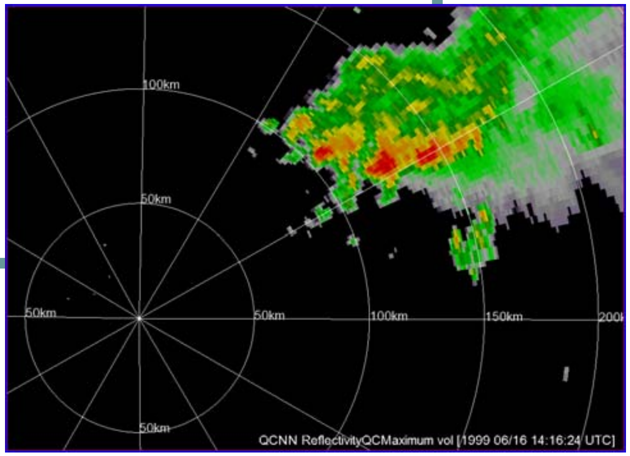
Quality Control Neural Network

Valliappa.Lakshmanan@noaa.gov

National Severe Storms Laboratory & University of Oklahoma

Information briefing to the NEXRAD Technical Advisory Committee, San Diego, CA

http://cimms.ou.edu/~lakshman/



Quality control

- Goal: clean up radar reflectivity data
  - AP/GC contamination
  - Biological targets
  - Terrain effects
  - Sun-strobes
  - Radar interference
  - Test patterns
  - Returns in clear air

Performance target

- Challenge: Errors are additive.
  - QC errors degrade quality of down-stream applications
  - Especially for algorithms that accumulate data over space/time
  - If a QC algorithm that is correct 99% of the time is used:
    - A national mosaic will have incorrect data 73% of the time
      - 130 radars
      - 0.99^130 = 0.27
    - A three-hour accumulation of precipitation (single radar) will be incorrect 30% of the time
      - 36 frames (assuming 5 minute volume scans)
      - 0.99^36 = 0.70
- Performance target
  - Keep 99.9% of good echoes (POD=0.999)
  - 99.9% of echoes that remain should be good (FAR=0.001)
  - Errors in previous example will be 12% and 4% if we can get 99.9% correct.

Existing quality control methods

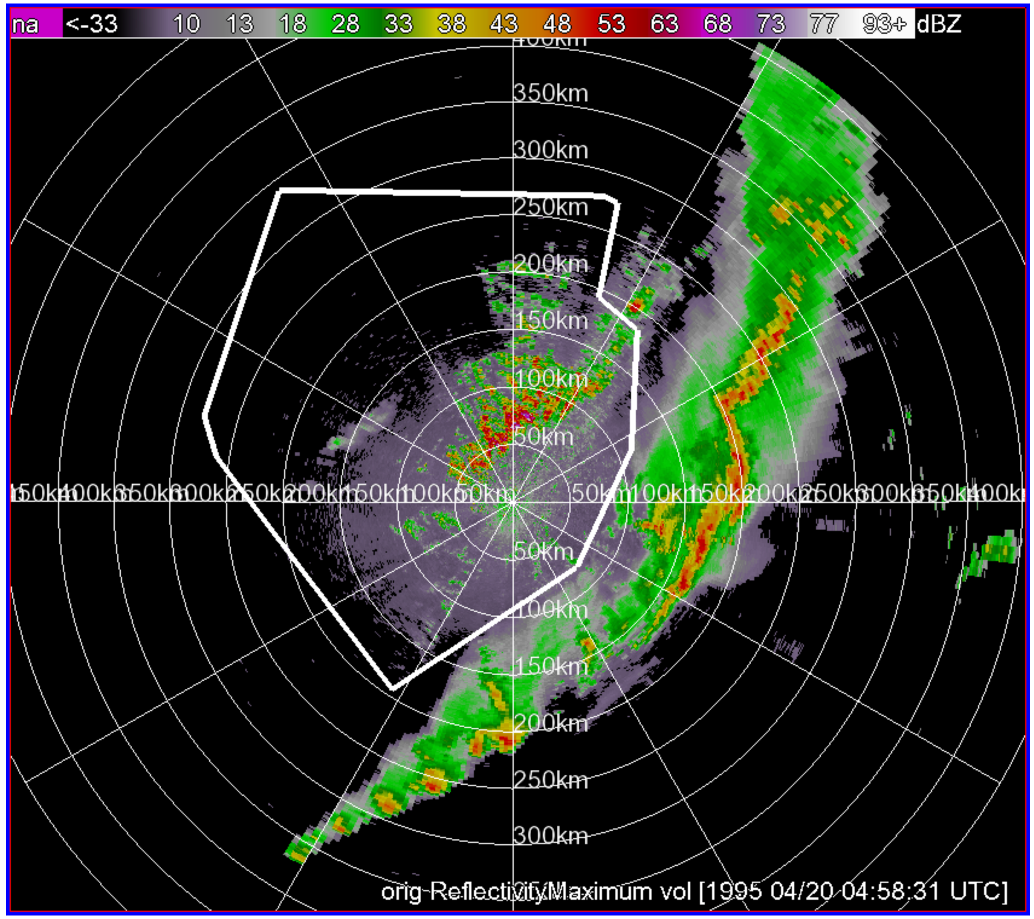
- An extensively studied problem.
  - Thresholding
  - Median filters for speckle removal
  - Vertical tilt tests (Fulton et al, 1998)
  - Echo top and texture features (Steiner & Smith 2002)
  - Artifact detection (DQA, Smalley et al, 2004)
  - Texture features on all 3 moments (REC, Kessinger et al, 2003)
- Drawbacks with existing methods
  - None of them is designed to be “omnibus”
  - Mostly operate like clutter filters -- gate-by-gate
  - Parameters and thresholds chosen through experiment
  - Hard to get 99.9% accuracy without rigorous statistical methods.

