An Alternative Interpretation of Bayesian Updating

Sahyun Hong and Clayton V. Deutsch

Bayesian Updating (BU) is a widely used technique to combine primary and multiple secondary variables. BU decomposes the updated estimate and variance into the combination of prior and likelihood. The prior term results from the primary variable estimation and likelihood term results from the secondary variable. Representative statistics, global mean and variance, are also considered in BU. However, from the traditional expression of BU it is not clear how the information is combined. An alternative expression of traditional BU is derived. The main advantage of new expression is that the updated term is clearly decomposed into the prior, likelihood and global term. The new expression may lead to developments in non-stationary Bayesian updating.

Introduction

Secondary data integration is an important research field in reservoir characterization. The aim of data integration is to reduce uncertainty in predicted variable and provide plausible reservoir model. Simple collocated cokriging is a conventional way to account for secondary information, however, Bayesian updating has been developed as an alternative expression to collocated cokriging. Main advantage of Bayesian updating is that it decouples the primary and secondary information, which enables us to perform sensitivity analysis transparently. Besides, the decoupled form provides more flexibility in likelihood estimation when non-linear relation between primary and secondary exists.

In this note, we proposed another interpretation of Bayesian updating. New form divided the influence of primary, secondary and global information, which can be clearly understood. Besides, we applied the new form of Bayesian Updating (simply noted as BU) to the non-stationary modeling.

Bayesian Updating

BU form is proposed by Doyen et al. (1996) and it has been widely used to incorporate secondary variables for the primary variable modeling. CCG developed a new version of BU and it has some advantages over original form of BU. New form of BU considers three steps: (1) secondary variable aggregation if some secondary variables show relatively high correlation and (2) likelihood is calculated using the aggregated secondary variables and primary is calculated by kriging, and (3) combine two terms, primary and likelihood based on Bayesian formula. Bayesian updated estimate and variance are derived from the following updating equations:

$$\left.\begin{aligned} Y_{BU}(\mathbf{u}) &= \frac{Y_P(\mathbf{u})\sigma_L^2(\mathbf{u}) + Y_L(\mathbf{u})\sigma_P^2(\mathbf{u})}{\sigma_P^2(\mathbf{u}) - \sigma_P^2(\mathbf{u})\sigma_L^2(\mathbf{u}) + \sigma_L^2(\mathbf{u})} \\ \sigma_{BU}^2(\mathbf{u}) &= \frac{\sigma_P^2(\mathbf{u})\sigma_L^2(\mathbf{u})}{\sigma_P^2(\mathbf{u}) - \sigma_P^2(\mathbf{u})\sigma_L^2(\mathbf{u}) + \sigma_L^2(\mathbf{u})} \end{aligned}\right\} \tag{1}$$

Subscripts P and L means primary and likelihood. Simple kriging is used to estimate $Y_P(\mathbf{u})$ and $\sigma_P^2(\mathbf{u})$ over entire grids. Detail derivation of BU equation (1) can be seen in the references (Doyen, 1996; Lyster and Deutsch, 2004; Neufeld and Deutsch 2004). BU estimate is the combination of prior and likelihood term weighted by their estimation variance.

Alternative Interpretation of BU

Bayesian updated variance can be rearranged as

$$\frac{1}{\sigma_{BU}^2} = \frac{\sigma_P^2 - \sigma_P^2\sigma_L^2 + \sigma_L^2}{\sigma_P^2\sigma_L^2} = \frac{1}{\sigma_L^2} + \frac{1}{\sigma_P^2} - 1 \tag{2}$$

To reduce notation complexity, we removed (**u**) although all terms are function of location **u**. Eq. (2) is equivalent to the below equation since primary variable used in BU are normal scored values with zero mean and unit variance

$$\frac{1}{\sigma_{BU}^2} = \frac{1}{\sigma_L^2} + \frac{1}{\sigma_P^2} - \frac{1}{\sigma^2} \tag{3}$$

$\sigma_{BU}^2, \sigma_L^2, \sigma_P^2$ are locally derived variances and σ^2 is variance of primary variable being constant 1 over the entire study area. The updated variance must be between 0 and 1 and this is straightforwardly checked.

$$\left.\begin{array}{l} 1 \le \dfrac{1}{\sigma_L^2} \le \infty \\ 1 \le \dfrac{1}{\sigma_P^2} \le \infty \\ 1 \le \dfrac{1}{\sigma^2} \le \infty \end{array}\right\} \Rightarrow 1 \le \frac{1}{\sigma_L^2} + \frac{1}{\sigma_P^2} - \frac{1}{\sigma^2} \le \infty \Rightarrow 0 \le \sigma_{BU}^2 \le 1$$

From Eq. (1), Bayesian updated estimate and variance has common denominator $\sigma_P^2 - \sigma_P^2\sigma_L^2 + \sigma_L^2$, which provides a clue to a different expression. Updated variance of original BU is equal to the following

$$\sigma_P^2 - \sigma_P^2\sigma_L^2 + \sigma_L^2 = \frac{\sigma_P^2\sigma_L^2}{\sigma_{BU}^2} \tag{4}$$

