

From: Sam <samuel.arey@epfl.ch>
Sent: Monday, December 20, 2010 5:49 AM
To: chris reddy <creddy@whoi.edu>
Cc: Sam <samuel.arey@epfl.ch>
Subject: LMRP GOR vs daily oil production rate

Hi Chris,

I did some statistics on the data that you sent me. I first plotted the data in matlab to convince myself that this is the same data set that is shown in Figure 1. By visual inspection I can see that it is the same data set as Fig 1. So, no problem there.

The squared correlation coefficient (r^2) for this data set is 0.10. By itself, this statistic suggests that there is really no significant relationship between LMRP GOR and daily production to the surface. To me, this is enough to end the story here. I think this tells us that there is not justification to draw a line through the data.

However, to give the benefit of the doubt to this possible relationship, I bootstrapped the linear regression using 10000 synthetic samples, using singular value decomposition as the linear regression engine. This is overkill, but it allows us to obtain solid numerical uncertainty values on the regression-computed slope and regression-computed intercept.

Here are the values.

The regression slope is -0.0140 and the regression intercept is 2405.

The regression slope has uncertainties as follows. At a confidence interval of one-sigma uncertainty in the slope, the slope values could range from -0.0022 to -0.0228.

For the moment let's ignore the uncertainty in the intercept and propagate only the uncertainty in the slope. Assume that the regression line passes through the centroid of the data, 14593 bbls and 2200 scf/bbls. This is a generous assumption but we will give the benefit of the doubt. If we "swivel" the regression line upon this axis (the data centroid) so that it explores the range demarcated by a 1-sigma uncertainty in the slope, then we can calculate the propagated uncertainty in the predicted daily oil production rate at the 1640 GOR line.

At a GOR value of 1640, the resulting predicted daily production rate values are:

daily production rate = 347800 bbls (upper bound) and

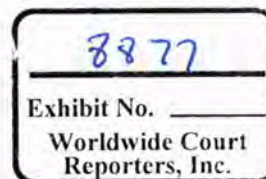
daily production rate = 33600 (lower bound).

This means that the daily production rate is constrained to within a factor of ten, using optimistic assumptions.

Again, these are only 1-sigma intervals. The 2-sigma (95% confidence) interval would generate an uncertainty in the daily production rate that is much larger still, and if we also included the uncertainty contributed from the regression intercept, then the uncertainty would be even yet larger.

Based on these numbers, I don't think we should include the plot in the paper. I think it will be most likely to get us shot down.

CONFIDENTIAL



WHOI-102091

Let me know what you think. Please feel free to forward to Rich and Jeff.

Sam

—
J. Samuel Arey
Joint Assistant Professor, EPFL/Eawag
Environmental Chemistry Modeling Laboratory

GR C2 544
ENAC-IIE-LMCE
Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne
Switzerland

<http://lmce.epfl.ch>
telephone: +41 (0) 21 693 8031
fax: +41 (0) 21 693 8070