Quality Control Neural Network

- The QCNN approach is novel in 4 ways
  - Compute local 2D and 3D features
    - Texture features on all three moments.
    - Vertical features on latest ("virtual") volume
    - Can clean up tilts as they arrive and still utilize vertical features.
  - Identify the crucial / best features
    - Called "feature selection"
  - Determine optimal way to combine the features
    - Called "training" the neural network
    - A non-linear statistical technique called ridge regression
  - Identify regions of echo and classify them
    - Called "segmentation"
    - Not range-gate by range-gate
    - Reduces random errors

Human truthing

- Looked at loops
- Examined radar data
- Other sensors
- Considered terrain, time of day, etc.
- Identified bad echoes



Feature selection

- We considered 60+ features
  - Included texture, image processing, tilt-test features suggested by various researchers.
  - Experimented with different definitions
    - echo-top to include tilts above 1-degree only.
    - Tilt-test based on tilt at 3km height
    - Echo tops based on elevation angle or physical height
- Removed the features one at a time
  - If cross-entropy on validation set remained unchanged, leave that feature out permanently
- Ended up with 28.

Input features (final 28)

- Lowest velocity scan
  - Value, local mean, local variance, difference, minimum variance
- Lowest spectrum width scan: value
- Lowest reflectivity scan
  - Mean, variance, difference, variance along radials, difference along radials, minimum variance, SPIN, inflections along radial
- Second lowest reflectivity scan
  - Mean, variance, difference, variance along radials, difference along radials, minimum variance
- In virtual reflectivity volume
  - Vertical maximum, weighted average, vertical difference, echo top, ht of maximum, echo size, in-bound distance to 3km echo top, out-bound distance to zero velocity
- Actually TWO neural networks
  - One for gates with velocity data
  - The other for gates with missing/range-folded velocity data

