Quality Control of Accumulated Fields By Applying Spatial and Temporal Constraints

Valliappa Lakshmanan[1,2], Madison Miller[1,3], Travis Smith[1,2*]

*Corresponding author: V Lakshmanan, 120 David L. Boren Blvd, Norman OK 73072; lakshman@ou.edu

[1]Cooperative Institutute of Mesoscale Meteorological Studies, University of Oklahoma; [2]National Oceanic and Atmospheric Administration / National Severe Storms Laboratory; [3] School of Meteorology, University of Oklahoma

ABSTRACT

Accumulating gridded fields over time greatly magnifies the impact of impulse noise in the individual grids. A quality control method that takes advantage of spatial and temporal coherence can reduce the impact of such noise in accumulation grids. Such a method can be implemented using the image processing techniques of hysteresis and multiple hypothesis tracking (MHT). These steps are described in this paper and the method is applied to simulated data to quantify the improvements and explain the effect of various parameters. Finally, the quality control technique is applied to some illustrative real-world datasets.

1. Motivation

Time accumulation of gridded fields is a common step in a variety of meterological applications. For example, precipitation totals are obtained from instantaneous rain-rates derived from measured quantities such as infrared temperature or radar reflectivity. "Tracks" of severe weather phenomena such as hail or low-level circulations are created by accumulating instantaneous calculations of the size of hail or the magnitude of shear over time.

Suppose a pixel at the location (x, y) experiences hail of magnitude $h(x, y, t)$ at time t. Then, the "hail track" ht at the pixel over the time period $(t - T, t)$ is given by:

$$ht_T(x, y, t) = \max_{t_k \epsilon (t-T,t)} h(x, y, t_k) \tag{1}$$

In other words, the hail track at any location over a time period is the maximum hailfall at the location over the time period (See Figure 1b).

Accumulating gridded fields over time greatly magnifies the impact of noise in the indi-

vidual grids. Even if a spurious value exists at only one time-step, the accumulation product will show its effect. Finally, simply as a demonstration, MHT-QC is applied to hail, shear and precipitation accumulation grids. Qualitatively, it is shown that the results are better – for a more detailed analysis of the effects of MHT-QC and other QC methods on a real data set, readers are directed to Miller et al. (2012).

If one were to create a long-term track from N grids, and a given pixel has a likelihood p of having an abnormally high value at any time step, then the likelihood of the pixel having an abnormally high accumulation value is $1-(1-p)^N$. This is because the likelihood of the pixel being "good" at any time step is $(1-p)$ and all the time steps need to be good for the pixel to be unaffected. The likelihood of all the timesteps being good, assuming that the timesteps are independent, is $(1-p)^N$. To put this in perspective, consider the error rate if one were to create a 24-hour hail track from 5-minute imagery. Even if p were to be as low as 0.001, the likelihood of any pixel in the 24-hour accumulation being affected is 0.76. On average, then, 76% of pixels in the final accumulation will have abnormally high values when only 0.1% of the pixels in any individual frame have such an abnormally high value. The quality control of accumulation fields, therefore, requires quality control that is more effective than techniques that apply QC to the component frames. It is necessary to apply quality control to the accumulation product in addition to the quality control of individual frames.

Precipitation totals are not affected as dramatically as hail tracks because the value at a single time step contributes only a small amount to the total. The related calculations become more involved to carry out in a close form since the noise probability is a function of the noise threshold which itself varies with the accumulation interval. Thus, unlike with

the maximum operator, one needs assumptions on the complete error distribution. In fact, if the precipitation error is distributed symmetrically about zero, the accumulation error can reflect a smaller bias than the instantaneous error. For simplicity in depiction of the results, therefore, we will employ the maximum as our accumulation operation of choice. However, the improvements in accumulation techniques we describe in this paper are useful even if the accumulation operator is an average or a summation. The degree to which they are useful depends on the actual error distribution.

The quality control method described in this paper was developed to carry out quality control of long-term accumulations of radar-derived azimuthal shear. Azimuthal shear, being a derivative of velocity, is a very noisy field and very susceptible to spurious values at individual time steps. Accumulations of azimuthal shear, therefore, are severely impacted by noise. For details on the application of the quality control method to azimuthal shear accumulations, called "rotation tracks", please see Miller et al. (2012).

Even though accumulation fields are affected to a greater extent than the individual components, quality control of accumulation products has typically focused only on improving the quality of the individual grids. For example, Grecu and Krajewski (2000) developed a neural network to remove anamalous propagation and ground clutter from rain rate products, but did not describe any methods to improve the quality of the accumulation itself. Similarly, the operational stages of precipitation products (rain gage, radar, blended, etc.) distributed by the National Weather Service do not involve quality control of the accumulation product beyond the quality control applied to the individual rain rate grids and some simple thresholding of low accumulation totals.

Yet, it stands to reason that in meteorological fields that are being accumulated, tem-

poral continuity can be profitably employed – echoes, hail estimates and high rain rates that appear at only one time step are unlikely to be legitimate and can be removed in the accumulation. This does rely on the spatial and temporal resolutoin at which these fields are being considered, so care should be taken that temporal continuity is a valid assumption to make. This paper describes the use of spatial and temporal association (spatial within each time step and temporal across time steps) to improve the quality of accumulation products.

There are two key insights that make the techniques adopted in this paper relevant to the problem of QCing accumulation fields. One is that in accumulation grids, temporal association does not need to be strictly causal. A causal system is one where results can be available in real-time, i.e., without having to use input data from the future. In an accumulation product that consists of N frames, it is possible for the temporal association to look ahead L frames and still remain causal as long as the look ahead is limited to the first $N - L$ frames. In other words, when the accumulation is being generated in real-time, the quality control is applied in a lagging sense – temporal continuity measures can be employed to remove spurious echoes in the first $N - L$ frames whereas the last L frames will suffer from noise that could not be removed from the individual frames. Thus, one can obtain improved quality for the majority of the frames while the product remains up-to-date.

The second insight is that temporal association in a non-causal setting can be performed in a more sophisticated manner than real-time storm tracking algorithms. In real-time storm tracking algorithms such as Johnson et al. (1998); Dixon and Wiener (1993); Lakshmanan and Smith (2010), the id assigned to storms is important because the id is used to examine long-term trends and is presented to human users of the system. If the ids are constrained to not change over time, one can not retrospectively change object associations. However,

when temporal association is employed purely for quality control purposes, the track itself is irrelevant. Object associations at previous frames can be freely changed without any untoward effects. Thus, rather than a single set of tracks, one can maintain multiple hypotheses, i.e., multiple sets of tracks and keep around echoes that are temporally correlated in any of these sets. Therefore, it is not neccessary for an object to be the "best" match to an object at the previous time step for it to be retained – it merely has to be one of the "K" best matches (in a global sense).

Multiple Hypothesis Tracking (MHT; Reid (1979)) is a well-known technique in video processing and missile tracking. In meteorology, it was employed by Root et al. (2012) to track storms in simulations of a fast scanning radar. However, because MHT breaks causality, and because storms typically last only a few time steps, it has not been employed in real-time storm tracking algorithms. As far as we know, this is the first time that MHT has been used in meteorology or related disciplines for quality control. Because MHT and the two algorithms that underpin MHT will probably be new to most readers of this paper, they will be explained in detail in Section 2. However, proofs of the mathematical theorems and derivations of the formulae will not be presented. Readers interested in mathematical proofs are directed to Nering and Tucker (1993); Reid (1979); Cox and Hingorani (1996); Bourgeois and Lassalle (1971); Murty (1968) and to the citations therein.

The rest of this paper is organized as follows. The quality control technique is described in Section 2. Input time steps are simulated and the improvement in quality of the accumulation product quantified in Section 3a. The technique is demonstrated on some real data in Section 3b.

2. Spatial and Temporal Association

The underlying assumption behind the quality control presented in this paper is that noisy observations tend to be "spiky" (a narrow peak) in space and/or time. Noise, we assume, is spatially smaller, and/or more sporadic than real signal. Of course, not all noise behaves this way. So, the technique described in this paper will remove only noise that matches these characteristics. It will also erroneously remove any true observations that are small or sporadic. Nevertheless, noise in many remotely observed fields does meet these characteristics and can be effectively removed.

There are two criteria that observations have to meet in order to be retained in the accumulation field: the observations need to be spatially and temporally coherent.

a. Spatial coherence

The first constraint that can be placed on pixels that will contribute to the accumulated field is that these pixels are part of valid objects, i.e., that they belong to spatially coherent entities. Grid points ("pixels") in the individual spatial grids are first grouped into "objects" and these objects are associated across time. Objects that are either too small or not part of a long enough track are assumed to be noise and pruned. Accumulation is carried out only on those pixels that belong to valid objects.

