

12-2010

# NEW TOOLS FOR MONITORING GAMMA CAMERA UNIFORMITY

Brad K. Lofton

Follow this and additional works at: [http://digitalcommons.library.tmc.edu/utgsbs\\_dissertations](http://digitalcommons.library.tmc.edu/utgsbs_dissertations)

 Part of the [Biological and Chemical Physics Commons](#), [Medical Biophysics Commons](#), [Other Biomedical Engineering and Bioengineering Commons](#), and the [Other Physics Commons](#)

---

## Recommended Citation

Lofton, Brad K., "NEW TOOLS FOR MONITORING GAMMA CAMERA UNIFORMITY" (2010). *UT GSBS Dissertations and Theses (Open Access)*. Paper 91.

This Thesis (MS) is brought to you for free and open access by the Graduate School of Biomedical Sciences at DigitalCommons@The Texas Medical Center. It has been accepted for inclusion in UT GSBS Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@The Texas Medical Center. For more information, please contact [laurel.sanders@library.tmc.edu](mailto:laurel.sanders@library.tmc.edu).

# NEW TOOLS FOR MONITORING GAMMA CAMERA UNIFORMITY

By

Bradley Kyle Lofton, B.S.

APPROVED:

---

Richard E. Wendt III, Ph.D., Supervisory Professor

---

William Erwin, MS.

---

S. Cheenu Kappadath, Ph.D.

---

Paul H. Murphy, Ph.D.

---

Lynn H. Vogel, Ph.D.

---

R. Allen White, Ph.D.

APPROVED:

---

George M. Stancel, Ph.D.

Dean, The University of Texas

Health Science Center at Houston

Graduate School of Biomedical Sciences at Houston

# NEW TOOLS FOR MONITORING GAMMA CAMERA UNIFORMITY

A

THESIS

Presented to the Faculty of  
The University of Texas  
Health Science Center at Houston  
and  
The University of Texas  
M.D. Anderson Cancer Center  
Graduate School of Biomedical Sciences  
in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

by

Bradley Kyle Lofton, B.S.

Houston, Texas

December 2010

## Acknowledgments

I would like to thank my advisor, Dr. Richard Wendt, for his encouragement in taking on this proof-of-concept project. I appreciate his confidence in me, and his latitude in allowing me to be creative and find my own path. I will value this throughout my career.

I would also like to thank my parents, Larry and Becky Lofton, for their support of me in these often difficult times. I will always treasure their love and encouragement.

To my beautiful wife Melissa, I literally could not have done this without you. You've kept me going and have supported me through your finances, your love, and your faith in me.

This thesis is dedicated to my grandmother, Elnora Curtis, who passed away of cancer while I was working on this project. I can only hope that this simple work will be the introductory statement of a career devoted towards the fighting of this disease through the foundations of physics.

And finally, “now to Him who is able to do far more abundantly than all that we ask or think, according to the power at work within us” (Ephesians 3:20). My God, You are faithful when I am not, and I pray that Your power will always be perfected in my weakness.

# NEW TOOLS FOR MONITORING GAMMA CAMERA UNIFORMITY

Bradley Kyle Lofton, BS

Supervisory Professor: Richard E. Wendt III, Ph.D.

Detector uniformity is a fundamental performance characteristic of all modern gamma camera systems, and ensuring a stable, uniform detector response is critical for maintaining clinical images that are free of artifact. For these reasons, the assessment of detector uniformity is one of the most common activities associated with a successful clinical quality assurance program in gamma camera imaging. The evaluation of this parameter, however, is often unclear because it is highly dependent upon acquisition conditions, reviewer expertise, and the application of somewhat arbitrary limits that do not characterize the spatial location of the non-uniformities. Furthermore, as the goal of any robust quality control program is the determination of significant deviations from standard or baseline conditions, clinicians and vendors often neglect the temporal nature of detector degradation (1).

This thesis describes the development and testing of new methods for monitoring detector uniformity. These techniques provide more quantitative, sensitive, and specific feedback to the reviewer so that he or she may be better equipped to identify performance degradation prior to its manifestation in clinical images. The methods exploit the temporal nature of detector degradation and spatially segment distinct regions-of-non-uniformity using multi-resolution decomposition. These techniques were tested on synthetic phantom data using different degradation functions, as well as on experimentally acquired time series floods with induced, progressively worsening defects present within the field-of-view. The sensitivity of conventional, global figures-of-merit for detecting changes in uniformity was evaluated and compared to these new image-space techniques.

The image-space algorithms provide a reproducible means of detecting regions-of-non-uniformity prior to any single flood image's having a NEMA uniformity value in excess of 5%. The sensitivity of these image-space algorithms was found to depend on the size and magnitude of the non-uniformities, as well as on the nature of the cause of the non-uniform region. A trend analysis of the conventional figures-of-merit demonstrated their sensitivity to shifts in detector uniformity. The image-space algorithms are computationally efficient. Therefore, the image-space algorithms should be used concomitantly with the trending of the global figures-of-merit in order to provide the reviewer with a richer assessment of gamma camera detector uniformity characteristics.

## Table of Contents

1. PROJECT HYPOTHESIS AND SPECIFIC AIMS.....	1
2. BACKGROUND .....	3
2.1 Gamma Camera Basics .....	3
2.1.1 Function and Modes of Operation.....	3
2.1.2 System Components.....	4
2.1.2.1 Collimators.....	5
2.1.2.2 NaI (TI) Scintillation Crystal and its Container .....	6
2.1.2.3 Photomultiplier Tubes (PMTs) .....	8
2.1.2.4 Light Guides and Optical Coupling .....	9
2.1.2.5 Shielded Housing .....	9
2.1.2.6 Summing and Positioning Circuits.....	9
2.2 Uniformity as a Measure of Performance .....	10
2.2.1 Uniformity Calibrations .....	11
2.2.2 Mechanisms of Non-Uniformity .....	12
2.2.2.1 PMT Drift and Position Dependence .....	12
2.2.2.2 Optical Causes of Uniformity Degradation.....	14
2.2.2.3 Non-linearity as a Cause of Non-uniformity.....	15
2.2.3 Uniformity Floods for Routine Quality Control .....	16
2.3 Introduction to Counting Statistics and Distribution Models.....	18
2.3.1 Characterizing Data.....	18
2.3.2 Predictive Models.....	20
2.3.3 Hypothesis Testing.....	21
3. REVIEW OF THE CURRENT STATE OF THE ART IN UNIFORMITY MONITORING .....	23
3.1 Conventional Quantitative Uniformity Analysis.....	23
3.1.1 The NEMA Uniformity Index.....	23

3.1.2	Corrected Relative Standard Deviation.....	25
3.2	Temporal Pixelwise Methods of Monitoring Uniformity in Image-space .....	25
3.2.1	Kalemis' Pixelwise Trend Analysis .....	25
3.2.2	Kalemis' Pixelwise Statistical Test.....	28
4.	CURRENT LIMITATIONS OF THE STATE OF THE ART .....	30
4.1	Limitations in Conventional Routine Uniformity Evaluation.....	30
4.2	Limitations in Kalemis' Pixelwise Techniques.....	32
5.	DESCRIPTION OF SOLUTIONS .....	34
5.1	Specific Aim 1 – Time Series Phantoms.....	34
5.1.1	Materials and Methods.....	35
5.1.1.1	Synthetic Uniformity Floods.....	35
5.1.1.2	Acquired Uniformity Floods – Altering Sensitivity via Regional Attenuation.....	37
5.1.1.3	Acquired Uniformity Floods – Altering Sensitivity via Energy Window Shifts	40
5.1.1.4	Retrospective Uniformity Floods – PMT Failure .....	40
5.1.1.5	Retrospective Uniformity Floods – NaI (Tl) Hydration.....	42
5.1.2	Results.....	42
5.1.3	Discussion .....	44
5.2	Specific Aim 2 –Time Series Image Analysis Methods .....	44
5.2.1	Materials and Methods.....	44
5.2.1.1	Temporal Analysis .....	45
5.2.1.2	Multi-resolution Spatial Decomposition .....	46
5.2.1.3	Multi-resolution Minimum Detectable Activity .....	50
5.2.1.4	Multi-resolution Two Sample Z and t-tests.....	51
5.2.1.5	Multi-resolution Strength-of-correlation.....	55
5.2.1.6	Sliding Window t-test (SWTT).....	56

5.2.2	Results and Discussion.....	58
5.2.2.1	Multi-resolution Minimum Detectable Activity (MDA) .....	58
5.2.2.2	Multi-resolution Two Sample Z and t-tests.....	64
5.2.2.3	Multi-resolution Strength-of-correlation.....	70
5.2.2.4	Sliding Window t-test .....	73
5.3	Specific Aim 3 – Statistical Process Control Techniques .....	78
5.3.1	Materials and Methods.....	78
5.3.1.1	Shewhart Control Chart for Individual Observations .....	79
5.3.1.2	CUSUM Control Chart .....	80
5.3.1.3	EWMA Control Chart.....	82
5.3.2	Results and Discussion.....	83
6.	CONCLUSIONS .....	94
7.	FUTURE WORK.....	98
7.1	Reviewer-based Studies .....	98
7.2	Algorithm Optimization.....	98
7.3	Extensive Retrospective Analysis .....	99
7.4	Study of Clinical Implications of Uniformity Defects.....	100
7.5	Software Integration.....	100
8.	BIBLIOGRAPHY .....	101
9.	APPENDIX.....	104
9.1	Summary of Results .....	104
9.2	Results at Various Time Points from Image-space Algorithms.....	111
9.3	NEMA Uniformity Trends Using Statistical Process Control .....	149



## List of Figures

Figure 1 – Gamma Camera Image Formation Chain.....	5
Figure 2 – Flood Image of Defective PMT.....	13
Figure 3 – Flood Image of Unstable PMT.....	13
Figure 4 – Crystal Hydration Artifact.....	14
Figure 5 – Optical Decoupling Artifact.....	15
Figure 6 – Dependency of PMT response on Event Position.....	16
Figure 7 – Histogram of a Flood (CFOV).....	19
Figure 8 – Superimposed Histograms of Flood and Random Poisson Deviates.....	20
Figure 9 – 3×3 Binomial Smoothing Kernel.....	24
Figure 10 – Time Series of Spatially Registered Floods.....	27
Figure 11 – Crystal Hydration Artifact in Flood Images of Different Count Densities.....	30
Figure 12 – Example Application of the Degradation Function.....	37
Figure 13 – Experimental Setup of Attenuation Disk Phantoms.....	39
Figure 14 – Retrospective Time Series: PMT Failure.....	41
Figure 15 – Retrospective Time Series: NaI (TI) Hydration.....	42
Figure 16 – Volumetric Rendering of a Registered Time Series of Floods.....	45
Figure 17 – Graphical Representation of Time Series Floods.....	46
Figure 18 – Statistical Segmentation Demonstrating MCE.....	48
Figure 19 – Graphical Representation of Spatial Decomposition.....	50
Figure 20 – Parametric Results Using 2 Different Thresholds.....	54
Figure 21 – Locations of Uniformity Defects in Synthetic Time Series.....	58
Figure 22 – Uniformity Defect Contrast Enhancement of Image-space MDA Parametric Map...	59
Figure 23 – Regions of Non-uniformity Segmented via Multi-resolution MDA.....	61
Figure 24 – Multi-resolution MDA Results for Off-peak Time Series, Day 50.....	62
Figure 25 – Crystal Hydration Uniformity Defects Identified Using Multi-resolution MDA.....	63
Figure 26 - Uniformity Defect Contrast Enhancement of Image-space t-test Parametric Map.....	65
Figure 27 – Detailed View of t-test Map Results for Hydration Spots.....	67
Figure 28 – Multi-resolution and Single Resolution Parametric Maps for Ts_vol_T6.....	68
Figure 29 - Uniformity Defect Contrast Enhancement of Image-space t-test Parametric Map.....	69
Figure 30 – False Positive Segmentation.....	70
Figure 31 – Multi-resolution Strength-of-correlation Results for Ts_vol_off-peak.....	71
Figure 32 – Multi-resolution Results for Ts_vol_T6, Day 121.....	71
Figure 33 – Strength-of-correlation Results for Hydration Time Series.....	72

Figure 34 – Sliding Window t-test Results for Synthetic_13, Day 81.....	73
Figure 35 – Synthetic_19 Sliding Window t-test Results and Original Floods, Days 48 and 84 ..	74
Figure 36 – Multi-resolution Parametric Results for Phantom Synthetic_19, Day 46 .....	77
Figure 37 – Shewhart Control Charts for Phantom Synthetic_15.....	84
Figure 38 – CUSUM and EWMA Control Charts for Crystal Hydration Time Series .....	86
Figure 39 – Multi-resolution Strength-of-correlation at Point of Shewhart Differential Uniformity Control Violation.....	87
Figure 40 - Multi-resolution MDA at Point of EWMA Differential Uniformity Control Violation .....	88
Figure 41 – Multi-resolution Strength-of-correlation Result at the Point of EWMA Integral Uniformity Control Chart Violation.....	89
Figure 42 – Parametric Results for Attenuation Disk Images at Day 40 .....	90
Figure 43 – Parametric Result for Synthetic_19, Day 40 .....	91
Figure 44 – Comparison of Multi-resolution Maps to Shewhart Control Chart for Off-peak Time Series .....	93
Figure 45 – Parametric Results for Attenuation Disk Phantom – Day 70 .....	95

## **List of Tables**

Table 1 – Pixel Dimensions for Common Acquisition Matrices .....	8
Table 2 – MDACC Protocol for Daily Extrinsic Floods.....	18
Table 3 – Calculated Attenuation for Flashing Materials .....	38
Table 4 – Summary of Time Series Phantoms.....	43
Table 5 – Summary of the Algorithm Performance on Phantoms .....	76
Table 6 – Summary of Observations .....	110

## Table of Equations

Equation 1 – Derivation of Septal Thickness for Parallel-hole Collimator .....	6
Equation 2 – Compton Scatter Photon Energy .....	7
Equation 3 – Uniformity Correction Coefficient Calculation.....	11
Equation 4 – Sample Mean and Variance.....	19
Equation 5 – NEMA Uniformity Calculation.....	24
Equation 6 – CRSD Calculation .....	25
Equation 7 – Derivation of LSR .....	26
Equation 8 – Chi-square Test Statistic.....	27
Equation 9 – Incomplete Gamma Function .....	28
Equation 10 – Kalemis Z-score Calculation .....	29
Equation 11 – Gaussian PDF .....	29
Equation 12 – Box-Muller Transform .....	36
Equation 13 – Derivation of HVL .....	38
Equation 14 – Spatial Decomposition Levels.....	49
Equation 15 – MDA Calculations.....	51
Equation 16 – Z-test Null Hypothesis.....	52
Equation 17 – Z-test Statistic.....	52
Equation 18 – 2 Sample t-test Statistic .....	53
Equation 19 – Student-t Probability Density Function .....	53
Equation 20 – Strength of Correlation Test Statistic .....	56
Equation 21 – Moving Average Filter Process .....	57
Equation 22 – t-test Statistic for SWTT.....	57
Equation 23 – Shewhart Moving Range Calculation.....	80
Equation 24 – Shewhart Control Limit Calculations.....	80
Equation 25 – CUSUM Statistic Calculations .....	81
Equation 26 – EWMA Statistic.....	82
Equation 27 – EWMA Control Limit Calculations .....	82