Then Eq. (4) replaces the denominator of Bayesian updated estimate,

$$Y_{BU} = \frac{Y_P\sigma_L^2 + Y_L\sigma_P^2}{\sigma_P^2 - \sigma_P^2\sigma_L^2 + \sigma_L^2} = \frac{Y_P\sigma_L^2 + Y_L\sigma_P^2}{\dfrac{\sigma_P^2\sigma_L^2}{\sigma_{BU}^2}} = \frac{Y_P\sigma_L^2}{\dfrac{\sigma_P^2\sigma_L^2}{\sigma_{BU}^2}} + \frac{Y_L\sigma_P^2}{\dfrac{\sigma_P^2\sigma_L^2}{\sigma_{BU}^2}}$$

$$Y_{BU} = \frac{Y_P}{\sigma_P^2}\sigma_{BU}^2 + \frac{Y_L}{\sigma_L^2}\sigma_{BU}^2$$

Both sides are divided by updated variance σ_{BU}^2 then we obtain,

$$\frac{Y_{BU}}{\sigma_{BU}^2} = \frac{Y_P}{\sigma_P^2} + \frac{Y_L}{\sigma_L^2} \tag{5}$$

Eq. (5) can be seen that updated term is the variance weighted sum of primary and likelihood. Global mean and variance can be similarly added as.

$$\frac{Y_{BU}}{\sigma_{BU}^2} = \frac{Y_P}{\sigma_P^2} + \frac{Y_L}{\sigma_L^2} - \frac{m}{\sigma^2} \tag{6}$$

Eq. (6) is equal to Eq. (5) because the mean (*m*) and variance (σ^2) of primary variable are 0 and 1.

Now, BU form is re-expressed as following. Notation (**u**) representing location is added again to complete the equation.

$$
\left.\begin{aligned}
\frac{1}{\sigma_{BU}^2(\mathbf{u})} &= \frac{1}{\sigma_L^2(\mathbf{u})} + \frac{1}{\sigma_P^2(\mathbf{u})} - \frac{1}{\sigma^2} \\
\frac{Y_{BU}(\mathbf{u})}{\sigma_{BU}^2(\mathbf{u})} &= \frac{Y_P(\mathbf{u})}{\sigma_P^2(\mathbf{u})} + \frac{Y_L(\mathbf{u})}{\sigma_L^2(\mathbf{u})} - \frac{m}{\sigma^2}
\end{aligned}\right\} \tag{7}
$$

This new interpretation of BU provides a decomposed format of BU depending on the information sources; local primary, local secondary and global. This form is better than the original BU Eq. (1) in terms of simplicity, clarity and potential applicability to non-stationary mean and variance, $m(\mathbf{u})$ and $\sigma^2(\mathbf{u})$.

Application to Non-stationary Modeling

Many regionalized earth science data exhibit local variability depending on local mean. Original BU Eq. (1) assumes mean and variance independent of location $\mathbf{u}$. Locally varying mean and variance can be better integrated using the new interpretation of BU Eq. (7). $m(\mathbf{u})$ and $\sigma^2(\mathbf{u})$ are used in Eq. (7) for non-stationary modeling instead of constant m and σ^2.

$$
\left.\begin{aligned}
\frac{1}{\sigma_{BU}^2(\mathbf{u})} &= \frac{1}{\sigma_L^2(\mathbf{u})} + \frac{1}{\sigma_P^2(\mathbf{u})} - \frac{1}{\sigma^2(\mathbf{u})} \\
\frac{Y_{BU}(\mathbf{u})}{\sigma_{BU}^2(\mathbf{u})} &= \frac{Y_P(\mathbf{u})}{\sigma_P^2(\mathbf{u})} + \frac{Y_L(\mathbf{u})}{\sigma_L^2(\mathbf{u})} - \frac{m(\mathbf{u})}{\sigma^2(\mathbf{u})}
\end{aligned}\right\} \tag{8}
$$

When considering the locally varying variance, the primary and secondary likelihood variance also should be rescaled using that local variance. New BU equation for non-stationary modeling is derived as,

$$
\left.\begin{aligned}
\frac{1}{\sigma_{BU}^2(\mathbf{u})} &= \frac{1}{\sigma^2(\mathbf{u})\sigma_L^2(\mathbf{u})} + \frac{1}{\sigma^2(\mathbf{u})\sigma_P^2(\mathbf{u})} - \frac{1}{\sigma^2(\mathbf{u})} \\
\frac{Y_{BU}(\mathbf{u})}{\sigma_{BU}^2(\mathbf{u})} &= \frac{Y_P(\mathbf{u})}{\sigma^2(\mathbf{u})\sigma_P^2(\mathbf{u})} + \frac{Y_L(\mathbf{u})}{\sigma^2(\mathbf{u})\sigma_L^2(\mathbf{u})} - \frac{m(\mathbf{u})}{\sigma^2(\mathbf{u})}
\end{aligned}\right\} \tag{9}
$$