Pixels are grouped into objects based on the data values of the pixels. Such grouping can not be based simply on thresholding the grid because selecting this threshold can be problematic. If the selected threshold is too high, the object will be missed in the early stages of its growth and in the late stages of its decay. On the other hand, if one selects a

low threshold so as to capture the entire lifetime of an object, it is likely that one starts to include noise (which is assumed to be weaker in intensity than valid objects). Lakshmanan et al. (2009) discuss these issues in greater detail and suggest the use of the Watershed algorithm (Beucher 1982) suitably modified to isolate storm cells. However, the modifications result in only the cores of storms being isolated and therefore, the enhanced watershed algorithm approach of Lakshmanan et al. (2009) can not be be used when accumulating fields like precipitation where one requires the entirety of the object. In this paper, we suggest the use of hysteresis to mitigate threshold selection problems while not being subject to size constraints.

Hysteresis as applied to object identification is simply the use of two thresholds. An object is defined to consist of a group of contiguous pixels with data values greater than T_1 that are connected to at least one pixel with a data value greater than T_2 where $T_2 > T_1$. Hysteresis works well at identifying objects as long as the objects are continuous in space and can be differentiated from their surroundings based on their data values. Object identification using hysteresis is commonly implemented using a recursive image processing algorithm known as region growing (Lakshmanan 2012). If the noise characteristics are such that individual pixels above T_2 are relatively likely, the input field can first be smoothed by a speckle-removing filter such as a median filter before object identification is carried out. Once objects have been identified, they can be classified as either noise or valid data based on their size.

b. Temporal coherence

The second constraint that can be placed on pixels that will contribute to the accumulated field is that the objects that these pixels are part of are temporally coherent, i.e., that they are long-lived. The length of time that the object is required to "live" is, of course, dependent on the data being accumulated and the phenomena being observed. If one would accept any object that occurs in any two consecutive frames, this requires that an object identified at a time frame t_0 has to be associated with an object identified either at time frame t_1 or at time frame t_{-1}. Because this is not causal, the accumulation grid will have to lag real-time by one time frame. If it is important that the accumulation grid be produced in real-time, based on the most current data, the very last frame of the accumulation will have to retain unassociated objects but all the previous frames can be quality controlled based on temporal coherence.

Given an object identification technique and a cost function for associating objects, it is possible to determine the optimal association between objects across time. It should be clear, then, that the performance of the QC method described in this paper depends quite heavily on how well objects are identified, represented and associated across time. The interested reader is directed to Lakshmanan et al. (2009) for a discussion on the impact of thresholds and smoothing on object identification, to Lakshmanan and Smith (2009) for discussions of object representation and attribute extraction, and to Lakshmanan and Smith (2010) for comparisions of different ways (including the use of cost functions) of associating objects.

1) Hungarian Method

The Hungarian method is an algorithm to find the best way to match each item in a set of items to an item in another set of items, where each match carries with a cost, a problem known in linear optimization as the assignment problem. In meteorology, it has been successfully used for storm tracking (Dixon and Wiener 1993). Given a set of objects at time t_{-1} and a set of objects at time t_0, the cost of assigning any object in t_0 to an object in t_{-1} is computed. The cost function can be as simple as the Euclidean distance between the object centroids or a complex, domain-specific function that incorporates conservation of physical properties. The result is a cost matrix where the rows correspond to objects at t_0 and the columns to objects at t_{-1}. Munkres (1957) described a pen-and-paper algorithm to reduce the matrix to a set of "starred" entries which form the optimal assignment. The algorithm was modified by Bourgeois and Lassalle (1971) for rectangular matrices so that the number of objects at the two time steps can be different and it is the description of Bourgeois and Lassalle (1971) that is summarized below. For a mathematical proof that this method works, readers are directed to Nering and Tucker (1993) or any other text on linear programming. In cases where there are multiple possible solutions to the assignment problem, the Bourgeois and Lassalle (1971) algorithm provides one of those solutions. For further implementation details on the algorithm, including how to adapt the pen-and-paper description to something amenable to a computer program, readers are directed to Lakshmanan (2012).

Given a cost matrix that represents the distance between every object at time t_0 to every object at time t_{-1}, the Hungarian method as adapted by Bourgeois and Lassalle (1971) for rectangular matrices consists of these steps:

i. For each row of the cost matrix, find the smallest element and subtract it from every element in its row.

ii. For every zero in the matrix that results from the previous step, if there is no starred zero in its row or column, star this zero. Repeat for each element in the matrix.

iii. Cover each column containing a starred zero. If all the columns are covered, the starred zeros now describe the final optimal assignment.

iv. For every noncovered zero, prime it and check if there is a starred zero in the row containing this newly primed zero. If not, move on to Step 5. If there is a starred zero Z in this row, however, cover this row and uncover the column containing Z. Finally, move on to Step 6.

v. Construct a sequence of alternating primed and starred zeros as follows. Let Z_0 represent the uncovered primed zero found in the previous step. Let Z_1 denote the starred zero in Z_0's column (if any). Let Z_2 denote the primed zero in Z_1's row. Continue until the sequence terminates at a primed zero that has no starred zero in its column. Unstar each starred zero of the sequence, star each primed zero of the sequence, erase all primes and uncover every line in the matrix. Return to Step 3.

vi. Find the smallest uncovered value in the matrix. Add this value to every element of each covered row, and subtract it from every element of each uncovered column. Return to Step 4 without altering any stars, primes, or covered lines.

In practice, it is necessary to specify an upper bound for the cost function so that objects that are too far apart are not assigned to each other. The time complexity of the Hungarian

method is $O(N^3)$ where N is the number of rows (or columns, which are assumed to be similar in order of magnitude).

2) K-best Solution

The Hungarian method provides the optimal assignment for every object in the current frame to every object in the previous frame. While this is desirable for tracking, this is not quite what is needed for quality control. We would like to retain any object that is reasonably likely to be associated with an object in the previous frame, especially because the optimal assignment based on two frames may not be the best assignment once we receive a third frame (See Figure 2).

In order to accomodate this desire to maintain multiple hypotheses of tracks, it is necessary to obtain not just the best assignment of objects between two frames but a few more reasonable assignments.

Murty (1968) proved the correctness of an algorithm to find the $k+1$th best assignment given the kth best assignment and whose complexity is linear in k. Starting with the Hungarian method (which yields the best set of assignments), one can find the second-best solution and starting with the second-best, the third-best and so on.

Although the algorithm was described as a way to find the K best solutions, it is an iterative algorithm and can, therefore, be used in a manner such that the number of hypotheses K is dynamic and not fixed. For example, one can stop looking for further assignments as soon as they are worse than, say, 110% of the optimal cost.

The K-best algorithm due to Murty (1968) is as follows. Suppose the kth best set of

assignments, $\widehat{S_k}$, is determined to consist of $< x, y >$ where the xth row of the cost matrix is assigned to the yth column. To obtain the $(k+1)^{th}$ best set of assignments from the cost matrix that results after the kth best set is found:

i. For each assignment $< x, y >$ in $\widehat{S_k}$:

   (a) Delete $< x, y >$ in the solution $\widehat{S_k}$. This can be done by setting the cost of $< x, y >$ to be above the upper bound for the cost function so that it will never be part of an optimal assignment.

   (b) Apply the Hungarian method to the resulting cost matrix and find a candidate optimal solution S_{k+1}.

ii. Choose the S_{k+1} that has the lowest cost. This is $\widehat{S_{k+1}}$, the $(k+1)^{th}$ best assignment.

iii. Delete every $\hat{x}$ and $\hat{y}$ entry in the cost matrix, retaining only $< \hat{x}, \hat{y} >$. where $< \hat{x}, \hat{y} >$ was the $< x, y >$ whose deletion resulted in $\widehat{S_{k+1}}$.

Murty's K-best algorithm is based on partitioning the solution space, and therefore, the time complexity of the algorithm improves with increasing k, i.e., the third-best solution is found faster than the second-best and the fourth faster than the third. However, as a broad generalization, one can state that the complexity of the K-best algorithm is kN, and this combined with the complexity of the Hungarian method makes the entire process a $O(N^4)$ operation. Recall, however, that N here is not the number of pixels in the grid but is only the number of objects in the grid. Therefore, the process of maintaining multiple possibilities is quite feasible computationally.

3) Multiple Hypothesis Tracking

Given a set of spatial grids over time, the K-best algorithm will lead to a combinatorial explosion – if there are K solutions that we wish to consider between every pair of time steps, then if there are N time steps, there are K^{N-1} solutions that one needs to consider (see Figure 3).

Reid (1979) addressed the issue of combinatorial explosion by pruning the set of hypotheses at each stage so that the K^2 hypotheses from three successive frames get immediately pruned to the K best. Thus, for example, the four hypotheses in the bottom half of Figure 3 would be pruned to just two and it is these two that would be used to find the best assignments when the fourth time frame is being processed. A framework that does this sort of pruning is referred to as a Multiple Hypothesis Tracking (MHT) framework. However, the brute-force method of finding all K^2 solutions and then pruning them to K is computationally inefficient especially because the Hungarian method is $O(N^3)$.

