

A cloud based cyberinfrastructure for automated model/data integration

Roelof Versteeg^{1*}, Haruko Wainwright², Nicola Falco², Doug Johnson¹, Erek Alper¹, Satish Karra³, Ye Zhang⁴, Peter Lichtner⁵, Reza Soltanian⁶, Kelly Wrighton⁷, Chris Henry⁸, Haiyan Zhou¹, Rebecca Rubinstein¹

¹Subsurface Insights, Hanover, NH

²Lawrence Berkeley National Laboratory, Berkeley, CA

³Los Alamos National Laboratory, Los Alamos, NM

⁴University of Wyoming, Laramie, WY

⁵OFM Research, Santa Fe, NM

⁶University of Cincinnati, Cincinnati, OH

⁷Colorado State University, Fort Collins, CO

⁸Argonne National Laboratory, Lemont, IL

Contact: roelof.versteeg@subsurfaceinsights.com

BER Program: SBR

Project: SBIR

Project Website: www.subsurfaceinsights.com

Project Abstract:

It is increasingly recognized that for both scientific, operational and regulatory reasons an in depth, near real time understanding of subsurface processes at thousands of sites will be required

Obtaining such an understanding will require an enabling cyberinfrastructure which can support data collection, ingestion, management and analysis as well as collaborative research and result delivery and information use at scale. Over the last several years, through both in house funding and under multiple DOE SBIR awards and funding Subsurface Insights has developed a cloud based cyber infrastructure and associated autonomous hardware for process understanding which increasingly enables this large scale understanding.

This cyber infrastructure (PAF – Predictive Assimilation Framework) leverages open source scientific software components developed by DOE scientists (PFLOTRAN, E4D, PyFlotran), and university scientists (ODM2, PEST, Landlab). It also takes full advantage of open source frontend libraries for graphing (Plotly, D3) and user interfaces (jQuery, React) and backend tools for data analysis (e.g. Scikitlearn and R and CobraPy). As part of this development Subsurface Insights has been contributing to the development of several of these open source codes.

PAF has been developed and demonstrated using data from both DOE funded projects and data from other research institutions and the private sector. PAF currently can ingest geochemical, geophysical, hydrological and remote sensed data and uses the PFLOTRAN reactive transport model for modeling and analysis. Both data