## Abbreviations

2D	Two-dimensional
3D	Three-dimensional
ADC	Analog to Digital Converter
CFOV	Central Field of View
CUSUM	Cumulative Summation
DU	Differential Uniformity
EWMA	Exponentially Weighted Moving Average
FBP	Filtered Backprojection
FP	False Positive
HVL	Half Value Layer
IU	Integral Uniformity
LUT	Lookup Table
LEHR	Low Energy High Resolution
MCE	Multiple Comparison Error
MR	Multi-resolution
MDACC	MD Anderson Cancer Center
NaI(Tl)	Sodium Iodide with additional Thallium impurity
NEMA	National Electrical Manufacturers Association
P-value	Probability Value
PDF	Probability Density Function
PMT	Photomultiplier Tube
QC	Quality Control
ROI	Region of Interest
SPECT	Single Photon Emission Computed Tomography
SWTT	Sliding Window t-test
TAC	Time Activity Curve
UFOV	Useful Field of View

## 1. PROJECT HYPOTHESIS AND SPECIFIC AIMS

Daily collimated uniformity floods are a mainstay of routine clinical quality control (QC) for gamma cameras and can be used to determine the operational performance of systems prior to patient imaging. Although there is no unanimously agreed upon quantitative threshold for evaluating daily flood uniformity, Hines, et al have suggested that the most important aspect of routine quality control is the monitoring of change from a baseline condition and that the criteria to judge such change must not burden the clinic without justification (2). Historically, uniformity calculations defined by the National Electrical Manufacturers Association (NEMA) have been used concomitantly with expert, but subjective visual interpretation in order to evaluate performance status. There is room for improvement in this process. NEMA calculations do not indicate positional information about the magnitude and location of the non-uniformities present. NEMA calculations also do not inherently measure the significance of a change in uniformity over time. Furthermore, visual interpretation is subject to inter-observer variation and does not provide reproducible criteria for monitoring changes in flood uniformity. An automated analysis capable of objectively detecting significant temporal changes that correspond to spatially isolated uniformity defects would be useful in routine QC. The ability to detect subtle, progressive degradations over time, prior to the camera's uniformity value exceeding a predetermined limit, would provide the clinic with the ability to proactively, rather than reactively, remediate performance deterioration and potentially minimize scanner downtime.

**This project's hypothesis is that an automated time series analysis of daily flood images will detect a spatial non-uniformity at least 5 days prior to a single daily flood exceeding a NEMA uniformity of 5%, given a degradation rate of less than 0.1% per day.**

To test this hypothesis, the following specific aims were investigated:

1. Assemble test data by computer simulation of non-uniformities, scanning of phantoms that mimic non-uniformities, and identification of actual camera data leading up to service events for non-uniformities.

*Methodology* – Using retrospective clinical uniformity defects as a guide, methods of mimicking non-uniformities that resemble those commonly observed were developed using both mathematically modeled flood images and physically measured daily uniformity floods. For each type of non-uniformity, the size and magnitude of the degradation present was varied along with the rate at which the uniformity degrades. A pool of spatially registered, clinically acceptable floods was acquired to serve as the baseline set of a time series of floods. For each image acquired, the uniformity was evaluated using the current clinical criteria.

2. Investigate and develop quantitative pixelwise time series analysis techniques that exploit the temporal nature of these effects to detect and locate progressive degradation of the detector uniformity and to display that information in two-dimensional parametric maps.

*Methodology* – Published state-of-the-art approaches of automatically quantifying flood uniformity in a time series of uniformity floods were investigated and implemented computationally. Novel analytical techniques using spatial decomposition and time series averaging were developed. These algorithms were then applied to the phantom and retrospective time series flood data described in Specific Aim 1.

3. Apply statistical process control techniques to current conventional methods of uniformity monitoring in order to characterize how these metrics change over time in response to the mimicked degradation processes.

*Methodology* – Control chart techniques for statistically monitoring process performance were investigated and developed. The image-space techniques' performance on each phantom were qualitatively compared against global parameter control charts by identifying the images that first depicted the regions-of-non-uniformity and comparing them to the time points at which the control limits were violated.

## 2. BACKGROUND

### 2.1 Gamma Camera Basics

#### 2.1.1 *Function and Modes of Operation*

Gamma cameras are nuclear counting devices that are used clinically to create images of *in-vivo* radiopharmaceutical distributions (3). Nuclear medicine is a functional imaging modality in that the clinical images produced by gamma cameras are intrinsically linked to the underlying physiological or biochemical activity within the patients. Patients are administered a radiopharmaceutical, which is a biochemical compound labeled with a photon (gamma and/or x-ray) or positron-emitting radionuclide (3). The compound has an affinity to the constituents of a specific physiological process, and, after an uptake time has passed during which the radiopharmaceutical may concentrate in various regions of the body under the influence of the biological function being examined, images are acquired through the detection of photons emitted from the resultant (radio)activity source regions (e.g. organ, tumors). The images may help to identify and locate a disease process or tissue injury, or they may help the physician evaluate physiological performance, as in the case of assessing the left blood flow within the muscle tissue from a multi-gated cardiac blood pool study. Nuclear medicine imaging is unique among imaging modalities in that the instrumentation is used with a wide variety of radiopharmaceuticals to image anatomical function as opposed to morphology, each of which may help indicate multiple pathological disorders (e.g., myocardial perfusion and tumor detection with the radiopharmaceutical sestamibi), making gamma camera imaging an extremely versatile modality (4).

Most modern gamma cameras are used in either a planar (projection) or a tomographic mode of operation. In planar imaging, the system's detectors remain static with respect to the patient couch, producing two-dimensional (2D) images of the internal radiopharmaceutical distribution while the patient either remains in a single position, or translates between the fixed detector heads. Single photon emission computed tomography (SPECT) is a tomographic imaging technique in which the detector heads rotate around the fixed patient couch and acquire individual projection images. These are then reconstructed into a volume data set or a stack of transaxial slices using either filtered backprojection (FBP) or iterative techniques. Tomographic imaging offers the ability to view anatomy transaxially, which eliminates the superimposition of organs and other structures that is an inherent characteristic of planar imaging. Both acquisition types are common in modern usage.

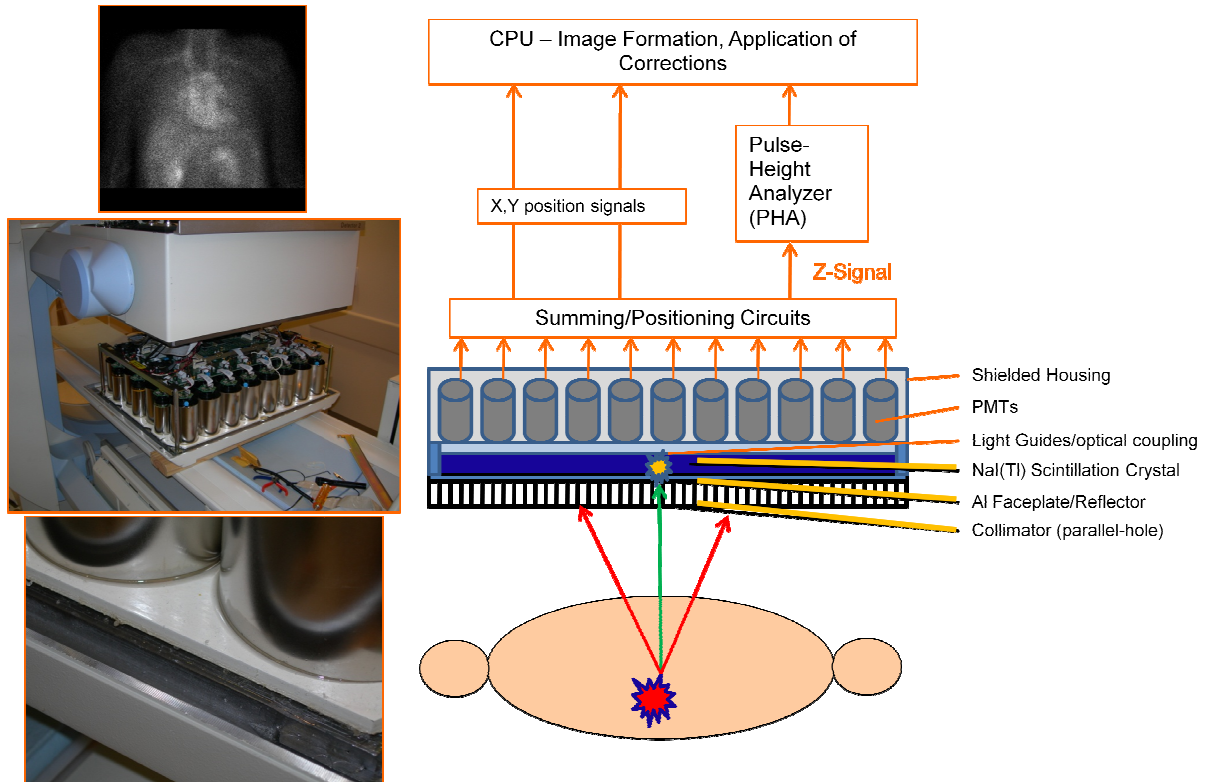
Gamma camera imaging is quantitative in many applications. Physicians are able to obtain functional, diagnostic information directly from the image pixel values derived in the acquisition. For example, the rate of clearance of a radiotracer in a target organ may be determined by creating a time-



activity curve (TAC) based upon the injected activity and the counts recorded within a region of interest (ROI) at different dynamic intervals (3). Myocardial perfusion imaging uses the amount of injected activity detected within time intervals over the cardiac cycle to quantify ejection fraction. Such quantification in imaging studies make paramount the requirement of optimal image quality, as diagnostic integrity depends on it. Properly functioning instrumentation is the foundation upon which a correct interpretation of the patient's condition is built. This requirement justifies a nuclear medicine clinic's having a robust quality control program in which QC image quality is used to assess the functionality of system components.

### **2.1.2 System Components**

Beginning at the source region of the patient in which the radiopharmaceutical has accumulated, photons of various energies from approximately 80-500 keV, depending upon the radionuclide used, are emitted. As those photons reach the detector, the collimator effectively localizes the signal by rejecting divergent photons. The accepted photons then pass through the aluminum faceplate and deposit energy within the NaI (Tl) crystal, where scintillation occurs and an amount of light that is proportional to the energy deposited by the incident photon is released. The light that reaches the photocathode of one of an array of photomultiplier tubes (PMTs) leads to a signal that is amplified to produce a signal pulse. The individual PMT signal pulses are summed in order to form the global "Z-signal" which is proportional to the energy deposited by the incident photon and hence may be used for energy discrimination. Spatial positioning is determined from the output pulses of the individual PMTs using so-called Anger logic, and image corrections are applied once the signal has been converted into a digital format (3).



**Figure 1 – Gamma Camera Image Formation Chain**

*The green arrow in the figure above represents an accepted photon, while the red arrows represent divergent photons that are rejected by the collimators. The images to the left of the diagram are of the detector/PMT interface (bottom), the PMT array extracted from the detector head (middle), and the resulting image from a planar cardiac blood pool study (top)*

### 2.1.2.1 Collimators

After leaving the patient, a photon that will ultimately contribute information to the image first reaches the collimator. The collimator is a passive device that defines the direction from which incoming photons will be accepted. Sources in the patient emit photons isotropically, meaning equally in all directions. In the absence of a collimator, photons from a point source would reach a broad area of the crystal face, producing an extremely blurred image. The collimator acts as a filter or sieve to block photons that are not incident upon the crystal in the desired direction. The most common collimator is parallel-hole, which admits only the photons that are normally (or near normally) incident on the detector. Through this *absorptive collimation*, only photons arriving essentially perpendicular to the detector will be able to reach the crystal. Only parallel-hole collimators are considered in this thesis, as they are the most commonly used collimator type and thus are the most appropriate when assessing overall system performance of gamma cameras. Parallel-hole collimators are usually fabricated by casting lead onto a preset pattern or by laying up stacks of

specifically-shaped lead foils (3). Parallel-holes are formed within the cast or layered foils, usually hexagonal in cross-section and typically 1-3 mm in diameter. The thickness of the lead septa separating these holes is dictated by the energy or energies of the desired photons to be detected (and the linear attenuation coefficient,  $\mu$ , of lead at that energy), the diameter of the holes being used,  $d$ , and the length of the septa,  $l$ . The septal thickness,  $t$ , can be derived by the following relationship (3):

$$t \geq \frac{6d/\mu}{l - \left(\frac{3}{\mu}\right)}$$

**Equation 1 – Derivation of Septal Thickness for Parallel-hole Collimator**

Generally speaking, the design of collimators (3) is an art of compromises among these physical parameters to achieve a desirable balance between sensitivity and spatial resolution for a particular energy of incident photons.

*2.1.2.2 NaI (Tl) Scintillation Crystal and its Container*

The function of the thin aluminum container that surrounds the NaI (Tl) crystal on all sides except that facing the PMTs is to protect the fragile, hygroscopic crystal from mechanical or chemical damage and moisture. The space between the aluminum and the crystal is normally filled with a highly reflective material such as TiO<sub>2</sub> in order to help increase the amount of scintillation light that can be collected from the NaI (Tl) crystal (3). The rear face of the crystal container is a glass plate with an index of refraction similar to that of NaI to allow transmission of the scintillation light, and the aluminum container and the glass face are hermetically sealed to isolate the crystal from moisture.

Photons that enter directly perpendicular to the detector, given a parallel-hole collimation scheme, and pass unattenuated through the aluminum “can”, have a high likelihood of interacting with the NaI (Tl) crystal and depositing energy there. NaI is an inorganic scintillator and when doped with trace amounts of thallium, activation centers may be formed in the crystalline lattice of the NaI (Tl). When energy is imparted to these activation centers, subsequent de-activation occurs, and energy proportional to that deposited by the incident photon is released in the form of ultraviolet light (3). Photons of the energy range useful to nuclear medicine imaging may interact with the crystal in two ways, namely photoelectric absorption and Compton scattering. In photoelectric absorption, the incoming photon transfers all of its energy to an orbital electron, thereby eliminating the photon, and ejecting the electron from its shell with energy equal to the difference between that of the incoming photon and the orbital binding energy of the electron. This type of interaction is proportional to the cube of the atomic number of the absorbing material and inversely proportional to the cube of the

incident energy of the photon, i.e., proportional to  $\frac{Z^3}{E^3}$ . In Compton scattering, the photon is not eliminated, but rather transfers only a portion of its energy to a loosely bound outer orbital electron (the recoil electron). The photon scatters at an angle  $\theta$  with a reduced energy that may be calculated from the following relationship (3, 5):

$$E_{Scatter} = \frac{E_{\gamma}}{(1 + (\frac{E_{\gamma}}{0.511})(1 - \cos(\theta)))}$$

**Equation 2 – Compton Scatter Photon Energy**

The energy of the recoil electron equals the difference in energy between the incident photon and the scattered photon. Compton scattering within the crystal degrades image quality when a scintillation event from the scattered photon of reduced energy occurs at a point somewhat removed from the site in the crystal of the original interaction. Compton scattering within the patient also degrades image quality. A divergent photon may be scattered within the patient and the resulting scattered photon of reduced energy may be perpendicular to the collimator/detector, resulting in an interaction in the detector at a distance away from the origin of the original photon emission. This phenomenon is addressed to some degree using energy discrimination, as will be discussed later. NaI(Tl) is a popular scintillator in nuclear medicine applications because it has a relatively high density ( $3.67 \text{ g cm}^{-3}$ ), and thus a high probability for a photoelectric event, and it can be manufactured relatively inexpensively into large slabs of a single continuous crystal. NaI(Tl) has an approximately 90% (for a 3/8'' crystal thickness) intrinsic detection efficiency for 140 keV photons (3) and is transparent to the wavelength of light produced (i.e., there is no self-absorption of its own scintillation light). Light emitted from the de-excitation process passes through the crystal and thence through a transparent thin glass layer coupled to the back of the crystal. On the other side of the glass are light guides or optical coupling to the PMTs.