Cox and Hingorani (1996) pointed out that Murty's K-best algorithm can be generalized to deal with triples $< x, y, l >$ where l is the time frame of the assignment. Murty's algorithm can therefore be used to maintain the set of hypotheses at a manageable number. Using Murty's method in combination with a termination condition that involves proceeding no more than a certain fraction above the optimal cost function, it is often unnecessary to actually find K new solutions at each time step. Cox and Hingorani (1996) also suggested the use of a "N-scan-back" algorithm such that ambiguities at time k are resolved by time $k + N$, so that one does not need to maintain multiple hypotheses for frames that are more than N scans old.

When used together, the Hungarian method, the triple-form of Murty's K-best algorithm and N-scan back algorithm (hereafter simply termed "MHT") can be used to impose a non-causal, fault-tolerant temporal coherence check on accumulation grids. This will be demonstrated in Section 3.

4) Splits and Mergers

Before we proceed to demonstrating the results of the spatial and temporal coherence checks, we'd like to address a question that might have arisen in the minds of many readers. How are splits and mergers handled in this method?

None of the techniques described above – the Hungarian method, Murty's algorithm or N-scan-back – deals explicitly with splits and mergers. However, this does not matter because our use of MHT is to apply temporal coherence checks, and both sides of a split or merger should remain amongst the K-best solutions and will not get pruned (see Figure 4). As time goes on, both sides of the split will normally attain temporal coherence and be retained in the accumulation grid regardless of which side of the split is chosen at the point of the "N-scan-back" decision as the final one.

3. Results and Discussion

In order to demonstrate the impact of the spatial and temporal coherence checks on an accumulation grid, we simulated a set of input grids and created an accumulation product as the maximum of the input grids over the entire length of the simulation. To the input

grids, simulated noise was added and the MHT-QC method applied to try and remove the noise. The QCed field is compared against the field that would have been obtained had the simulated noise not been added. By comparing against the "true" accumulation, it is possible to quantify the noise pixels and real data removed. Finally, simply as a demonstration, MHT-QC is applied to hail, shear and precipitation accumulation grids. Qualitatively, it is shown that the results are better – for a more detailed analysis of the effects of MHT-QC and other QC methods on a real data set, readers are directed to Miller et al. (2012).

a. Simulation

The simulation was carried out as follows. One hundred time frames of a spatial grid were simulated and accumulated using the maximum operator in three ways: (1) raw, without any quality control (2) applying only the spatial coherence criterion based on a minimum size for the objects and (3) applying both the spatial coherence criterion and a temporal coherence criterion using MHT. In the individual spatial grids, true objects and noisy objects were simulated and added. The resulting accumulations were compared against a hypothetical method that would retain only the true objects in the accumulation.

The individual spatial grids were simulated to have a size of 500x500 and in these grids, true objects and noise were placed randomly. The true objects were simulated to last 10 +/- 5 frames and move at a speed of 10 +/- 5 pixels per time frame, i.e., the lifetime of the objects was chosen from a Normal distribution with mean of 10 and standard deviation of 5. The peak intensity of the objects was simulated to be 60 +/- 15 (with the data range clamped between 0 and 100) with the peak intensity being reached at the half-life of the

objects. In other words, the objects intensified over the first half of their lifetime and decayed over the second half of their lifetime. Spatially, the intensity of the objects was simulated to fall off exponentially with distance from the center of the object. Because the intensity of the object changes over its life time, its size also changes. At every time step, 5 true objects were simulated. The probability of noise at any pixel in the grid at any time step was chosen to be 0.00001 and noise objects were simulated to have half the intensity of real objects, i.e. their intensity was chosen from a Normal distribution of 30 +/- 7.5. The noise objects were not persistent over time and did not, therefore, exhibit any temporal variation in intensity.

In addition to randomly placing noise objects, we also simulated noise close to real objects so as to cause confusion during tracking. The movement of true objects was also simulated to exhibit a strong turn once during their life time and at the time of the turn, spurious objects were placed all around the true object.

Two successive time frames of the simulated input grids are shown in the top row of Figure 5. Some of the objects in the grids are real and the others are noise. Note that the spurious objects surrounding the real object make it difficult to track the real object (based on center position alone). An accumulation of the simulated input grids without any quality control is shown in Figure 5c. An accumulation using only the true objects is shown in Figure 5d. This is the result of an ideal QC method. Quality control using spatial and temporal criteria is shown in Figure 5e. A post-processed clean copy of the same frame as Figure 5b where small or unassociated objects have been removed is shown in Figure 5f. The spatial coherence criterion here does not remove much noise because the threshold on size is set very low, so that most of the noise objects also qualify. Using the same size criterion, but adding a temporal continuity criterion of 3 time frames in one of the five best hypotheses

results in the accumulated field shown in Figure 5f. Because the spatial criterion did not result in any improvement, all the improvement here is due to the MHT.

The following parameters were varied in order to examine their effect on the MHT-QC method:

i. The minimum size threshold before an object was accepted as a valid object was varied within the range 1 to 100.

ii. The minimum temporal length threshold was varied from 1 to 8. Because the average lifetime of the objects is 10 frames, the larger threshold should result in more true objects being removed.

iii. The number of hypotheses was varied from 1 to 10. With only one hypotheses, MHT reduces to the Hungarian method.

The errors and skill scores are computed by comparing the accumulation field to the result of the ideal accumulation where none of the true objects are removed and all the noise is eliminated. This comparision is done pixel-wise.

At a size threshold of 1, none of the objects are removed whereas at 100, nearly all the noise objects are removed, but so are most of the true objects (See Figure 6a,b). Therefore, one can see the variation in the impact of the spatial coherence criterion as the size threshold is increased. When the size threshold is very low, there is no decrease in error (i.e. the spatial coherence criterion has no impact), but as it is increased, the error initially decreases (as noise is removed with little impact on true objects) and then starts to increase as good objects also start to get impacted. The probability of detection (of true objects) decreases as the size threshold is increased, but the likelihood that a noise object is falsely retained

also decreases. Because of the countervailing balance of these two effects, the mean square error (MSE) decreases as the size threshold is increased, but only to a certain point after which it starts to increase. The graphs in Figure 6a,b are generated with neither simulation of spurious objects nor pruning based on temporal coherence.

Because none of the objects are removed due to the spatial constraint when the size threshold is 1 pixel, this number can be used to validate the impact of employing just temporal coherence (See Figure 6c). Shown are the mean square errors (MSEs) for different track length thresholds and different numbers of hypotheses. The graph for a track length threshold of 2 (marked as "len=2" in the graph) indicates that an object is presumed valid if it is associated with an object in an earlier frame or an object in a later frame. In general, the MSE is reduced by increasing the number of hypotheses, but only up to a point. At that point, due to the incorporation of noise in the simulation, spurious objects start being part of some hypothesis, causing the MSE to start increasing again. It is clear that, for the simulation, the ideal value for the number of hypotheses is 4. This corresponds to the number of spurious objects along the direction of movement. Thus, in a realistic situation, one needs to choose as the number of hypotheses the number of "ties" one would reasonably expect to resolve at any individual time step.

The result of applying the MHT QC technique to the simulated inputs is shown in Figure 5f. The image corresponds to the best set of parameters: a size threshold of 10, a track length threshold of 3 and using 4 hypotheses.

b. Real Data

The MHT-based QC algorithm described in this paper was developed in order to apply quality control in the creation of NSSL's "rotation tracks" product. The full quality control of the rotation tracks product, of which MHT-based QC was just one step, is described by Miller et al. (2012). One problem with applying this QC technique to rotation tracks was the inability to quantify its effect using measures such as MSE. This is because mesocylones are a radar-inferred phenomenon, i.e., there is no ground truth of mesocylones. Consequently, there is no easy way to identify whether any strong shear was incorrectly removed by the algorithm or whether any spurious shear was incorrectly retained. In the absence of an objective way to create a contigency matrix, quantification of the skill of the MHT-based QC is not possible. Therefore, in this paper, in an effort to quantity the effect of the MHT QC technique, data were simulated, noise added and the QC technique applied and compared against the simulated data before noise was added. This allowed us to explore the effect of the various parameters on the MHTQC.

In order to illustrate the MHT QC algorithm's effect on real data, and its applicability beyond rotation tracks, it was applied to several real-world datasets. The examples shown in this section are merely illustrative, as we did not tune the various parameters. Further research is required to choose the best hysteresis thresholds and MHT settings for specific datasets and applications (see Figures 7, 8 and 9).

An example of a 3-hour hail accumulation from 1 km grids of Maximum Expected Size of Hail (MESH) grids created every 5 minutes is shown in Figure 7. The spatial constraint was to apply hysteresis, defining valid clusters as contiguous pixels above 5 mm connected to a

pixel with hail size of 8 mm and with a minimum size of 15 km^2. The temporal constraint was implemented using MHT with 5 hypotheses, looking ahead one frame, coasting two frames if necessary, and with clusters associated only if the centroids are within 42 km^2 of each other.

