

Paradigm for Research at Intevep

Daniel Joseph

September 2001

After many years of working with different research teams at Intevep I have developed some ideas about how research might be more effectively focused on problems of production of PSVSA affiliates. I also think that these procedures will produce a higher quality of research since it provides a mechanism for identifying real problems and for discarding approaches that don't work. I divide the research into two parts, data and simulators. These parts interact, but simulators must be tested against data, whereas research results from data alone can be obtained by intelligent interrogation without the use of simulators or models. The use of data at Intevep is very sporadic and way below where it should be. Simulators are the modern way to predict data; they are based on models that should be modified or discarded when they don't work.

Data

We may divide this into **data structure**, **data acquisition** and **data processing**.

The *data structure* is the way data is arranged; mainly quantities relevant to production, analysis and modeling are listed on spreadsheets in a data bank. Intelligent selection of data is maybe the first and arguably the most important aspect of research. The data structure is used to compare performance of different units used in production, to test ideas about which dimensionless groups are important, to create correlations and construct models.

Data acquisition. Any source of data is good but data from field operations, from experiments and from the literature are the main sources. Probably data from field operations should have a priority; it would be perhaps the hardest to get but it represents the real world of the PSVSA business.

Data processing involves the set of techniques used to process data for comparison, correlations, models and model validation. You don't need to construct models to do research. A tremendous amount of information is hidden in data and the research that needs to be done ought to be directed at how to ferret out this hidden information. Maybe we can find ways to understand and control sand and water by an intelligent comparison of different wells and reservoirs. It is a challenge.

In the oil industry there is a great effort to create mechanistic models of different production processes. The models are believed to be more fundamental than correlations. I can say that these models are not from first principles; they all use correlations and in some cases they lead to results which are less reliable than the underlying correlations. Every model *must* be validated against real data.

Data fitting is an element of data processing that I have surprisingly not seen used at Intevep. When one variable is functionally related to another, the plots of data scatter around the functional relationship. There are standard regression software programs that you use with spread sheets to get the "best" fit from data; the best fit comes with a variance measurement which tells you how good the fit is, or rather is a measure of the scatter. Usually it is a good idea

to identify the functional relationship by data fitting before you try to fit to a model. Research workers should have regression software and know how to use it.

Dimensionless groups. The construction of dimensionless groups is an important paradigm in fluid mechanics. When you get the right groups you have achieved the minimal description and possess the rules for scale up. For example, most research workers working with fluids at Intevep know that the flow of an incompressible fluid like water or oil is determined by a Reynolds number $Re = Ud\rho_f/\eta$ where U is a velocity, d is a length, ρ_f is the fluid density and η is its viscosity. The separate values U , d , ρ_f and η enter into dynamics only in the combination Re . This is a tremendous simplification. Multiphase systems are more complicated and more than one dimensionless group controls this system. For example, in my lecture on power laws in liquid-solid flows I showed that a sedimentation Reynolds number $R_G = \rho_f(\rho_f - \rho_p)gd^3/\eta^2$ was also needed to describe the fact that heavy particles in a slurry must be pulled down by gravity g , and $(\rho_f - \rho_p)gd^3$ is the buoyant weight for particles of diameter d and density ρ_p . There are 6 quantities, U , d , ρ_p , ρ_f , g and η but only two combinations of these 6 enter into dimensionless description.

Every process in the oil industry has a simplified description in terms of dimensionless parameters but the right parameters are to be determined.

When you know the equations governing a system the dimensionless parameters come out of scaling the equations. If you had a good mechanistic model you might try for a dimensionless description by scaling. But even if you have no equations, a thoughtful interrogation of the database using dimensional analysis could lead to the identification of relevant parameters.

One value of spreadsheets in a database is that dimensionless combinations of data can be created and functional relations between these combinations can be tested with a click on the mouse.

Correlations. Correlations can be regarded as data fits in the most primitive case. These are the functional relations between variables that I have mentioned before.

Power law correlations are of the form say,

$$Re = aR_g^n$$

where a is a prefactor and n the exponent. Such an expression was obtained for the criterion of lift-off of a heavy particle resting on the floor of a channel filled with fluid in a plane Poiseuille flow. By processing data from about 20 numerical experiments in log-log plots we found that power law correlations are ubiquitous in nature. The subject of power law correlations has been discussed in some recent papers¹ and to discuss them in detail here would tire me out. A power law is one kind of functional relationship between variables which occurs frequently and can be revealed by plotting data in log-log plots. Research workers should look for correlations in log-log plots.

Simulators and models. Simulators are numerical packages which implement mathematical models of production processes. “Eclipse” and “Stars” are simulators which use models of flow

¹ D.D. Joseph , 2001. Power law correlations for lift from direct numerical simulation of solid-liquid flow. *Int. J. Multiphase Fluids*. Accepted.

N.A. Patankar, D.D. Joseph, J. Wang, R. Barree, M. Conway, M. Asadi, 2001. Power law correlations for sediment transport in pressure driven channel flows, *Int. J. Multiphase Flow*. Submitted.

in pipelines, wells, porous media, etc. The simulators are based on models which should be validated against data. Research workers should build models for simulators. To do this they should understand what is inside the simulator and not use it as a black box. Mostly research workers at Intevep do not understand what is inside and do use the simulator as a black box. Maybe Intevep should have people lecture on the models used in a simulator to understand how the models might be improved.

Summary. You can obtain important information by intelligent interrogation of data. Being smart is identifying the *right data* and the way it should be interrogated.