A Mathematical Framework for Uncertainty Quantification in Multimodal Image Analysis via Probabilistic Clustering Models

Maximillian G. Chen, David J. Stracuzzi, and Michael C. Darling

Sandia National Laboratories, Albuquerque, NM

Sandia National Laboratories

Background

- Most statistical and machine learning models only provide point estimates of parameters. However, statistical models are fundamentally stochastic: predicted and inferred values are random variables and inherently uncertain.
- Uncertainty quantification (UQ) provides a measure of sufficiency of the available data and the selected modeling approach for answering a question of interest.
- If we have multiple datasets from different sources , we need to be able to determine which datasets are useful in analysis and decision-making.

Consensus clustering (ensemble clustering): determination of an *overall (consensus) clustering or partition* of the observations in a dataset that agrees the most with the *source-specific clusterings*.

Goal: Develop a mathematical framework for relating the *uncertainty of the overall consensus clustering* to the *uncertainty of the source-specific clusterings* and apply it to the problem of segmenting multiple types of imagery over the same scene (multimodal imagery).

Bayesian Consensus Clustering (BCC) [4, 5]

1. Assumptions

- Source-specific clusterings adhere loosely to the overall consensus clustering.
- Source-specific and overall clusterings all have K clusters
- Data from M sources: $\mathbb{X}_1, \dots, \mathbb{X}_M$ (each data source may have disparate structure)
- Each data source is available for a common set of N objects
- X_{mn}: data m for object n
- Probability model for each data source: $f_m(X_n|\theta_m)$
- Each X_{mn}, $n = 1, \dots, N$, is drawn independently from a K-component mixture distribution specified by the parameters $\theta_{m1}, \dots, \theta_{mK}$.
- $L_{mn} \in \{1, \dots, K\}$: component corresponding to X_{mn}
- $C_n \in \{1, \dots, K\}$: overall mixture component for object n
- The source-specific clusterings $\mathbb{L}_m = (L_{m1}, \dots, L_{mN})$ are dependent on the overall clustering $\mathbb{C} = (C_1, \dots, C_N)$:

$$P(L_{mn} = k|C_n) = \nu(k, C_n, \alpha_m)$$

where α_m adjusts the dependence function ν.

- Data $\mathbb{X}_m$ are independent of $\mathbb{C}$ conditional on the source-specific clustering $\mathbb{L}_m$.

1. Assumptions (cont.)

- Conditional model:

$$P(L_{mn} = k|X_{mn}, C_n, \theta_{mk}) \propto \nu(k, C_n, \alpha_m) f_m(X_{mn|\theta_{mk}})$$

- We assume ν has the simple form

$$\nu(L_{mn}, C_n, \alpha_m) = \begin{cases} \alpha_m, & \text{if } C_n = L_{mn} \\ \frac{1-\alpha_m}{K-1}, & \text{otherwise} \end{cases}$$

where $\alpha_m = P(L_{mn} = C_n)$.

- Assume a Dirichlet(β) prior for $\Pi = (\pi_1, \dots, \pi_K)$, where $\pi_k = P(C_n = k)$
- Probability that an object belongs to a given source-specific cluster:

$$P(L_{mn} = k|\Pi) = \pi_k \alpha_m + (1 - \pi_k)\frac{1 - \alpha_m}{K - 1}$$

- Conditional distribution of $\mathbb{C}$:

$$P(C_n = k|\mathbb{L}, \Pi, \alpha) \propto \pi_k \prod_{m=1}^{M} \nu(L_{mn}, k, \alpha_m)$$

- Joint marginal distribution of $\mathbb{L}_1, \dots, \mathbb{L}_m$:

$$P(\{L_{mn} = k_m\}_{m=1}^{M}|\Pi, \alpha) \propto \sum_{k=1}^{K} \pi_k \prod_{m=1}^{M} \nu(k_m, k, \alpha_m)$$

2. Conjugate Prior Estimation

Data:

- $\mathbb{X}_i$ has a normal-gamma mixture with cluster-specific mean and variance

$$X_{mn}|L_{mn} = k \sim N(\mu_{mk}, \Sigma_{mk})$$

- μ_{mk} is a D_m dimensional mean vector, where D_m is the dimension of the data source m
- Σ_{mk} is a $D_m \times D_m$ diagonal covariance matrix, $\Sigma_{mk} = \text{Diag}(\sigma_{mk1}, \dots, \sigma_{mkD_m})$
- Prior distribution for θ_{mk}: D_m dimensional normal-inverse-gamma distribution

$$\theta_{mk} = N\Gamma^{-1}(\eta_{m0}, \lambda_0, A_{m0}, B_{m0})$$

where η_{m0}, λ_0, A_{m0}, and B_{m0} are hyperparameters.