The thickness of the crystal used for the acquisition of experimental data for this project was 3/8'' thick. This is the most common thickness for modern clinical gamma cameras. Thicker crystals may be used, and may improve the detection efficiency of the system, but spatial resolution is slightly degraded because the spherically radiating light is spread among more PMTs, causing each to gather a slightly smaller portion of the total.

Experimental data was acquired on a Siemens e.cam gamma camera (Siemens Medical Solutions, Malvern, PA). The UFOV of the Siemens e.cam detector is  $532 \times 386 \text{ mm}^2$ . The linear dimensions of a digitized image pixel vary according to the acquisition matrix size. The useful field-of-view

(UFOV) is the portion of the detector that is free from edge effects. The central field-of-view (CFOV) is defined as the central 75% of each respective (x, y) dimension of the UFOV. Common acquisition matrix sizes are listed for this system:

<b>Image Matrix</b>	<b>Linear Dimension of a Single Pixel (zoom factor = 1)</b>
<b>1024x1024</b>	0.6 mm
<b>512x512</b>	1.2 mm
<b>256x256</b>	2.4 mm
<b>128x128</b>	4.8 mm
<b>64x64</b>	9.6 mm

**Table 1 – Pixel Dimensions for Common Acquisition Matrices**

Pixel size calibration is performed by the vendor at the time of installation and should be tested semiannually (2). The pixel size (mm/pixel) is calibrated by acquiring an image of an array of point sources that are separated from each other by known distances and calculating the distance in pixels between the centroids of each pair of point sources (3).

The point-spread-function (PSF) of a gamma camera describes its spatial resolution and is measured as the full-width at half maximum (FWHM) of the image of a point source. A typical intrinsic FWHM for modern gamma cameras is ~4 mm. Pixels that are closer together than the FWHM of the PSF are correlated by virtue of the intrinsic blurring that occurs.

### 2.1.2.3 *Photomultiplier Tubes (PMTs)*

The scintillation crystal is a frequency shifter. It converts photons to visible light, which is then converted to electrical charge by photomultiplier tubes. Light from the scintillation crystal enters the PMTs, striking the photocathode, which is coated with a photoemissive substance such as Sb-Cs that ejects electrons as a result of the photoelectric effect. The number of electrons emitted in response to an incident photon is proportional to its wavelength (i.e., its energy). The PMT contains a chain (or stages) of electrodes called dynodes, and each dynode is at a potential difference of 100-150 V with respect to its neighboring prior stage dynode. The photoelectrons are drawn by electrostatic force towards the first dynode, which is at a positive potential of 200-400 V with respect to the photocathode. When incoming electrons strike a dynode, more electrons than were incident are ejected from the surface of the dynode in a process called secondary emission, thereby creating an amplification of charge. The electrons emitted from the dynode are then drawn electrostatically toward the next stage of the dynode structure, which is held at a potential 100-150 V higher than the

first stage. This continues for 9-12 stages. Each subsequent stage is maintained at a progressively higher potential and thus the number of electrons that were emitted from the photocathode is greatly multiplied, significantly amplifying the incoming optical signal. The amount of current collected at the terminal anode resulting in a voltage pulse across a reference load is proportional to the intensity of light (i.e. the number of photons) incident upon the photocathode and is proportional to the energy deposited by the photon that interacted with the crystal and produced the scintillation light. The PMT array requires a stable, high voltage power supply (3). A minor deviation in potential at any one of the stages in the dynode complex can result in a distortion of the amplitude of the resultant voltage pulse, which leads to an incorrect estimate of the deposited energy. Individual PMTs are mounted within thin cylinders of magnetic shielding material to reduce the influence of external magnetic fields (including that of the earth) upon the gain of the PMT, and hence upon the measured energy calibration.

#### *2.1.2.4 Light Guides and Optical Coupling*

The PMTs may have gaps between them. Plastic light guides (pipes) positioned between the PMTs and the glass face of the NaI(Tl) crystal assembly, and optically couple to both using optical coupling gel or adhesive, help to direct light emitted from the back of the crystal away from the gaps and towards the more sensitive regions of the PMTs (3). In the absence of light guides, the PMTs are coupled directly to the back glass surface of the crystal. This gel minimizes the reflection of light at the optical interface. Light guides can also help to minimize the non-linear effects that source positioning has on PMT response.

#### *2.1.2.5 Shielded Housing*

The entire detector assembly is contained in a shielded housing made of lead, which protects the back and sides of the detector assembly from external sources of ionizing radiation so that, ideally, the only detected events would be from photons that pass through the collimator.

#### *2.1.2.6 Summing and Positioning Circuits*

The signals from all of the PMTs are summed together to form the “Z-signal”, which is proportional in amplitude to the total light output detected from the scintillation event. An energy discrimination window may be applied directly to the Z-signal, which allows events that have the wrong energy (such as those arising from photons that underwent Compton scattering within the patient) to be rejected. An ideal situation would require a window with a width equal to 10% of the emitted photon energy (since the energy resolution of NaI(Tl) is about 10% at 140 keV, which is the photopeak of  $^{99m}\text{Tc}$ ). In practice a wider window is required to account for variations in light output

across the large NaI(Tl) crystal in the gamma camera. An undesirable consequence of the wider window is that more scattered events will also be accepted, which degrades image resolution and contrast (3). Alternatively, photopeak offsets of the regional spectra across the FOV of the detector from the Z-signal may be stored as a lookup table (LUT), and thus energy discrimination windows may vary spatially across the detector, allowing them to be set to a narrower width.

The spatial positioning of a scintillation event within the detector FOV involves what is essentially a weighted average of the response to the events from each of the PMTs. The amount of light detected by a PMT is greater when the interaction event in the crystal is near the center of the PMT, although this effect is non-linear as a function of distance from the center. In the original gamma camera design, X and Y position signals were generated by calculating the weighted response of individual voltage readings from each PMT in the detector using analog circuitry, in which the weighting terms were embodied in the values of resistors or capacitors. On modern digital detectors, a similar processing occurs digitally, once the analog signals from each PMT are digitized using analog-to-digital converters (ADCs). This positioning determination is commonly called Anger logic, named after the inventor of the gamma camera, Hal Anger (3).

## **2.2 Uniformity as a Measure of Performance**

Uniformity is a fundamental performance characteristic of all gamma cameras, and uniformity tests are perhaps the most sensitive method of monitoring change in the detector (6). Uniformity performance has significant implications for both SPECT and planar image quality (7-10) and has the ability to detect the consequences of degradation in multiple detector components, including those that arise from the source, collimators, scintillation crystal, PMTs, and electronics (11), making it an extremely comprehensive measure of performance. Uniformity acquisitions essentially test the detector's ability to create a uniform, homogeneous image signal given a spatially uniform fluence of radiation incident upon the detector FOV (7). In addition to testing uniformity performance, high-count, non-uniformity-corrected calibration floods are acquired in order to generate uniformity correction maps for the detector. The spatial response of the detector may be corrected for the spatial non-uniformities measured in the high-count flood image, and thus uniformity acquisitions serve both to test the performance of the detector and to correct its response. Uniformity tests and calibrations produce a flood field image, which may be used to visually inspect for the presence of non-uniformities. As is later discussed in this thesis, floods may be quantified in order to provide a figure-of-merit for uniformity performance.

### 2.2.1 Uniformity Calibrations

Uniformity calibrations are used to correct for spatial count differences in the image due to detector efficiency variation across the FOV, and are critical in clinical imaging because the uniformity requirements in SPECT, more so than in planar imaging, are very demanding (9, 10). As discussed in section 2.1.1, transaxial images are reconstructed from tomographically acquired planar projections, which are then backprojected across the image matrix. Each pixel maintains equal weighting in the backprojection. Therefore, if a pixel or group of pixel values has an artificially different number of counts due to some non-uniformity of system performance, it will be the same in every projection and will manifest itself as a ring in the reconstructed transaxial image (9). To correct for defects in the uniformity of the planar projections, uniformity corrections are multiplicatively applied to each pixel. Because ring artifacts are only visible when their magnitude is above the noise level in the image, their visibility is a function of the count density of each planar projection and thus of the type of study being performed. Because there are a wide variety of studies performed in nuclear medicine, it has been recommended that calibration floods used to derive uniformity correction images be acquired with a root mean squared (RMS) error of less than 1% in each pixel (7, 9, 10). This means that if a uniformity calibration flood was acquired in an image matrix consisting of  $64^2$  square pixels, a total of 30-million counts would need to be acquired in order to yield a count density of 10,000 counts per pixel (cpp). This is, of course, dependent upon the UFOV of the detector, or the usable image region, and the size of the image matrix at which corrections are applied. For example, a  $128^2$  calibration image matrix would thus require that 120-million counts be acquired in order to satisfy the 1% RMS error requirement.

The calibration factors, or coefficients, are derived by acquiring the high-count calibration flood, computing the mean pixel value within a central region of UFOV of the flood and dividing that value by each pixel value at each x, y location in the high count flood. Mathematically this operation is represented in Equation 3:

$$map_{x,y} = \frac{\overline{pv}_{Cal}}{pv_{x,y}}$$

**Equation 3 – Uniformity Correction Coefficient Calculation**

The result is a 2D matrix of uniformity correction coefficients whose values are all approximately equal to one. This calibration map is stored in a matrix size specified by the vendor and rebinned to the size of the image that is to be corrected by multiplication (either static planar or SPECT projections). Almost all clinical studies utilize uniformity corrections, as do the daily floods used for



quality control (QC) testing purposes. While uniformity corrections improve image quality by optimizing image uniformity, they can act to mask or obscure the symptoms of detector non-uniformity. This may make an argument for acquiring uniformity test floods without corrections being applied in order to sample the true response of the detector. It should be noted that the above calibration is performed for the intrinsic response of the detector as well as for the collimated response, and thus two separate calibration maps are applied to the raw image. For each collimator set used clinically, a separate collimator correction map must be generated.

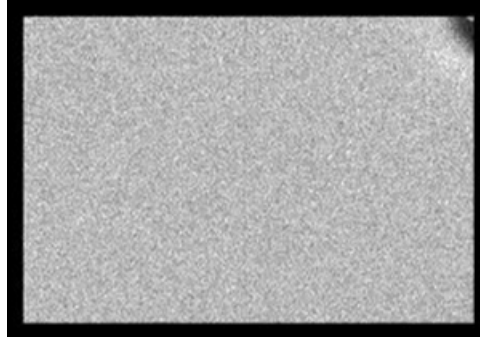
## **2.2.2      *Mechanisms of Non-Uniformity***

Detector non-uniformities are the result of two causes: 1) non-uniform detection efficiency, and 2) spatial non-linearity (3). Non-uniform detection efficiency is perhaps the most common cause of flood non-uniformity and is often the consequence of subtle differences in the pulse height signal of individual PMTs in the detector.

### **2.2.2.1      *PMT Drift and Position Dependence***

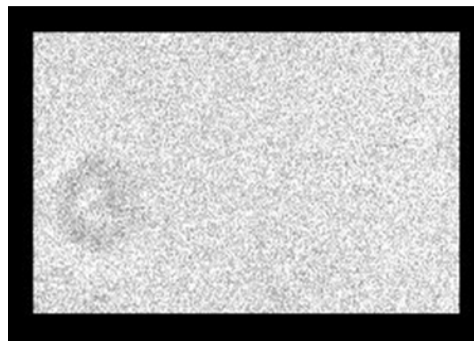
Recall that the amplitude of the Z-signal pulse is proportional to the light emitted by the scintillation event at that corresponding spatial location in the detector and is therefore proportional to the energy deposited by the incident photon. If a PMT is malfunctioning and producing a pulse height that is not proportional to the energy of the incident photon, or if the pulse height is proportional the energy but miscalibrated, events may appear to have energy outside of the selected energy window and this will cause the event to be rejected. A malfunctioning PMT will also cause a distortion in the energy detected in the surrounding area, thus affecting regions outside of the immediate vicinity of the defective PMT (6). Changes in response may be the result of drift in the detector main high voltage supply and/or the voltage gain of the individual PMTs. The preamplifier is the PMT component most likely to vary in response. Signal amplitude may also drift as the PMTs age (3). A recent study of PMTs used in positron emission tomography (PET) demonstrated an average gain drop of 11% after 100 days of operation (12). PMT response is also very temperature dependent and one author noted a two-fold increase in integral uniformity (discussed in the next chapter) over a 9° (C) increase in room temperature (6). It is thus standard practice to allow a detector to equilibrate at its operational temperature before detector calibrations or clinical scanning is performed. A common remedial action taken to correct PMT drift is detector tuning. In many modern gamma cameras, tuning is a semi-automated process that involves exposing the detector to a uniform fluence of radiation and running an iterative algorithm that inspects the count ratio of two narrow energy windows set on the high side of the photopeak (to minimize the effects of scatter) for

each individual PMT in the array. If the count ratio has changed significantly from a stored reference value, the preamplifier gain for the individual PMT is automatically adjusted in order to restore the count ratio to its proper value (3). Stable PMT gains help to ensure a uniform response to incident radiation. Some examples of how PMT drift appears in a flood are shown in Figure 2 and Figure 3.



**Figure 2 – Flood Image of Defective PMT**

*PMT in top right corner of FOV has failed. Image courtesy of WD Erwin, MS*



**Figure 3 – Flood Image of Unstable PMT**

*Drifting PMT in left portion of detector is causing artifact. Image courtesy of WD Erwin, MS*

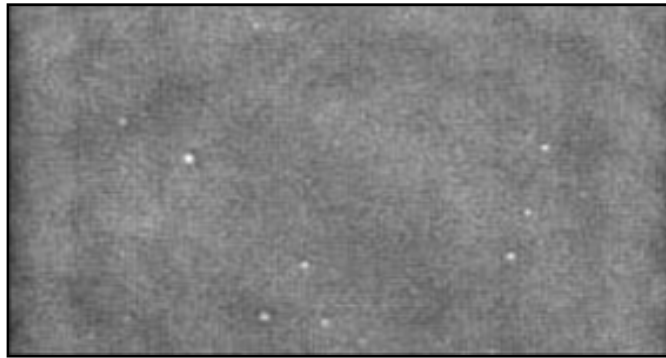
It should be noted that while PMT response may drift slowly over time, eventually to the point of severe degradation, PMTs can also fail catastrophically due to mechanical trauma or overheating. In these cases, drift may not necessarily be an indicator of PMT degradation, but rather failure can occur abruptly and unpredictably.

