Water fraction uncertainty

Mark Hagemann

7/10/2019

Abstract

In this report I consider the task of estimating pixel water fraction with uncertainty. I show that using a Bayesian estimator with a non-informative conjugate prior distribution results in a water fraction uncertainty estimate that is larger than that obtained using the current estimation method and that agrees more closely with empirical errors from simulated data. The error in the current method arises from substituting an imprecise estimate of water fraction into an equation that computes water fraction uncertainty as a function of true water fraction. The corrected uncertainty estimate constitutes a simple linear scaling of the current method's estimate, making implementation of the corrected estimate a simple fix. The Bayesian estimator also results in an updated estimate of water fraction, although the fix is slightly more complicated to implement and smaller in magnitude than the correction to water fraction uncertainty.

Derivation

Note: here I use the k, θ (shape, rate) parameterization for both the gamma and inverse gamma distributions. Elsewhere these are sometimes written using α, β (shape, scale) paraemters, where $\alpha = k$ and $\beta = \frac{1}{\theta}$

We begin with the assertion that interferometric power follows a gamma distribution with shape parameter, k, equal to the number of rare looks and rate parameter, θ, related to water fraction, α, as follows.

$$\theta = \frac{\alpha \mu_w + (1 - \alpha)\mu_l}{k}$$

where μ_w and μ_l are the expected power from water and land, respectively, and treated as known constants.

The estimation of α is therefore equivalent to the problem of estimating the θ parameter from a single observation generated from a $Gamma(k, \theta)$ distribution. This is straightforward using Bayes rule, which gives

$$f(\theta|x) \propto f(x|\theta)\pi(\theta)$$

where x is the observed power. By selecting a noninformative conjugate prior distribution $\pi(\theta) = \frac{1}{\theta}$, we obtain the following posterior distribution:

$$f(\frac{1}{\theta}|x) \sim \text{Gamma}(k, \frac{1}{x})$$

Equivalently, the posterior for θ has an $inversegamma$ distribution:

$$f(\theta|x) \sim \text{Inv-Gamma}(k, \frac{1}{x})$$

By properties of the inverse gamma distribution, we obtain the expected value and variance of θ:

- $E[\theta] = \frac{x}{k-1}$
- $Var(\theta) = \frac{x^2}{(k-1)^2(k-2)}$

Converting to water fraction, α, and substituting N_l for k we obtain estimates of α and its uncertainty:

- $\hat{\alpha}_{new} = E[\alpha] = \frac{\frac{N_l}{N_l - 1} x - \mu_l}{\mu_w - \mu_l}$
- $\sigma^2_{\alpha,new} = Var(\alpha) = \frac{N_l^2 x^2}{(\mu_w - \mu_l)^2 (N_l - 1)^2 (N_l - 2)}$

Comparison to current estimates

The current method estimates α and its variance as follows:

$$\hat{\alpha}_{cur} = \frac{x - \mu_l}{\mu_w - \mu_l}$$

and

$$\sigma^2_{\alpha,cur} = \frac{x^2}{N_l(\mu_w - \mu_l)^2}$$

Note that these look similar to the expected value and variance presented in the previous section, particularly if N_l is large. However, for the SWOT pixel cloud, $N_l = 7$, so the difference is significant. $\hat{alpha}_{cur}$ is actually the maximum likelihood estimator for α, but unlike $\hat{\alpha}_{new}$, $\hat{alpha}_{cur}$ is a biased estimator.

The change from $\sigma^2_{\alpha,cur}$ to $\sigma^2_{alpha,new}$ is a simple scaling:

$$\sigma^2_{\alpha,new} = \frac{N_l^3}{(N_l - 1)^2 (N_l - 2)} \sigma^2_{\alpha,cur}$$

For $N_l = 7$, the adjustment factor works out to 1.9056, or for standard deviation, $\sigma_{\alpha,new} = 1.3804\sigma_{\alpha,cur}$.

Empirical results

In order to validate $\sigma_{\alpha,new}$, a simulated pixel cloud from the Sacramento river was modified by applying the simple scaling of the `water_frac_uncert` variable (corresponding to σ_α) described in the previous section. The pixel clouds–modified and unmodified–were processed through RiverObs along with a GDEM-derived validation-truth pixel cloud to produce node-level estimates of area with uncertainty. The unmodified pixel cloud was also run through RiverObs using the "simple" area aggregation method, which does not use the water fraction estimate (or its uncertainty).

