheologies using rheometry and fit a particular fluid rheology model to the obtained data
2. Determine associated uncertainties using polynomial chaos
3. Determine particle trajectories with numerical simulation and experimental work
4. Comparison of numerical with experimental results
5. Model calibration & refinement
6. Determine global uncertainties and conclusions

### Fluids rheology & uncertainties

The rheology of the two PAC solutions [Concentration = g/L; 2; 4] is represented with a power-law model [Rheology model = Power-Law] according to Eq. (1).

$$\mu = A\dot{\gamma}^{n-1} \quad (1)$$

The coefficients A and n are obtained from rheometer measurements of (Time and Rabenjaminantsoa, 2012), which represents a data flow from a previously conducted

experimental result to the CFD model parameters (See Figure 2 : Experimental runs, results, model calibration). The uncertainties in such measurements should always be stored along with the results, and for the measurements presented here they vary between 1-3 % for the viscosity [Rheometric uncertainty interval = %; 1, 3]. However, as indicated by Figure 4, the largest uncertainty here is that the chosen power-law model does not represent a perfect fit to the rheometric data. To assess the uncertainty introduced by the power law model, we vary the number of points that are used to fit the power law to the data in order to obtain a one-parameter family of power laws that may be used for each fluid. Following the pseudospectral collocated generalized polynomial chaos approach (Xiu, 2010) [Uncertainty analysis = Polynomial Chaos, (Xiu, 2010)], an approximation to the particle path can be found through the series

$$[x(t, Z), y(t, Z)] = \sum_{k=0}^n [\hat{x}(t), \hat{y}(t)] L_k(Z) \quad (2)$$

where  $L_k$  is the k-th order Legendre polynomial.

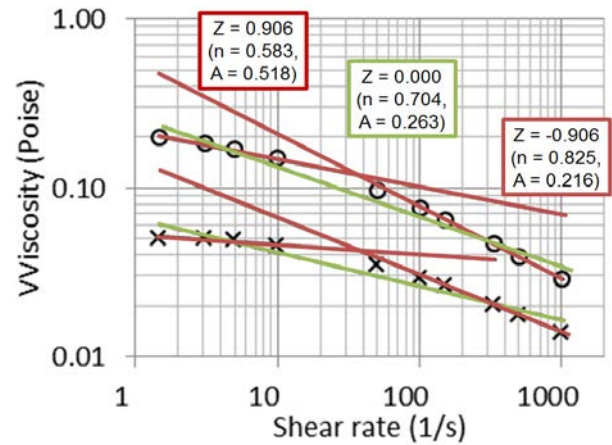


Figure 4: Rheology of PAC 2 (x) and PAC 4 (o); green line is best power-law fit, red lines are extremes used to assess uncertainty.

It is outside the scope of this paper to discuss the method in detail, but the coefficients  $\hat{x}$  and  $\hat{y}$  are found by exploiting the property that the Legendre polynomials are orthogonal under the uniform distribution and approximating the orthogonality integral through a high order quadrature.

For this case study we have opted to use five points in the quadrature, resulting in five simulations per fluid to find the approximate solution to the stochastic problem. It is paramount to the accuracy of the method to use the zeros of the 5th order Legendre polynomial in this case. For PAC 4 the parameters corresponding to these zeros are given in Figure 4. Once the parameters in the expansion (2) are found it is trivial to extract mean value, standard deviation (SD), or even the probability density function for the particle path.

### Particle trajectories

#### Numerical simulation

Numerical simulations were conducted with FLUENT using an Eulerian-Lagrangian concept. The fluid flow as the continuous phase is considered [Assumptions = Incompressible; Time-Independent; Isothermal; Laminar]. Using the Discrete Phase Model (DPM), the particle is subject to gravity and drag only and is additionally coupled to the fluid just one-way. In order to obtain fully-established velocity profiles the entrance length of the computational domain was increased to 2 m. Initially a 2D and finally a 3D Cartesian staggered grid of 1 mm cell size were found to produce grid-independent solutions after [Iterations = 300]. Numerical