Note that there are several updrafts in the southeast quadrant of the domain in Figure 7a, which depicts data from Oklahoma on 16 July 2009. However, only one of those updrafts actually persists (Figure 7b) and is used in the accumulation. However, the movement of the storm causes even the location of the short-lived updraft to later get filled. Several short-lived features in the southwest of the domain (one of which is shown circled in Figure 7c) are also removed in the three-hour accumulation. By comparing against hail reports collected by telephone surveys (Ortega et al. 2009), it was verified that the removed echoes were really spurious and that suppressing these in the final accumulation was correct.

A similar process is at work in Figure 8 which shows accumulation of low-level shear over time. Circulations which are not persistent over time are removed and not used in the accumulation. The resulting QCed rotation tracks are less noisy than simply accumulating the raw azimuthal shear. Figure 8 demonstrates, through the use of a post-event damage survey, that signatures corresponding to real tornadoes were not removed. It should be noted that not all non-tornadic circulations have been removed; the aim of the QC technique was simply to remove temporally incoherent circulations and in that, the QC method was successful. The application of this technique to the quality control of rotation tracks is described in detail by Miller et al. (2012).

Finally, the MHT-based QC technique of this paper is illustrated on precipitation fields in Figure 9. The one-hour accumulation from radar-derived precipitation grids at a resolution of

1 km $\times$ 1km $\times$ 5 minutes was quality controlled by applying spatial and temporal constraints. The spatial constraint was to define valid clusters as contiguous pixels with rain rates above 0.8 mm/hr, connected to a pixel with rain rate above 1 mm/hr and with a minimum size of 100 km^2. The temporal constraint was applied using MHT, 10 hypotheses, looking ahead one frame, coasting a maximum of one frame, and associating clusters only if their centroids were within 22 km of each other. The impact of temporal coherence checks on precipitation is quite minimal because the value of the accumulation grid is an average over time, thus lowering the impact of temporally incoherent noise. There are, however, differences in the precipitation accumulation fields having to do with spatial coherence checks (such as the radial spike in the southeast quadrant and the low precipitation totals in the east-central region). Comparing the rainfall at the nearest mesonet station indicates that these precipitation objects have been correctly removed. Thus, the MHT-based technique of this paper can contribute to the quality of precipitation accumulation, even if the improvement is not to the extent as that obtained on severe weather accumulation fields.

It might appear that this quality-control technique is quite general and broadly applicable, but this is not the case. In order to carry out QC by insisting on temporal coherence, it is necessary to be able to look ahead. Otherwise, new initiation will always get removed. Looking ahead implies that the QC technique can only operate in a lagging sense, i.e., it can never operate on the latest time step of a sequence. Therefore, the quality control technique described in this paper is useful only in scenarios where QC of the latest time step is not important. This restriction eliminates many practical applications in meteorology. One situation where it is acceptable to not apply the QC to the latest time step is when creating accumulation fields because one can QC the majority of frames that go into the

accumulation product, while leaving the latest frame(s) unQCed.

4. Summary

Because accumulating gridded fields over time greatly magnifies the impact of impulse noise in the individual grids, it is important to be able to apply some quality control to accumulations beyond what is done on the individual grids. This can be achieved by applying applying spatial and temporal coherence constraints. In this paper, spatial coherence constraints were imposed using hysteresis and temporal coherence constraints using MHT. The resulting QC method was applied to simulated data to quantify the improvements and explain the effect of various parameters. Finally, the quality control technique was applied to some real-world datasets and the resulting improvements were illustrated.

Acknowledgements

Funding for the authors was provided under NOAA-OU Cooperative Agreement NA17RJ1227. We wish to thank Kiel Ortega for providing the SHAVE data used in Figure 7, Kiel Ortega, Brandon Smith and Gabe Garfield for the damage survey shown in Figure 8 and the Oklahoma Mesonet for providing the surface observations used in Figure 9.

REFERENCES

Beucher, S., 1982: Watersheds of functions and picture segmentation. *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Proc.*, Paris, 1928–1931.

Bourgeois, F. and J. Lassalle, 1971: An extension of theMunkres algorithm for the assignment problem to rectangular matrices. *Communcations of the ACM*, **14 (12)**, 802–804.

Cox, I. and S. Hingorani, 1996: An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. *IEEE Trans. on Pattern Anal. and Mach. Intell.*, **18 (2)**, 138–150.

Dixon, M. and G. Wiener, 1993: TITAN: Thunderstorm identification, tracking, analysis and nowcasting – a radar-based methodology. *J. Atmos. Ocean. Tech.*, **10**, 785–797.

Grecu, M. and W. Krajewski, 2000: An efficient methodology for detection of anamalous propagation echoes in radar reflectivity data using neural networks. *J. Atmospheric and Oceanic Tech.*, **17**, 121–129.

Johnson, J., P. MacKeen, A. Witt, E. Mitchell, G. Stumpf, M. Eilts, and K. Thomas, 1998: The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Wea. Forecasting*, **13**, 263–276.

Lakshmanan, V., 2012: *Automating the Analysis of Spatial Grids: A Practical Guide to Data Mining Geospatial Images for Human and Environmental Applications.* Springer, 320 pp., iSBN: 978-94-007-4074-7.

Lakshmanan, V., K. Hondl, and R. Rabin, 2009: An efficient, general-purpose technique for identifying storm cells in geospatial images. *J. Ocean. Atmos. Tech.*, **26 (3)**, 523–37.

Lakshmanan, V. and T. Smith, 2009: Data mining storm attributes from spatial grids. *J. Ocea. and Atmos. Tech.*, **26 (11)**, 2353–2365.

Lakshmanan, V. and T. Smith, 2010: An objective method of evaluating and devising storm tracking algorithms. *Wea. Forecasting*, **25 (2)**, 721–729.

Miller, M., V. Lakshmanan, and T. Smith, 2012: An automated method for depicting mesocyclone paths and intensities. *Wea. Forecasting*, **1**, 1.

Munkres, J., 1957: Algorithms for the assignment and transportation problems. *J. Society for Industrial and Applied Math.*, **5 (1)**, 32–38.

Murty, K., 1968: An algorithm for ranking all the assignments in order of increasing cost. *Operations Res.*, **16 (3)**, 682–687.

Nering, E. and A. Tucker, 1993: *Linear Programs and Related Problems.* Academic Press, Inc., Boston, 275-318 pp.

Ortega, K., T. Smith, K. Manross, K. Scharfenberg, A. Witt, A. Kolodziej, and J. Gourley, 2009: The severe hazards analysis and verification experiment. *Bulletin of the American Meteorological Society*, **90**, 1519–1530.

Reid, D., 1979: An algorithm for tracking multiple targets. *IEEE Trans. on Automatic Control*, **24 (6)**, 843–854.

Root, B., M. Yeary, and T. Y, 2012: Performance analysis of multiple hypothesis tracking on simulated storm cells. *J. Oce. Atmos. Tech.*, **1 (1)**, InPress.

List of Figures

1 Examples of accumulated grids over South Texas on May 10, 2012. (a) Instantaneous estimate of the Maximum Expected Size of Hail (MESH) (b) MESH accumulated over 2 hours. (c) Instantaneous radar-derived rate of precipitation (d) Rainfall accumulation over 2 hours. 29

2 Why Multiple Hypothesis Tracking is needed: (a) The optimal assignment between two objects in the first frame and three objects in the second frame. (b) The second best assignment. (c) Once the third time frame is available, it is clear that the second hypothesis was actually the better choice. 30

3 Combinatorial explosion results as a result of multiple hypothesis tracking: (Top) The two best (K=2) hypotheses for assigning objects between two frames (Bottom) When the third time frame arrives, there are two best hypotheses leading from the original two best. Thus, there are four (K^2) hypotheses to be considered. 31

4 MHT gracefully handles the problem of splits and mergers since both sides of a split (or merger) will remain among the K-best assignments. Top: Optimal assignment Bottom: Second-best hypothesis. As long as $K >= 2$, both hypotheses will be retained, and both objects used in the accumulation. 32

5 (a,b): Two successive time frames of the simulated input grids. (c) Accumulation of all 100 grids of the sequence. (d) Ideal accumulation that does not incorporate any noise. (e) Accumulation when the individual time frames are QCed by insisting that objects last at least 3 frames in one of the 4 best hypotheses. (f) Individual time frame QCed to remove small or unassociated objects. 33

6 Impact of the hysteresis and MHT criteria on the quality of the accumulated grids. (a) Impact of spatial coherence. As the size threshold is increased, the mean square error initially drops but then starts to increase. (b) Both the Probability of Detection (POD) of true objects and the False Alarm Rate (FAR) at which noise objects are wrongly identified fall when the size threshold is increased. (c) Impact of temporal coherence. The mean square error initially drops as the number of hypotheses are increased, but then starts to increase. 34