- It follows that
  - $\frac{1}{\sigma^2_{mkd}} \sim \text{Gamma}(A_{m0d}, B_{m0d})$
  - $\mu_{mkd} \sim N(\eta_{m0}, \frac{\sigma^2_{mkd}}{\lambda_0})$ for $d = 1, \dots, D_m$

2. Conjugate Prior Estimation (cont.)

Conjugate prior distributions:

- $\alpha_m \sim \text{TBeta}(a_m = 1, b_m = 1, \frac{1}{K})$ (prior for α_m is uniformly distributed between $\frac{1}{K}$ and 1)
- $\pi \sim \text{Dirichlet}(\beta_0 = (1, 1, \dots, 1))$ (prior for Π is uniformly distributed on the standard $(M-1)$-simplex
- $\theta_{mk} \sim N\Gamma^{-1}(\eta_{m0}, \lambda_0, A_{m0}, B_{m0})$

Conditional posterior distributions (iteratively sampled via MCMC):

- $\Theta_m|\mathbb{X}_m, \mathbb{L}_m \sim p_m(\theta_{mk}|\mathbb{X}_m, \mathbb{L}_m)$ for $k = 1, \dots, K$

$$\theta_{mk} \sim N\Gamma^{-1}(\eta_{mk}, \lambda_k, A_{m0}, B_{m0})$$

- $\mathbb{L}_m|\mathbb{X}_m, \Theta_m, \alpha_m, \mathbb{C} \sim P(k|X_{mn}, C_n, \theta_{mk}, \alpha_m)$ for $n = 1, \dots, N$, where

$$P(k|X_{mn}, C_n, \theta_{mk}, \alpha_m) \propto \nu(k, C_n, \alpha_m) f_m(X_{mn}|\theta_{mk}).$$

- $\alpha_m|\mathbb{C}, \mathbb{L}_m \sim \text{TBeta}(a_m + \tau_m, b_m + N - \tau_m, \frac{1}{K})$, where τ_m is the number of samples n satisfying $L_{mn} = C_n$.
- $\mathbb{C}|\mathbb{L}_m, \Pi, \alpha \sim P(k|\Pi, \{L_{mn}, \alpha_m\}_{m=1}^{M})$ for $n = 1, \dots, N$, where

$$P(k|\Pi, \{L_{mn}, \alpha_m\}_{m=1}^{M}) \propto \pi_k \prod_{m=1}^{M} \nu(k, L_{mn}, \alpha_m)$$

- $\Pi|\mathbb{C} \sim \text{Dirichlet}(\beta_0 + \rho)$, where ρ_k is the number of samples allocated to cluster k in $\mathbb{C}$

Multimodal Image Segmentation Uncertainty Analysis

Problem Statement

- Use the *variance* as our measure of uncertainty
- Uncertainty of the overall consensus clustering: $\text{Var}[P(C_n = k)]$
- Uncertainty of the source-specific clusterings: $\text{Var}[P(L_{mn} = k_m)]$
- Question: Is there a function f such that

$$\text{Var}[P(C_n = k)] = f(\text{Var}[P(L_{mn} = k_m)]_{m=1}^{M})?$$

Overall Clustering Uncertainty ($M = 2$)

$$\text{Var}[P(C_n = k|\mathbb{L}, \Pi, \alpha)] \propto \text{Var}[\pi_k \prod_{m=1}^{M} \nu(L_{mn}, k, \alpha_m)]$$

For $M = 2$,

$$\text{Var}[P(C_n = k|\mathbb{L}, \Pi, \alpha)] \propto \text{Var}\{\pi_k[\nu(L_{1n}, k, \alpha_1)][\nu(L_{2n}, k, \alpha_2)]\}$$

$$\propto \begin{cases} \text{Var}[\pi_k \alpha_1 \alpha_2], & \text{if } L_{1n} = k, L_{2n} = k \\ \text{Var}[\pi_k \alpha_1 \frac{1-\alpha_2}{K-1}], & \text{if } L_{1n} = k, L_{2n} \neq k \\ \text{Var}[\pi_k \frac{1-\alpha_1}{K-1} \alpha_2], & \text{if } L_{1n} \neq k, L_{2n} = k \\ \text{Var}[\pi_k \frac{1-\alpha_1}{K-1}\frac{1-\alpha_2}{K-1}], & \text{if } L_{1n} \neq k, L_{2n} \neq k \end{cases}$$