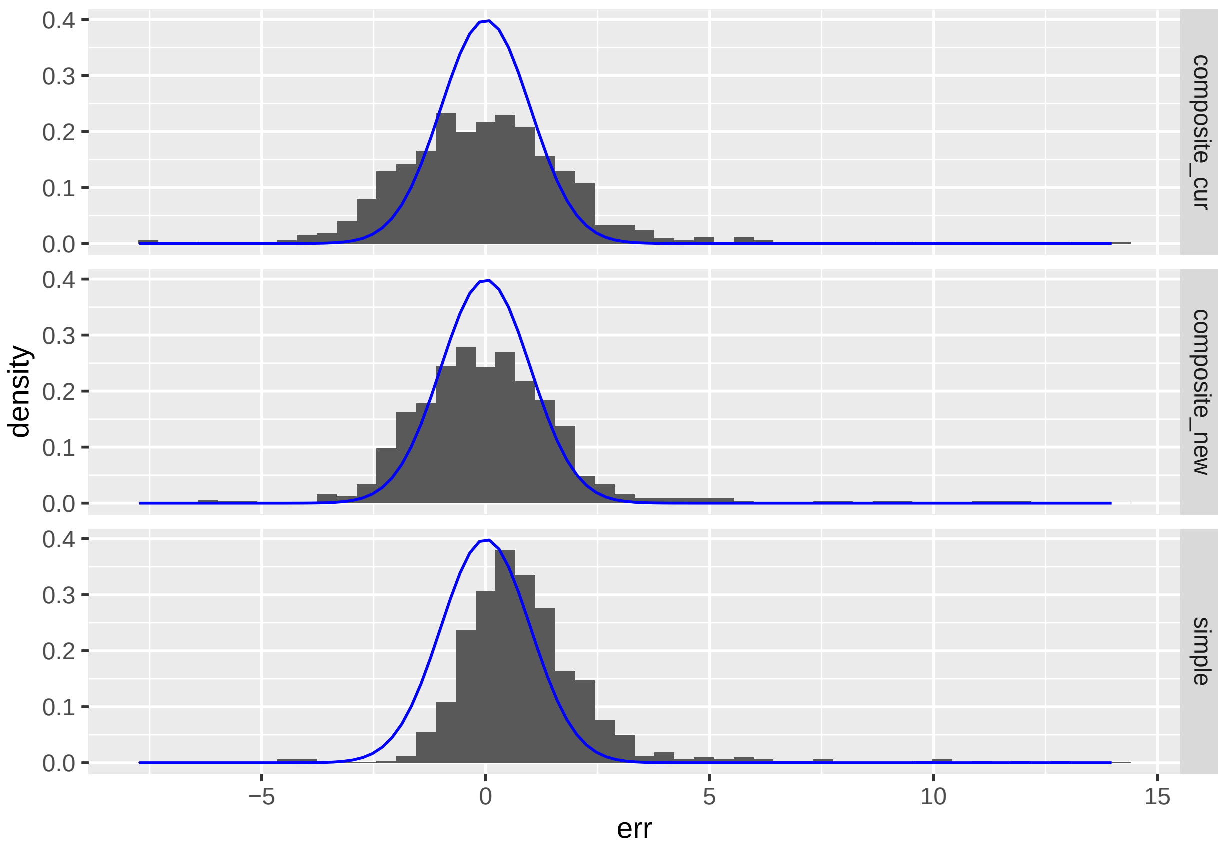
Results are presented as distributions of scaled node area errors across all 736 nodes in the simulation. The scaled errors, ϵ, are calculated as follows

