



## Record Filtering and Flagging

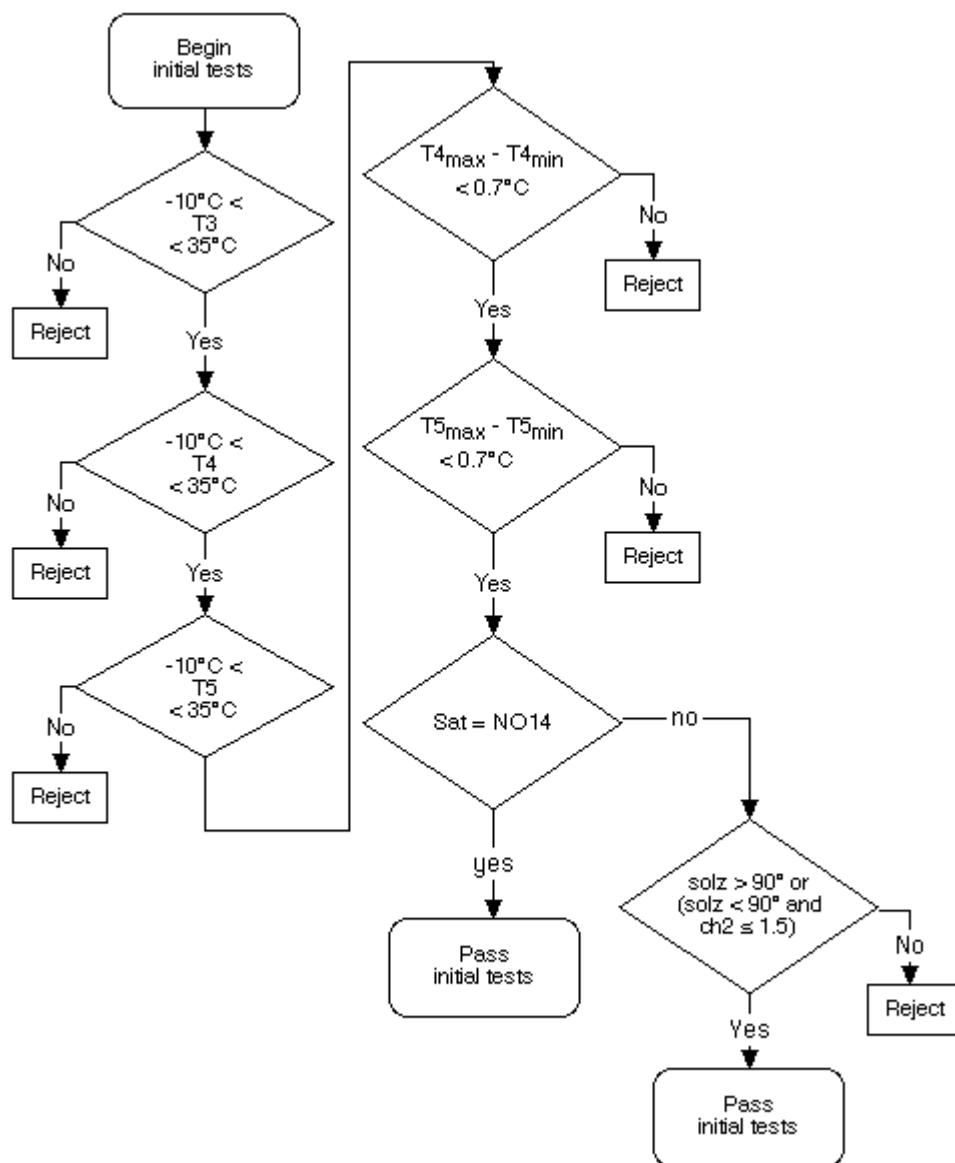
A large number of matchup records have obvious problems (for instance, gross cloud contamination) and cannot be used for algorithm testing. For that reason, a sequence of filters was implemented to exclude most cloud-contaminated matchups. The tests include two stages: (a) a set of initial tests common to matchups from all satellites, and (b) a decision tree involving tests derived separately for data from each AVHRR. The two types of tests are briefly discussed in this section. We note that in Version 19 of the PFMDB, we are including the majority of records that fulfilled the space/time matchup criteria, with the only exception of records with missing or geophysically unreasonable *in situ* SSTs. This differs from earlier PFMDB versions in which many records were excluded. The inclusion of most of the data has the purpose of allowing interested users to develop their own cloud-flagging procedures. Nevertheless, we flag those records identified as potentially cloud-contaminated by the tests described below. We used only matchups flagged as cloud-free for algorithm coefficient estimation.