Since π_k, α_1, and α_2 are all dependent,

$$\begin{aligned} \text{Var}(\pi_k \alpha_1 \alpha_2) &= \text{Cov}(\pi_k^2, \alpha_1^2 \alpha_2^2) + (\text{Var}(\pi_k) + [E(\pi_k)]^2)(\text{Var}(\alpha_1 \alpha_2) \\ &\quad + [E(\alpha_1 \alpha_2)]^2) - [\text{Cov}(\pi_k, \alpha_1 \alpha_2) + E(\pi_k)E(\alpha_1 \alpha_2)]^2 \\ \text{Var}(\alpha_1 \alpha_2) &= \text{Cov}(\alpha_1^2, \alpha_2^2) + (\text{Var}(\alpha_1) + [E(\alpha_1)]^2)(\text{Var}(\alpha_2) + [E(\alpha_2)]^2) \\ &\quad - [\text{Cov}(\alpha_1, \alpha_2) + E(\alpha_1)E(\alpha_2)]^2 \end{aligned}$$

$$\begin{aligned} \text{Var}(\pi_k \alpha_1 \alpha_2) &= \text{Cov}(\pi_k^2, \alpha_1^2 \alpha_2^2) + (\text{Var}(\pi_k) + [E(\pi_k)]^2)(\text{Cov}(\alpha_1^2, \alpha_2^2) \\ &+ (\text{Var}(\alpha_1) + [E(\alpha_1)]^2)(\text{Var}(\alpha_2) + [E(\alpha_2)]^2) - \underbrace{[\text{Cov}(\alpha_1, \alpha_2) + E(\alpha_1)E(\alpha_2)]^2}_{[E(\alpha_1 \alpha_2)]^2} \\ &+ [E(\alpha_1 \alpha_2)]^2) - [\text{Cov}(\pi_k, \alpha_1 \alpha_2) + E(\pi_k)E(\alpha_1 \alpha_2)]^2 \\ &= \text{Cov}(\pi_k^2, \alpha_1^2 \alpha_2^2) + (\text{Var}(\pi_\mathbf{k}) + [E(\pi_k)]^2)(\text{Cov}(\alpha_1^2, \alpha_2^2) \\ &+ (\text{Var}(\alpha_\mathbf{1}) + [E(\alpha_1)]^2)(\text{Var}(\alpha_\mathbf{2}) + [E(\alpha_2)]^2)) \\ &- [\text{Cov}(\pi_k, \alpha_1 \alpha_2) + E(\pi_k)E(\alpha_1 \alpha_2)]^2 \end{aligned}$$

Simulation One Results

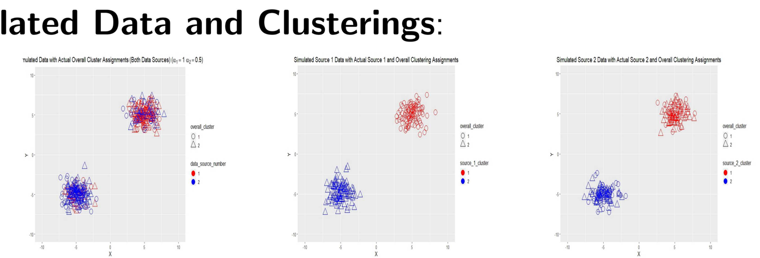
Generate simulated datasets $\mathbb{X}_1 : 2 \times 200$ and $\mathbb{X}_2 : 2 \times 200$:

1. Let $\mathbb{C}$ define two clusters, where $C_n = 1$ for $n \in \{1, \dots, 100\}$ and $C_n = 2$ for $n \in \{101, \dots, 200\}$.
2. Set $\alpha_1 = 1$ (perfect relationship) and $\alpha_2 = 0.5$ (no relationship).
3. For $m = 1, 2$ and $n = 1, \dots, 200$, generate $L_{mn} \in \{1, 2\}$ with probabilities $P(L_{mn} = C_n) = \alpha$ and $P(L_{mn} \neq C_n) = 1 - \alpha$.
4. For $m = 1, 2$, draw values X_{mn} from a $N_2([5, 5]', I_2)$ distribution if $L_{mn} = 1$ and from a $N_2([-5, -5]', I_2)$ distribution if $L_{mn} = 2$.