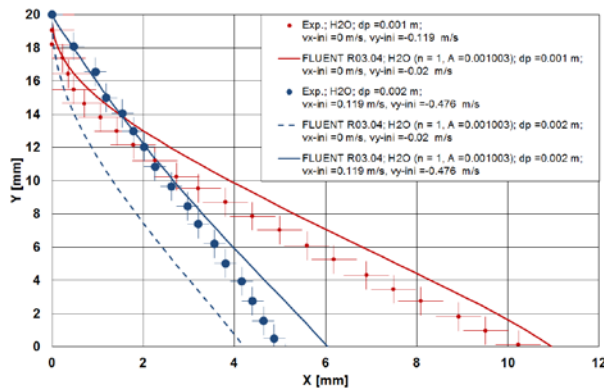
schemes used were the SIMPLE method for pressure-velocity coupling, implicit backwards Euler for temporal discretization and QUICK (Leonard) for spatial discretization.

#### Experimental work

Experimental studies were conducted using a small-scale flow loop where the test section (1.5 m in Figure 3) is transparent in order to obtain particle trajectories using PIV measurement techniques. This was done using a high-speed camera [Camera = SpeedCam MiniVis], [Frame rate = fps; 2500], [Resolution = Pixels; 512, 512] and a subsequent data processing using an in-house Matlab code based on a time-resolved digital Particle Imaging Velocimetry (PIV) lab tool developed by Thielicke and Stamhuis (2015) and Particle Tracking Velocimetry (PTV) lab tool developed by Ptalano and Wernher (2013).

#### Results

Both simulations and experiments were conducted for multiple combinations of inlet velocities [ $u_{avg} = \text{m/s}; 0.042, 0.085$ ], and above mentioned particle diameters as well as fluid types. Figure 5 shows experimentally and numerically obtained trajectories for  $\text{H}_2\text{O}$ .

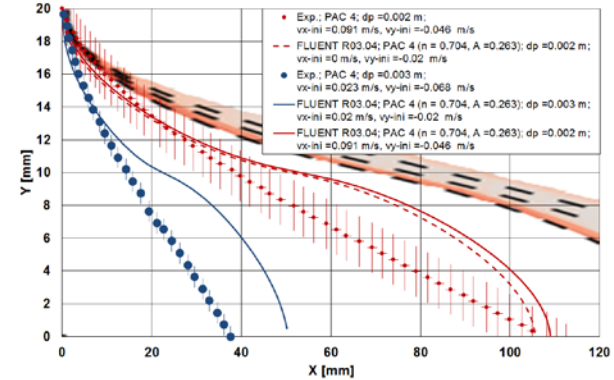


**Figure 5: Trajectories for  $\text{H}_2\text{O}$ ;  $v_{avg} = 0.085 \text{ m/s}$ ;  $\pm \text{SD}$**

The smaller particle is being transported further downstream resulting in an experimentally obtained [Impingement = mm; 10.34] (red circles). The numerical [Impingement = mm; 11.03] (solid red line) is just outside the SD of the experimentally obtained value. The bigger particle settles faster with an [Impingement = mm; 5] (blue circles). The experimentally obtained particle trajectories for  $\text{H}_2\text{O}$  show different shapes very close to the injection point. The trajectory for the larger particle shows a remarkably larger x-velocity-component compared to the smaller particle's trajectory. Figure 5 shows two different numerically obtained trajectories for the larger particle: One where the initial particle velocity is very small (dashed blue line) and one where the initial particle velocity was calibrated according to experimental data (solid blue line) with the respective numerically obtained values outside the experimental SD.

In the case of the two PAC fluids multiple simulation runs using the different combinations of model coefficients obtained with the Polynomial Chaos Approach were conducted. Figure 6 shows experimentally and numerically obtained trajectories for PAC, with the Polynomial Chaos results using 2D CFD represented by the reddish area for the  $d_p = 2 \text{ mm}$  case. Because the respective experimental trajectory (red circles) shows an [Impingement = mm; 107.4] a second CFD run using a 3D domain [Calibration 2 = Dimensions, 3] was conducted; the results of which could reproduce the experimental result within its SD (red dashed line). As with the  $\text{H}_2\text{O}$  case the CFD model was calibrated regarding the initial particle velocities (red solid line). The bigger particle ( $d_p = 3 \text{ mm}$ ) settles faster with an [Impingement = mm; 36.09] (blue circles). The respective

numerical trajectory is far off the experimental SD (blue solid line).