7 Illustration of the MHT-QC algorithm of this paper on a hail accumulation over Oklahoma starting at 21:00 UTC on 16 July 2009. (a) Individual frame before QC (b) Same frame after QC based on spatial and temporal coherence (c) Raw accumulation without any QC (d) QCed accumulation (e) Actual hail reports on that day 35

8 Illustration of the MHT-QC algorithm of this paper on a two-hour accumulation of low-level rotation over Oklahoma on 24 May 2011. (a) Individual frame before QC (b) Same frame after QC based on spatial and temporal coherence (c) Raw accumulation without any QC (d) QCed accumulation (e) Actual damage observed on the ground. 36

9 Illustration of the MHT-QC algorithm of this paper on a one-hour accumulation of radar-derived precipitation over Oklahoma on 13 April 2012. (a) Individual frame before QC (b) Same frame after QC based on spatial and temporal coherence (c) Raw accumulation without any QC (d) QCed accumulation (e) Actual rainfall measured on the ground (in mm) over the time period. 37

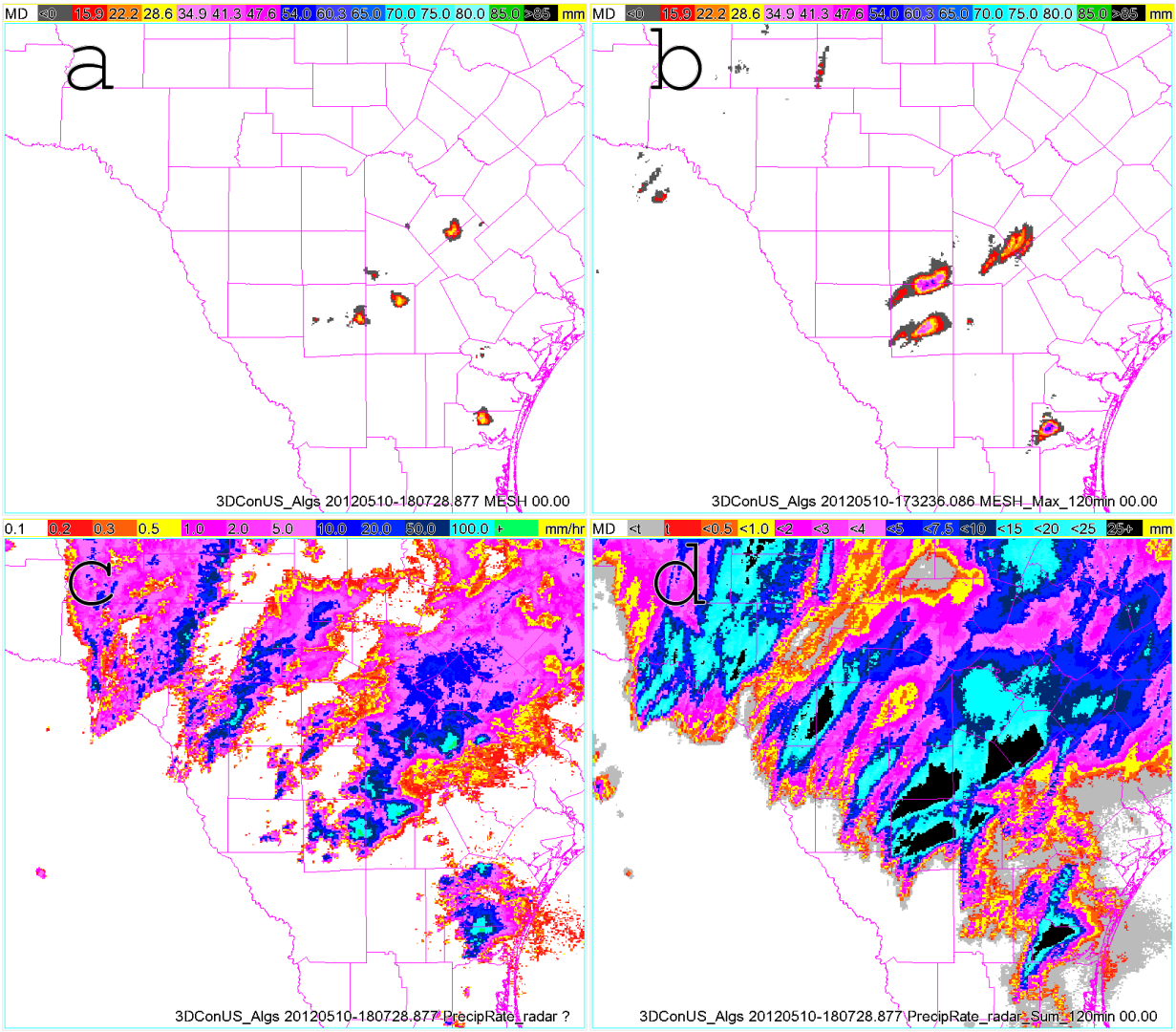

FIG. 1. Examples of accumulated grids over South Texas on May 10, 2012. (a) Instantaneous estimate of the Maximum Expected Size of Hail (MESH) (b) MESH accumulated over 2 hours. (c) Instantaneous radar-derived rate of precipitation (d) Rainfall accumulation over 2 hours.

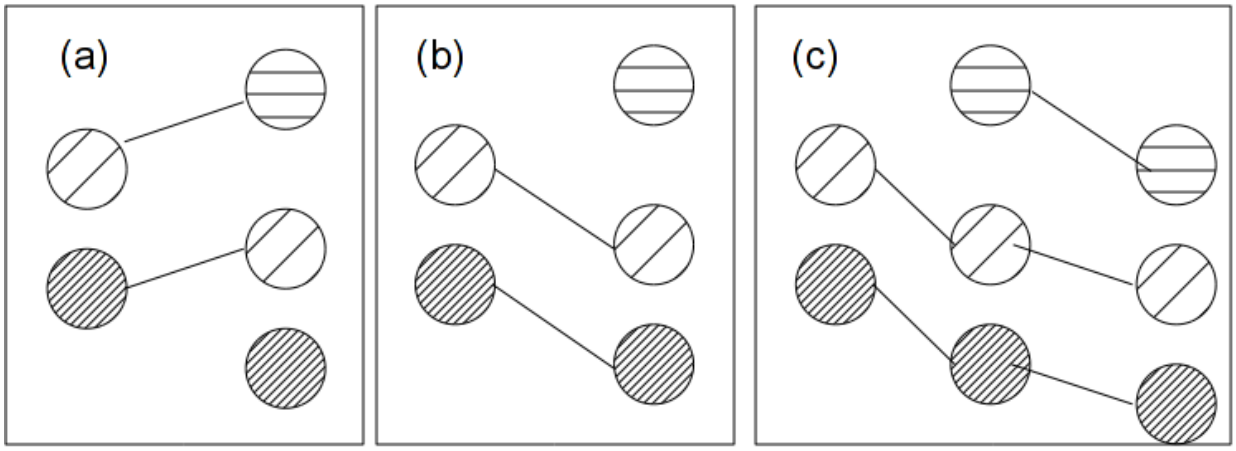

Fig. 2. Why Multiple Hypothesis Tracking is needed: (a) The optimal assignment between two objects in the first frame and three objects in the second frame. (b) The second best assignment. (c) Once the third time frame is available, it is clear that the second hypothesis was actually the better choice.