Example

Reservoir quality data was used to apply BU equation. 137 sample data were transformed into normal space. Data locations are shown in the figure on the next page. Obviously a data trend exists along with NW (low values)-SE (high values) direction. Although data trend is mitigated by strong conditioning, the trend still remains. The large scale trend of primary variable is shown. Global kriging with all conditioning data (137 samples) was used for the trend modeling. A secondary variable is shown. The correlation between primary and secondary is 0.55. The local variability is a function of local mean through the proportional effect. It is not the goal of this note to build mathematical relation between local mean and variance in order to describe proportional effect. It should be noted that data variability in high valued areas is highest. The local standard deviation increases as the local mean changes. A regression fit of $\sigma(\mathbf{u}) = 0.11(m(\mathbf{u})+0.65)^2+0.23$ was considered in the non-stationary BU of Eq. (9). Bayesian updating was applied to the primary, secondary and locally varying mean and standard variation.

Conclusions

New interpretation of Bayesian updating is introduced in this short note. Traditional and new expressions are mathematically equivalent; however, new form of BU has some advantages rather than old one. Because new interpretation decomposed the prior, likelihood and global information, it can be clearly understood how information source is combined. Besides, global mean and variance can be modified as locally varying value, which has great potential to apply to the non-stationary modeling.

References

- Neufeld, C. and Deutsch, C. V., 2004, Incorporating Secondary Data in the Prediction of Reservoir Properties Using Bayesian Updating. *Centre for Computational Geostatistics*.
- Doyen, P. M., den Boer, L. D., and Pillet, W. R., 1996, Seismic porosity mapping in the Ekofisk field using a New form of Collocated cokriging. *Society of Petroleum Engineers*, SPE 36498.

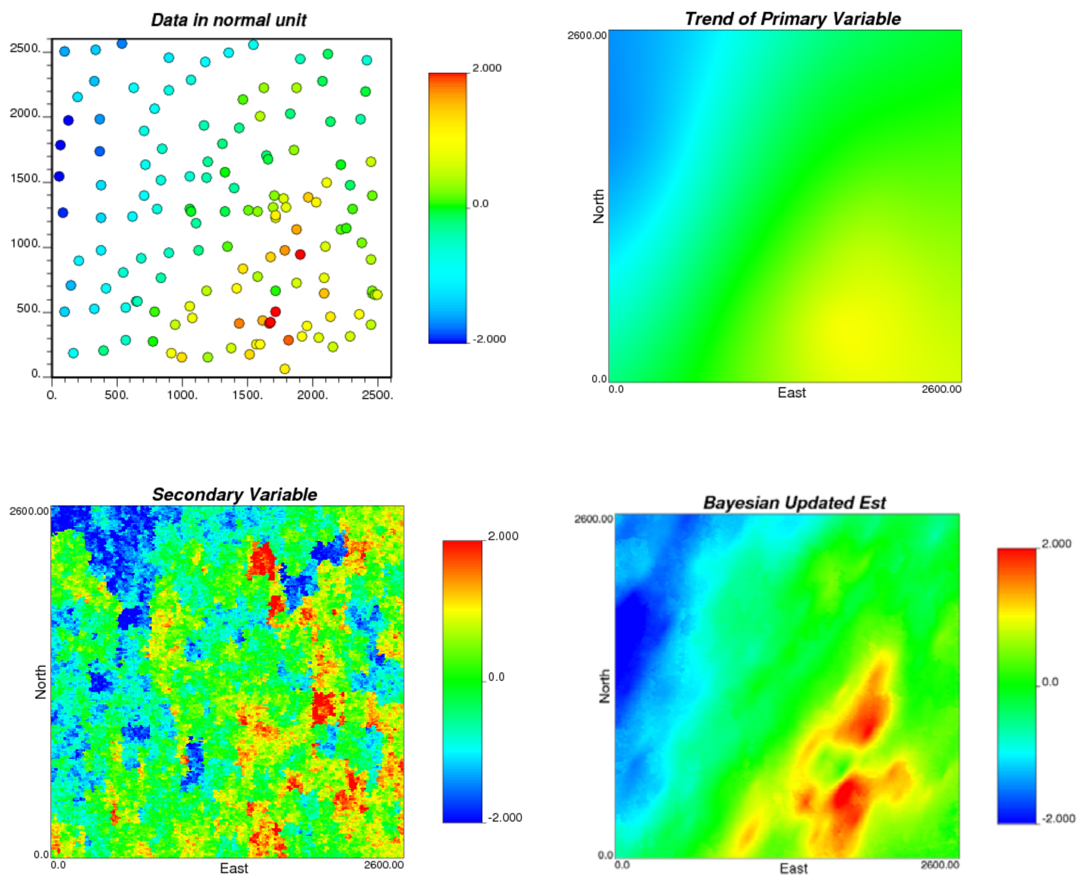

Figure: Data locations, trend of primary variable, secondary variable (0.55 correlation to primary) and updated map.