### Initial Tests

The sequence of initial tests, common to matchups from all satellites, is shown in Figure 2. The variable names used in the figure are listed in section ["Description of the Fields in the Pathfinder Matchups"](#). The filters include absolute thresholds for brightness temperature values in channels 3–5 that intend to exclude very anomalous values that may result from digitizer errors. Spatial uniformity tests are intended for further cloud identification. These tests involve thresholds for differences between the minimum and maximum brightness temperature values for channels 4 and 5 inside a 3x3 pixel extraction box.

The spatial homogeneity threshold ( $0.7^{\circ}\text{C}$ ) for both channels was selected by applying various thresholds (from  $0.1^{\circ}\text{C}$  to  $1.5^{\circ}\text{C}$ ) and plotting the mean of SST residuals (*in situ* SST minus satellite-derived SST) as a function of the homogeneity threshold (plot not shown). It was observed that the mean of the residuals (which can be considered as a measure of bias) was fairly close to zero (a desirable characteristic!) for homogeneity thresholds lower than  $0.7^{\circ}$ ; for higher thresholds, the bias was increasingly negative. Daytime matchups ( $\text{solz} < 90^{\circ}$ ) for which central values of channel 2 aerosol optical thickness (ch2) were higher than 1.5, were excluded.

**Figure 2.** Initial tests applied to identify cloud-contaminated matchups. These tests are common to matchups from all AVHRRs. The boxes labelled "Reject" indicate that a matchup will be flagged as potentially cloud-contaminated. Note, however, that these records are still included in the distributed data.



## Decision Trees

Earlier versions of the PFMDB included cloud flagging tests that had been defined after extensive interactive examination of the data. Although these tests fulfilled their objective of excluding cloud-contaminated matchups from the algorithm estimation process, they were overly conservative, rejecting a large amount of potentially usable matchups. Furthermore, the selection of tests had to be repeated for matchups from each AVHRR, as calibration changes from sensor to sensor could potentially invalidate the use of the same tests for all matchups. For these reasons, a new methodology was developed in Version 19 of the PFMDB for the second stage of the cloud flagging step. This methodology is based on the tree models described by Clark and Pregibon (1992) and by Venables and Ripley (1994). There are two main advantages in the new method: the cloud flagging tests are selected objectively, and the number of potentially useful matchups rejected is lower.

Briefly, the classification trees are based on binary recursive partitioning, whereby a data set is successively split into increasingly homogeneous subsets. In the present context, tree models can find the best way to predict membership in one of two groups ("cloud-contaminated" or "cloud-free", also referred to as "bad" and "good" matchups) as a function of a set of predictor variables which may contain information about cloud contamination (e.g., differences in brightness temperatures between channels).

A tree is "grown" from a training sample for which the actual classification of all records is known. Matchups in the training set are defined as "bad" or "good", indicating whether they are potentially contaminated by clouds or not. Good matchups were defined as those which showed absolute values

of SST residuals (with respect to a first-guess satellite SST) lower than 2°C. Note that in studies of cloud identification using objective classification methods, the training sample is usually classified by an operator who visually examines satellite imagery and defines whether a location is cloud-covered. In this case, the AVHRR satellite data were extracted prior to the development of the cloud tests, so the definition based on SST residuals was chosen as a reasonable alternative. Most large SST residuals are caused by potential cloud contamination, but other factors (e.g., anomalous atmospheric or oceanic conditions) may affect SST retrieval as well.

The first-guess satellite SSTs used to compute residuals and, subsequently, to classify matchups in the training sets was different for the various AVHRR matchups. For the main NOAA-9 sequence, NOAA-11, and NOAA-9 gap data, the first guess SST was computed using the Pathfinder algorithm (see documentation on algorithm) and coefficients estimated using cloud flagging tests described in the documentation for Version 18 matchups. For the first year of NOAA-14 (1995), the first-guess SST was computed using the operational coefficients published by NOAA-NESDIS (separate coefficients for day and night). In subsequent NOAA-14 years, the first-guess SST was computed using the Pathfinder algorithm and the monthly coefficients from the previous year. The choice of a first-guess SST is not all that crucial, provided no significant bias exists in the first-guess residuals.