Analysis Details:

- Run BCC to obtain overall and source-specific clsuterings of two clusters each. ($K = 2$)
- 10,000 MCMC iterations using the R package bayesCC
- Point estimates of clustering probabilities: MAP estimates

Simulated Data and Clusterings:

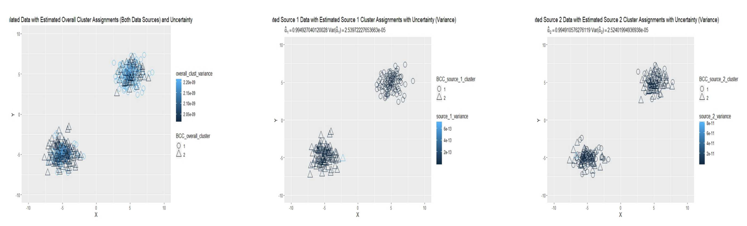


Overall Source 1 Source 2

BCC Estimated Clusterings and Uncertainty:

$\hat{\alpha}_1 = 0.994927$ $\text{Var}(\hat{\alpha}_1) = 2.5397 \times 10^{-5}$

$\hat{\alpha}_2 = 0.9949106$ $\text{Var}(\hat{\alpha}_2) = 2.52402 \times 10^{-5}$



Overall Source 1 Source 2

Simulation Two Results

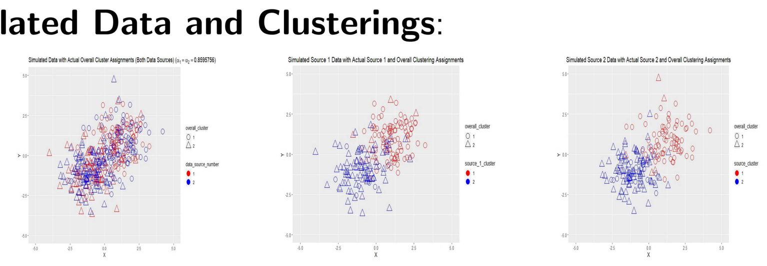
Generate simulated datasets $\mathbb{X}_1 : 2 \times 200$ and $\mathbb{X}_2 : 2 \times 200$:

1. Let $\mathbb{C}$ define two clusters, where $C_n = 1$ for $n \in \{1, \dots, 100\}$ and $C_n = 2$ for $n \in \{101, \dots, 200\}$.
2. Draw α from a Uniform(0.5,1) distribution. Let $\alpha_1 = \alpha_2 = \alpha$. The true $\alpha = 0.8595756$.
3. For $m = 1, 2$ and $n = 1, \dots, 200$, generate $L_{mn} \in \{1, 2\}$ with probabilities $P(L_{mn} = C_n) = \alpha$ and $P(L_{mn} \neq C_n) = 1 - \alpha$.
4. For $m = 1, 2$, draw values X_{mn} from a $N_2([1.5, 1.5]', I_2)$ distribution if $L_{mn} = 1$ and from a $N_2([-1.5, -1.5]', I_2)$ distribution if $L_{mn} = 2$.

Analysis Details:

- Run BCC to obtain overall and source-specific clsuterings of two clusters each. ($K = 2$)
- 10,000 MCMC iterations using the R package bayesCC
- Point estimates of clustering probabilities: MAP estimates

Simulated Data and Clusterings:

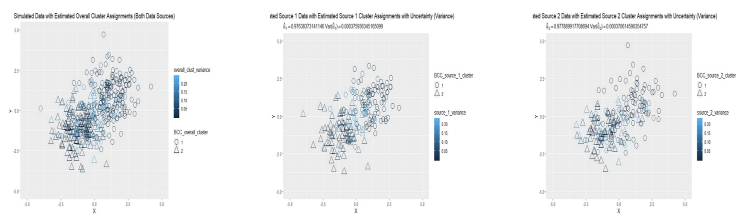


Overall Source 1 Source 2

BCC Estimated Clusterings and Uncertainty:

$\hat{\alpha}_1 = 0.97638$ $\text{Var}(\hat{\alpha}_1) = 0.0003759$

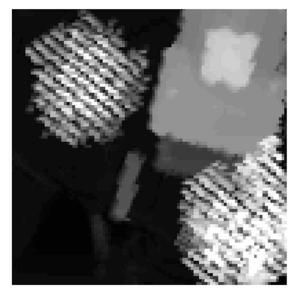
$\hat{\alpha}_2 = 0.9778899$ $\text{Var}(\hat{\alpha}_2) = 0.0003706$



Overall Source 1 Source 2

Philadelphia Imagery Data Results

Optical (RBG Values)



Lidar (Height of Objects)

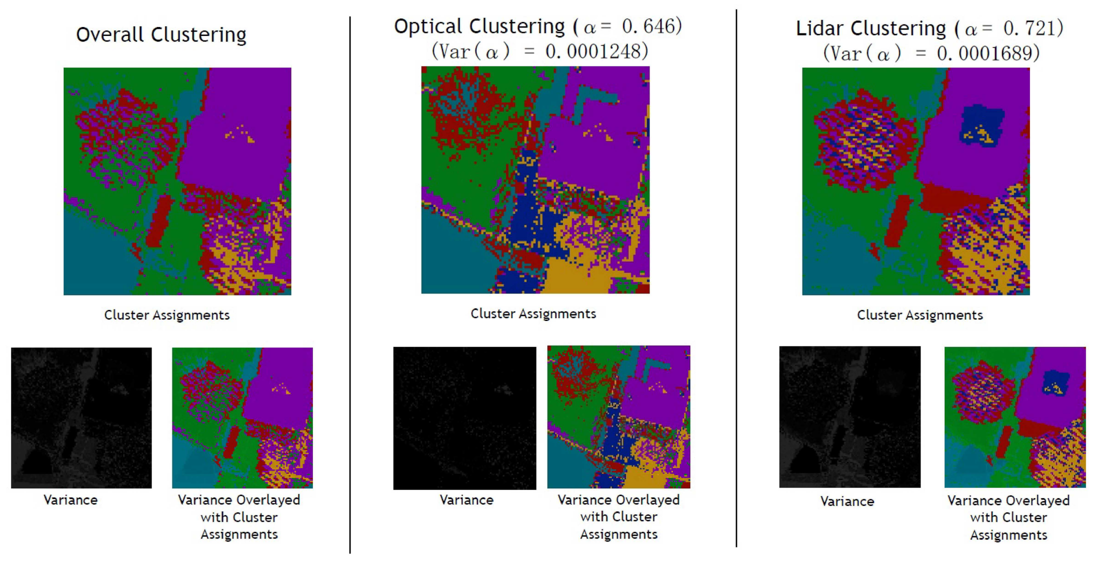
Analysis Details:

- Each image contains 10,000 pixels (100×100). ($N = 10,000$)
- Run BCC to obtain overall and image-specific clusterings, each with six clusters. ($K = 6$)
- 1,000 MCMC iterations using the R package bayesCC
- Point estimates of clustering probabilities: MAP estimates

BCC Estimated Clusterings and Uncertainty:

$\hat{\alpha}_{\text{optical}} = 0.646$ $\text{Var}(\hat{\alpha}_{\text{optical}}) = 0.0001248$

$\hat{\alpha}_{\text{lidar}} = 0.721$ $\text{Var}(\hat{\alpha}_{\text{lidar}}) = 0.0001689$



Conclusion:

- Lidar image has more influence on the overall clustering.
- Can we conclude that the lidar image is more valuable than the optical image?

Conclusion

The **uncertainty for the overall clustering is directly proportional to the uncertainties for the adherence of the source-specific clusterings to the overall clustering**, which affect the results of the source-specific clusterings.