**Figure 6: Trajectories for PAC 4;  $v_{avg} = 0.085 \text{ m/s}$ ;  $\pm 3\text{SD}$**

#### Discussion of preliminary results

For both the Newtonian and non-Newtonian case the experimentally obtained trajectories indicate different initial particle velocities. A rough estimate using the gradient of the first two sample points of each experimental series revealed some considerable scatter. The so obtained numerical values were then used as improved initial particle velocity in the simulations [Calibration 3 = m/s;  $v_{x-init} 0.119$ ;  $v_{y-init} -0.476$ ], leading to slightly improved predictions in the case of  $\text{H}_2\text{O}$  (Figure 5, solid blue line) but not on the PAC case (Figure 6, solid red line). The initial particle velocities seem to have a major influence on the produced trajectories in the Newtonian case ( $\text{H}_2\text{O}$ ). In the non-Newtonian case (PAC) with a high effective viscosity the impact of the initial velocity is significantly smaller than for  $\text{H}_2\text{O}$ . Thus, initial particle velocities are a major uncertainty in the CFD model with an [Approximate uncertainty interval = m/s; -0.09,-0.47;-0.03,-0.06;-0.01,-0.06] for  $\text{H}_2\text{O}$ , PAC 2 and PAC 4 respectively.

Another major uncertainty is the power-law rheology model and its fit to the experimental data. Even though the 2D CFD results based on the power-law coefficients obtained with the Polynomial Chaos Approach do not result in correct trajectories the results can be used to estimate the [Approximate uncertainty interval = mm; 136, 163] of impingement with an [Approximate uncertainty range = mm; 27].

For the lower part of the channel, using the PAC fluids, a significant mismatch between the trajectories is found; numerically obtained trajectories show a S-shape whereas experimentally obtained trajectories show a more-or-less straight line [Mismatch = True], [Polynomial order of experimentally obtained trajectory = 2], [Polynomial order of simulated trajectory = 3]. The difference in the trajectories in the lower half of the channel is considered to be due to the fact that with the current modeling approach the local flow field of the particle is unresolved. Hence, the computed fluids viscosity, purely based on the main fluid flow velocity and the associated shear-rate, will not be sufficient to represent the local viscosity experienced by the particle. As the particles sediments through the liquid, we expect that the local strain will reduce the viscosity further, reduce the drag force and enhance the sedimentation.

Based on the preliminary results a description of necessary further work formatted as attribute-value pairs is given as follows:

- [Further work 1 = Redesign the particle inlet geometry in order to reduce natural scatter of direction and magnitude of the initial x- and y-velocity components and the associated uncertainty.]
- [Further work 2 = Generate a more sophisticated model which accounts for the local shear around the particle and

the respective local viscosity in order to improve the shape of the computed trajectory.]

- [Further work 3 = Use a more sophisticated rheology model, i.e. the Carreau-Yasuda instead of the power-law model, to improve the model of the rheometric data and reduce the associated uncertainty.]
- [Further work 4 = Use the Polynomial Chaos Approach on other important but uncertain parameters such as the flow rate.]
- [Further work 5 = Finalize 3D simulations and compute an overall uncertainty of the model using the [Validation metric = Regression, (Oberkampf and Barone, 2006)] on both the impingement distance and the trajectory.]

### Framework design considerations

Following a data-centric approach resulted in some interesting experiences that we discuss in this section.

*Preparing a case description for the database:* The project team understood and supported the documentation of the case using templates and metadata and the need for an iterative improvement process (four iterations). It was beneficial to have six participants in the case: one solution architect guiding the organization and structure of the case and providing the templates for description of the data and information and five analysts: two modelers, two experimentalists and one person dealing with uncertainties.

During the project some analysts occasionally showed the tendency to take experience-based shortcuts and use the quickest way to conclude some phases, without describing the whole procedure. Here the templates proved beneficial as they forced the analysts to document their approaches and findings. Different mind-sets of "modelers" and "experimentalists" will always challenge the multi-expertise projects. It always takes some time to establish "a common dictionary". It is important that the agreed dictionary becomes a part of the case description (dictionary metadata table). We suggest using two level of dictionaries (lookup tables): a *domain specific dictionary*, which is not part of the particular use-case data structure as it is of generic nature, and a *case-specific dictionary*, which is an integral part of the case data structure as it includes the case-specific definitions, terms, abbreviations, symbols, etc. While the first may have thousands of terms, the second one should be limited just to the nomenclature used in the case. That will also allow for some case-specific definitions.