For each AVHRR, a training set was selected by sampling from all nighttime matchups that had passed the initial tests. The rationale for using only night data in growing the tree was to prevent stray sunlight from influencing expected associations among channels. To avoid potential biases in the trees, the training set was built using a weighted sampling procedure to ensure approximately comparable numbers of "good" and "bad" matchups. Also, the size of the training set was selected so that a significant number of independent data remained for subsequent independent validation of the estimated tree. The size of the training set was approximately 30% of nighttime matchups passing the initial tests, and ranged between about 1000 for NOAA-9 gap data, to about 20,000 for NOAA-11 data.

An important aspect in objective classification is to decide which variables or features are to be used as predictors when growing the classification tree. We used (a) homogeneity values for channels 3, 4 and 5 (i.e., maximum minus minimum brightness temperature in a 3x3 box), (b) pairs of differences between channels 3, 4 and 5, (c) satellite zenith angle intervals, and (d) the differences between the observed brightness temperature in one channel and the value that would be expected based on the brightness temperature for another channel. Initially we wanted to grow a tree which could be applied to both day and nighttime data and so channel 1 and 2 brightness temperature was not included. Also, note that the quantities used are of a spectral nature (e.g., differences between channels). The trees were grown prior to the incorporation to the PFMDB of pixel values for a 5x5 extraction box. These values, first available in Version 19, open up opportunities for future extensions to the classification trees via the incorporation of textural metrics. Beginning with NOAA-14 1996, the availability of new parameters in the Version 19 matchups enabled a decision tree to be grown using the following additional parameters; channel 2 brightness temperature, sunside, glint, day or night.

A classification tree was grown for each AVHRR using a training set which included the predictor quantities listed above. During 1987-1995 only nighttime data was used in the training set, beginning with NOAA-14 1996 both day and night data were included. The tree software finds a large number of splits in the training sets until the final nodes either reach a certain degree of purity, or a minimum size. These criteria, however, usually yield not only a very large and complex tree but, most importantly, a tree that may classify the training sample correctly, but will not perform as well for independent data sets (this can be viewed as "overfitting"). Because the trees were to be used to flag clouds in the larger matchup data sets, we attempted to reduce (or "trim") the size of the tree, while retaining adequate performance (defined as the proportion of data classified correctly).

There are automated tree-trimming procedures which find an optimal tree size while retaining performance. In the particular software we used, however, these procedures consider all types of error as equally important. This was not acceptable in our case, because one of the two possible types of tree error was considered more serious than the other type. The first type of error is classifying a matchup as "good" when it is truly "bad". This means we are allowing cloud-contaminated matchups to be used in algorithm estimation, an undesirable action. On the other hand, the second type of error (classifying a "good" matchup as "bad") was not considered as serious, as there were many additional matchups. Because the tree software could not handle the unequal misclassification cost functions, we chose to trim the tree interactively, attempting to obtain a tree as parsimonious as possible while minimizing the amount of "bad" matchups classified as "good". The trees for each AVHRR are shown in Figures 3-6.