$$\epsilon = \frac{\hat{A} - A}{\hat{\sigma}_A}$$

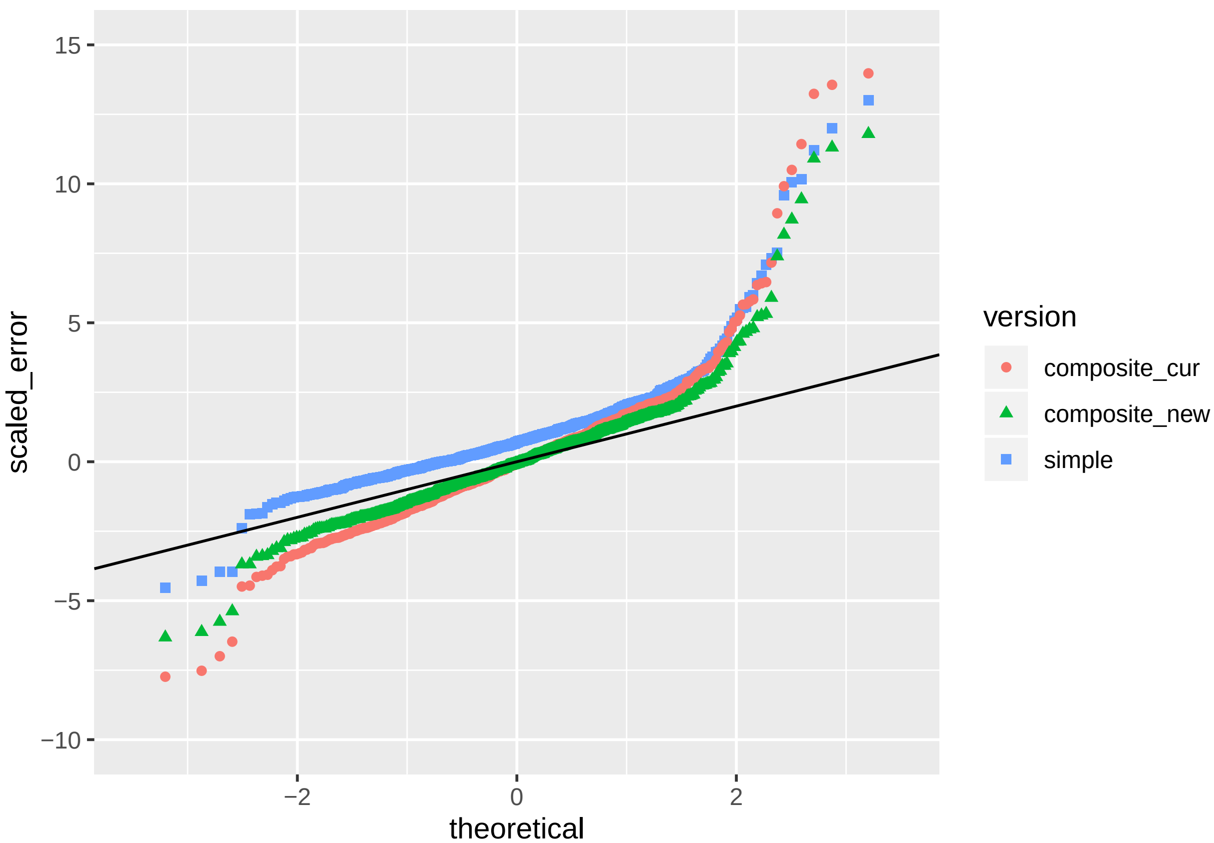
where $\hat{A}$ and A are the estimated and GDEM-true node areas and $\hat{\sigma}_A$ is the estimated uncertainty associated with $\hat{A}$. If the uncertainty estimates are correct, then the RMSE of ϵ over all nodes should be approximately 1, and if $\hat{A}$ is an unbiased estimate of A, then the mean of ϵ should be approximately 0. However, these may not necessarily be the case even if the water fraction are is correct, as other sources of error may confound the statistics, particularly in the tails of the distribution (as previously shown due to classification error and problematic truth data). Instead of relying on statistics, then, I instead present the distributions visually using quantile-quantile (QQ) plots and histograms.

Simple area aggregation

The "simple" method of aggregating pixel areas to produce node areas does not use the water fraction estimate, and therefore the uncertainty estimate, $\hat{\sigma}_A$ does not use σ_α. The histogram of ϵ with theoretical $N(0, 1)$ curve superimposed reveals scaled errors to be slightly biased and somewhat right-skewed, but generally having the spread that the model would predict. The current composite-area estimation is without bias, but its spread is significantly wider than theory would expect if $\hat{\sigma}_A$ were accurate. This difference in behavior between the simple and composite aggregation methods suggests that water-fraction uncertainty is being underestimated in the composite case. Applying the adjustment to water-fraction uncertainty results in a distribution of ϵ that is closer to what theory would expect–having a tighter spread and remaining unbiased–but the spread is still too large, suggesting that there is yet unaccounted-for uncertainty in the area estimate, likely in the water-fraction component.



Normal QQ plots of the ϵ distributions reveals the same behavior in finer detail, as a function of the scaled errors ranks in the distribution, mapped onto theoretical quantiles of a standard normal distribution. Points lying above the 1:1 line reflect ϵ values that are larger than theory would predict for that part of the distribution, and vice-versa for points below the line. For example, the sharp upward slope on the right side of the plot reflects a heavy upper tail to the ϵ distribution for all 3 versions of the node data. This heavy-tailed behavior is a separate phenomenon to the issue at hand, which is reflected in the slope of the QQ plot points roughly between -2 and 1.5 on the horizontal axis. A slope of 1–parallel to the 1:1 line, as the "simple" case exhibits–constitutes a positive validation of $\hat{\sigma}_A$, although the vertical offset from the line indicates an unaccounted-for bias in $\hat{A}$. A slope larger than 1 reflects roughly the factor by which $\hat{\sigma}_A$ is underestimated. This slope is approximately 1.7 for the current estimation method, 1.4 for the new method, and 1.1 for the simple method.



Conclusion

Estimated node area uncertainty is improved using a Bayesian estimator that accounts for imprecision in estimated water fraction when calculating water fraction uncertainty. The correction amounts to a simple scaling of the current method's estimated water fraction uncertainty based on the number of looks. Validation of the new method shows that it improves the performance of the uncertainty estimate, but that further unaccounted-for uncertainty remains.

Reference

Yang, Ruoyong, and James O. Berger. A catalog of noninformative priors. Institute of Statistics and Decision Sciences, Duke University, 1996. http://www.stats.org.uk/priors/noninformative/YangBerger1998.pdf