Non-uniform detection efficiency is also caused by spatially-dependent sensitivity differences that occur across the FOV of the detector, most notably differences in detection sensitivity between regions centered over the PMTs and regions in-between individual PMTs in the array (3). As noted above, this is remedied in part by the application of energy corrections, in which the FOV is segmented into square sub-regions and a variable scale factor that is applied to the overall Z-signal at each individual photopeak location within each sub-region of the detector. This helps to correct for

naturally occurring differences in the local pulse height spectra within the FOV (3) by accounting for the individual offsets in the locations of the photopeaks as a function of position within the FOV.

#### 2.2.2.2 *Optical Causes of Uniformity Degradation*

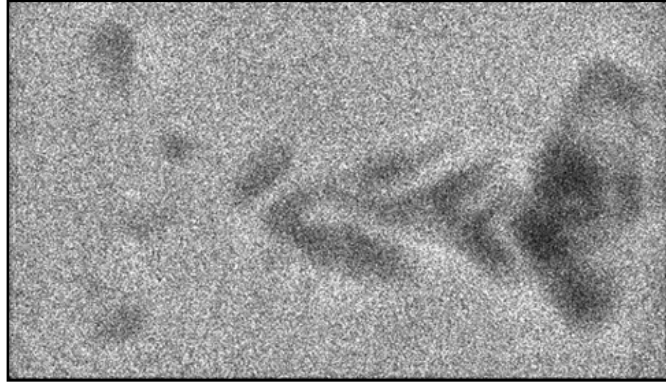
The uniformity of detection efficiency may also be degraded by changes in the optical properties of the NaI (Tl) crystal. NaI (Tl) is hygroscopic and therefore must be hermetically sealed from the ambient environment. Crystal hydration occurs when moisture infiltrates this sealed container and reacts with the NaI (Tl), causing the crystal to discolor. This leads to a loss of detected scintillation light as it is absorbed or scattered within the hydrated portion of the crystal. The loss of light will cause an artificial reduction in the height of the signal pulse from the corresponding PMTs and hence cause an apparently lower energy ascribed to the incident photon energy deposition (13). Hydration spots often appear as small circular regions within the UFOV, and are normally observed by acquiring a flood using a energy window shifted to the lower side of the photopeak. This will cause the hydration regions to appear as bright spots within a less intense background.



**Figure 4 – Crystal Hydration Artifact**

*Image courtesy of WD Erwin, MS*

Another optical artifact that rarely occurs on newer systems is optical decoupling, in which the PMTs become decoupled from the optical coupling material used to buffer the interface between the photocathodes and either the light guide or the transparent glass backing affixed to the NaI(Tl) crystal; or the glass backing decouples from the crystal. This decoupling causes a loss of scintillation light, resulting in event rejection. The cause may be a mechanical shift that results in a mismatch with the coupling substrate, or it may be the result of desiccation of the coupling material over time. The artifact may appear localized or dispersed across the UFOV. It is also possible for the optical coupling gel to discolor over time, which reduces the amount of light that reaches the PMT.



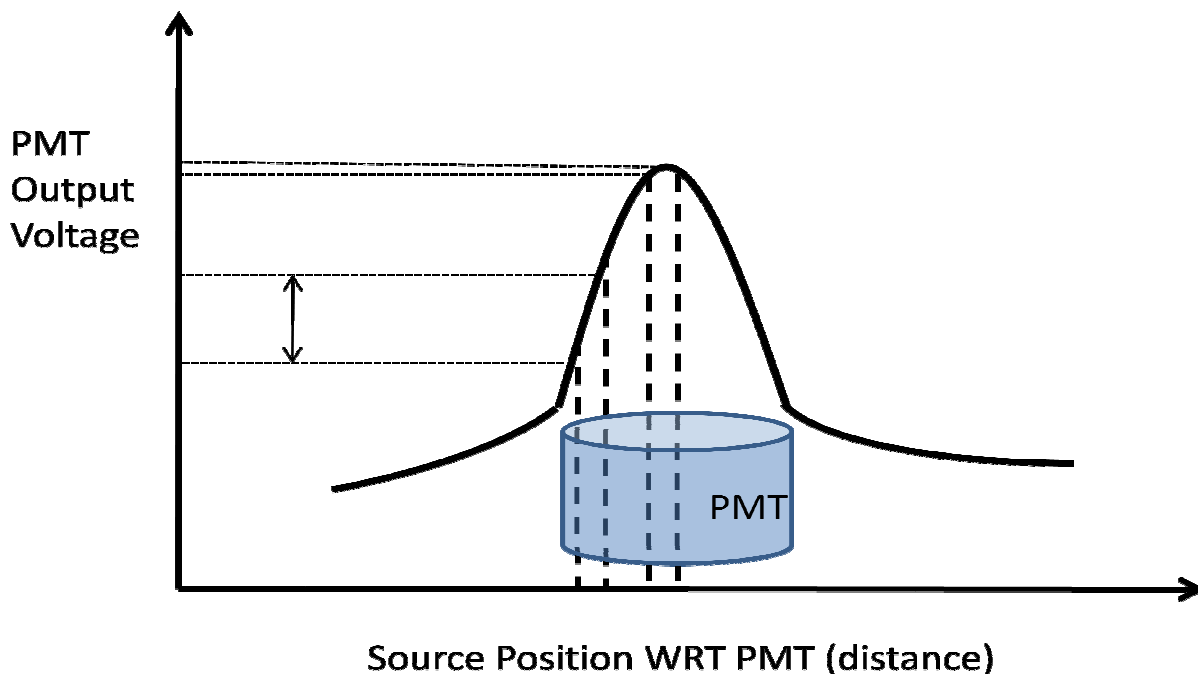
**Figure 5 – Optical Decoupling Artifact**

*Image shows decoupling of the crystal from the glass window.*

*Image courtesy of WD Erwin, MS*

### 2.2.2.3 *Non-linearity as a Cause of Non-uniformity*

The other primary cause of image non-uniformity, aside from non-uniform detection efficiency that may be caused by any one of the mechanisms described above, is spatial non-linearity in the detector response. Spatial positioning distortion results from the non-linear signal output of the PMT as a function of the location of the input scintillation event relative to the PMT. If the location of the event moves some distance  $x$  from the outer edge region of the corresponding PMT, a large change in signal results. If the source shifts the same distance within the central region of the PMT, a much smaller change in signal is observed. The net effect is that events that occur under the central region of the PMT are pulled in toward the central region because the events positions cannot be readily distinguished based on signal differences, causing a ‘hot spot’ to appear in the image in areas where the PMTs are located. This process may be illustrated as follows:



**Figure 6 – Dependency of PMT response on Event Position**

This non-linear response to light output may be mitigated with the use of physical or digital light guides, which help to make the output voltage change more linearly as a function of the distance of the event from the center of the PMT. Additionally, linearity corrections are also applied to the detector by acquiring a flood with a lead slit or orthogonal-hole pattern positioned between the source and detector. The known distances between the slits or holes in the lead are used to generate an x, y offset correction map, thus calibrating the position-dependent distortion that occurs near the PMT locations. There is an energy dependence associated with these phenomena due to the increase in the signal pulse height potentially causing a sharper voltage gradient near regions under the outer portions of the PMT. While some vendors utilize proprietary software to address this dependence, others require that uniformity corrections, described above, be acquired separately for each radionuclide that may be used clinically.

### **2.2.3 Uniformity Floods for Routine Quality Control**

Daily uniformity flood tests are the most common task performed as part of an ongoing QC program in nuclear medicine (6, 7, 14), in large part because they are relatively easy to acquire and they offer a high sensitivity in detecting performance issues for the myriad of components that comprise the detector complex. It is recommended by many experts that either intrinsic floods, which are performed without collimators applied, or extrinsic floods, acquired with collimators, be acquired

on a daily basis, prior to any patient imaging (6-8, 11, 14, 15). Healthcare accrediting bodies such as The Joint Commission and the American College of Radiology (ACR) require that uniformity floods be acquired daily for sites to meet accreditation criteria. The specific protocol parameters are less well-defined, but the general consensus is that the count density in uniformity floods should be higher than those used for patient imaging, so that any defects in the UFOV may be detected in the flood before they have become so prominent that they are visible in patient images. However, as higher count densities mean longer acquisition times, there is a compromise that must be made between sensitivity and clinical practicality, and it is important that routine QC acquisitions not introduce an unwarranted burden on the clinic (2). In a recent study, Zanzonico recommended that a total of 10-15 million-counts be acquired in the daily flood and that the flood may be acquired either intrinsically or extrinsically (14). The benefit of extrinsic acquisitions is that extra time need not be spent removing and then remounting the collimators each day. Also, a sealed  $^{57}\text{Co}$  sheet source, which does not require preparation, may be used to acquire the extrinsic flood image. Acquiring collimated floods has the additional benefit of enabling the reviewer to check for defects that may be present in the collimators themselves, which are most often a result of some physical damage.  $^{57}\text{Co}$  is often used for QC floods because it has a relatively long half-life and a photopeak similar in energy to that for  $^{99\text{m}}\text{Tc}$  (122 keV vs. 140 keV), which is the most common radionuclide used for clinical imaging. More modern gamma camera platforms utilize automated QC acquisitions in which a retractable source of either  $^{57}\text{Co}$  or  $^{153}\text{Gd}$  (100 keV) protrudes from the patient couch and sweeps across the UFOV of the detector under the control of the operating software of the camera.

The daily flood protocol at The University of Texas MD Anderson Cancer Center (MDACC), the institution at which this research was conducted, includes acquiring extrinsic floods on four days of the work week and an intrinsic flood on the fifth day. For the e.cam gamma cameras, a  $^{57}\text{Co}$  sheet source is positioned in between the two detectors of the dual-headed system, and a total of 10-million counts are acquired for each head. The Siemens Symbia systems, acquire a total of 10-million counts as well, but using an automated retractable  $^{153}\text{Gd}$  rod source that is used to generate planar images from a circular-orbit tomographic acquisition. The most common collimator used in the clinic is parallel-hole, low-energy high-resolution (LEHR), and thus most extrinsic daily floods utilize these collimators. The protocols on the two gamma camera systems are summarized in Table 2:

Parameter	Symbia	e.cam
Total Counts	10 M	10 M
Radionuclide	<sup>153</sup> Gd	<sup>57</sup> Co
Source Delivery	Automated (retractable rod)	sheet source
Image Matrix	256x256	1024x1024
Corrections	Uniformity, Energy, Linearity	Uniformity, Energy, Linearity

**Table 2 – MDACC Protocol for Daily Extrinsic Floods**

In summary, the reason that daily extrinsic floods are the focus of this thesis is that they offer a great deal of sensitivity for detecting non-uniformities, and they supply the largest sample size for time series analysis, as they are the most often performed routine clinical gamma camera QC performance test.

### 2.3 Introduction to Counting Statistics and Distribution Models

Fundamentally, daily QC floods are simply statistically independent counting experiments; each conducted using essentially the same experimental conditions (assuming that the gamma camera's operating characteristics are not changing). In a temporal sense, each flood represents a separate measurement of the same process. This assumption can be made in the spatial sense, also, in that each pixel value in a single flood is a separate measurement of the same nuclear decay process, assuming source uniformity and correct positioning of the source. In this sense, floods lend themselves readily to simple statistical analyses that have been used to test for the presence of error in radiation detection instrumentation. Because these simple analytical methods and predictive assumptions make up the core of the solutions that are presented in this thesis, the necessary statistical background is qualitatively covered in this section. All formulas specific to the analytical solutions presented later are provided in the Methods and Materials of each section.

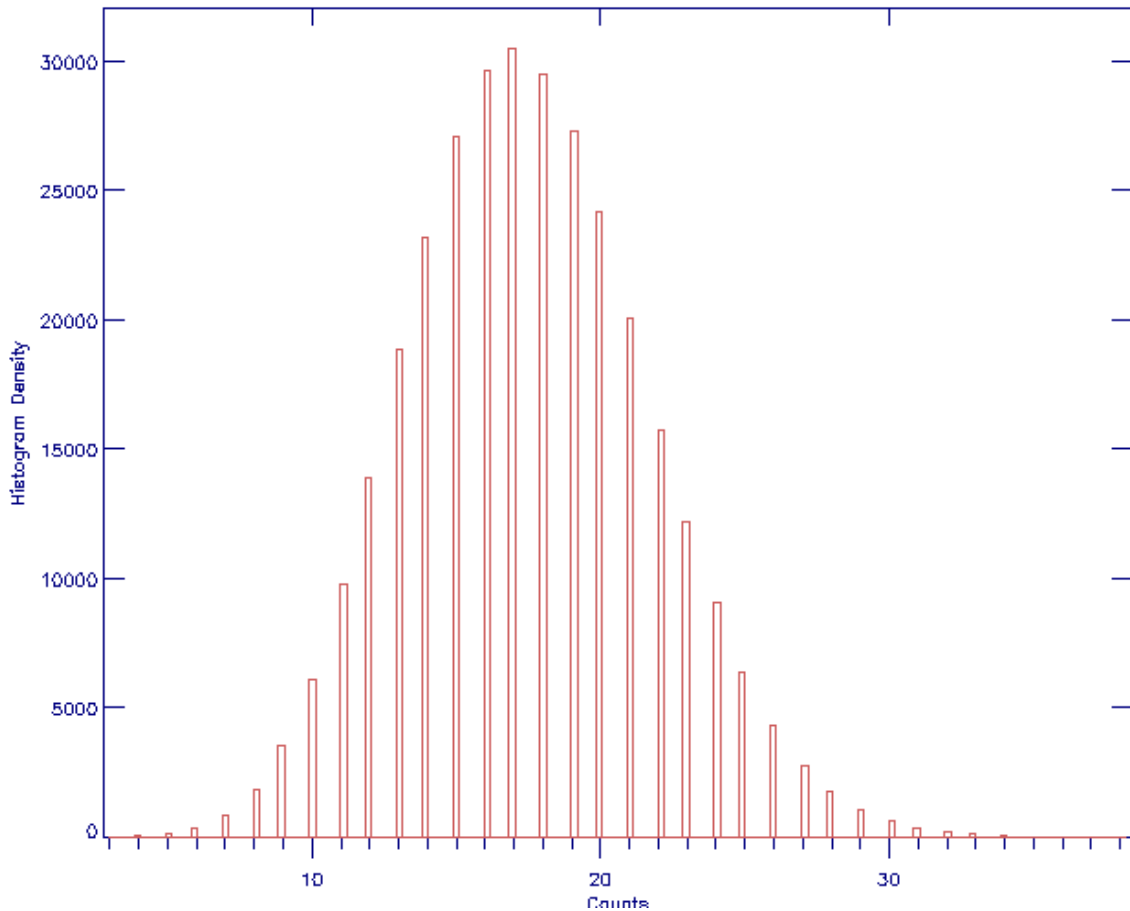
#### 2.3.1 *Characterizing Data*

Stochastic data can be characterized by a histogram, or frequency or probability distribution function (PDF), of the data set. This function is defined as the number of occurrences of each value in the collective data divided by the total number of measurements. This characterization is most useful when plotted with the ordinate being the PDF and the abscissa being the total range of values present in the data set. Two descriptive parameters that result from any data set are the location, or mean value, and the sample variance of the data. The location is typically the point of peak amplitude in the histogram and the variance is the measure of internal fluctuation or spread present within the

data (16). Because an infinite number of samples is not possible to acquire, the sample mean ( $\mu$ ) and sample variance ( $s^2$ ) of the data are estimates or predictions of the true values, and are referred to as the experimental mean and sample variance. Returning to the single flood example, these parameters are mathematically defined below, where  $pv_i$  refers to the pixel value at the  $i$ th location in the flood image and  $N$  is the number of pixels.