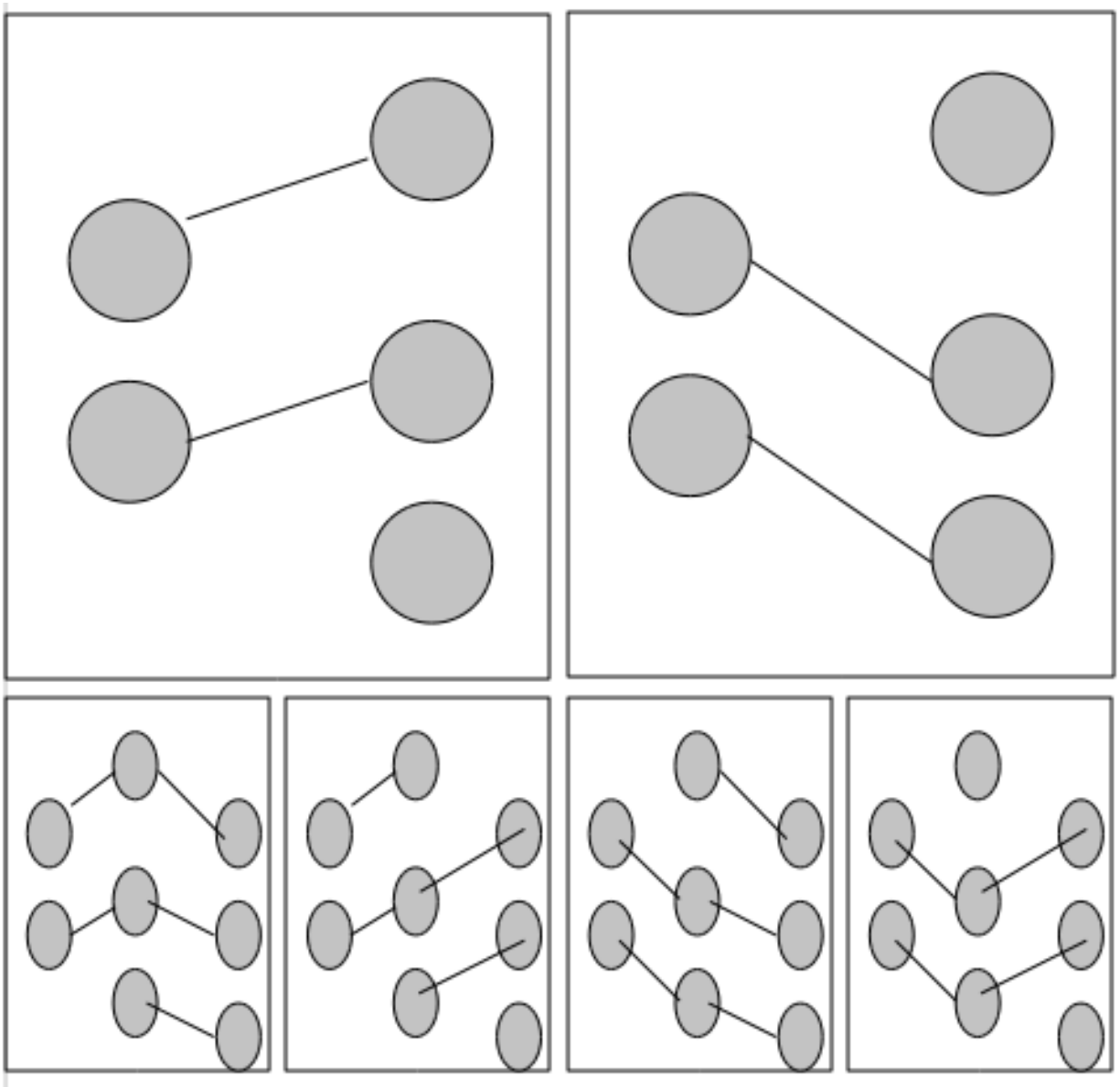

FIG. 3. Combinatorial explosion results as a result of multiple hypothesis tracking: (Top) The two best (K=2) hypotheses for assigning objects between two frames (Bottom) When the third time frame arrives, there are two best hypotheses leading from the original two best. Thus, there are four (K^2) hypotheses to be considered.

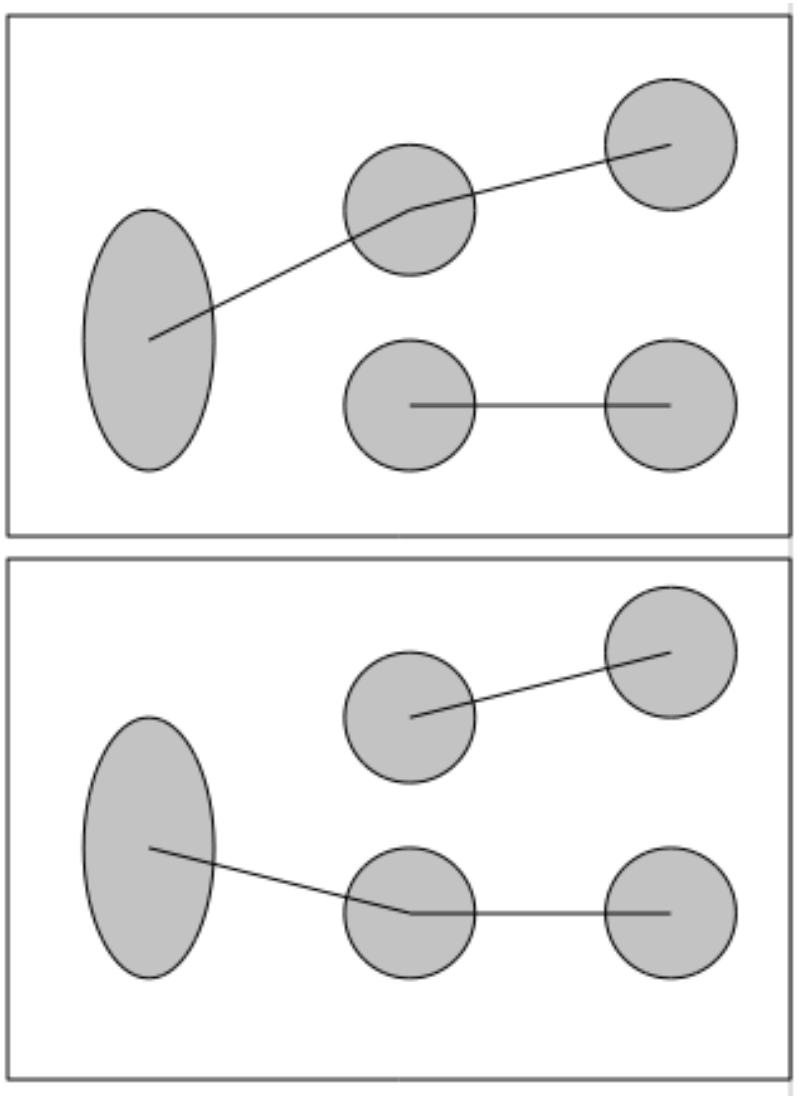

FIG. 4. MHT gracefully handles the problem of splits and mergers since both sides of a split (or merger) will remain among the K-best assignments. Top: Optimal assignment Bottom: Second-best hypothesis. As long as $K >= 2$, both hypotheses will be retained, and both objects used in the accumulation.

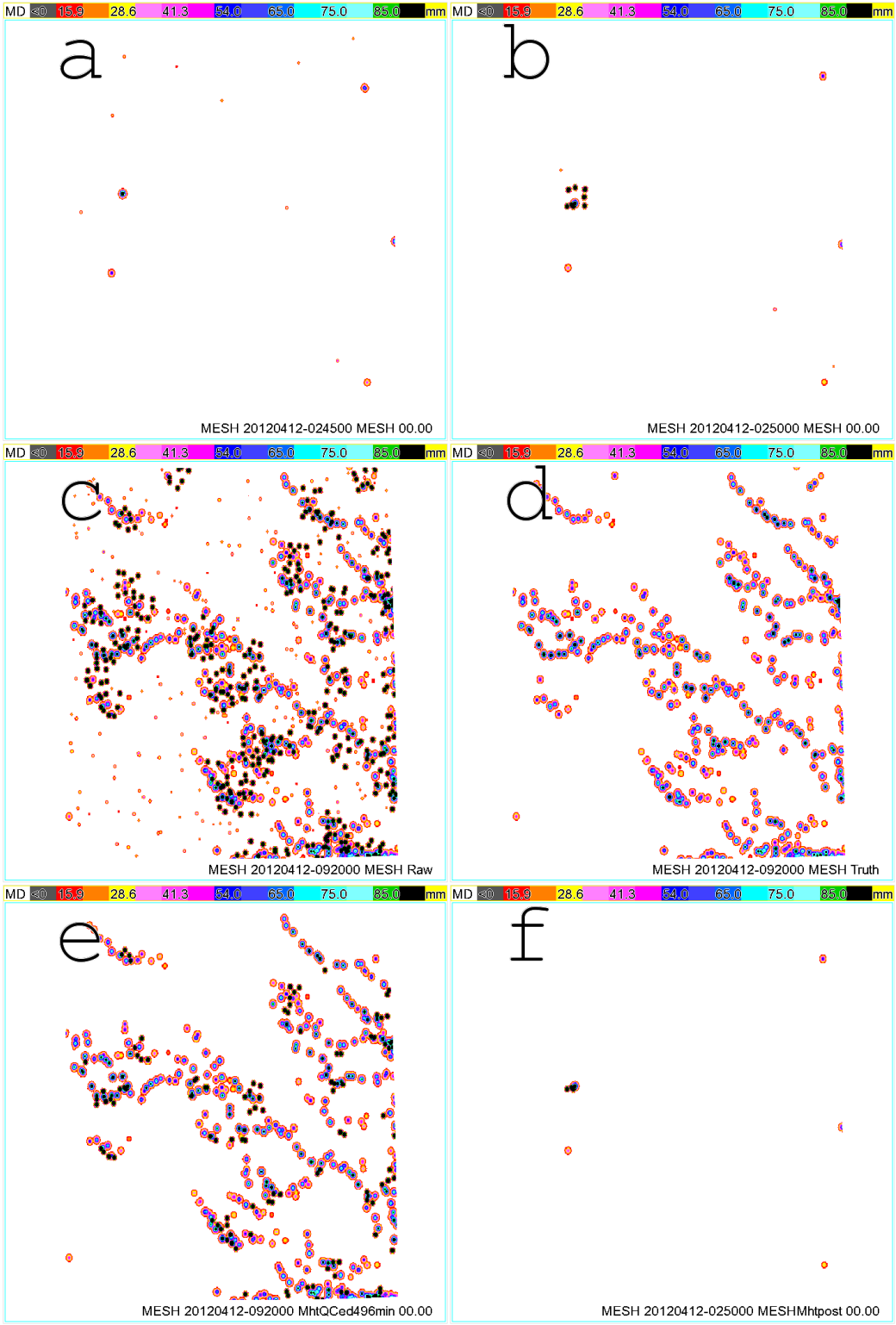

FIG. 5. (a,b): Two successive time frames of the simulated input grids. (c) Accumulation of all 100 grids of the sequence. (d) Ideal accumulation that does not incorporate any noise. (e) Accumulation when the individual time frames are QCed by insisting that objects last at least 3 frames in one of the 4 best hypotheses. (f) Individual time frame QCed to remove small or unassociated objects.

