Operational Concept for Obs-Based Forecasting of Clear Sky Condition

Timothy J. Hall, Carl N. Mutchler, Greg J. Bloy, Rachel N. Thessin, Stephanie K. Gaffney, and J. J. Lareau

The Aerospace Corporation, Engineering Technology Group, Chantilly, Virginia

Introducton

A range of military, civil and commercial activities benefit from cloud-free sky conditions. Passive optical or thermal systems, such as those on unmanned aerial systems, need a cloud-free-line-of-sight in order to sense their targets. Solar energy available to fuel photovoltaic power plants is strongly modulated by clouds due to their ability to reflect incoming shortwave radiation. Additionally, electricity demand on power grids is correlated to the amount of solar irradiance. Very short-range (i.e., up to 6-hr) sky condition forecasts are useful to decision makers for these applications.

In this paper, we present a concept for an approach to global, obs-based sky condition (i.e., total cloud amount) forecasting. This system would utilize meteorological satellite (METSAT) data and data-assimilation derived analyses to build feature databases for training advanced prediction algorithms capable of generating very short-range, probabilistic forecasts. Such a system is potentially useful for global, point-specific forecasts – particularly in data-sparse regions. Results from a separate thread of our investigation are presented that serve as a demonstration of this concept. More extensive findings based on the algorithm research are included in two companion conference papers.

Concept Components

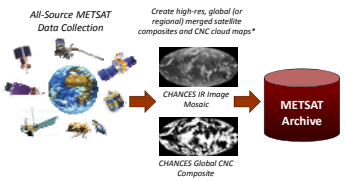
- Foundational data
- *Global, merged satellite composites & cloud/no cloud (CNC) maps*
- *Data assimilation-based, gridded analyses*
- Feature database for point or area of interest
- Prediction algorithm training
- Prediction algorithm initialization with "current" information
- Probabilistic sky condition forecast production

Application Frame of Reference

- To provide context for our concept – a hypothetical military decision maker scenario was developed
- Military commander manages a unique (i.e., high-demand, low availability), high-flying, high-cost unmanned aerial surveillance system (UAS)
- UAS's sensors are ineffective with cloud between sensor and target
- Commander must make frequent, short-fused decisions based on probabilistic forecasts to send (or not) the UAS to surveil locations in a large, data-sparse region
- Forecast system must provide probability of a clear sky condition over specific targets at 1-hr increments for a window of 1-5-hr from the time a decision is made
- Decision maker is assumed to act in order to minimize cost/operational risk and maximize benefit (i.e., info. collected)
- Support from a human weather forecaster is not present

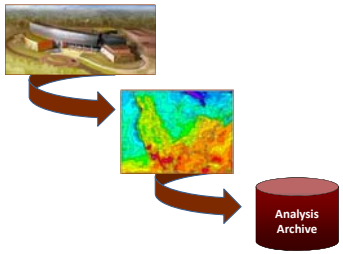
Foundational Data

METSAT Imagery & Data Assimilation-Based Analyses

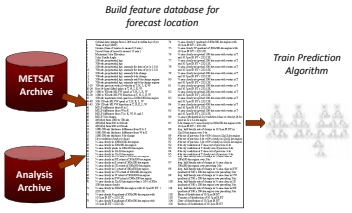


CHANCES project (Reinke et al. 2003, Vonder Haar et al. 1995) developed and demonstrated the production of high-resolution, merged METSAT composites and addressed challenges including registration, navigation, quality control, and cross-platform calibration critical for obs-based forecasting.