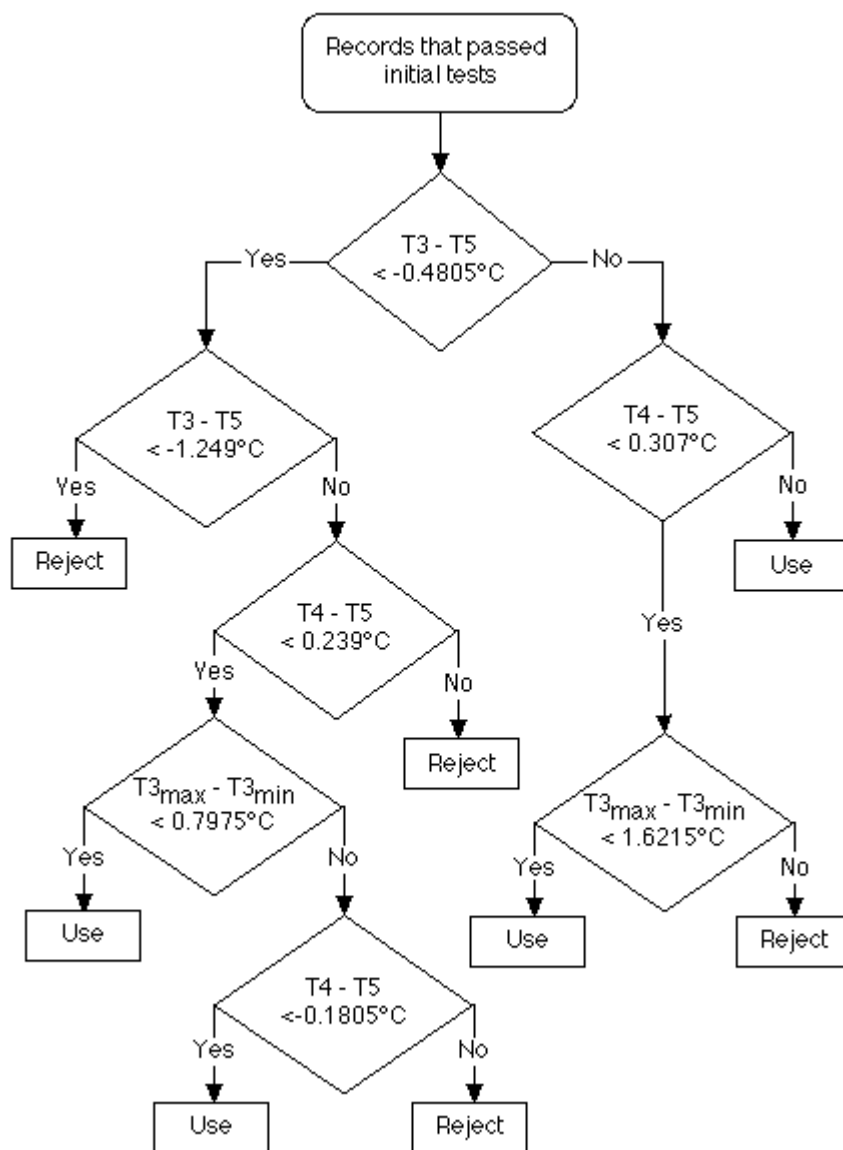
The values used at each split are generated automatically by the tree software to ensure that all records fall on either side of a split, and thus may have several decimals. We list the tests as derived, but we note that the number of decimal values shown is higher than what would be allowed by the sensor digitization.