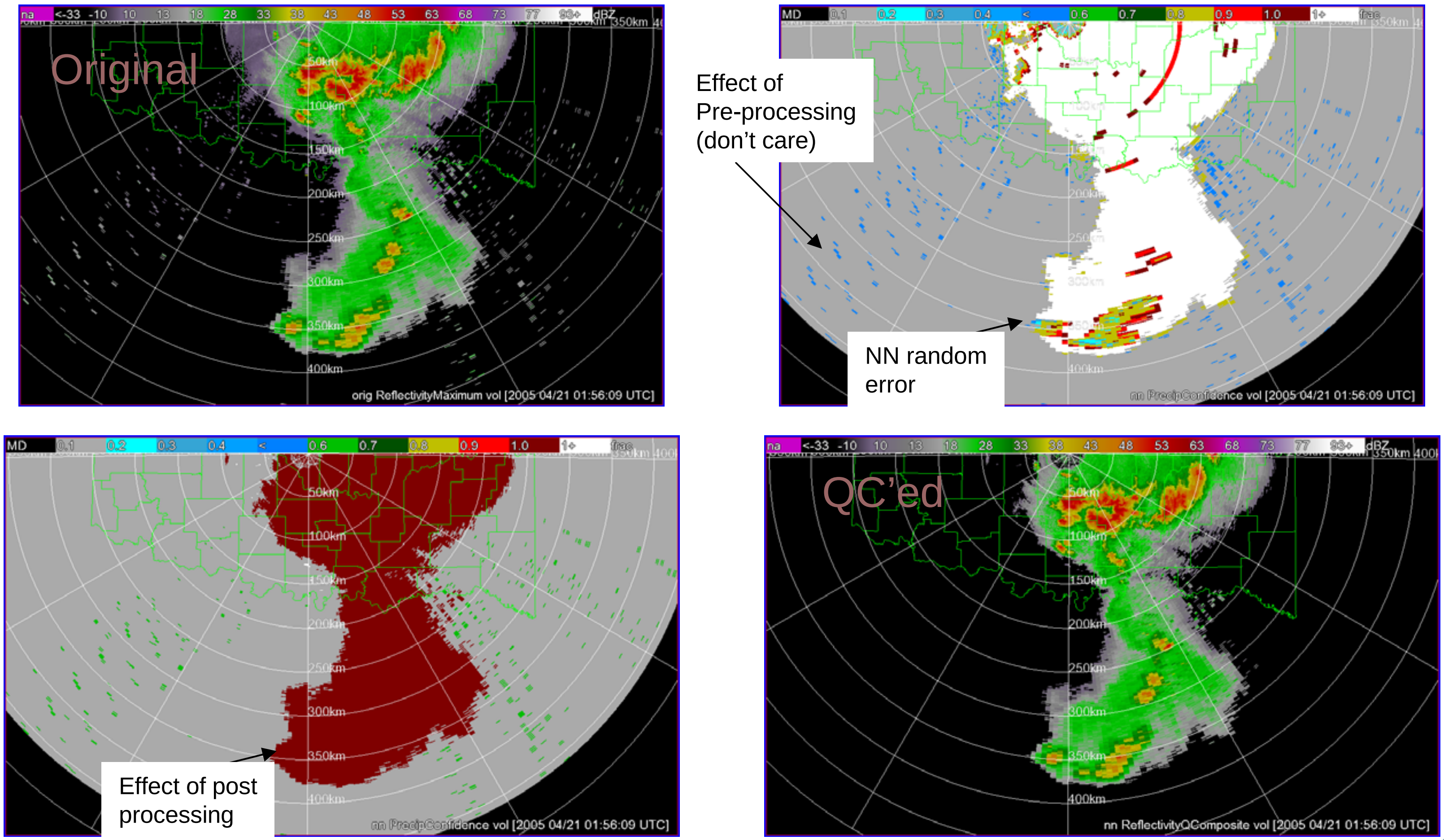
Pre-processing

- The NN is presented with “hard” and “significant” cases only.
  - Pre-classify “obvious” cases
    - Echo-top above 3 km
    - Velocity gates with exactly zero velocity
    - Reduces chance of sub-optimal optimization.
    - Also remove bad radials or test patterns
      - Not local features
      - NN is trained on local features only
  - Mark some range-gates as “don’t care”
    - Range-gates near the edges of an echo
    - Local statistics are poor in such regions

Spatial post-processing

- A range-gate by range-gate classification is subject to random errors
  - This can cause severe degradation in data quality
  - Need to perform quality control on echo regions
- Identify echo regions (“blobs”)
  - Using segmentation (Lakshmanan et. al 2003)
  - Average the NN classification on these blobs
  - Blob stays in or blob goes out.

Effect of pre & post processing



Our REC comparison

- We used the operational REC (ORPG Build 8)
  - We wanted to compare our research algorithm against the operational one.
  - Could have also used the DQA
- The operational REC is only for AP
  - But we compared on all types of QC problems.
  - Our goal was a omnibus QC algorithm
  - So, our numbers should not be used to gage how well the REC performs
  - Such a comparison would use only AP cases

Targeting a QC application

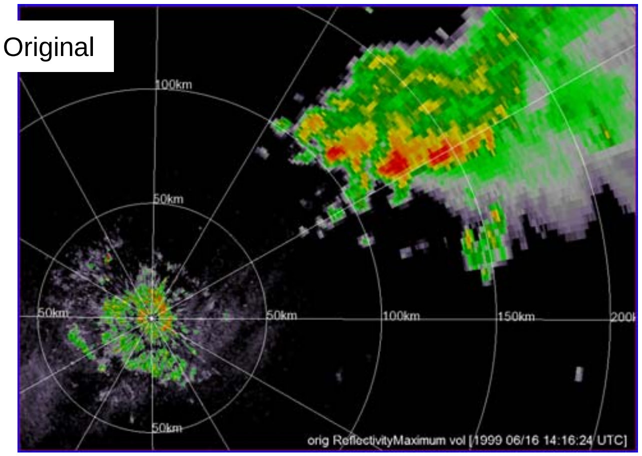
- Can improve QC performance by targeting
  - Targeting can be by region, season, reflectivity values, etc.
  - Example of seasonal targeting
    - Require blobs to have at least 0 dBZ in winter or 20 dBZ in summer.
    - Very useful in removing clear-air return.
  - Targeting limits search space for optimization: "If you don't need to be broad, you can be deep".
- Implications of such targeting:
  - Need to provide both edited and unedited data streams.
  - End-user can perform more targeted QC if needed.
- But targeting the type of echo is not a good idea
  - Should not have separate algorithms for AP, insects, artifacts etc.
  - What users want is an "edited" data stream
  - So, the user will use rules of thumb to combine
  - The QC algorithm should take care of combining in an optimal way.

How to assess skill?

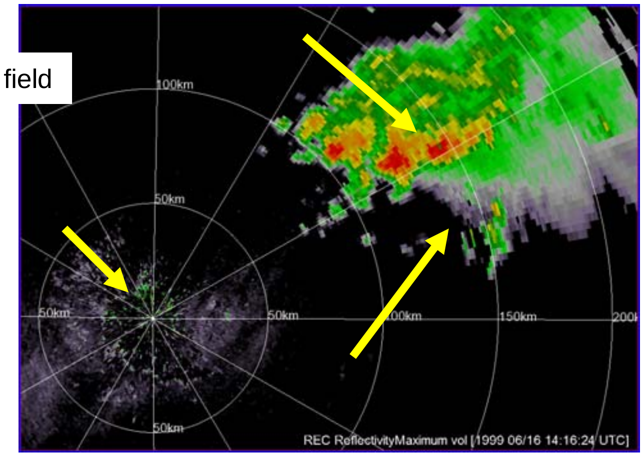
- Use other radars/sensors
  - Dual-pol hydromet classifier
    - Kessinger et. al (REC)
  - Rain gages
    - Robinson et. al (2001)
  - Limited by bias, coverage and availability of the other sensor
- We scored against human-edited data
  - On independent test cases.