The templates were useful for good description of the case, and in particular media-to-metadata mapping tables (each information/media item, e.g. text, image, schema, formula was accompanied by a descriptor metadata-table with attribute-value keys). These metadata tables were later used to insert the information in the database tables. Metadata is integral part of the queries for later retrieval of cases.

We are evaluating several solutions for mapping the geometry descriptions, various design information items, (e.g. technical drawings), formulas (equations sets) and other heterogeneous media to their metadata representatives. They are important for future automation of the case management and retrieval.

*Retrieval of the cases from the database:* We anticipate the need for several levels of queries: (a) queries for the case (based on the domain-specific metadata), (b) search for subcases and their descriptors (by use of case-specific metadata) and finally (c) search for the case results and various raw data.

*Organization of the data/information:* It is important to cater for differences in topology, structure and organization of the cases, i.e. ensure the standardization of data/information and metadata containers, while allowing the flexibility in their hierarchy and relations. Our ambition goes definitely towards a database management system, if possible in a standard way,

including interworking with major analytical (experimental and modeling) tools and FWs.

## 6. DISCUSSION AND CONCLUSION

Assessment of the uncertainty in predictions is a critical part or pragmatic modeling. We categorize the uncertainty into two classes: (1) Known uncertainty. Here the uncertainty lies into model input and numerics and is possible to be assessed with well-defined procedures. (2) Unknown uncertainty. This is the uncertainty that lies into the model concepts and physics. Only new experiments can help to reduce this uncertainty.

The presented case study illustrates the importance of addressing the a priori unknown uncertainty. This is done by exploitation of experimental data as an integral part of pragmatic modeling: Because of the experimental data the numerical model was improved with respect to the initial velocities and the constitutive equation of the fluid and vice versa the experimental setup received improvement because the simulations revealed the importance of the initial particle velocity. In addition it was clarified that the physical model for the drag-force is unsatisfactory for the non-Newtonian fluids tested, as significant deviations between experimental and theoretical particle trajectories were found. This allows us to go back and improve the physical model in order to reduce the uncertainty.

We believe that almost all the six phases in a typical industrial (pragmatic) analytical process illustrated in Figure 1 can to some extent get standardized, e.g.: (1) problem and context identification, (3) architecture of the analytical framework, (5) standard ways and criteria to evaluate the solution, and (6) standards for communicating the results and analytical context (for which they are valid, and usable). We can standardize the structure of the processes, the data, the tools that are used, the quality assurance methods, as well as establish standards for how the input, output / results and analytical context are presented and described. In particular the standardization of data types and flows is of major importance and both a challenge and major opportunity. Establishing a template-based database including all data, metadata, dictionaries and findings will enhance the interoperability and reuse of models as well as consistency of various models and/or experiments. This paper tries to contribute to these ideas, by taking the metadata and data / information perspective and providing some considerations how such a solution could look like.

## ACKNOWLEDGEMENTS

The project [Advanced Wellbore transport Modelling \(AdWell\)](#) with its sponsor, the Research Council of Norway and its partners Statoil, GDF SUEZ E&P Norge, IRIS, UiS, NTNU and SINTEF are gratefully acknowledged for funding and supporting part of this work.

## REFERENCES

- AIAA, (2005), "AIAA R-101A-2005 Recommended Practice - The CFD General Notation System - Standard Interface Data Structures." American Institute of Aeronautics and Astronautics.
- AIAA, (1998), "Guide: Guide for the Verification and Validation of Computational Fluid Dynamics Simulations (AIAA G-077-1998(2002))." American Institute of Aeronautics and Astronautics, Inc., Reston, VA.
- ANSYS - CFD,  
<http://www.ansys.com/Products/Simulation+Technology/Fluid+Dynamics> [Accessed June 9, 2015].
- ANSYS - STRUCTURAL,  
<http://www.ansys.com/Products/Simulation+Technology/Structural+Analysis> [Accessed June 9, 2015].
- ANSYS WORKBENCH PLATFORM,  
<http://www.ansys.com/Products/Workflow+Technology/ANSYS+Workbench+Platform> [Accessed June 9, 2015].