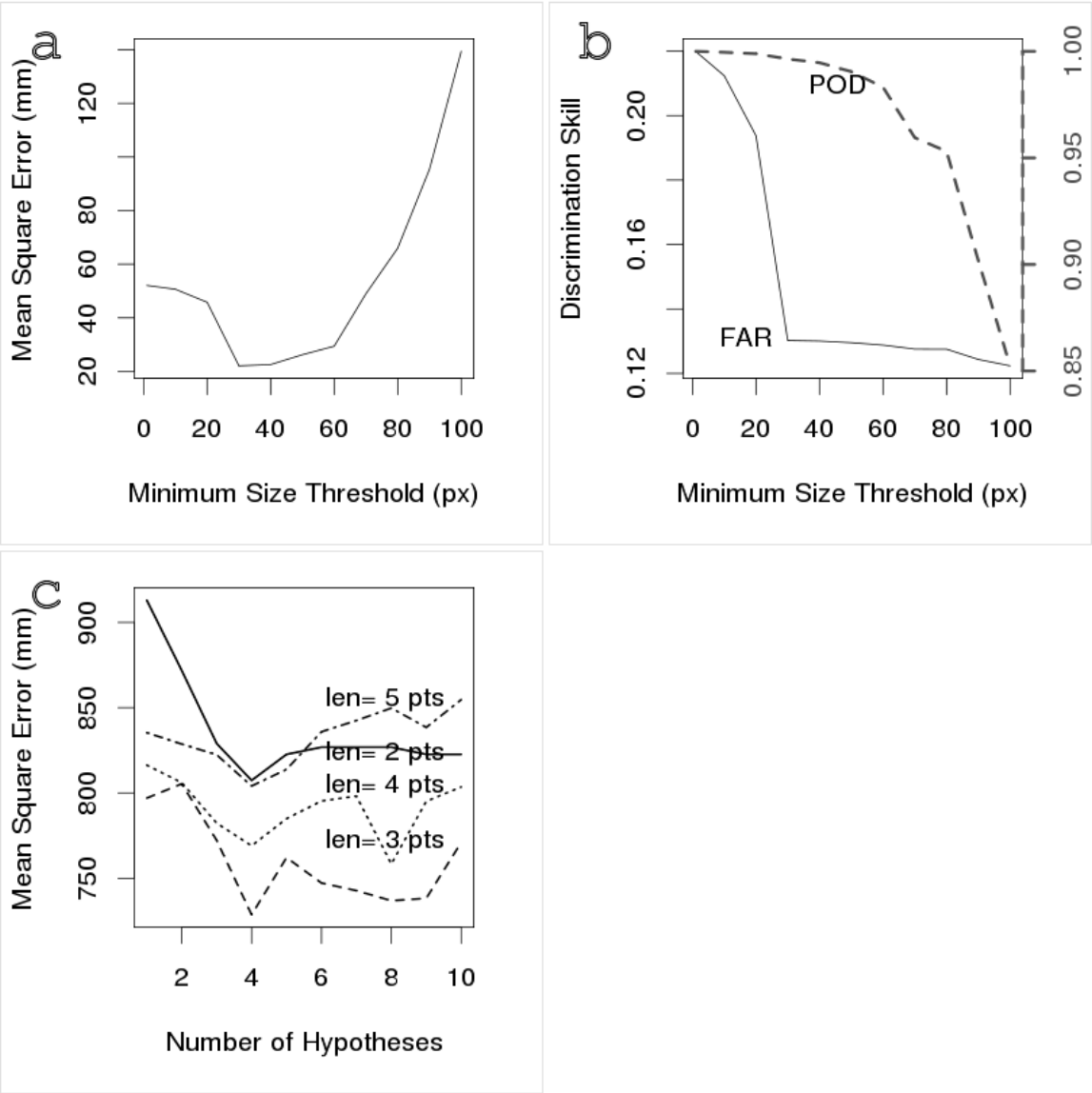

FIG. 6. Impact of the hysteresis and MHT criteria on the quality of the accumulated grids. (a) Impact of spatial coherence. As the size threshold is increased, the mean square error initially drops but then starts to increase. (b) Both the Probability of Detection (POD) of true objects and the False Alarm Rate (FAR) at which noise objects are wrongly identified fall when the size threshold is increased. (c) Impact of temporal coherence. The mean square error initially drops as the number of hypotheses are increased, but then starts to increase.

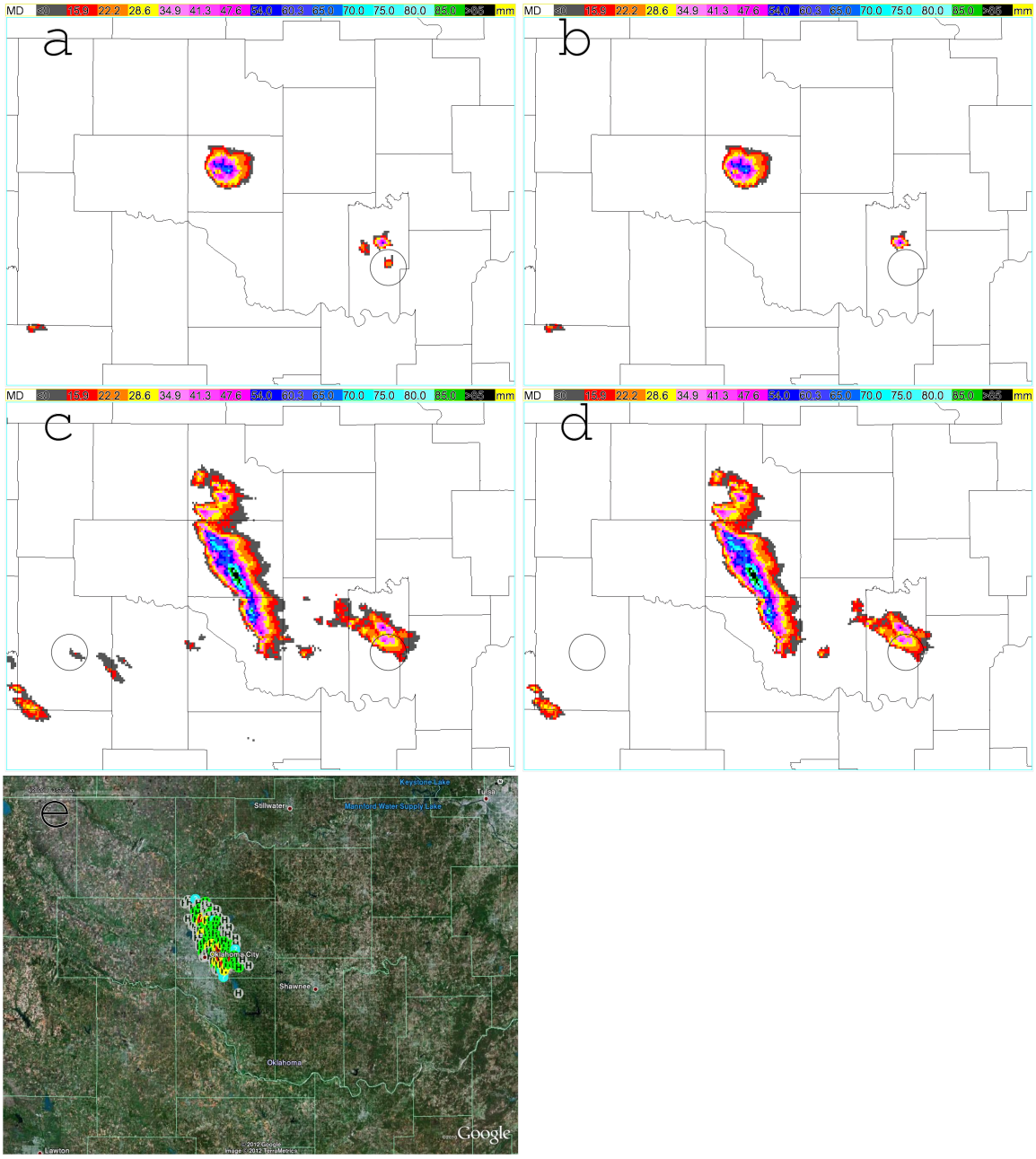

Fig. 7. Illustration of the MHT-QC algorithm of this paper on a hail accumulation over Oklahoma starting at 21:00 UTC on 16 July 2009. (a) Individual frame before QC (b) Same frame after QC based on spatial and temporal coherence (c) Raw accumulation without any QC (d) QCed accumulation (e) Actual hail reports on that day