Test case: AP + precipitation

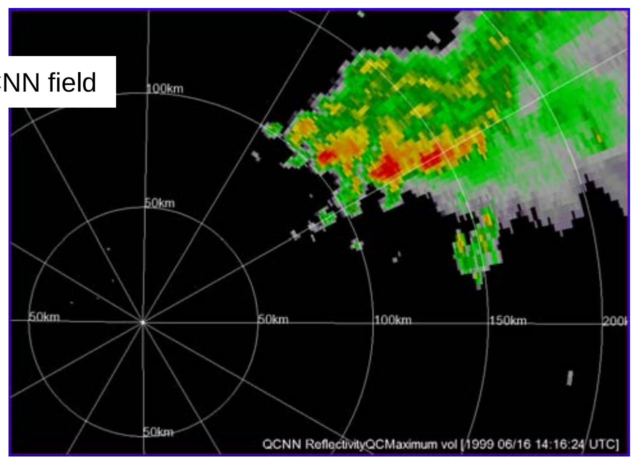
Original



REC field



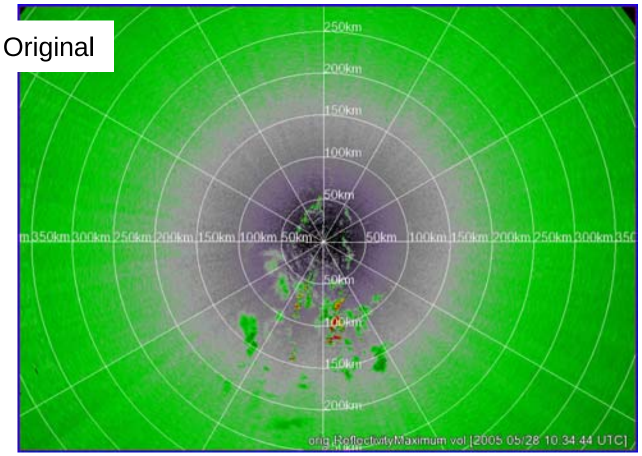
QCNN field



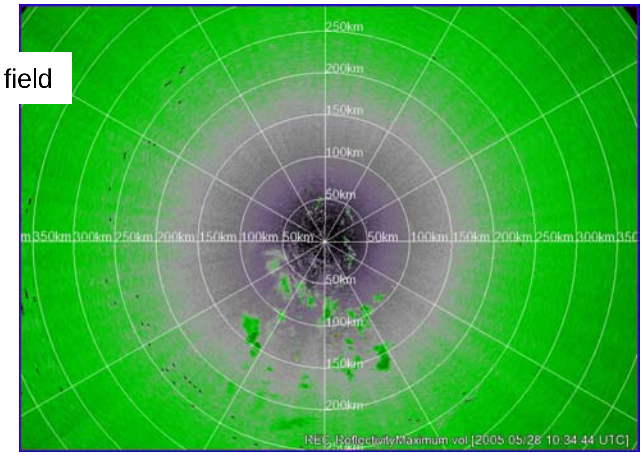
- Near-perfect correlation of QCNN field with human truthing.
- REC is mostly correct
  - Random errors within precip
  - Not all non-precip removed