### ***Classifying "growing" data sets***

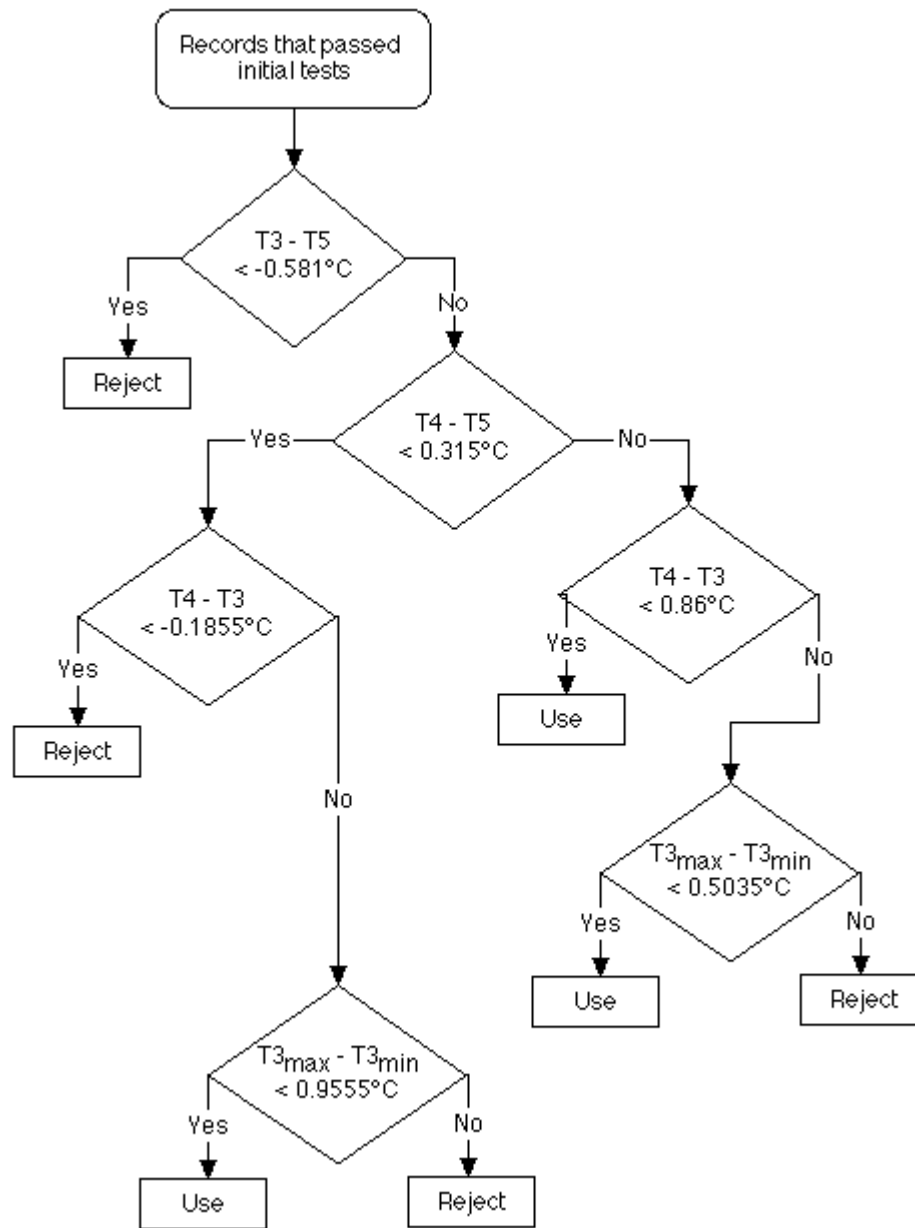
When a data set is "closed" (that is, we are processing only historical data), it is easy to grow a single tree with all the matchups available (of course, properly divided into training and validation sets). In contrast, what should be done when a data set is growing, as new matchup years are added? This is the case for NOAA-14. We have made two decisions: (a) we will not re-classify data for a given year, even if a new tree is developed for that year, and (b) once a tree is grown with sufficient matchups, it will not be changed until its performance is not satisfactory. These two policies are reflected in the following actions:

1. For the first portion of NOAA-14 (1995), we grew a tree with whatever matchups were available. See Figure 6a.
2. When more data became available for NOAA-14 (1996), we grew another tree with BOTH 1995 and 1996 data (see Figure 6b). However, this tree was only applied to 1996 data. That is, the matchups for 1995 were classified with the 1995 tree and were not re-classified afterwards.
3. As new matchup years became available (e.g., 1997), we tested the performance of the tree grown with 1995-1996 matchups. When the performance of the classification is similar to that of the previous year, we used the 1995-1996 tree to classify 1997 matchups. The same procedure has been adopted for 1998 data (i.e., the 1995-96 tree was applied to 1998 matchups).

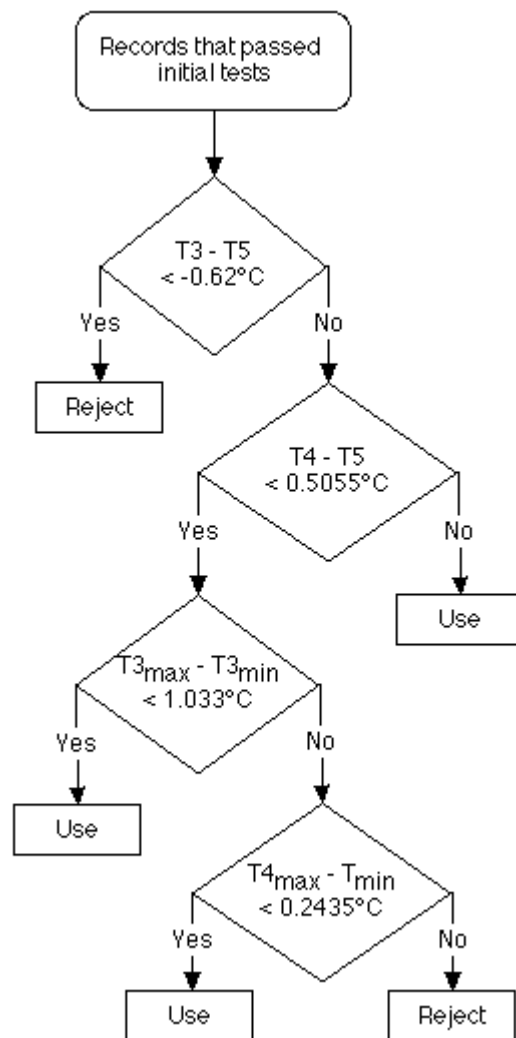
**Figure 3.** Classification tree for NOAA-9 data (1985–1988). Matchups that (a) pass the initial tests and (b) are classified as usable by the tree, are used for algorithm coefficient estimation and validation.