Data assimilation-based analyses

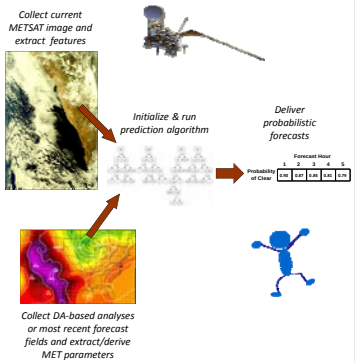


Build Feature Database & Train Prediction Algorithm



**CHANCES images used by permission of CSU/CIRA*

Collect Current Data & Run Prediction Algorithm



Concept Demonstration

- Five advanced forecast algorithms were applied to produce probabilistic forecasts for targets in six AORs at intervals from 1-5 hr
- Here, we present results from one of those algorithms (Random Forest) along with forecasts from three benchmark methods for comparison
  - These results are representative of the performance achieved by several of the algorithms tested across all of the AORs
- A half-hourly, tailored database of 78 features was used to train the Random Forest algorithm
- *Period of record was 1 May 03 - 29 Jun 08 (first 3 yr reserved for training)*
- ~ 50% of features were cloud structural features extracted from METSAT imagery
  - *GOES-12 & GOES-10 (at 4-km resolution)*
  - *Cloud-no cloud (CNC) maps created using Bispectral Composite Technique (Jedlovec et al. 2008)*
- ~ 50% of features were meteorological parameters from DA-based analyses
  - *Variables were extracted or derived from archived analysis fields generated by NCEP's Eta model data assimilation system (EDAS)*

Map showing the Ft Hood location (indicated by the triangle) at 31.14°N 97.71°W. The black square surrounding Ft Hood represents the outer extent of the 1000 x 1000-km region in which METSAT data were used as a source for features.

Concept Demo. (cont'd)

Random Forest Algorithm

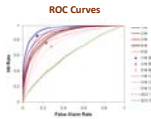
- Random Forest was developed by Leo Breiman to improve the performance of decision tree algorithms.
- Random Forest creates an ensemble of decision trees by training on a random redistribution of the training set.
- Each distribution is generated by randomly drawing N samples (with replacement), where N is the size of the training set. A tree is grown on a fixed-size subset of features randomly drawn on each round.
- The algorithm outputs the class that is the mode of the output by the individual trees.

Baseline Forecast Methods

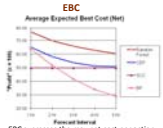
- *Basic persistence (BP):* 0% or 100% probability of clear at forecast time based on the initial sky condition.
- *Conditional-expectancy of persistence (CEP)*: The probability of clear at each forecast interval (1, 2, 3, 4, and 5 hr) given the current (or initial) sky condition (i.e., cloudy or clear) based on training data events within ± 1 hr and ± 30 d of that time of day and day of year.
- *Satellite Cloud Climatology (SCC):* The unconditional, prior probability of clear conditions for obs within ± 1 hr and ± 30 d (in the training datasets), calculated as the percentage of total occurrences that were clear at that time of day, day of year.

Performance Assessment

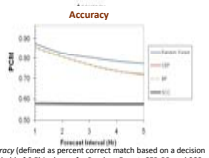
Overall performance potential was assessed using relative operating characteristic (ROC) analysis, sharpness, accuracy, expected best cost (EBC), and reliability.



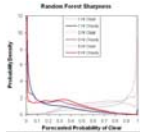
ROC curves for Random Forest, CEP, and SCC are shown above. RDC points for BP are plotted as triangles. Better performance potential is indicated by curves more toward the upper left of the diagram.



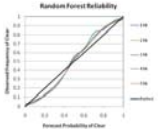
EBC expresses the max net cost per action taken or not taken. It is converted to profit in this figure by letting c = -100.



Accuracy (defined as percent correct match based on a decision threshold of 0.5) is shown for Random Forest, CEP, BP, and SCC.



Sharpness is the tendency for a forecast algorithm to correctly forecast extreme probability values. In this chart, visually connecting the same colored solid and dashed lines provides a visual indication of sharpness.



Reliability is equivalent to bias and answers the question of how well the predicted probabilities of an event correspond to the observed frequencies. Random Forest had a tendency to under-forecast above 0.5 and over-forecast below 0.5.