Hardware test pattern

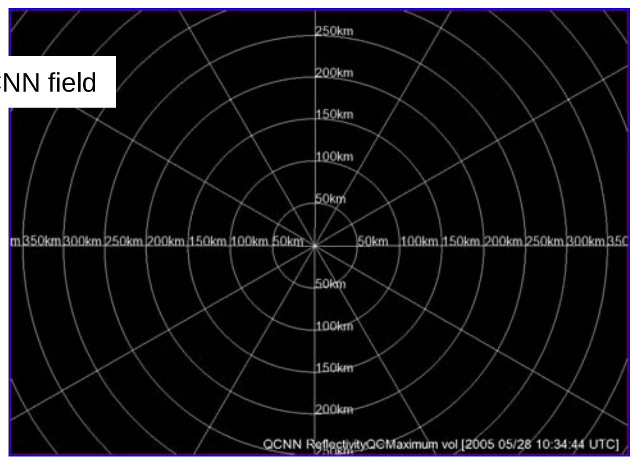
Original



REC field



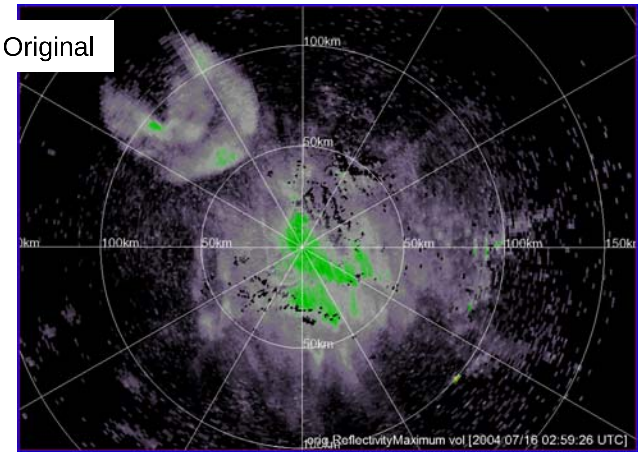
QCNN field



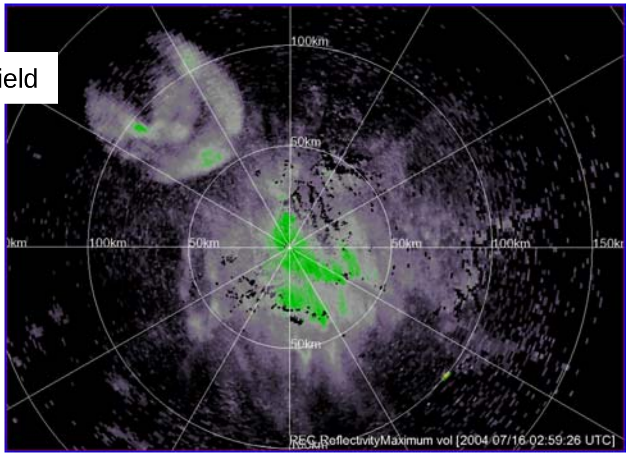
- Perfect correlation of QCNN with human truthing.
- The REC was not designed for this

Extreme biological contamination

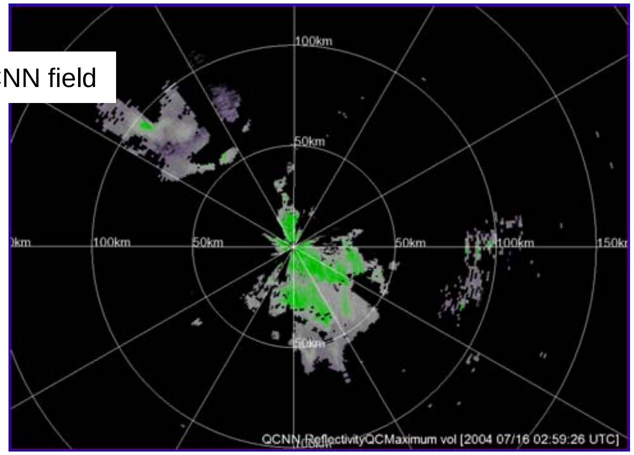
Original



REC field

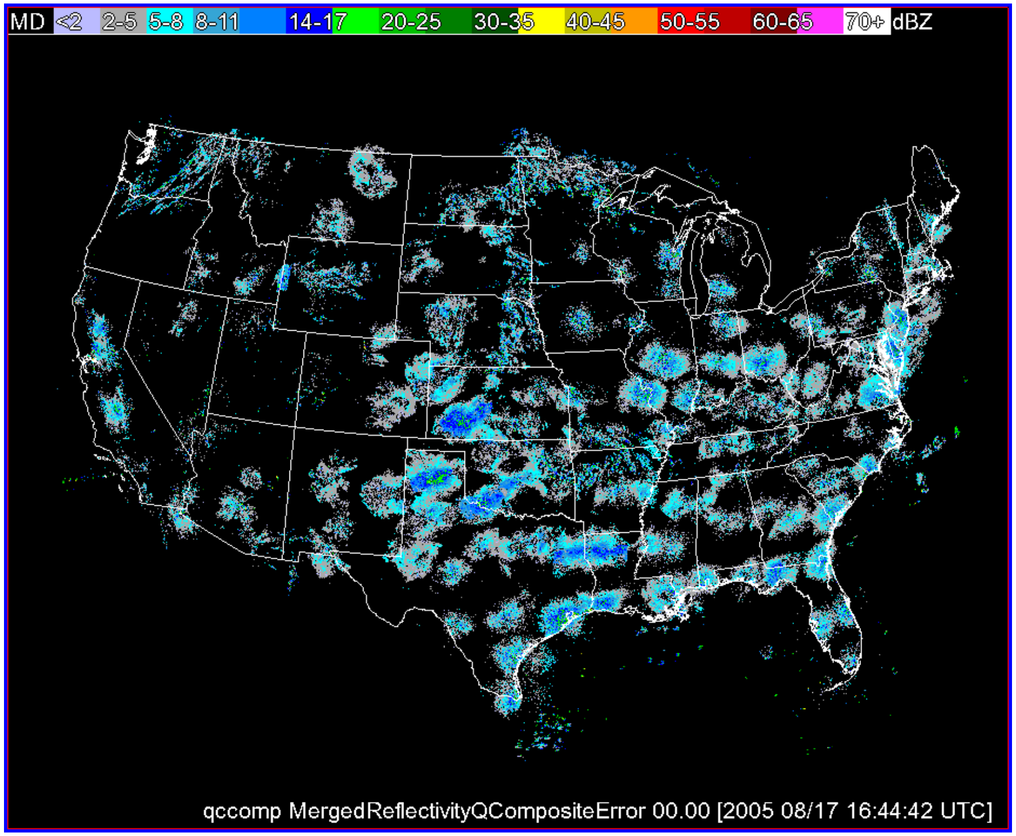


QCNN field



- QCNN had never encountered this kind of data during training.
- The REC was not designed for this
- Multi-sensor QC is useful
  - Use satellite/surface observations
  - Don't have statistics yet