$$\mu = \frac{1}{N} \sum_{i=1}^N pv_i \quad s^2 = \frac{1}{N-1} \sum_{i=1}^N (pv_i - \mu)^2$$

**Equation 4 – Sample Mean and Variance**

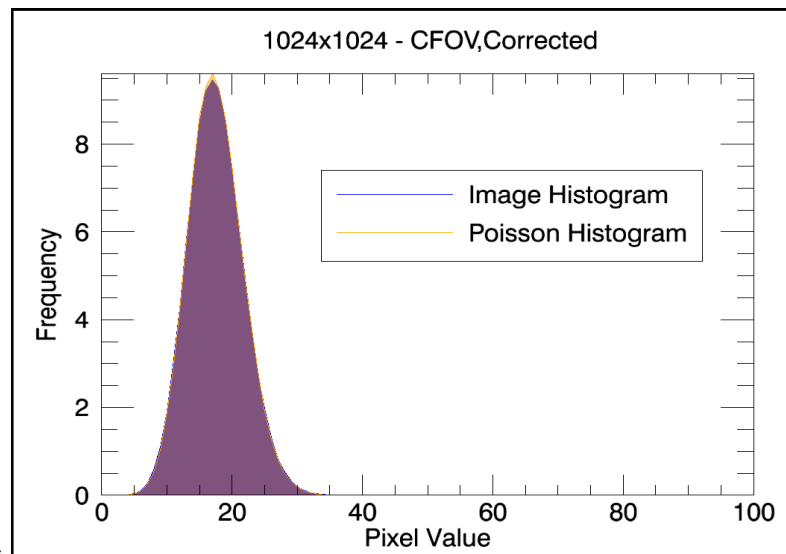


**Figure 7 – Histogram of a Flood (CFOV)**



### 2.3.2 Predictive Models

The histograms of the sample data might converge to a known distribution function, if there were a sufficient number of measurements to provide a better estimate of the mean and variance. These known distributions are models of expected behavior common in processes such as nuclear decay. The PDFs take on a well-characterized shape and the relationship between the sample parameters are well understood. They demonstrate the probability of occurrence based on the sample values present. The PDFs of most consequence in nuclear medicine are the Poisson and Gaussian distributions. The reader may refer to (17) for a more comprehensive discussion as well as the equations for the PDFs of these models. The parameters from multiple, separate sample distributions can be tested against one another to see if those distributions are significantly different or from the same statistical process. The process of nuclear decay is governed by a probability of ‘success’ (one disintegration) in a time interval  $t$  being equal to  $1-\lambda t$ , and it turns out that this is nicely modeled as a Poisson process, in which characteristically the variance is equal to the mean (17). When the mean value of the sample data approaches 20, the Poisson and Gaussian models are virtually indistinguishable, and the variance of the Gaussian process may be approximated by the mean. This relationship is described by the Central Limit Theorem (16). Graphically, the pixel distribution in a sample flood can be plotted as a histogram, and the histogram of randomly generated Poisson deviates based upon the same mean value are overlaid in the same plot:



**Figure 8 – Superimposed Histograms of Flood and Random Poisson Deviates**

It is evident from the Figure 8 that the two distributions are virtually identical because the superimposed distributions are aligned. The histogram does seem to indicate that the assumption of a

Poisson process is, in fact, a valid one because the histogram of the randomly generated Poisson deviates is almost perfectly aligned with the flood histogram, thus the PDF shapes are identical.

Considering each pixel to be a single measurement of  $N$  counts at a particular  $x, y$  location in the FOV, the uncertainty in the measurement is described by the sample standard deviation, which is the square root of the variance. The uncertainty of the measurement,  $N$ , which is the best approximation to the mean value, may be expressed using the signal-to-noise ratio (SNR), which is simply the ratio of the counts to the standard deviation (18). For example, if a pixel in a 10-million count flood has a value of 17 counts; the SNR would be given as  $\frac{17}{\sqrt{17}} = 4.12$ . If 200-million counts were acquired in the flood, the SNR would be  $\frac{17*20}{\sqrt{17*20}} = \frac{340}{\sqrt{340}} = 18.44$ . A four-fold increase is seen in the SNR of the flood pixel. By increasing the cpp, the statistical noise in the pixel value has been reduced. This has major implications in imaging, for by reducing the uncertainty in the pixel values, the measurement becomes more sensitive to true, non-random changes in the sample pixel populations.

### 2.3.3 Hypothesis Testing

Assumptions about frequency distributions can be evaluated by calculating *test statistics* from the sample parameters of single or multiple sample distributions. PDFs may be generated from random test statistics and will exhibit their own unique shape and characteristics. Examples of these distributions include the Chi-squared and Student t-distributions. Once a test statistic is calculated from a sample distribution, one may then test the likelihood or probability that the sample statistic is described by the PDF of the test distribution. This is often done by evaluating the PDF using the test statistic as a limit of integration. Integration of the PDF over the sample limits of integration will yield a *probability value* (p-value), or solution to the PDF. This p-value may be described as the probability that a *random* test statistic will take on a value as least as extreme as the *sample* test statistic calculated from the sample data. This p-value has several implications. It provides a meaningful indicator that the parameters calculated from the sample match those *predicted* by the assumed distribution model, as is the case for the Chi-squared test. In other cases, the p-values indicate whether or not the sample parameters calculated from two or more separate sample distributions are statistically different from one another. This is a special case known as *hypothesis testing*, where the user forms the null hypothesis that two different parameters, say the experimental means, are no different from one another, and the alternative hypothesis that they are different. Several different test statistics may be generated for such hypotheses, depending on which parameters are already known and which parameters must be approximated. For example, a Student t-test will be used later in this study to test the hypothesis that the means from two samples are equal, given that

the variances for each sample are assumed to be different and are unknown (and therefore must be approximated by calculating the sample variance of each). Had the variances been known to be true values and equivalent to one another, the more appropriate test would have been the two-sample Z-test (16).

The p-values calculated have several generally accepted 'critical' values, most notably values of 0.05 and 0.01. P-values below 0.05 are often referred to as *statistically significant*, and indicate that a random test statistic has less than a 5% chance of being less than or equal to the sample statistic calculated. Stated more generically, the chance that the calculated sample parameter was different by random chance is less than 5%. In this sense, p-values provide the reviewer with very meaningful indicators of the validity of the assumptions made about a data set.

### 3. REVIEW OF THE CURRENT STATE OF THE ART IN UNIFORMITY MONITORING

#### 3.1 Conventional Quantitative Uniformity Analysis

Historically, several figures-of-merit have been identified to characterize uniformity performance for a detector. These figures-of-merit can be considered *global* metrics, in that they provide a single value that characterizes uniformity over the entire FOV (either the UFOV or the CFOV). It was mentioned in section 2.2 that routine QC floods are often inspected visually for any uniformity defects, but a quantitative value that describes performance helps to reduce the subjectivity associated with observer review. For this reason, many clinics use the uniformity indices described below to evaluate performance. In practice, these values provide a quick means of objectively characterizing and assessing uniformity performance prior to patient imaging. Currently there is not an endorsed method of trending these values over time, although doing so has been previously advocated (8).

##### 3.1.1 *The NEMA Uniformity Index*

The National Electrical Manufacturers Association (NEMA) has defined a measure of uniformity for a single flood image that has gained widespread clinical acceptance in part because it is relatively straightforward and explicit in its implementation. While the original protocol specifies certain acquisition parameters and count densities that may not be practical for routine QC acquisitions, the general pre-processing steps and calculation provide a NEMA-derived means of using this figures-of-merit for daily floods (19). NEMA-derived implies that while the precise acquisition conditions may deviate from the published NEMA protocol (e.g. extrinsic as opposed to intrinsic, no lead mask used, etc.), the fundamental setup as well as the final uniformity calculations are similar.

At MDACC, flood images are acquired using 10-million counts total (14), in an acquisition matrix that depends on the gamma camera system (Table 2 – MDACC Protocol for Daily Extrinsic Floods). The flood images are then rebinned down (i.e., resized by summing together adjacent pixel groups to form larger square pixels and thus an image matrix that covers the same physical area with fewer, larger pixels) to yield a pixel size that has a linear dimension of about 6.4 mm (20). The NEMA protocol allows  $\pm 30\%$  latitude for this dimension, and in this implementation, a linear dimension of 7.8 mm is used. The reason that pixels are rebinned is to improve the counting statistics for each pixel used in the subsequent calculation. Sub-groups of pixels are effectively summed together in order to preserve the total counts within the UFOV. The percent uncertainty, or expected error, in the value of each pixel in the UFOV is reduced by increasing the number of counts per pixel

(cpp). For example, doubling the cpp from 17 to 34 reduces the percent uncertainty by a factor of the square root of two.

$$\frac{\sqrt{17}}{17} (100\%) = 24.25\% \quad \frac{\sqrt{34}}{34} (100\%) = 17.15\%$$

The rebinned image is then convolved with a 2D 3-pt × 3-pt binomial smoothing kernel defined in Figure 9. Smoothing the image further reduces the amount of uncertainty or noise in the pixel array prior to performing the NEMA uniformity calculation.

$$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

**Figure 9 – 3×3 Binomial Smoothing Kernel**

The NEMA equation is then calculated over the pixels in the flood.

$$\frac{Max-Min}{Max+Min} (100\%)$$

**Equation 5 – NEMA Uniformity Calculation**

Two types of uniformity are measured: integral and differential. Both use Equation 5. Integral uniformity is calculated by determining the maximum and minimum pixel values within the entire FOV, for both the UFOV and the CFOV. Differential uniformity, on the other hand, applies equation 5 locally to every possible row and column of five pixels within the UFOV or CFOV. The differential uniformity is the largest value of equation 5 among all of the length-5 rows and height-5 columns in the UFOV or CFOV (20). These four uniformity indices are calculated to quantitatively characterize the flood acquired in routine QC. Uniformity values in the CFOV can never be higher than values in the UFOV, as the CFOV is by definition a subset of the UFOV. Also, integral uniformity values are always equal to or larger than differential values, as integral uniformity takes all pixels within each FOV into consideration.

In 2008, Zanzonico recommended that NEMA values should not exceed 5% for floods acquired using between 10 and 15-million counts (14), and this is the standard threshold to which other approaches will be compared in this thesis.

### 3.1.2 *Corrected Relative Standard Deviation*

Another quantitative measure of flood uniformity is the Corrected Relative Standard Deviation (CRSD) (21). The flood image is analyzed as acquired using this technique. The literature does not indicate a preferred matrix size that should be used. Mathematically, the CRSD is defined as follows:

$$CRSD = \frac{\sqrt{V_t - \bar{p}\bar{v}}}{\bar{p}\bar{v}}, \text{ where } V_t = \frac{\sum_{i=0}^N (pv_i - \bar{p}\bar{v})^2}{N}$$

#### **Equation 6 – CRSD Calculation**

In the above equation, N is the total number of pixels present,  $\bar{p}\bar{v}$  is the average pixel value across the FOV, and  $pv_i$  is the pixel value at the *i*th location in the image matrix. In effect, this calculation attempts to isolate pixel fluctuations due to non-uniformity (NU) from the expected noise or uncertainty in pixel values due to the Poisson process (15). This method has been advocated over NEMA uniformity analysis by some authors because it is less sensitive to random fluctuation over time and thus is more sensitive to true shifts in detector response due to non-uniformity (15, 22, 23).

## 3.2 **Temporal Pixelwise Methods of Monitoring Uniformity in Image-space**

Kalemis, et al, were the first to propose using pixelwise image analysis to locate uniformity defects (1) and to use the resulting two-dimensional (2D) parametric map in place of conventional global figures-of-merit. Their methods include using either a single high-count baseline control image or a spatially registered time series of images as the basis for comparison. While this method has not gained wide acceptance for clinical use, it represents a novel approach to quantitatively deciphering spatio-temporal trends in flood non-uniformity, and has provided the basis for the methods developed and described here.

### 3.2.1 *Kalemis' Pixelwise Trend Analysis*

This method involves compiling and ordering a time series of individual flood images acquired over time. The floods must have been acquired with the same acquisition parameters (i.e. same total counts acquired, same sources used to acquire image, same collimator set) in order to minimize potential sources of uncertainty other than Poisson counting statistics in the analysis. The images are sorted with respect to time, and a linear model is fit to the pixel values over time at each (x, y) pixel location in the image volume (where the axes of the volume are x- and y- pixel locations and time)

(1). This method was implemented using least squares regression (LSR), where  $pv_{x,y}(t)$  is the observed pixel value at time point  $t$ :

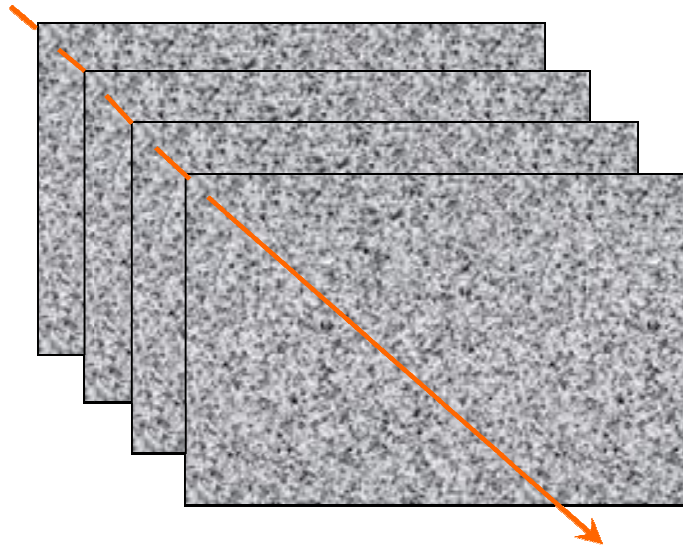
$$\begin{aligned}
 pv_{x,y}(t) &= \beta_0 + \beta_1 t + \varepsilon \\
 \varepsilon &= pv_{x,y} - \beta_0 - \beta_1 t \\
 S_r &= \sum_{i=1}^n \varepsilon^2 = \sum_{i=1}^n (pv_{x,y} - \beta_0 - \beta_1 t_i)^2
 \end{aligned}$$

Minimizing the sum of squares (24):

$$\begin{aligned}
 \frac{\partial S_r}{\partial \beta_0} &= -2 \sum pv_{x,y} - \beta_0 - \beta_1 t_i = 0 \\
 \frac{\partial S_r}{\partial \beta_1} &= -2 \sum (pv_{x,y} - \beta_0 - \beta_1 t_i) t_i = 0 \\
 \beta_1 &= \frac{n \sum pv_{x,y} t_i - \sum t_i \sum pv_{x,y}}{n \sum t_i^2 - (\sum t_i)^2} \\
 \beta_0 &= \overline{pv_{x,y}} - \beta_1 \bar{t}
 \end{aligned}$$

#### Equation 7 – Derivation of LSR

$\beta_0$  represents the value at time 0, and  $\beta_1$  represents the first order rate-of-change of the pixel intensity. Epsilon describes the random variation associated with the model, assumed to be independent and identically distributed around zero (1). Each pixel location is assumed to consist of a normally distributed set of values over time, and in a non-degraded case it is assumed that pixel values will vary only by what is predicted by the Poisson model and that the mean value will not change ( $\beta_1 = 0$ ). Under these assumptions, the intercept of the fit will be the mean of the sample distribution, and thus,  $\beta_0 = \mu$ . The orientation of the time series floods in this analysis is illustrated in Figure 10.