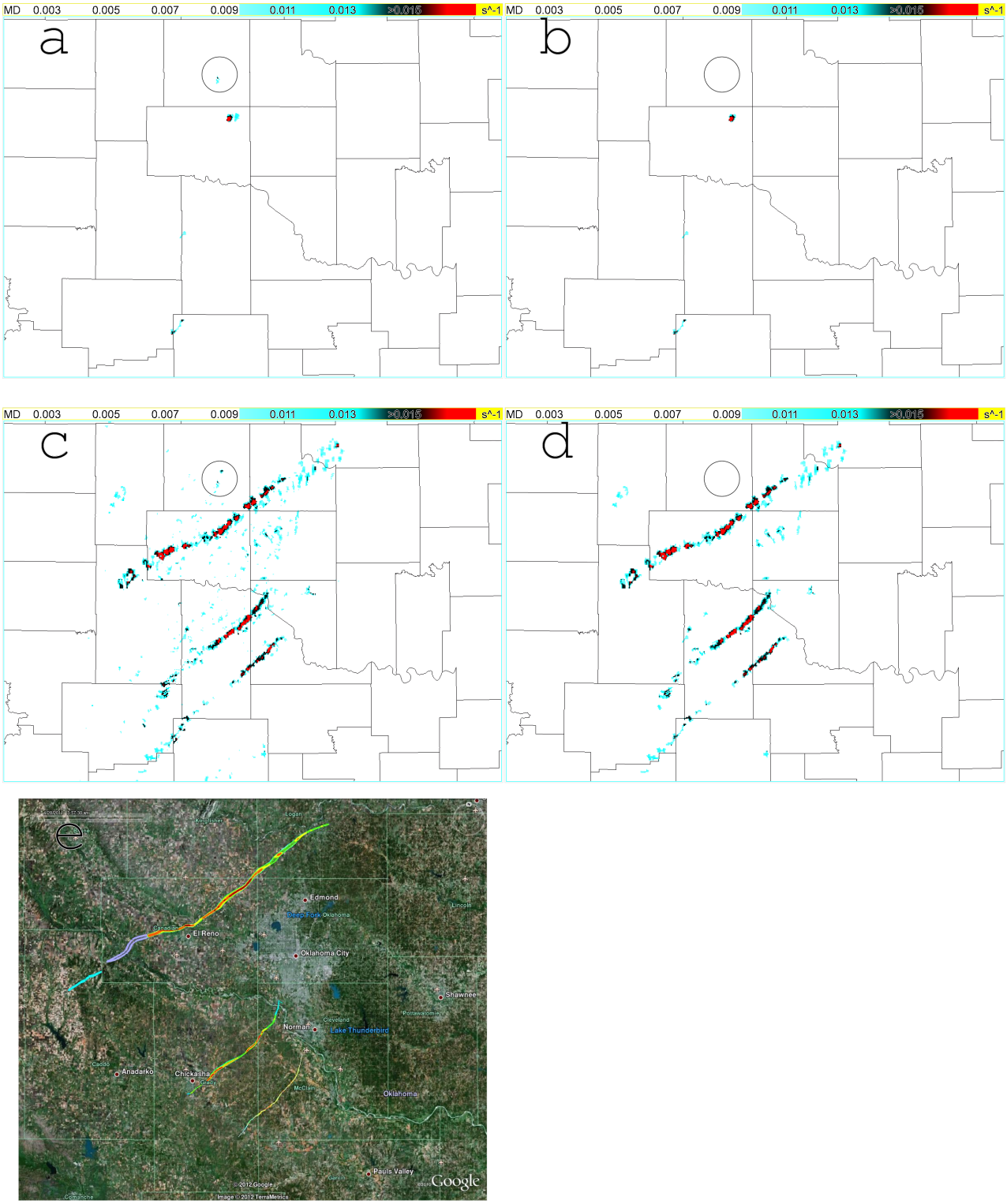

FIG. 8. Illustration of the MHT-QC algorithm of this paper on a two-hour accumulation of low-level rotation over Oklahoma on 24 May 2011. (a) Individual frame before QC (b) Same frame after QC based on spatial and temporal coherence (c) Raw accumulation without any QC (d) QCed accumulation (e) Actual damage observed on the ground.

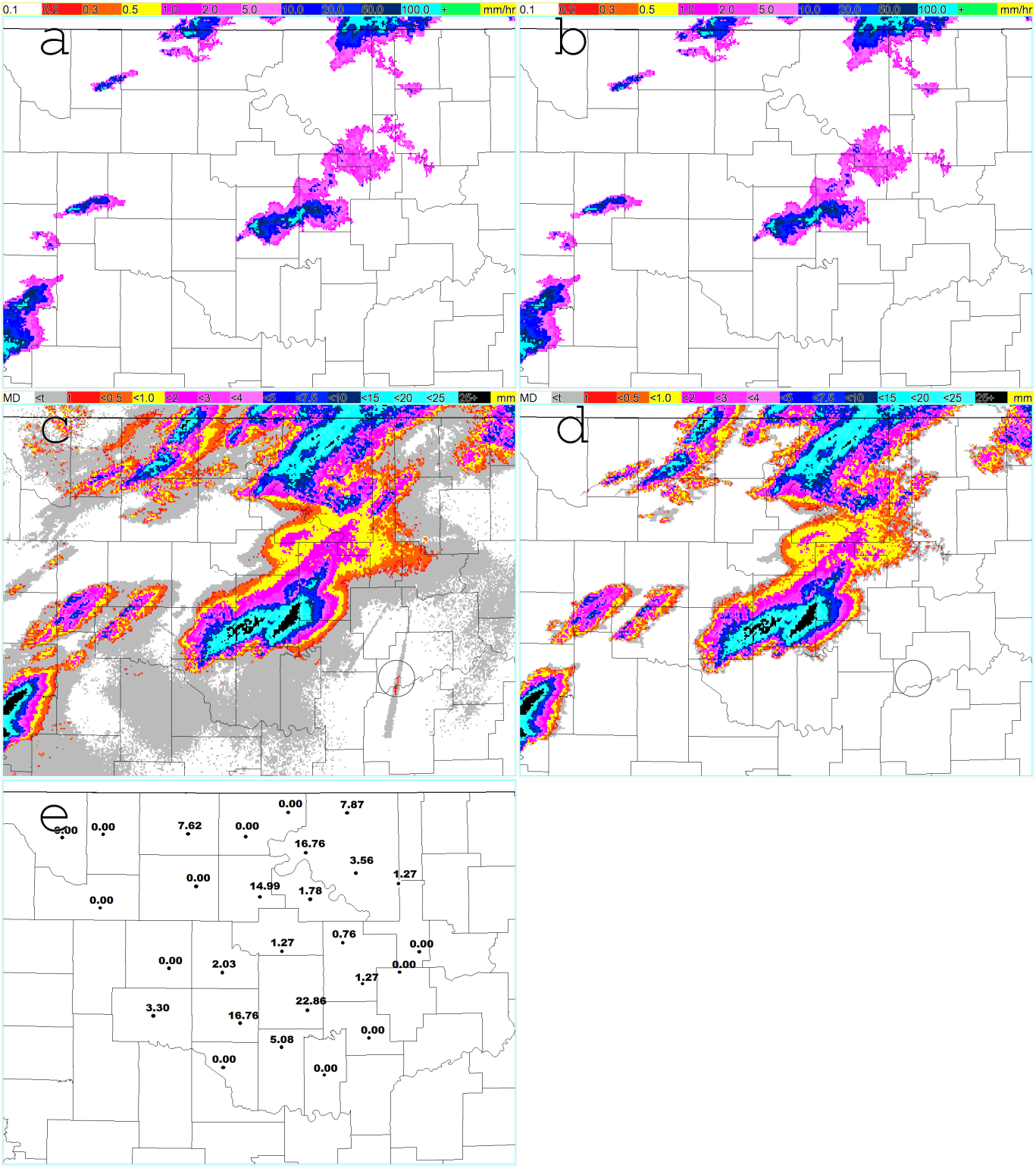

FIG. 9. Illustration of the MHT-QC algorithm of this paper on a one-hour accumulation of radar-derived precipitation over Oklahoma on 13 April 2012. (a) Individual frame before QC (b) Same frame after QC based on spatial and temporal coherence (c) Raw accumulation without any QC (d) QCed accumulation (e) Actual rainfall measured on the ground (in mm) over the time period.