Impact of QCNN on bloom …



- Bloom is a common problem
  - Not addressed by any operational QC algorithm
  - Left: Echoes removed by the QCNN.
- Clear-air echoes have meteorological value
  - But nuisance for many downstream automated applications
  - McGrath et. al (2002) found that most MDA false alarms occur in clear-air situations.
  - Mazur et. al, (2003) found that using the QCNN, 92% of MDA false alarms in clear-air could be removed without impacting the probability of detection.
  - Stationary echoes impact motion estimates.
- Another reason to provide both edited and unedited data streams

Performance Assessment

Radar-only QCNN with no seasonal targeting

POD: Probability of detection of "good" echo (fraction of good echo retained)

FAR: Fraction of echoes in final product that are "bad"

CSI: Critical success index

HSS: Heidke skill score

Product	Data range	Measure	No QC	REC	QCNN
Composite	$> 0\ dBZ$	CSI	0.61 +/- 0.06	0.59 +/- 0.057	0.86 +/- 0.011
		FAR	0.39 +/- 0.06	0.4 +/- 0.06	0.02 +/- 0.0072
		POD	1 +/- 0	0.96 +/- 0.0031	0.88 +/- 0.0088
		HSS	0.89 +/- 0.02	0.88 +/- 0.019	0.98 +/- 0.0016
Composite	$> 10\ dBZ$	CSI	0.68 +/- 0.071	0.66 +/- 0.069	0.96 +/- 0.0083
		FAR	0.32 +/- 0.071	0.32 +/- 0.073	0.02 +/- 0.007
		POD	1 +/- 0	0.94 +/- 0.0023	0.92 +/- 0.0039
		HSS	0.93 +/- 0.017	0.93 +/- 0.016	0.99 +/- 0.0011
Composite	$> 30\ dBZ$	CSI	0.92 +/- 0.02	0.84 +/- 0.014	1 +/- 0.00072
		FAR	0.08 +/- 0.02	0.09 +/- 0.011	0 +/- 0.00057
		POD	1 +/- 0	0.92 +/- 0.0065	1 +/- 0.00029
		HSS	1 +/- 0.00064	0.99 +/- 0.00052	1 +/- 0
Composite	$> 40\ dBZ$	CSI	0.91 +/- 0.023	0.8 +/- 0.013	1 +/- 0.00038
		FAR	0.09 +/- 0.023	0.1 +/- 0.0074	0 +/- 0.00039
		POD	1 +/- 0	0.88 +/- 0.0088	1 +/- 0
		HSS	1 +/- 0.00016	1 +/- 0.00018	1 +/- 0
VIL	$> 0\ kg/m^3$	CSI	0.53 +/- 0.16	0.48 +/- 0.13	1 +/- 0.0011
		FAR	0.47 +/- 0.16	0.49 +/- 0.15	0 +/- 0.00053
		POD	1 +/- 0	0.9 +/- 0.0078	1 +/- 0.00084
		HSS	0.97 +/- 0.0091	0.97 +/- 0.0085	1 +/- 0
VIL	$> 25\ kg/m^3$	CSI	1 +/- 0.0022	0.65 +/- 0.033	0.99 +/- 0.0027
		FAR	0 +/- 0.0022	0.19 +/- 0.025	0 +/- 0.0022
		POD	1 +/- 0	0.76 +/- 0.026	1 +/- 0.00075
		HSS	1 +/- 0	1 +/- 0	1 +/- 0

Visual quality 95 to 97%

Effect on precip algorithms 99.9 to 100%

Effect on severe weather algorithms 99.9 to 100%

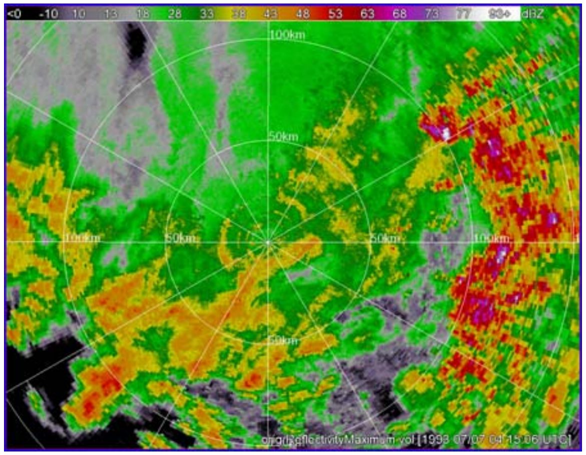
Summary

- QC errors, for algorithms that accumulate in space/time, are additive.
  - So the QC algorithm has to be near-perfect.
- The QCNN approach is "evidence-based":
  - A timely (virtual volume) method for computing features
  - A formal method of selecting features
  - An optimization procedure (ridge regression) to combine these features
  - Classify regions of echo, not range-gate by range-gate
- QCNN is an omnibus algorithm
  - Designed to handle AP/GC, radar test patterns, interference, sun strobes, clear air return
  - In shallow precipitation, strong convection, snow from all over the CONUS
  - Constantly adding new training cases (both good and bad echoes)
  - Better performance possible if the algorithm is targeted or if we use multi-sensor
- The approach can be adapted to ORDA
  - Collect enough training cases for both good and bad data
  - Possibly identify new features for new problems observed
- Technical details of algorithm
  - http://cimms.ou.edu/~lakshman/Papers/qcnnjam.pdf
  - Paper submitted to J. Applied Meteorology

