VALIDATION OF THE CLOUD AND CLOUD SHADOW ASSESSMENT SYSTEM FOR IKONOS IMAGERY (CASA-I2 VERSION 1.3)

GDA Corp. has developed an innovative system for **C**loud **A**nd cloud **S**hadow **A**ssessment (**CASA**) in Ikonos imagery. The system relies on spectral (VNIR), spatial and contextual information present in the image, and hierarchical self-learning logic to provide automated, per-pixel detection of clouds and cloud shadows. Average runtime per scene, on a standard 2GHz Pentium development computer, is 3 to 7 minutes with limited algorithm/code optimizations to date.

The current version of CASA-I2 has been validated on 200 Ikonos scenes for various areas, seasons, atmospheric and ground conditions. Each scene was visually inspected to assess, separately, percent dense cloud cover, percent light, transparent cloud and haze cover, and percent of total cloud and light cloud / haze cover. For each scene, two independent assessments of cloud cover were made. Then, results were compared and cases of significant disagreement were resolved by scene re-evaluation simultaneously by both operators. The distribution of cloudy and light cloud / hazy scenes within the dataset is presented in **Table 1**.

% Cloud Cover	Percent of Scenes		
	Dense Cloud Cover	Light Cloud / Haze Cover	Total Cloud / Haze Cover
0 to 10%	65%	82%	56%
0 to 30%	80%	95%	74%
0 to 50%	85%	100%	81%
0 to 70%	86%	100%	84%
0 to 90%	93%	100%	87%
Max Cover	100%	53%	100%

Table 1: Distribution of cloud and light cloud / haze contaminated scenes in the validation dataset

The results of the validation comparing CASA-I2 outputs to the visual assessments consistently show the high accuracy of CASA-I2 with respect to dense cloud detection (**Figure 1**). Comparison of the "truth dataset" to the cloud cover values reported in Ikonos metadata is presented in **Figure 2**. CASA-I2 results have also shown a relatively good identification of light, transparent clouds and haze (**Figure 3**), even though CASA was not specifically designed for light cloud / haze detection. Consequently, an overall assessment of combined, cloud and haze contamination of the imagery has shown rather promising results (**Figure 4**).