Future Work

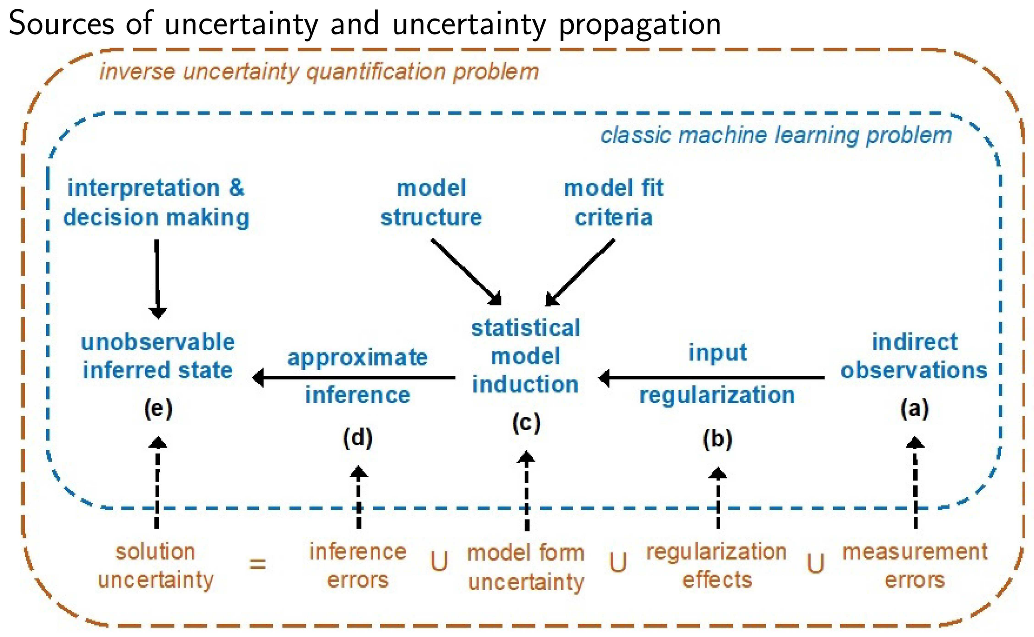
- Deriving expressions in closed form, particularly with the covariances
- Extending to the case of any number of M data sources
- BCC Implementation: Nonparametric distributions, different numbers of clusters and semantic meanings of clusters for each data source and overall clustering
- Analogous derivations for other uncertainty measures (standard deviation, entropy)
- Frequentist approach to consensus clustering and uncertainty quantification

Ongoing Uncertainty Quantification Work

- Empirical multimodal image analysis using Gaussian mixture models and nonparametric mixture models [6, 2]
- Supervised classification in image analysis [7]
- Seismic onset detection [8]
- URL classification [1]
- Visualizing clustering and uncertainty results for time-dependent data [3]

Future Uncertainty Quantification Research Directions

- Tying in cluster separability measures to uncertainty measures
- UQ and creating robust ground-truth evaluation benchmarks
- Extension from pixel-level to image-level UQ (uncertainty of cluster boundarries)
- Semi-supervised probabilistic clustering and uncertainty quantification
- Sources of uncertainty and uncertainty propagation



References

- Toward uncertainty quantification for supervised classification, author=Darling, Michael C. and Stracuzzi, David J. , year=2018, institution=Sandia National Lab.(SNL-NM), Albuquerque, NM (United States); Sandia National Lab.(SNL-CA), Livermore, CA (United States). Technical report.
- Maximillian G. Chen, Michael C. Darling, and David J. Stracuzzi. Multimodal image analysis and uncertainty quantification via nonparmaetric probabilistic clustering. Poster presented at the Conference on Data Analysis, March 2018.
- Maximillian G. Chen, Kristin M. Divis, J. Dan Morrow, and Laura A. McNamara. Visualizing clustering and uncertainty analysis of multivariate time-series data. Submitted to Statistical Analysis and Data Mining: The ASA Data Science Journal, June 2018.
- Eric F. Lock and David B. Dunson. Bayesian consensus clustering. *Bioinformatics*, 29(20):2610–2616, 2013.
- Eric F. Lock and David B. Dunson. Supplement to "bayesian consensus clustering". *Bioinformatics*, 2013.
- David J. Stracuzzi, Maximillian G. Chen, Michael C. Darling, Matthew G. Peterson, and Charles Vollmer. Uncertainty quantification for machine learning. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States); Sandia National Lab.(SNL-CA), Livermore, CA (United States), 2017.
- David J Stracuzzi, Michael C Darling, Maximillian G Chen, and Matthew G Peterson. Data-driven uncertainty quantification for multi-sensor analytics. In *SPIE Defense + Security, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX, 10635. Orlando, FL*. SPIE. Also available as Sandia Report 2018-3541C., 2018.
- Charlie Vollmer, Matthew Peterson, David J. Stracuzzi, and Maximillian G. Chen. Using data-driven uncertainty quantification to support decision making. In *Statistical Data Science Workshop*, 2017.

SAND2018

Contact: mgchen@sandia.gov