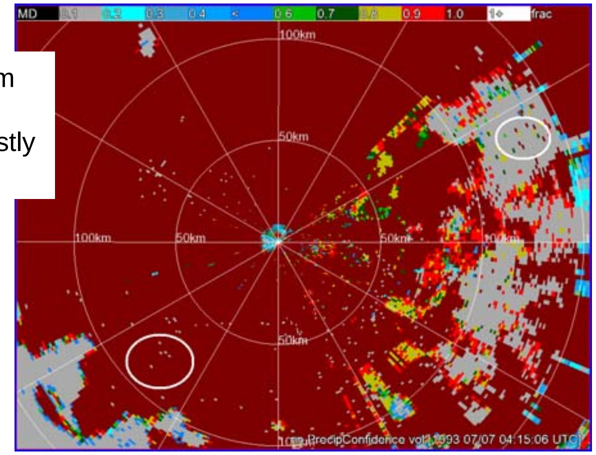
Supplementary slides

Quality Control Neural Network

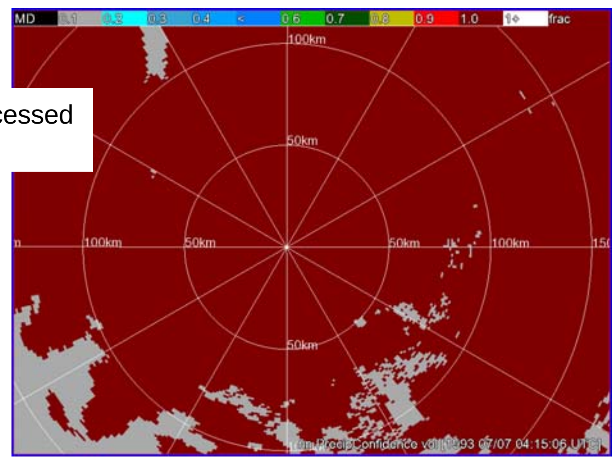
Spatial post-processing …



Random errors, but mostly correct

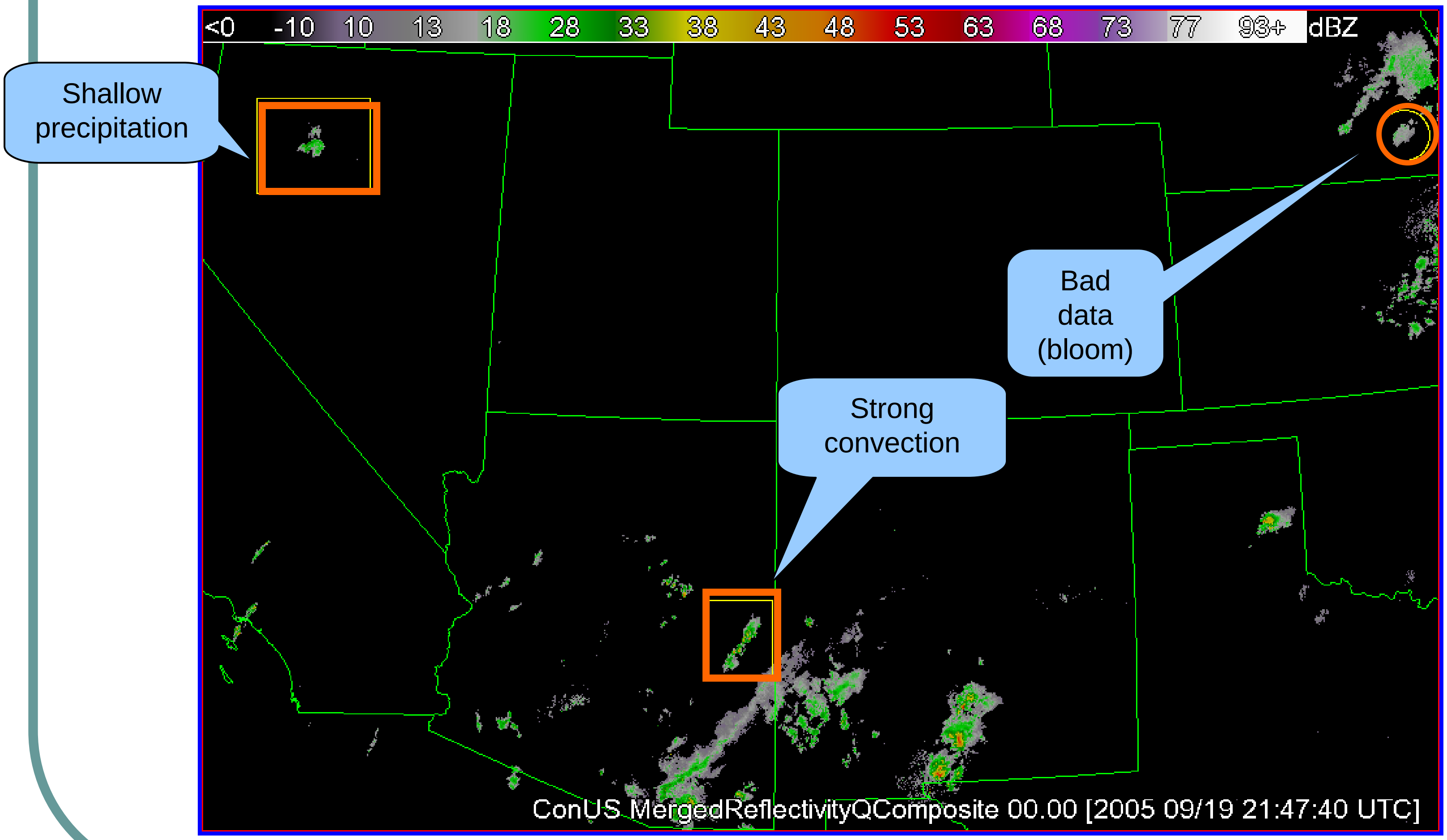


Post processed Field

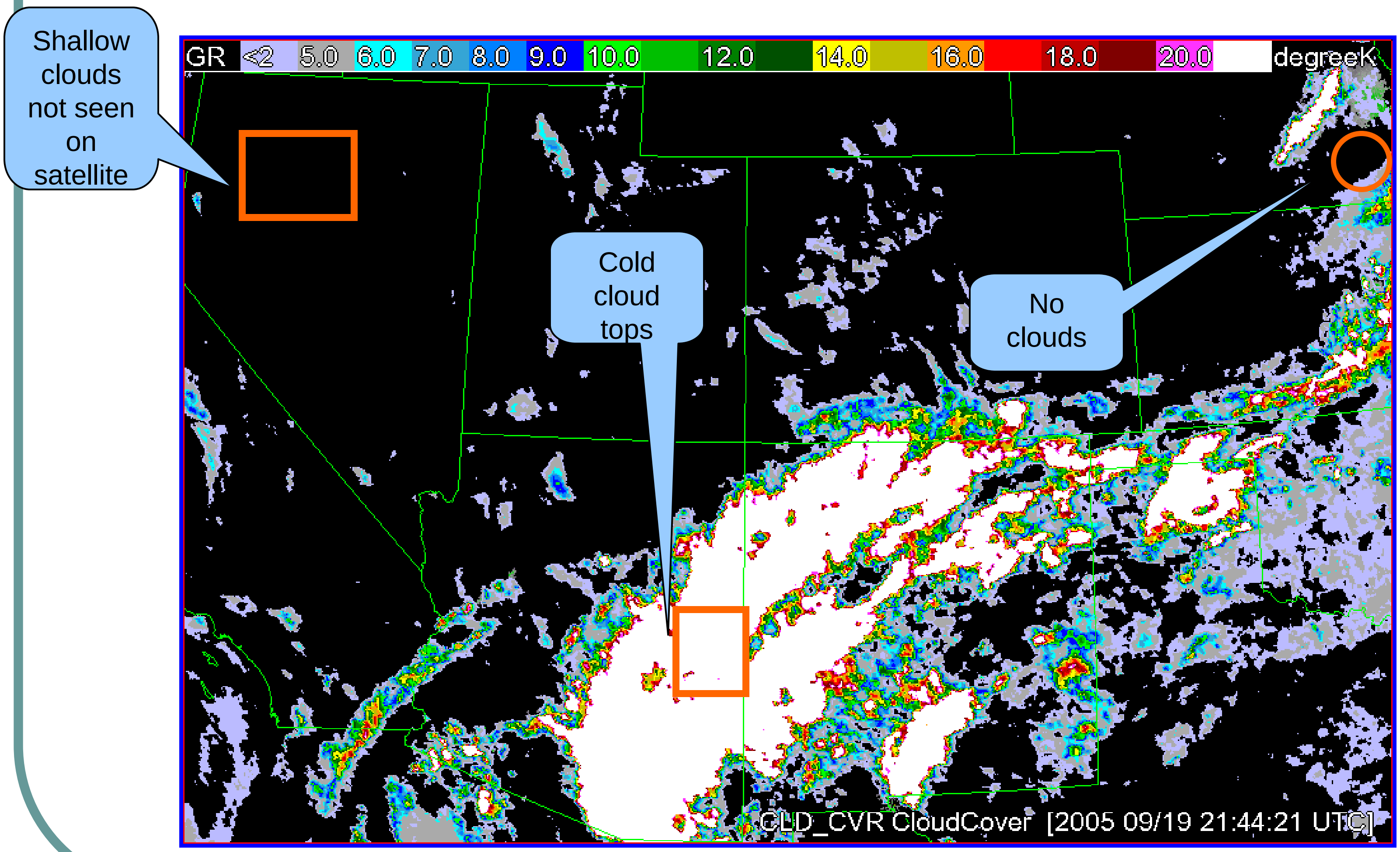


- Spatial post-processing can itself cause problems
- Here, AP embedded inside precipitation is not removed because of spatial post-processing.
- A vector segmentation approach might help here.

Radar-only QC



Cloud cover (T_surface – T_IR)



Multi-sensor QC

Good data wrongly removed

Good data correctly retained

Bad data correctly removed