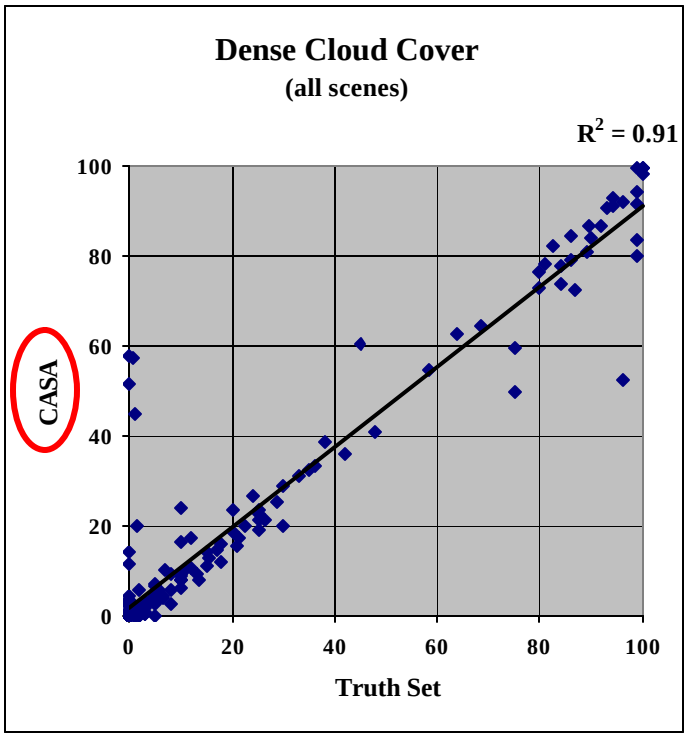

Figure 1: Cloud Detection: CASA vs. Truth Dataset

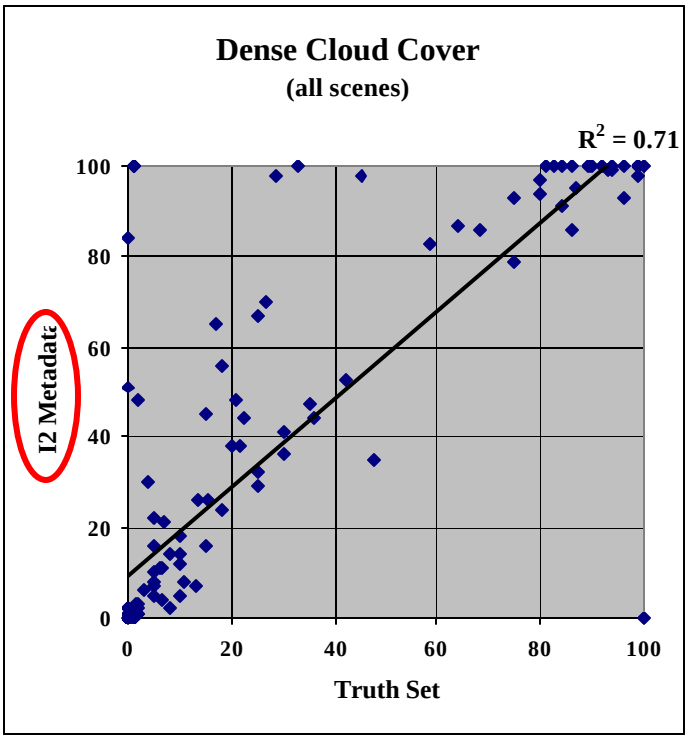

Figure 2: Cloud Detection: Ikonos Metadata values vs. Truth Dataset

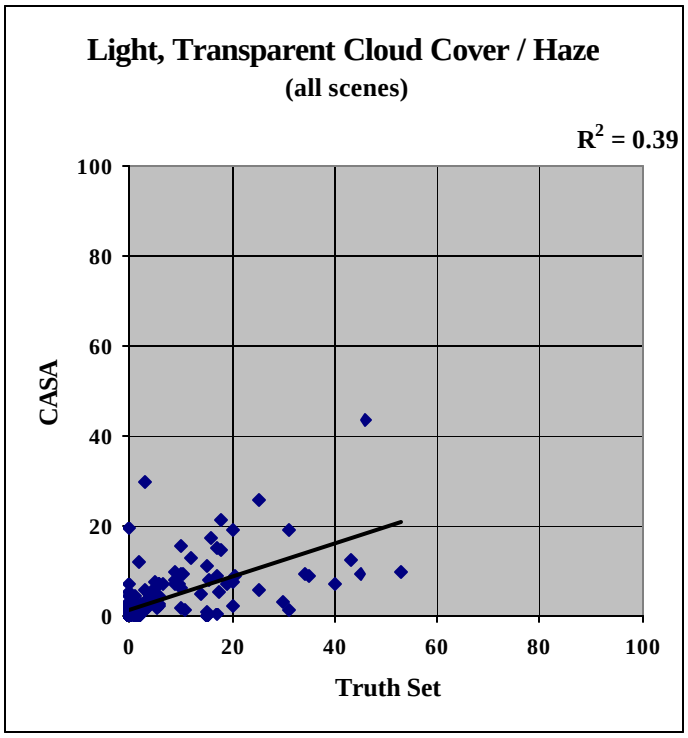

Figure 3: Haze Detection: CASA vs. Truth Dataset

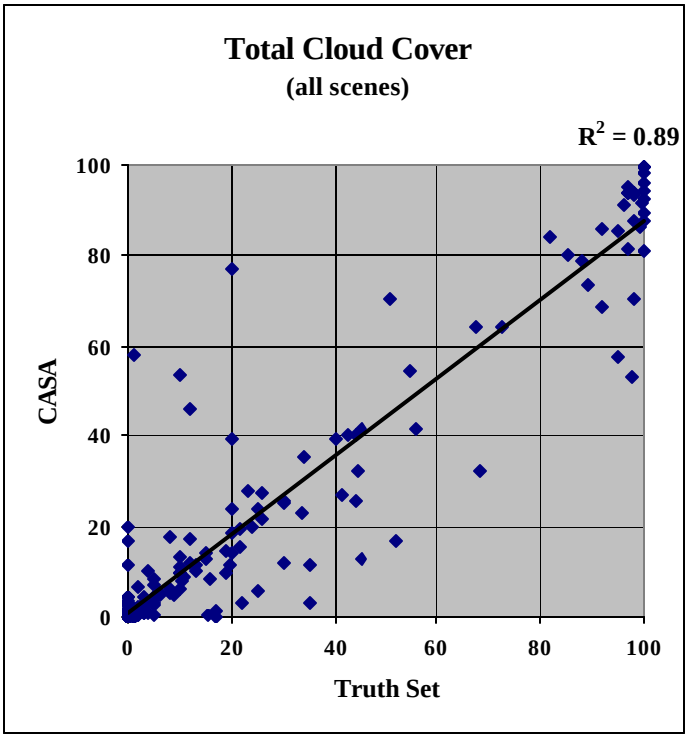

Figure 4: Combined Cloud and Haze Detection: CASA vs. Truth Dataset

The correlation values between the CASA-I2 results and the "truth" dataset are as follows:

- Dense clouds: 95.5%
- Light, transparent clouds and haze: 62.1%
- Combined, clouds and haze: 94.4%

The correlation values between the Ikonos metadata and the "truth" dataset (dense cloud cover) values is 84.2%.