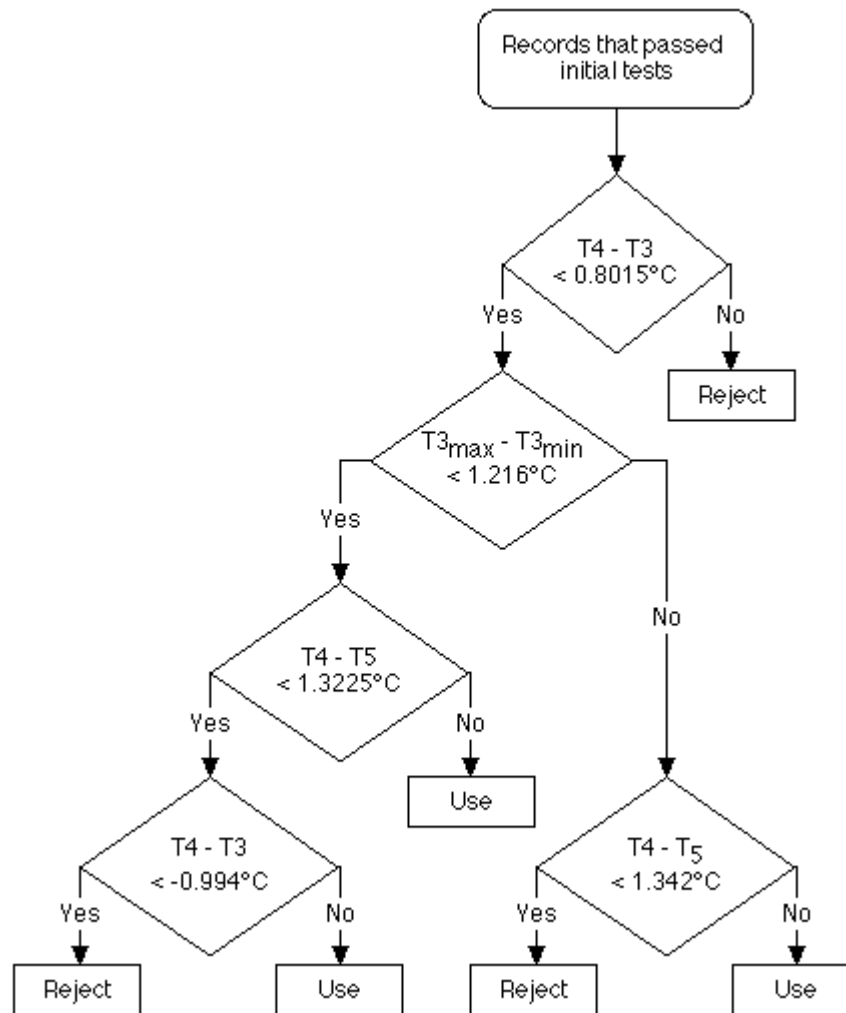
**Figure 4.** Same as Figure 3, but for NOAA-11 data (1988–1994).