**Figure 10 – Time Series of Spatially Registered Floods**

The value at each pixel location is expected to be the fit parameter of the linear model (1):

Kalemis then proposed three cases in which could be rejected (1):

- 1.
2. Pixels do not follow a linear trend
3. , where is the observed variance with-respect-to time

To test the null hypothesis, Kalemis suggested generating three parametric maps conveying the results of these three cases. The reviewer could then visually assess the stability of the detector from the calculated maps. The slope of the linear model at each pixel location, , is depicted in the first map, denoted the s-map. To test the goodness-of-fit of the linear model (case 2) to the time series pixels at each location, Kalemis proposed using a Chi-squared test statistic, historically used to calculate the probability that the observed variance is consistent with the variance predicted by the assumed distribution model (see Predictive Models). The Chi-squared statistic is essentially a ratio of the sample variance to the sample mean; and because in a Poisson distribution the two are assumed to be equal, the ratio's departure from unity is a direct indication of the statistical difference between the observed and expected distributions (17). The Chi-square statistic may be calculated for  $n$  different pixel values in the time series as (17):

–

**Equation 8 – Chi-square Test Statistic**



where  $pv_i$  is the pixel value at point  $i$  in the time series of pixel values at location  $x, y$  and  $\mu_e$  is the expected pixel value calculated from the linear model. The probability that the value calculated from the linear model will generate a value of  $X^2$  or greater may be calculated directly using the incomplete gamma function (25):

$$P(a, k) \equiv \frac{\int_0^k e^{-\omega} \omega^{a-1} d\omega}{\int_0^{\infty} e^{-\omega} \omega^{a-1} d\omega}$$

### Equation 9 – Incomplete Gamma Function

where  $a = \frac{n-2}{2}$  and  $k = \frac{X^2}{2}$  (1). This probability map, or *p-map*, will give the probability that a *random* Chi-square value using the same degrees of freedom (DOF),  $n-1$ , will be greater than or equal to the one calculated. In effect, it indicates the departure of the sample variance from the sample mean, indicating how accurately a linear model is describing the time series data. In the implementation of this method represented in this work, the *p*-maps were noisy and difficult to visually interpret, and thus the Chi-squared parametric maps were interpreted directly, rather than by calculating a *p*-value.

Finally, the sample variance of the pixel values as a function of time are calculated, again using the expected value calculated from the model:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (pv_i - \mu_e)^2$$

For this thesis, the coefficient of variation,  $\sqrt{s^2}/\mu_e$ , is used, rather than a direct sample variance in order to normalize the sample variance to the mean. This modification was suggested, but not implemented, by Kalemis, et al (1).

### 3.2.2 Kalemis' Pixelwise Statistical Test

In another method proposed by Kalemis, et al, the time series approach was replaced by a pixelwise comparison between a high count and thus high signal-to-noise ratio (SNR) baseline flood and the sample, a routine daily flood (1). This method of analysis was not evaluated as part of this project due to some perceived limitations discussed in the following section, and the reader may find more information regarding this technique in (1, 26). The technique is summarized in the steps below.

1. Rebin both image matrices to a size that yields a linear pixel dimension approximately equal to the point-spread-function (PSF) of the detector.
2. Scale the sample flood to the same count density as that of the baseline flood image using its background ratio (26).

- The corresponding pixels at each  $x, y$  location in the two images, denoted  $\mu_{test}$  and  $\mu_{baseline}$  below, may be statistically compared by calculating a simple z-score test statistic, assuming a Poisson process (i.e., one in which the sample variance equals the sample mean). The values for  $f$  in the equation below are the scaling factors used for the pixel values:

$$z = \frac{\mu_{test} - \mu_{baseline}}{\sqrt{(f_{test}^2)(\mu_{test}) + (f_{baseline}^2)(\mu_{baseline})}}$$

**Equation 10 – Kalemis Z-score Calculation**

- Calculate the probability that a random variable is less than or equal to the test statistic by integrating the Gaussian PDF.

$$p(z) = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} e^{-k^2/2} dk$$

**Equation 11 – Gaussian PDF**

The p-value represents the statistical difference between the two pixels, given the null hypothesis that the scaled values do not differ.

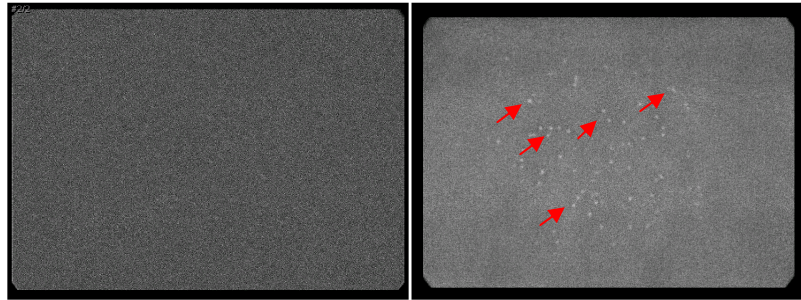
- Thresholding the resulting 2D parametric map containing all of the p-values for each  $x, y$  pixel location will isolate only those pixels that fall below the predetermined level of significance. As described in 2.3.3, common thresholds are  $p < 0.05$  and  $p < 0.01$ .
- Kalemis recommends reducing the rate of *false positives* (FPs), or pixels that fail the statistical test although they are outside of a true uniformity defect in the UFOV, by setting two different threshold values. If two separate pixels are identified below the lower threshold, then the pixels connecting the two distinct regions are tested to see if they are below the upper threshold. If they are not, only one pixel is kept in the map. If they are spatially connected via the second threshold, then both pixels are kept, and the sub-regions are joined together using image *dilation* (27). This method of FP control is referred to as *reef correction*.

## 4. CURRENT LIMITATIONS OF THE STATE OF THE ART

### 4.1 Limitations in Conventional Routine Uniformity Evaluation

As described in 2.2.3 and 3.1, conventional uniformity evaluation typically involves acquiring a daily flood at a high enough count density to detect significant changes in performance, even though such count densities may exceed those that would be acquired using patient protocols (23). This detection is accomplished by visually inspecting the flood for apparent uniformity (or other) defects by reviewing the quantitative figures-of-merit to assess uniformity performance based on a predefined upper threshold (14). The quality of this visual inspection, however, is dependent upon the expertise of the reviewer, and it is therefore a subjective analysis, especially in a setting where multiple reviewers inspect images.

The sensitivity of uniformity defect visual identification is affected by the count density in the image. Consider the images below depicting the crystal hydration artifact described in 2.2.2.2. The image on the left is of a 10-million count extrinsic daily flood and the image on the right is from the same detector using the same setup, but with 200-million counts used to acquire the image. The hydrated regions can be clearly identified in the high count, high SNR flood on the right, but not in the image on the left that uses the common, routine count density.



**Figure 11 – Crystal Hydration Artifact in Flood Images of Different Count Densities**

While there is less relative uncertainty in pixels that have more counts, acquiring those counts imposes a clinical burden, in that meaningfully higher count floods take substantially longer to acquire. This should be avoided so as not to impose an undue burden on the clinic (2).

Count density affects not only visual uniformity defect identification, but also the NEMA uniformity figures-of-merit. Floods with higher count density have lower relative uncertainty due to improved statistics, and therefore have lower NEMA values. As with visual detectability, NEMA uniformity is more sensitive at higher count densities. Several studies have been published reporting

on the effect that higher count densities have on NEMA uniformity (15, 22, 28). The consensus is that lower count densities result in more predicted uncertainty in the NEMA values due to a larger random noise component, and the uncertainty decreases with more total counts (15). Tenhunen, et al, have modeled the noise-free NEMA integral uniformity for one of their gamma cameras, and their plot of observed uniformity values as a function of count density demonstrates an obvious dependence of integral uniformity on total counts acquired, especially below 20-million counts (28). The implications that higher count densities have on NEMA uniformity values are clear, but the optimum count density to use in order to maximize visual and quantitative sensitivity to uniformity defects, while minimizing clinical burden by using the shortest necessary acquisition times is not well understood. While a 5% uniformity threshold is a reasonable limit to apply for floods acquired with between 10 and 15-million counts (14), the clinical significance of this value is not clear, and it seems that because higher count densities are expected to lower the NEMA values, there should be a lower threshold for floods with more total counts. Moreover, Young, et al, have suggested that figures-of-merit that take into account the entire distribution of pixels, rather than simply the extreme values, are more reproducible as a function of count density and less subject to random noise fluctuations (15). These metrics seem to perform similarly across the range of count densities encountered in clinical and QC acquisitions (15). For this reason, they advocate the use of the CRSD (see 1.4.2). However, it seems that by calculating a figure-of-merit over the entire distribution of pixels, rather than looking at the extrema, there is a potential loss of sensitivity to subtle uniformity defects because the extreme pixels are effectively averaged out when the sample set includes a large number of background pixels. In practice, this figure-of-merit is not used regularly in a clinical setting, and most gamma camera vendors do not provide this as an automated calculation in their QC tools, whereas NEMA calculation methods are provided on most modern gamma cameras.

This presents another limitation: quantitative evaluation of uniformity is currently threshold-driven, rather than trend-driven. Currently, routine QC is monitored in a pass or fail manner by visually inspecting for noticeable artifacts and checking to see if the NEMA values are below 5%, but most clinics do not currently track how these numbers change temporally, perhaps because few equipment vendors have included trending capabilities in their clinical QC software. Moreover, there is not a good understanding of how these metrics should be expected to change, or the expected uncertainty present in the uniformity calculation. Halama and Madsen argue that uniformity values should be trended, and that the gamma camera should be operating at the exact same level of performance as when it was first installed (11), but they do not define the appropriate method for doing this. Predictive trends may be harvested from a time series inspection of these metrics, and this may help us to determine a shift or change in performance. However, there is not a generally

accepted method for doing this using these global figures-of-merit in clinical QC. The assumption that uniformity degradation is a temporal process is valid (2, 12, 13), and therefore it is logical to monitor flood uniformity over time, i.e., in the temporal dimension of the volumes of data that are being discussed here.

Perhaps the most significant limitation in current uniformity monitoring methods is that global figures-of-merit such as NEMA and CRSD are non-specific measures of performance. Inherently, they do not provide information regarding the locations, extents, or numbers of regional non-uniformities present in the flood. This is useful information when trying to diagnose the causes and severity of uniformity defects present in the FOV. A method of automatically identifying the spatial locations of uniformity defects would be clinically beneficial.

To summarize, conventional methods of monitoring uniformity are count-density dependent, prone to inter-observer variation, non-specific in identifying uniformity defects, and lack temporal information regarding uniformity performance over time. Kalemis attempted to address some of these limitations in his methods (see 3.2), but there is substantial room for improvement.

## **4.2 Limitations in Kalemis' Pixelwise Techniques**

Kalemis' results suffer from some practical limitations. First, the maps created from his pixelwise linear models are still somewhat subjective in their interpretation, in that the uniformity defects are not segmented from the background pixel values. This makes smaller uniformity defects difficult to distinguish from the background noise. Kalemis stated that some of the maps may be difficult to interpret and that a large number of floods are needed in the time series of images in order to help distinguish true uniformity defects from the background (1).

Kalemis' second technique seemed to result in improved segmentation of uniformity defects, but he recognized that performing a pixelwise statistical test and thresholding the resulting map based on a level of significance would result in apparent defects being misidentified outside of the true uniformity defect. This phenomenon is referred to as multiple comparison error (MCE). Considering  $N$  total pixels in the parametric map, this represents  $N$  different statistical tests that were performed. If the threshold level of significance is set to  $p = 0.01$ , a false positive rate (FPR) given by  $FPR = N \times p$ , is expected. Because there are so many pixels in the parametric map, an appreciable number of them will be statistically different purely by chance. This is undesirable clinically, because it may lead a reviewer to believe that there is a performance issue when in reality the detector is operating as expected. To reduce the FPR, Kalemis recommended using a reef correction technique. Improved results were obtained with this method, but it requires a prior knowledge of the size of the uniformity defect that is expected to be detected in order to group together adjacent pixels that are below the

dual thresholds (1). A method in which no *a priori* knowledge of uniformity defect size is needed and would result in a more sensitive clinical performance. Furthermore, while Kalemis' method does use a baseline image as a basis for comparison, it is not innately temporal, in that it does not look at larger time series distributions of pixels from day to day, but instead scales single flood images up to a higher count density and compares only two images. Inspecting a continuous time series of pixel values would improve the sensitivity to detecting subtle changes.

## 5. DESCRIPTION OF SOLUTIONS

Developing image-space segmentation algorithms for spatially registered time series floods will exploit the temporal nature of detector degradation in order to detect statistically significant changes in detector uniformity from a baseline condition. These algorithms will characterize the spatial extent of the uniformity defects by decomposing the time series image volume into multiple resolution levels. The benefits of such algorithms are an automated, objective, and reproducible means of monitoring uniformity over time. In addition, these methods will provide a predictive component to uniformity monitoring, in that *observed* pixel values are tested against an *expected* statistical behavior and the results of these tests will be the identification of performance degradation prior to such degradation's being manifested in any one single QC flood. In addition to these image-space algorithms, the global figures-of-merit may be monitored using methods of time series statistical process control in order to improve their sensitivity in detecting sustained shifts in uniformity performance.

The clinical impact of these methods will be to reduce the impact that unforeseen (and thus unscheduled) scanner downtime, due to significant changes in uniformity performance, has on patient scheduling and imaging service revenue. By improving the sensitivity and spatial specificity of uniformity measurements, both physicists and clinicians are kept better informed of potential changes in scanner performance.

This is a proof-of-principle investigation into improving the sensitivity in monitoring clinical gamma camera uniformity. The goal of the project was to develop novel approaches to using time series QC flood images to detect non-uniformities that may be indicative of gamma camera performance defects and test these methods in a manner that reflects clinical feasibility. The methods were developed as an alternative to the current state-of-the-art.

