# Information content analysis for physical SST retrieval

**Prabhat K. Koner**
**Andy R. Harris & Eileen Maturi**

# History of Inverse Model

- Forward model: $Y = KX$; $d\mathbf{Y} = \mathbf{K}d\mathbf{X}$
- Inverse: $d\mathbf{X} = \mathbf{K}^{-1}d\mathbf{Y}$ (measurement error)
- Lengendre (1805) Least Squares:

$$\mathbf{X} = \mathbf{X}_{ig} + (\mathbf{K}^{T}\mathbf{K})^{-1}\mathbf{K}^{T}d\mathbf{Y}_{\delta}; \quad dY_{\delta} = \mathbf{Y}_{\delta} - \mathbf{Y}_{ig}$$

- Last 30~40 years: $\delta X \leq \kappa\, \delta E; \kappa = \mathrm{cond}(K)$

$$\mathbf{X} = \mathbf{X}_{ig} + (\mathbf{K}^{T}\mathbf{K} + \lambda\mathbf{R})^{-1}\mathbf{K}^{T}d\mathbf{Y}_{\delta}$$

- MTLS: $[u\,\sigma\,v] = [K \quad dY_{\delta}]$; $\lambda = (2\log(\kappa)/\|d\mathbf{Y}_{\delta}\|^{2})\,\sigma^{2}_{end}$
- OEM: $\mathbf{X} = \mathbf{X}_{a} + (\mathbf{K}^{T}\mathbf{S}_{e}^{-1}\mathbf{K} + \mathbf{S}_{a}^{-1})^{-1}\mathbf{K}^{T}\mathbf{S}_{e}^{-1}(\mathbf{Y}_{\delta} - \mathbf{Y}_{a})$
- Averaging kernel/Model Resolution Matrix

$$\mathbf{A} = \left\{(\mathbf{K}^{T}\mathbf{S}_{e}^{-1}\mathbf{K} + \mathbf{S}_{a}^{-1})^{-1}\mathbf{K}^{T}\mathbf{S}_{e}^{-1}\right\}\mathbf{K}; \quad \mathbf{MRM} = \left\{(\mathbf{K}^{T}\mathbf{K} + \lambda\mathbf{R})^{-1}\mathbf{K}^{T}\right\}\mathbf{K}$$

# Information Content

- Based on Shannon & Weaver (1949) study
- Rodgers stated (p. 34-37, 2000): information of measurement is the changing of entropy of the state space before and after measurement.

$$H = S(p_1) - S(p_2)$$

- After simplification:

$$H = -\frac{1}{2}\ln|I - A|$$

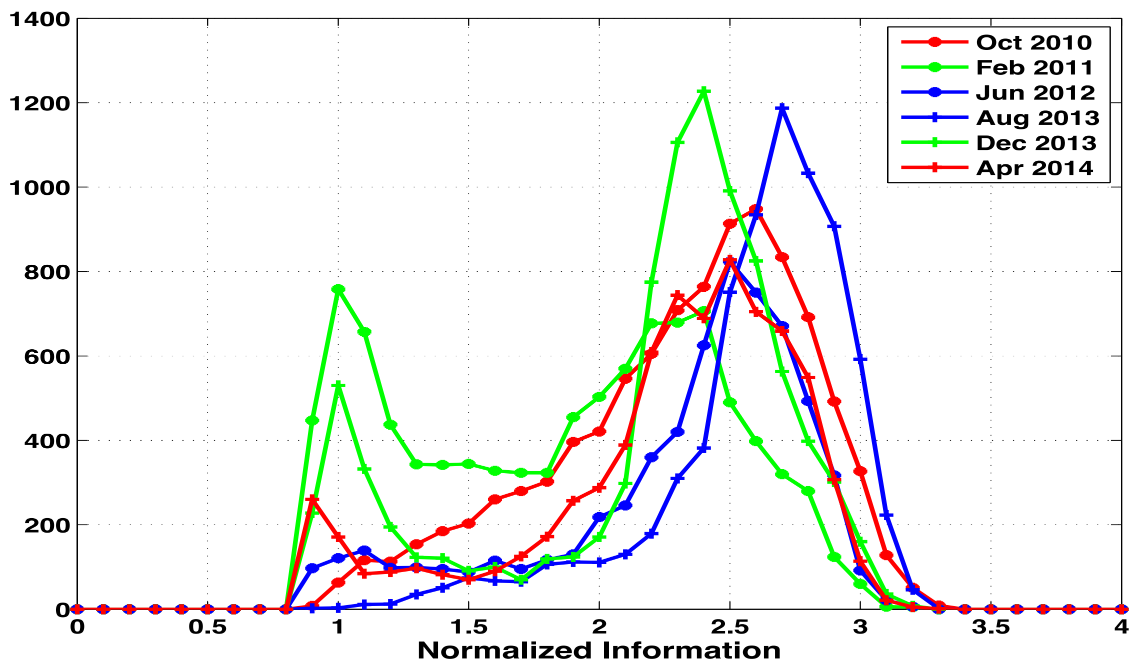
- **For LS, A=I, H=0!**

# Data and Forward model specifications

- Forward model using ver. CRTM2.1
- Monthly match up data with buoy
- iQUAM quality control data
- Using GFS ancillary data (NRT operational)
- Bayesian Cloud detection
- Night time scenarios
- Skin-bulk adjustment of 0.17K
- OEM error covariance:
  - GOES13 (3.9 11 13.4): 0.05 0.053 0.06
  - MTSAT1 (3.7 11 12): 0.18 0.15 0.18
  - MTSAT2 (3.7 11 12) : 0.09 0.11 0.2
  - CRTM error: 0.25 .15 .15
  - a-priori error: [1 15%tcwv)

# Normalized Information for SST retrieval from GOES13 using OEM

- NI=H/min(m,n)




❑ One measurement cannot produce more than one piece of information.

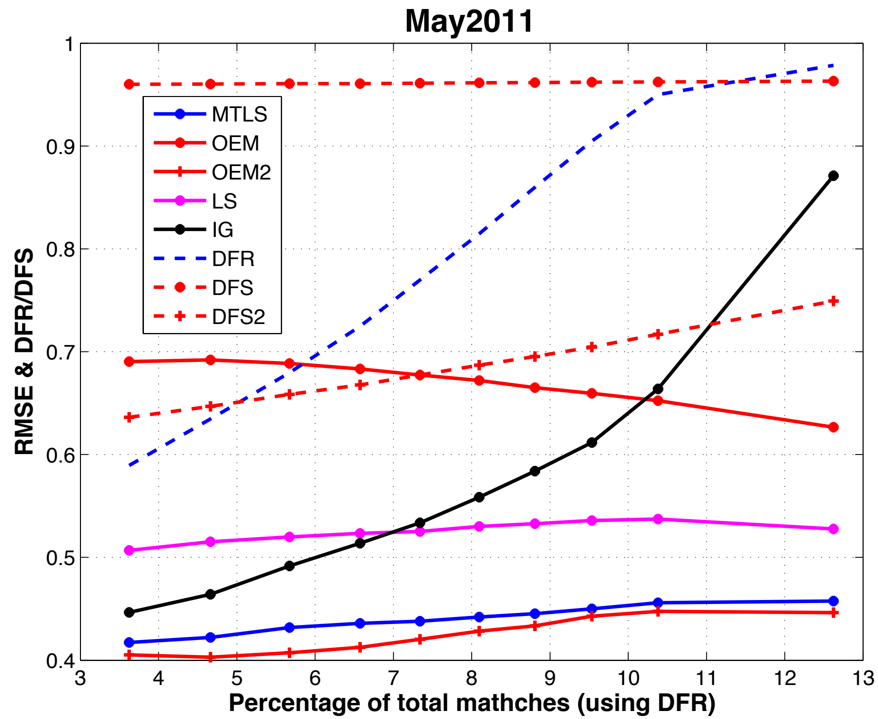
# Degree of Freedom

$$\mathbf{A} = \left\{ (\mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{e}}^{-1} \mathbf{K} + \mathbf{S}_{\mathbf{a}}^{-1})^{-1} \mathbf{K}^{\mathrm{T}} \mathbf{S}_{\mathbf{e}}^{-1} \right\} \mathbf{K}; \quad \mathbf{MRM} = \left\{ (\mathbf{K}^{\mathrm{T}} \mathbf{K} + \lambda \mathbf{R})^{-1} \mathbf{K}^{\mathrm{T}} \right\} \mathbf{K}$$

$$DFS_{nor} = trace(\mathbf{A}) / \min(m,n); \quad DFR_{nor} = trace(\mathbf{MRM}) / \min(m,n)$$