**Figure 5.** Same as Figure 3, but for NOAA-9 gap data (1994–1995).

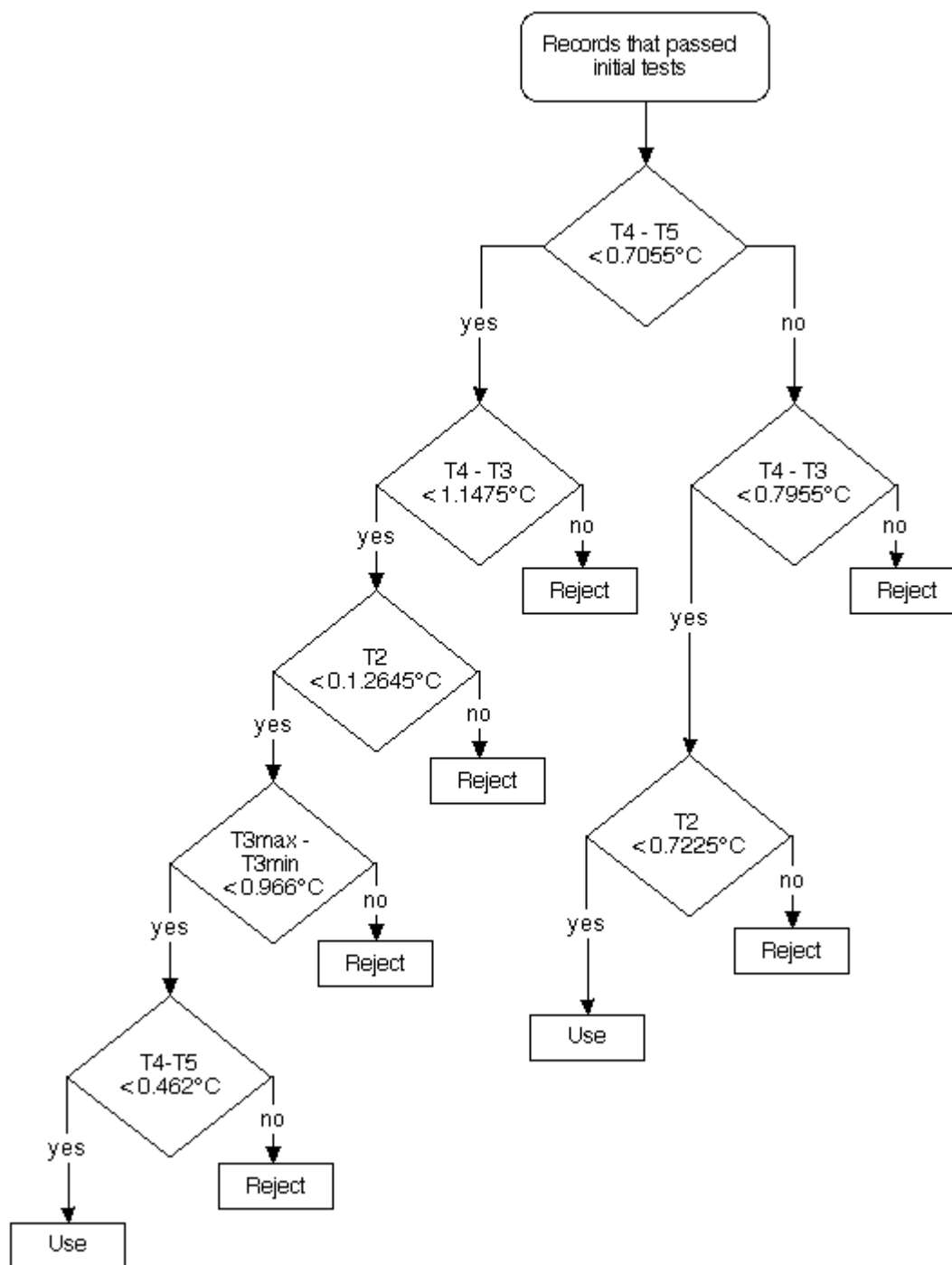


**Figure 6a.** Same as Figure 3, but for NOAA-14 data (February–December 1995).



**Figure 6b.** Same as Figure 3, but for NOAA-14 data (January 1996 to present). Note that this tree was grown with both 1995 and 1996 data, and is applied to NOAA-14 matchups from 1996 onwards.




[↑ Matchups Home](#)
[← Previous Section](#)
[→ Next Section](#)
[↑ | MAIN HOME PAGE |](#)

Page last Updated: Saturday, June 30, 2001 at 6:29 PM

Contact: Guillermo Podestá ([gpodesta@rsmas.miami.edu](mailto:gpodesta@rsmas.miami.edu)),  
 Telephone: +1.305.361.4142