As can be seen, CASA-I2 statistically correlates with visual cloud estimates very closely. **Figure 5** displays a summary of the CASA-I2 *vs.* visual assessment differences (error level) for the validation dataset. Overall, CASA is found to be within 10% of the visual estimate for more than 90% of all images tested (**Table 2**).

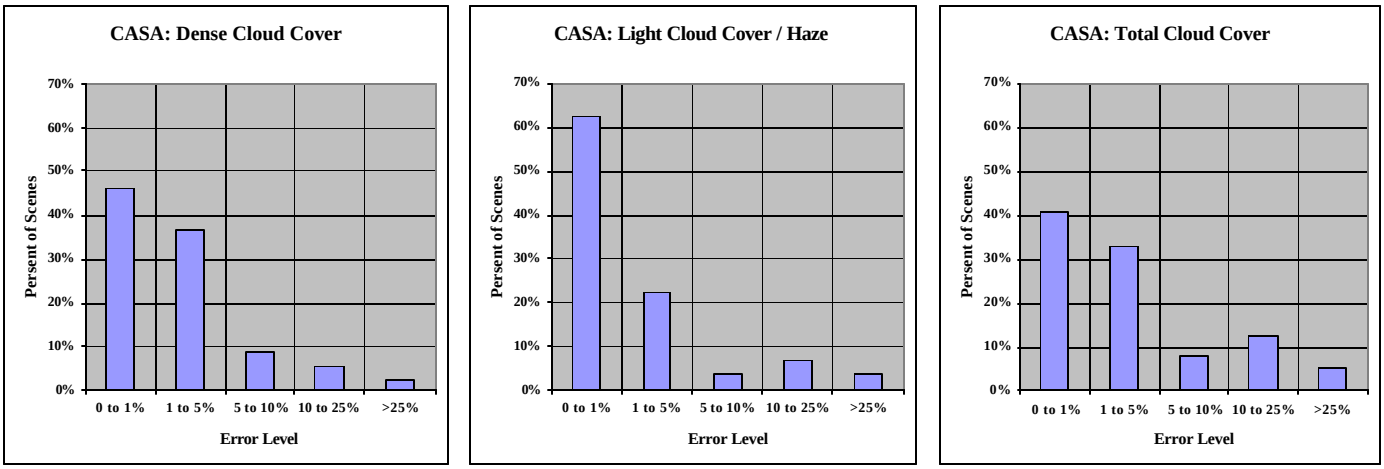

Figure 5: Summary of CASA vs. visual (truth) estimate of cloud cover: Differences by level of error

Error Level	Percent of Scenes		
	Dense Cloud Cover	Light Cloud / Haze Cover	Total Cloud / Haze Cover
0 to 1%	46%	63%	41%
0 to 2%	62%	76%	52%
0 to 5%	83%	85%	74%
0 to 10%	**92%**	89%	82%
0 to 15%	95%	93%	87%
0 to 25%	98%	96%	95%
Max Error	58%	43%	57%

Table 2: CASA vs. visual (truth) estimate of cloud cover: Differences by level of error

Initial implementation of the CASA-I2 version necessarily focused on accuracy over speed. However, performance is still quite reasonable (typically 3 to 7 minutes on a reasonably complex Ikonos image on a standard desktop PC). Also, it should be noted that while care has been taken to develop a computationally efficient implementation of CASA, there are many steps that could be taken to improve its performance.

Regardless of algorithm improvements, as with any fully automated system, there will always be cases where CASA may miss existing clouds or cloud parts and/or falsely label non-cloud objects as clouds. To aid identification of CASA results with potentially questionable quality of cloud detection, GDA Corp. is providing a quality flag in the textual output for each processed image. The flag grades CASA results as "good", "fair" or "poor" on the basis of (i) an internal CASA assessment of probabilities that detected

features are indeed clouds and (ii) the use of ancillary land cover, cloud probability, snow/ice probability datasets.

Furthermore, for situations where increased per pixel accuracy is desired, a user can request the generation of additional CASA spatial outputs to aid in editing CASA cloud masks. This would allow the user to improve the accuracy by manually correcting CASA output images. In addition to the standard cloud / cloud shadow mask, the user would be able to request various spatial outputs including: (i) a raster output depicting different cloud categories, (ii) raster outputs providing IDs for each individual cloud, separately for each cloud category, (iii) a raster output providing IDs for each individual cloud shadow, and (iv) raster with each cloud and/or cloud shadow being enlarged to a user-specified number of pixels/meters. These additional outputs give the image analyst more information with which to make decisions on individual potential cloud objects. The analyst's job would be simplified by the ability to remove/preserve either individual objects (based on their IDs) or object categories.

For further details please contact:

GDA Corp.
Innovation Park at Penn State University
200 Innovation Blvd.
Suite 234
State College, PA 16803
tel: 814-237-4060
fax: 814-237-4061
email: dmitry@gdacorp.com