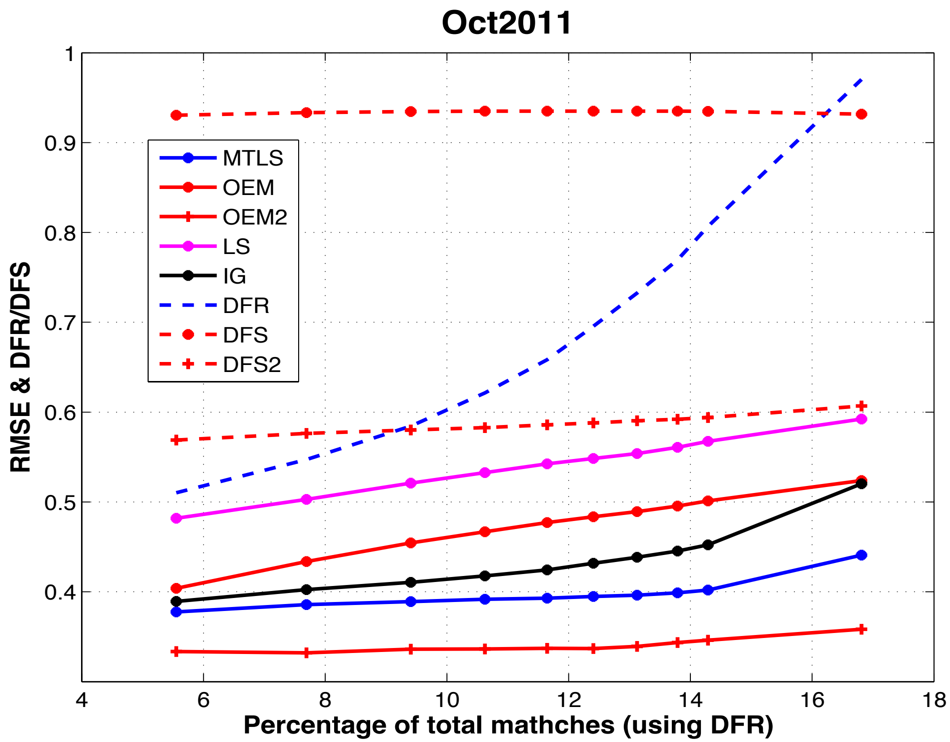
- **Normalized DFS/DFR of LS is one.**

# DFS/DFR and Retrieval error for GOES13



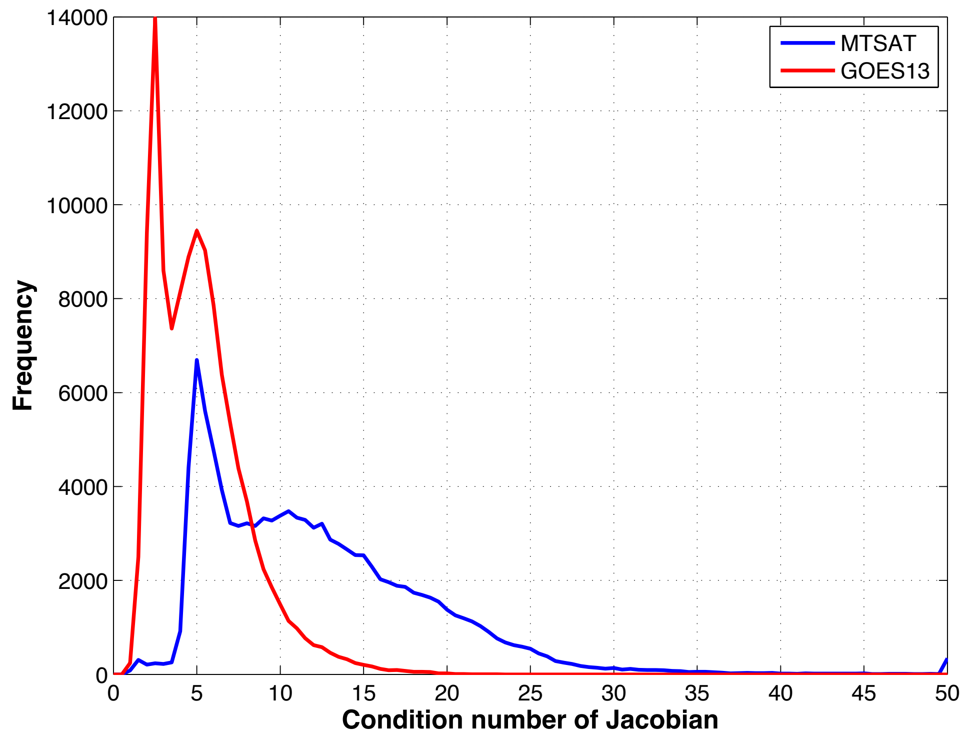

- Retrieval error of OEM higher than LS
- More than 75% OEM retrieval contains high error than a priori error.
- DFR of MTLS is high when a priori error is high
- The retrieval error of OEM is comparable when *a priori* perfectly known, but DFS of OEM is much lower than the same of MTLS.

# DFS/DFR and Retrieval error for MTSAT



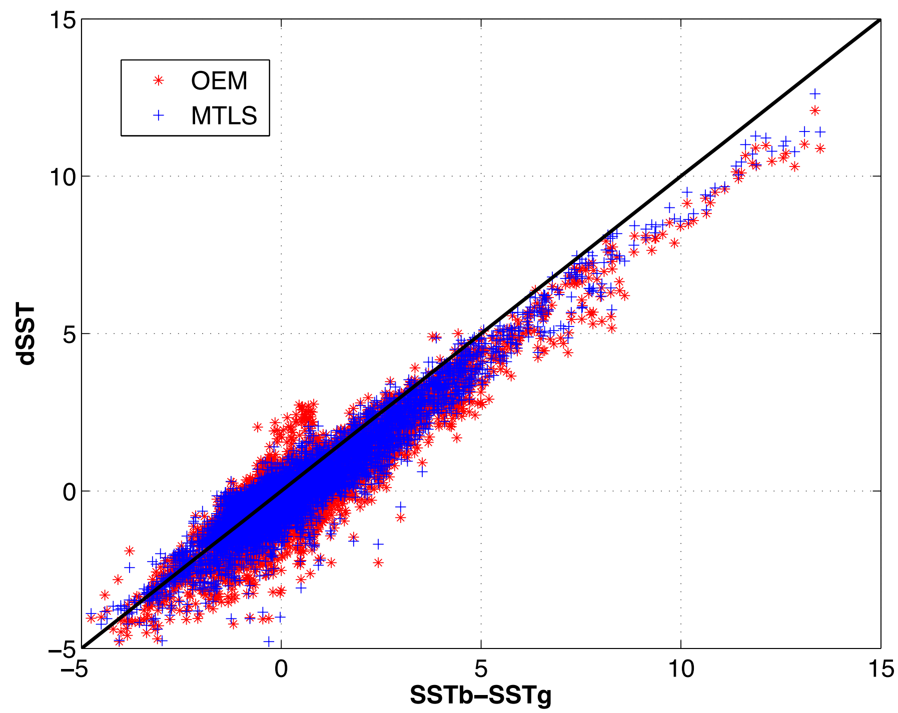

- LS error is higher than OEM
- OEM error is lower than MTLS when *a priori* perfectly known
- 100% OEM error higher than a priori error

# Distribution of Condition number




- ❑ Condition number of most of the GOES13 is lesser than 5
- ❑ Condition number of most of the MTSAT is higher than 5

# Innovation for GOES13




- MTLS regularizes more when signal-to-noise ratio is low and does not affect the sensitivity much
- In the other hand, when signal is high, it regularizes less and retains high DFR.
- For OEM, however, this mechanism relies on a fixed scheme.

# Summary and conclusions

- Developmental history of inverse algorithms and sensitivity study.

- In our study, MTLS shows the best performance

- This study also shows that for majority of cases, OEM solutions contain higher error than that of a priori.
- Additionally, whether OEM outperforms LS or vice versa depends on the condition number of the problem in hand. (discussed theoretically at the beginning, and shown practically)

- Sensitivity study shows that: a low DFR/DFS does not necessarily mean a more accurate product. In other words, DFR alone is inadequate to characterize the true sensitivity.

- The success of MTLS is attributed to its data-driven regularization, i.e., when IG error is high, regularization is low and vice versa.

# THANKS!