- ASME, (2009), "Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer: ASME V V 20." American Society of Mechanical Engineers (ASME).
- BABUSKA, I., AND ODEN, J. T., (2004), "Verification and validation in computational engineering and science: basic concepts." *Computer Methods in Applied Mechanics and Engineering* 193:4057–4066.
- BULATEWICZ, T., ANDRESEN, D., AUUVENSHINE, S., ET AL., (2014), "A distributed data component for the Open Modeling Interface." *Environmental Modelling & Software* 57:138–151.
- CFD GENERAL NOTATION SYSTEM, <http://cgns.sourceforge.net/> [Accessed June 9, 2015].
- DE LA CRUZA, R., CALMETA, H., AND HOUZEAUX, G., (2011), "Implementing a XDMF/HDF5 Parallel File System in Alya." DAKOTA, <https://dakota.sandia.gov/> [Accessed June 9, 2015].
- DONN, M., (2001), "Tools for quality control in simulation." *Building and Environment*:673–680.
- ERDEMIR, A., GUESS, T. M., HALLORAN, J., ET AL., (2012), "Considerations for reporting finite element analysis studies in biomechanics." *Journal of Biomechanics* 45:625–633.
- FREIBERG, M., STRIFFLER, A., AND PUPPE, F., (2012), "Extensible Prototyping for pragmatic engineering of knowledge-based systems." *Expert Systems with Applications* 39:10177–10190.
- FRENKLACH, M., (2007), "Transforming data into knowledge—Process Informatics for combustion chemistry." *Proceedings of the Combustion Institute* 31:125–140.
- GAYER, M., KORTELAINEN, J., AND KARHELA, T., (2010), "CFD modelling as an integrated part of multi-level simulation of process plants: semantic modelling approach." *In Proceedings of the 2010 Summer Computer Simulation Conference* pp. 219–227. Society for Computer Simulation International.
- GOPSILL, J. A., MCALPINE, H. C., AND HICKS, B. J., (2013), "A Social Media framework to support Engineering Design Communication." *Advanced Engineering Informatics* 27:580–597.
- GREENWALD, M., (2004), "Beyond benchmarking—how experiments and simulations can work together in plasma physics." *Computer Physics Communications* 164:1–8.
- HELTON, J. C., JOHNSON, J. D., AND OBERKAMPF, W. L., (2004), "An exploration of alternative approaches to the representation of uncertainty in model predictions." *Reliability Engineering & System Safety* 85:39–71.
- HINSEN, K., (2011), "A data and code model for reproducible research and executable papers." *Procedia Computer Science* 4:579–588.
- HU, K., DING, Y., ZHANG, X., ET AL., (2012), "A Scalable Infrastructure for Online Performance Analysis on CFD Application." *Chinese Journal of Aeronautics* 25:546–558.
- ISO/TC 184/SC 4, (2012), "ISO 14306:2012 - Industrial automation systems and integration - JT file format specification for 3D visualization." International Organization for Standardization (ISO).
- JARETEG, A., (2014), "Coupling of Dakota and OpenFOAM for automatic parameterized optimization." Chalmers University of Technology.
- KLEMENT, M., DÉD, T., ŠAFRÁNEK, D., ET AL., (2014), "Biochemical Space: A Framework for Systemic Annotation of Biological Models." *Electronic Notes in Theoretical Computer Science* 306:31–44.
- LLOYD, S., GAVAGHAN, D., SIMPSON, A., ET AL., (2007), "Integrative Biology — the challenges of developing a collaborative research environment for heart and cancer modelling." *Future Generation Computer Systems* 23:457–465.
- LOMBARD, M., AND YESILBAS, L. G., (2006), "Towards a framework to manage formalised exchanges during collaborative design." *Mathematics and Computers in Simulation* 70:343–357.
- MALEY, K., (2012), "STAR-CCM+ / Optimate." *In Singapore*.
- MALLERON, N., ZAOUL, F., GOUTAL, N., ET AL., (2011), "On the use of a high-performance framework for efficient model coupling in hydroinformatics." *Environmental Modelling & Software* 26:1747–1758.