In order to evaluate the proposed solutions, the following hypothesis was defined:

**Automated time series analysis of daily flood images will detect a spatial non-uniformity at least 5 days prior to a single daily flood exceeding a NEMA uniformity of 5%, given a degradation rate of less than 0.1% per day.**

### 5.1 Specific Aim 1 – Time Series Phantoms

*Develop time series phantoms that mimic progressive gamma camera image non-uniformities commonly observed in daily quality control floods, including drifting photomultiplier tubes (PMTs). Also, identify retrospective time series flood images to use as input to proposed analytical algorithms.*

### 5.1.1 *Materials and Methods*

Three different classes of time series phantoms were developed to mimic progressive degradation: synthetic (computed simulations), acquired (physically produced simulations), and retrospective (actual camera QC data). The synthetic data sets were designed to mimic daily extrinsic floods. They were created mathematically using a random number generator. A degradation function was multiplied by the uniform flood to produce the degraded images. For the acquired phantom time series, uniformity defects of increasing magnitude were introduced into the UFOV. For the retrospective time series, gamma cameras were selected that had recently exhibited non-uniform behavior and data from a series of acquisitions prior to the determination that a problem existed were retrieved from the MDACC picture archive and communications system (PACS). All physical data were acquired using the protocol employed for daily QC imaging at MDACC.

Common to all three types of phantoms was a twenty-flood baseline replicate set of floods. The term *baseline*, in this case, means that these floods were acquired (or created) at a time point in which the system was functioning properly, without any known defects. Halama and Madsen argued that the gamma camera should operate continuously as it did at installation (11), and in this phantom data set, these initial time points would provide the statistical basis of comparison. Twenty was chosen as the baseline size because the summation of twenty images acquired at the daily flood count density (10-million counts) equals the count density of a single, extrinsic calibration flood (200-million counts). In clinical practice, this set of twenty statistically independent baseline floods would be acquired one right after the other in a minimum time span, mitigating any degradation due to time. They could then be summed and used for the derivation of extrinsic uniformity calibration factors without having to acquire both the baseline replicates and a separate high-count correction flood.

#### 5.1.1.1 *Synthetic Uniformity Floods*

The Poisson process within the flood field was expected to remain stationary over time (11), therefore, each time series of synthetic floods was modeled as being independent observations of the same source decay process, and each pixel could be considered as an independent counting experiment, thus the uncertainty was independent of surrounding pixels (29). All synthetic datasets were composed of 120 total images, that is, 100 sample observations in addition to the 20 baseline replicates. The synthetic images were generated as a  $1024^2$  array of a uniform distribution of values on the interval  $[0, 1)$  and transformed into a Gaussian PDF by the Box-Muller relationship (30). As a simplified example, consider two uniform random deviates,  $U_1$  and  $U_2$ . One may transform these into independent Gaussian random variables with a mean of 0, and a standard deviation of 1 using the following relationship (30):



$$Z_0 = \sqrt{-2\ln U_1} \cos(2\pi U_2)$$

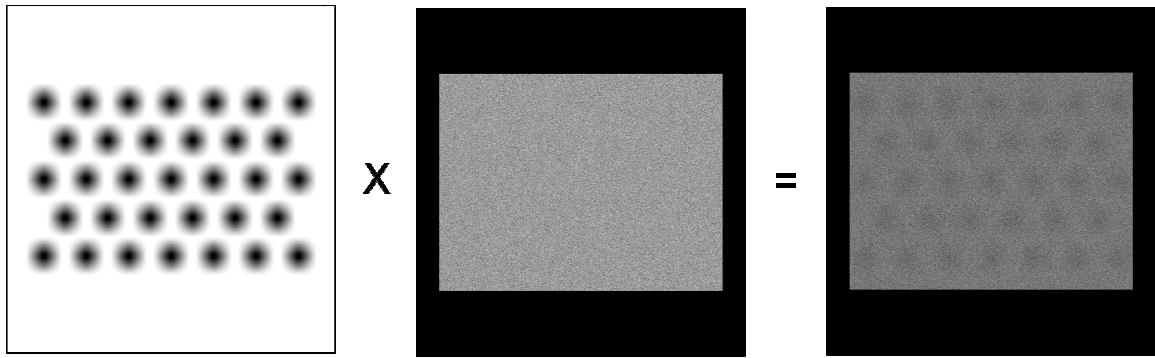
$$Z_1 = \sqrt{-2\ln U_1} \sin(2\pi U_2)$$

**Equation 12 – Box-Muller Transform**

A program was written to distribute Gaussian random deviates into a  $1024^2$  array using the Box-Muller transform. Each synthetic flood was then multiplied by a predefined mask of the UFOV of the e.cam detector (see 2.1.2.2) in order to define the characteristic rectangular UFOV. Ten-million total counts were in the rectangular UFOV. This process was performed repeatedly using different seeds in the random number generator in order to produce multiple statistically independent floods.

The first synthetic time series created was a control set of uniform flood images. The purpose in doing this was to provide a standard for the analysis programs, in order to observe the identification of false positives within the floods. This set included a 20-flood baseline set and 100 additional flood images that served as the time series.

For the next synthetic time series phantoms, a multiplicative linear degradation function was applied to each flood image after the baseline replicate set was created. The degradation function increased the magnitude of the uniformity defects as a function of time. The degradation function essentially consisted of a 2D binary mask depicting the locations of the PMTs within a Siemens e.cam detector. The diameters of the PMTs within the mask were equivalent to the center-to-center distance between the PMTs in a Siemens e.cam detector (~8 cm). In addition to this time series phantom, the diameter was reduced to approximately 7 mm to create another time series phantom with a very small region of non-uniformity. The number of regions ‘activated’ within the mask could be changed as well. Some of the phantom data contained only one 8 cm uniformity defect at a particular PMT location, while other phantom sets used a different uniformity defect of a different diameter located elsewhere in the UFOV. The magnitudes of the uniformity defects were established as a percent reduction in counts over the specified areas with respect to time. A few different rates of degradation were simulated. The maximum and minimum range of NEMA integral uniformity values is provided in Table 4 – Summary of Time Series Phantom. Prior to multiplying by the flood, the uniformity defect masks were blurred using a smoothing kernel in order to minimize the sharp edge transition between the background pixels and pixels within the uniformity defect in order to make it appear more subtle.



**Figure 12 – Example Application of the Degradation Function**

*The degraded flood on the right hand side of the equation provides an example of the maximum degradation applied, i.e., the last flood in the time series*

#### 5.1.1.2 Acquired Uniformity Floods – Altering Sensitivity via Regional Attenuation

In order to mimic non-uniformities in the detector of a clinical system, a means of introducing variable sensitivity changes with respect to background pixels in sub-regions of a time series of flood images was devised. Such changes were constrained to localized regions of the detector, as are many non-uniformities, and thus are not periodic across the UFOV (6). The artificial non-uniformities introduced into the UFOV needed to be small enough in magnitude to go virtually undetected in the results of the common NEMA analysis conducted on routine QC floods, and to have minimal visual detectability. Uniformity defects were simulated by placing attenuating material between the sheet source and the collimator. The magnitude of the degraded region was controlled by adjusting the number of attenuating layers in the stack. The size of the non-uniform region was also modified in the image by adjusting the diameter of the attenuation layers. To be able to create finely graded steps of attenuation, a thin attenuating material was needed that could be manipulated into various shapes and sizes, yet it needed to be dense enough so that the number of stacked attenuators needed to achieve certain non-uniformity levels would not introduce a significant scatter component into the floods. Localized non-uniformities were created by attenuating the incident radiation from a flood source using stacked disks of sheet aluminum. Commercial aluminum flashing is readily available and has a physical density of  $2.7 \text{ g cm}^{-3}$  (31) and a thickness of 0.027 cm. A single layer produces an attenuation that is undetectable in routine NEMA uniformity measures, while a thin stack (~0.135 cm) achieves NEMA uniformity values that exceed 3%. Aluminum provided a wider dynamic range than other attenuators, such as steel flashing or lead, and could be added incrementally in order to manipulate the flood uniformity by smaller amounts of attenuation.

The half value layer (HVL) is the thickness of attenuator needed to reduce the intensity of radiation beam intensity by 50%. It is related to the linear transmission coefficient,  $\mu$ , by the following derivation:

$$I(t) = I(0)e^{-\mu t}$$

$$\frac{1}{2} = e^{-\mu t}$$

$$\ln\left(\frac{1}{2}\right) = -\mu t$$

$$HVL = 0.693/\mu$$

**Equation 13 – Derivation of HVL**

The HVL of aluminum for broad beam geometries is 1.8 cm (31), which, according to the above relationship results in a linear attenuation coefficient of  $0.39 \text{ cm}^{-1}$ . Plugging this value into  $e^{-\mu t}$ , the percent transmission expected for each additional layer of the 0.027 cm thick Al flashing was calculated. Steel (0.4 cm HVL (31)) was also used as an investigational material and is added to the table for comparative purposes.

Steel Flashing			Al Flashing		
Thck (cm)	% Transmission	% Loss	Thck (cm)	% Transmission	% Loss
0.00	100.00	0.00	0	100.00	0.00
0.027	95.46	4.54	0.027	98.95	1.05
0.054	91.13	8.87	0.055	97.91	2.09
0.080	87.00	13.00	0.082	96.88	3.12
0.107	83.05	16.95	0.110	95.87	4.13
0.134	79.28	20.72	0.137	94.86	5.14
0.161	75.69	24.31	0.164	93.87	6.13
0.188	72.25	27.75	0.192	92.88	7.12
0.214	68.97	31.03	0.219	91.91	8.09
0.241	65.84	34.16	0.247	90.94	9.06
0.268	62.86	37.14	0.274	89.99	10.01

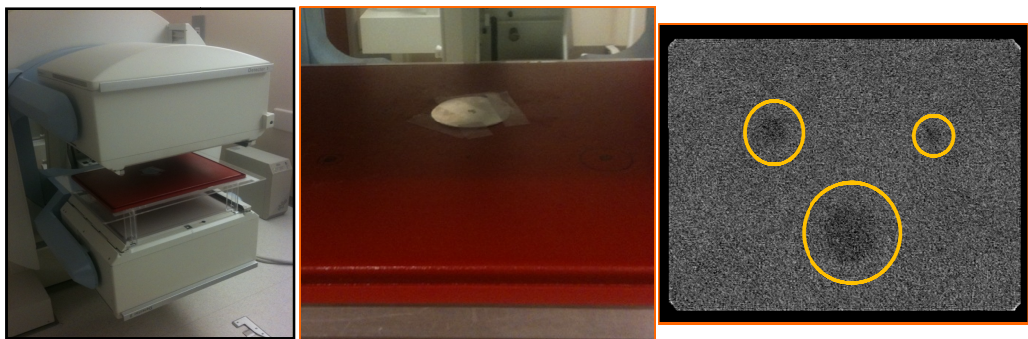
**Table 3 – Calculated Attenuation for Flashing Materials**

One layer of Al flashing produces approximately a 1% transmission loss, and thus each additional layer of flashing further reduces the beam intensity by approximately 1%. For seven layers of

flashing, the beam intensity from the sheet source was expected to be reduced to ~93% over the area of the detector overlain by the attenuation layer.

Attenuation disk phantoms (ADPs) of variable diameters were fashioned from aluminum flashing. The diameters initially chosen were 8, 4, and 2 cm, and 10 disks of each size were prepared. The disks were positioned at different locations on top of the  $^{57}\text{Co}$  sheet source (SN: BM01-15, calibration activity: 15 mCi, calibration date: 04-Feb-2009) that was used for all flood acquisitions in this project. The source was positioned at a source-to-collimator distance (SCD) of 6- $\frac{3}{4}$  inches. The reason for placing a gap between the source and detector was to blur the edges of the aluminum disks in the image. For high-resolution collimators, the collimator resolution (in FWHM) degrades from 0.2 cm at 0 cm SCD to more than 1.2 cm at 17 cm SCD (3). Thus, the resolution of the edges would be blurred by the projected aluminum disks, thereby minimizing the sensitivity of the NEMA differential uniformity calculation.

After acquiring the set of baseline, uniform images, an additional ten uniform images were acquired, and the seven sets of ten replicate images were acquired, each with an additional ADP added in the UFOV. This was repeated with gradually thicker packs of aluminum disks until a distinct upward trend was observed in the NEMA uniformity values. There were 100 floods in this time series including the twenty baseline replicates. The figure below illustrates the positioning of the source on the gamma camera and the locations of the three aluminum disks used for this experiment.



**Figure 13 – Experimental Setup of Attenuation Disk Phantoms**

*Left –  $^{57}\text{Co}$  sheet source positioned. Center – 8 cm attenuation disk placed on sheet source.*

*Right – High count flood depicted locations of 2, 4, and 8 cm attenuation disks.*

#### 5.1.1.3 *Acquired Uniformity Floods – Altering Sensitivity via Energy Window Shifts*

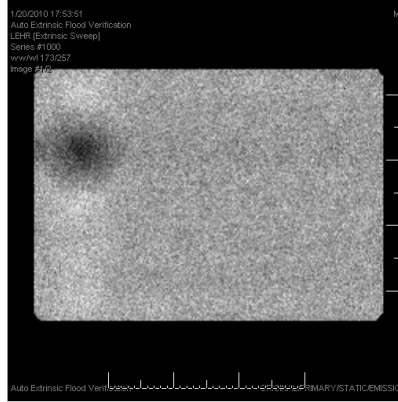
Another commonly observed clinical flood artifact is a "tubey" pattern of reduced counts centered over several of the photomultiplier tubes in the flood image. Such a pattern is often the result of shifts in the gains of individual PMTs, which distorts the local energy spectra and thus alters the sensitivity of the detector in the vicinity of the PMTs (see PMT Drift and Position Dependence). One option for inducing this uniformity defect was to adjust the voltage gain of a single PMT in order to degrade the uniformity; however, this was not feasible due to the reluctance of the camera manufacturer (who was also the vendor of the service contract) to misadjust a properly functioning system. Another option was to acquire the flood images with variable shifts in the energy window used in order to induce non-uniform detection efficiency across the field-of-view.

Off-peak images were acquired by shifting the energy window incrementally and acquiring 10 replicate floods per window shift. The time series comprised off-peak replicate sets spliced together with sets of 10 uniform images acquired from a control set of forty floods used to generate the parameters for the synthetic floods. All images were acquired in a  $1024 \times 1024$  matrix with flood corrections turned on. The off-peak shifts used were -6%, -4%, -3%, 3%, and 4% from the photopeak of the  $^{57}\text{Co}$  spectra (122 keV). The order of the time series included 20 uniform baseline replicates,  $10 \times -4\%$  shift,  $10 \times 3\%$  shift,  $10 \times -6\%$  shift, 10 more uniform control floods,  $10 \times 4\%$  shift, and finally  $10 \times -3\%$  shift, for a total of 80 floods in the time series.

#### 5.1.1.4 *Retrospective Uniformity Floods – PMT Failure*

The purpose of the retrospective flood data was to apply the algorithms (discussed later) in a more clinically relevant scenario in order to assess the feasibility of time series flood analysis. It was necessary to observe potential trends detected by the algorithms in order to verify whether or not their performance could be seen for both simulated and historical data sets. This would help to strengthen the case for using these methods clinically.