- MENDENHALL, M. R., CHILDS, R. E., AND MORRISON, J. H., (2003), "Best practices for reduction of uncertainty in CFD results." *In 41st AIAA Aerospace Sciences Meeting*. MODEFRONTIER, <http://www.esteco.com/modelfrontier> [Accessed June 9, 2015].
- MOOIJ, W. M., BREDEVELD, R. J., DE KLEIN, J. J. M., ET AL., (2014), "Serving many at once: How a database approach can create unity in dynamical ecosystem modelling." *Environmental Modelling & Software* 61:266–273.
- NLOPT, <http://ab-initio.mit.edu/wiki/index.php/NLOpt> [Accessed June 9, 2015].
- NOVERE, N. L., FINNEY, A., HUCKA, M., ET AL., (2005), "Minimum information requested in the annotation of biochemical models (MIRIAM)." *Nat Biotech* 23:1509–1515.
- OBERKAMPF, W. L., AND BARONE, M. F., (2006), "Measures of agreement between computation and experiment: Validation metrics." *Journal of Computational Physics* 217:5–36.
- OBERKAMPF, W. L., DELAND, S. M., RUTHERFORD, B. M., ET AL., (2002), "Error and uncertainty in modeling and simulation." *Reliability Engineering & System Safety* 75:333–357.
- PORTO, <https://github.com/NanoSim/Porto> [Accessed June 9, 2015].
- ROACHE, P. J., (1998), "Verification and Validation in Computational Science and Engineering." Hermosa Publishers.
- ROBERT W. WAITERS, AND LUC HUYSE, (2002), "Uncertainty Analysis for Fluid Mechanics with Applications." National Aeronautics and Space Administration (NASA).
- ROY, C. J., AND OBERKAMPF, W. L., (2011), "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing." *Computer Methods in Applied Mechanics and Engineering* 200:2131–2144.
- RUSSI, T., PACKARD, A., AND FRENKLACH, M., (2010), "Uncertainty quantification: Making predictions of complex reaction systems reliable." *Chemical Physics Letters* 499:1–8.
- SANKARARAMAN, S., AND MAHADEVAN, S., (2015), "Integration of model verification, validation, and calibration for uncertainty quantification in engineering systems." *Reliability Engineering & System Safety* 138:194–209.
- SILCO, <https://wci.llnl.gov/simulation/computer-codes/silo> [Accessed June 9, 2015].
- SLOTNICK, J., KHODADOUST, A., ALONSO, J., ET AL., (2014), "CFD Vision 2030 Study: A Path to Revolutionary Computational Aerodynamics." National Aeronautics and Space Administration (NASA), Hampton, Virginia.
- STRANDENES, H., (2013), "Implementation of an Efficient Storage Strategy for Fluid-Particle Simulations in OpenFOAM." *In NTNU, Trondheim*.
- SURRIDGE, B. W., BIZZI, S., AND CASTELLETTI, A., (2014), "A framework for coupling explanation and prediction in hydroecological modelling." *Environmental Modelling & Software* 61:274–286.
- SWAYNE, D., YANG, W., RIZZOLI, A., ET AL., (2010), "Linking Data, Models and Tools: An Overview."
- TECPLLOT, INC., (2011), "Challenges of Analyzing Parametric CFD Results."
- TIME, R. W., AND RABENJAFIMANANTSOA, A. H., (2012), "An Experimental Study of Gas Lift Efficiency in Polymer Systems." *Annual Transactions of the Nordic Rheology Society* 20:59–69.
- VARNELL-SARJEANT, J., AMSCHLER ANDREWS, A., LUCENTE, J., ET AL., (2015), "Comparing development approaches and reuse strategies: An empirical evaluation of developer views from the aerospace industry." *Information and Software Technology* 61:71–92.
- XIU, (2010), "Numerical Methods for Stochastic Computations: A Spectral Method Approach." Princeton University Press.
- ZIGH, G., AND SOLIS, J., (2013), "Computational Fluid Dynamics Best Practice Guidelines for Dry Cask Applications." Office of Nuclear Regulatory Research.
- ZORIC, J., JOHANSEN, S. T., EINARSRUD, K. T., ET AL., (2014), "On pragmatism in industrial modeling." *In CFD 2014 - 10th International Conference on Computational Fluid Dynamics in the Oil & Gas, Metallurgical and Process Industries* (S. T. Johansen and J. E. Olsen, eds.) pp. 1–16. SINTEF, Trondheim.