In January of 2010, a distinct artifact was observed in the daily QC flood acquired on a clinical gamma camera:

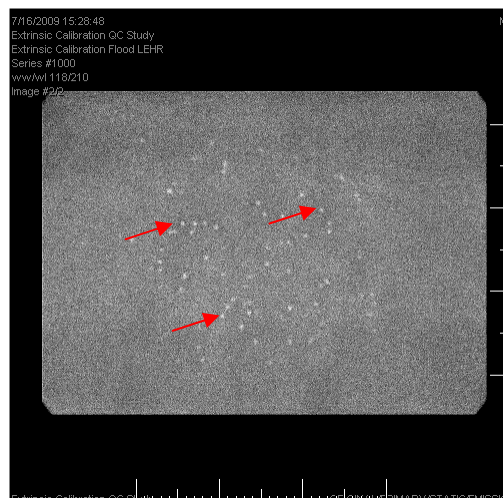


**Figure 14 – Retrospective Time Series: PMT Failure**

As can be seen in the image above, there was a significantly non-uniform region in the upper left quadrant of the UFOV. The camera was therefore deemed unusable for clinical imaging, and service personnel were dispatched in order to replace a defective photomultiplier tube. The camera was down for 3 business days while this repair was made. This clinical event presented an opportunity to use the detection algorithms developed in this thesis retrospectively on clinical data in order to determine if this type of commonly encountered performance degradation could have been identified earlier by analyzing a time series of registered floods of the same acquisition parameters acquired over time on the system than by waiting for a complete failure. As it was not clear when exactly to begin the retrospective analysis, it was decided that analyzing daily QC flood images over the course of the entire year prior to the date of this failure would be sufficient because a year's worth of flood images marks the number of routine QC images that accumulate between annual performance evaluations, when detector operation is tested for any subtle performance changes. This would in theory provide a normal baseline from which to establish the expected pixel response and from that to determine significant deviations, if there were any, leading up to the failure of the PMT. It is noteworthy that the acquisition conditions surrounding these retrospective images are different from the ones used to acquire the phantom data described above. While the stop condition of 10 million total counts was the same as for the datasets above, this gamma camera utilizes an automated QC routine in which a  $^{153}\text{Gd}$  rod source protrudes from the patient couch and is tomographically imaged and the tomographic images summed to produce a single  $256 \times 256$  static image matrix for each detector (see Table 2 – MDACC Protocol for Daily Extrinsic Floods). Thus, the equivalent of a static flood is generated from a tomographic study, and this image is assumed to depict the uniformity properties of the collimated detector rather than creating the flood by imaging a sheet source. Only images acquired with LEHR collimators were incorporated into the time series.

#### 5.1.1.5 Retrospective Uniformity Floods – NaI (TI) Hydration

The second retrospective time series was of a gamma camera detector on which NaI (TI) crystal hydration was observed in a 200-million count, high signal-to-noise ratio annual extrinsic calibration flood. The spots of hydration measured between 3 and 6 mm in diameter and were visually observed only because of their favorable contrast-to-noise ratio in the high count flood image (average contrast-to-noise ratio (CNR) = 1.84 over three regions-of-interest). The NEMA uniformity values for the flood in which the hydration spots were observed yielded results well below the clinical guideline values (CFOV – integral: 4.17%; differential: 2.02%), even without uniformity corrections applied (which is the standard practice when acquiring calibration floods).



**Figure 15 – Retrospective Time Series: NaI (TI) Hydration**

*Note: This image is a 200-million count calibration flood. This artifact was not observed in the daily QC floods. Of note are the bright, white dots that appear throughout the UFOV*

Because this artifact went unobserved in daily collimated floods, it was decided that a retrospective analysis beginning after the last extrinsic calibration flood was acquired up to the time point around the high SNR flood demonstrating hydration would be sufficient.

#### 5.1.2 Results

Each of the time series phantoms has been given a short, descriptive name in order to provide a direct reference to them throughout this thesis. Also, because the baseline images are used as the static basis of comparison, to be acquired prior to clinical use, they were not counted in the time series descriptions. Descriptions of the time series phantoms are tabulated below.

Time Series Phantom Alias	Description
<b>ts_vol_control</b>	<sup>57</sup> Co; 40×1024 <sup>2</sup> matrix; This volume was used to derive parameters for synthetic floods. Used as baseline replicates for ts_vol_offpeak; NEMA IU Range: 2.34 – 3.29%
<b>ts_vol_3disk</b>	<sup>57</sup> Co; 80×512 <sup>2</sup> matrix; corrections applied; linear count loss using 8,4,2 cm attenuation disks beginning in flood 11 up to 7 total disks; NEMA IU Range: 2.33 – 4.27%
<b>ts_vol_offpeak</b>	<sup>57</sup> Co; 60×1024 <sup>2</sup> matrix; ts_control replicates spliced with off-peak replicates (order: 10×-4%, 10×3%, 10×-6%, 10×0%, 10×4%, and 10×-4%); NEMA IU Range: 2.34 – 4.96%
<b>ts_vol_T6</b>	<sup>153</sup> Gd; 157×256 <sup>2</sup> matrix; PMT failure observed at time point 157; NEMA IU Range: 2.96 – 19.56%
<b>ts_vol_S5</b>	<sup>153</sup> Gd; 99×256 <sup>2</sup> matrix; crystal hydration observed in high SNR calibration flood; NEMA IU Range: 2.83 - 4.02%
<b>synthetic_control</b>	100×1024 <sup>2</sup> matrix; no degradation present; NEMA IU Range: 2.01 – 2.94%
<b>synthetic_13</b>	100×1024 <sup>2</sup> matrix; single 8cm uniformity defect; Degradation Rate: 0.1%; NEMU IU Range: 2.12 – 6.61%
<b>synthetic_15</b>	100×1024 <sup>2</sup> matrix; single 8cm uniformity defect; Degradation Rate: 0.2%; NEMU IU Range: 2.65 – 12.51%
<b>synthetic_17</b>	100×1024 <sup>2</sup> matrix; single 7mm uniformity defect; Degradation Rate: 0.1%; NEMU IU Range: 2.04 – 3.01%
<b>synthetic_19</b>	100×1024 <sup>2</sup> matrix; single 7mm uniformity defect; Degradation Rate: 0.2%; NEMU IU Range: 2.04 – 3.16%

**Table 4 – Summary of Time Series Phantoms**



### **5.1.3 Discussion**

These phantoms are used to demonstrate the algorithms. Different sizes of uniformity defects were sampled, along with different rates of degradation, and the data sets included both synthetic and actually acquired images to test the feasibility of the algorithms developed below.

Of note in Table 4, are the ranges of calculated NEMA uniformity values shown for the synthetic phantoms with smaller uniformity defects and for the retrospective hydration time series. The size of the uniformity defect has a direct bearing on the calculated global values, with a disparate range noted for the 8 cm uniformity defects. This is explained as a by-product of the rebinning that occurs prior to making the NEMA calculation (see The NEMA Uniformity Index). Rebinning will effectively average out smaller uniformity defects in the detector, making the calculation less sensitive to monitoring changes in a small defect in uniformity.

## **5.2 Specific Aim 2 –Time Series Image Analysis Methods**

*Develop and evaluate quantitative pixelwise time series analysis techniques that take advantage of the temporal nature of these uniformity defects and exploit them for the purposes of spatially detecting progressive degradation, and representing them in two-dimensional parametric maps.*

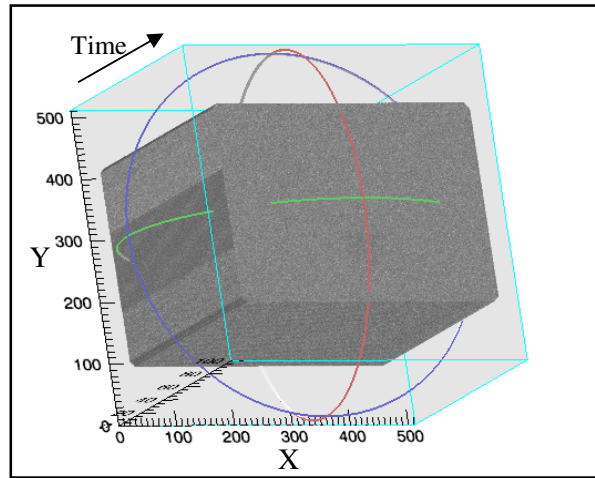
### **5.2.1 Materials and Methods**

The second and third specific aims demonstrate methods of exploiting the temporal nature of detector degradation in order to detect or isolate regions within the gamma camera UFOV where a statistically significant change from a baseline condition has occurred prior to the degradation being manifest in any single daily flood. One of them utilizes time series image-space, and the other uses time series of conventional global figures-of-merit. Image-space refers to the three dimensional domain of pixel values over time. The advantage of the image-space techniques over those of global figures-of-merit is that they provide a means of spatially isolating the uniformity defect and characterizing both its location and spatial extent.

Since the size, rate, or numbers of uniformity defects present in the UFOV are unknown *a priori*, methods were developed that took all of these factors into account. Beginning with an ordered time series volume of floods, analytical methods were developed to characterize temporal and spatial degradation simultaneously. All computational solutions were developed using the Interactive Data Language (IDL, ITT Visual Systems, Boulder, CO).

### 5.2.1.1 Temporal Analysis

Consider a three-dimensional (3D) representation of a generic time series of floods:

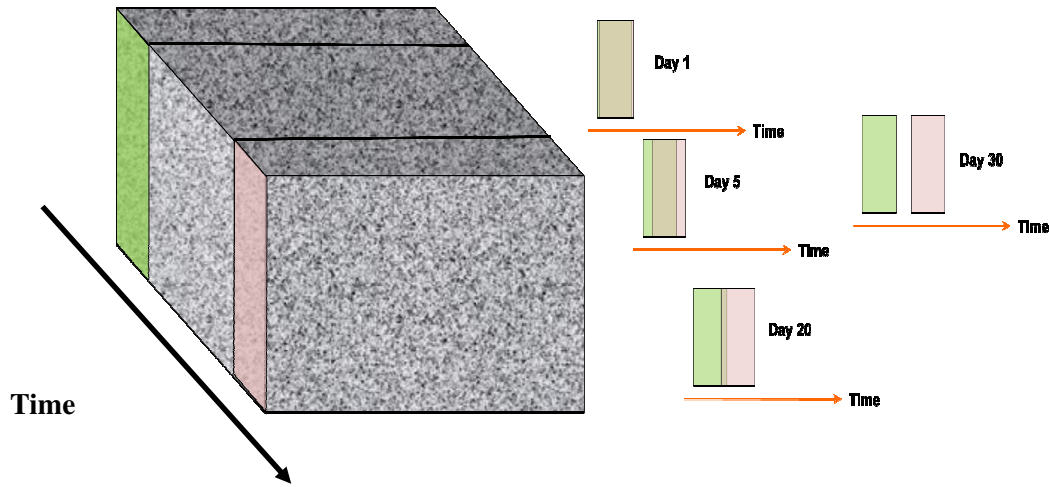


**Figure 16 – Volumetric Rendering of a Registered Time Series of Floods**

Figure 16 depicts multiple  $512^2$  floods sequentially ordered in time to form a volumetric flood space. Because the images were synthetically generated or had been acquired using the same detector without any spatial offsets applied, the UFOVs in the floods were already aligned and inherently registered spatially. The time dimension was analyzed both over the entire range of images present in the time series and over an advancing fixed window subset of images. Using the entire time range available, including the individual floods present in the baseline set, improves the characterization of the model parameters as more images are acquired in time (1) and it provides a trend over all time points available. However, while using larger sample sizes improves the confidence in the trend by improving the SNR, this approach introduces the possibility of averaging out subtle discontinuities or shifts in the underlying pixel process, thus limiting the algorithms' sensitivity to detecting shifts early on in the degradation process (32). For example, if pixels are trending consistently over time, and then a marked shift occurs in the most recent values, it may not necessarily impact the rate of change over the entire time series.

To minimize this potential effect, a method was developed to window the time series using a fixed size,  $w$ , which is the same length as the number of floods in the baseline replicate set. Windowing, in this case, implies extracting a subset of images from the time series, such as the most recent twenty images. The algorithms that utilize time-windowing identify the sub-volume consisting of  $w$  images, bin a window of  $w$  more images that advances one day forward with each new flood in the time series, and perform some statistical tests between those sub-volumes (in a fashion described

in detail later in this chapter). In other words, the algorithms identify the baseline images,  $w$  in number, and then analyze the  $w$  most recent images as new data are added to the data set.



**Figure 17 – Graphical Representation of Time Series Floods**

*Representation of Baseline sub-volume (Green) is stationary while the Time Window (Rose) advances as more floods are acquired.*

The algorithms presented in this chapter use both temporal approaches: some use temporal fixed-window binning while others use the entire time series range. This dual approach was taken to characterize any trends that may be present over the entire time series of floods as well as to maintain more sensitivity to sudden changes in flood pixel values. In either approach, incorporating the time dimension into the automatic detection of uniformity defects improved the sensitivity of the method by making a larger number of counts available with which to characterize the process as a function of time. Rather than having a single 10-million count flood from which to evaluate uniformity, there were multiple samples from the same Poisson process. An example is the improvement in SNR observed in a single pixel of a 200-million count flood over the SNR in the same pixel in a 10-million count flood (see 2.3.2).

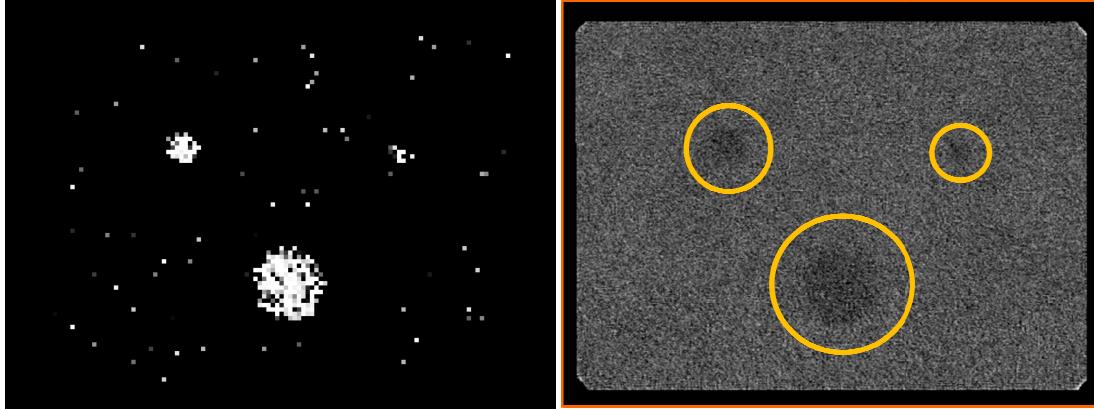
#### 5.2.1.2 Multi-resolution Spatial Decomposition

In addition to factoring in the temporal dimension, a method to spatially locate uniformity defects within the UFOV in an automated manner was needed. Image segmentation accomplished this effectively, for it is the process of dividing an image into regions with similar properties (32). Local pixel values spatially connected by common statistical characteristics are grouped together, or smoothed, to a single value, leaving pixels that do not conform as distinct, or segmented, regions

within the background. In many 2D image applications, an initial spatial window size is selected, and image pixels within this window are inspected using a statistical operator at various locations within the FOV. The limitation of this approach, however, is that it is not known *a priori* what window size to use. Choosing too large a window may result in smaller distinct regions being averaged out and passed over by the operator (32). Choosing too small a window and repeating a statistical analysis over many spatial regions in the FOV may lead to multiple comparison error (MCE) (see 4.2), which the random identification of statistically different pixels is resulting from the sheer number of times that the operator was repeated (1).

As an example of MCE, consider a parametric image derived using the phantom `ts_vol_3disk` (Figure 18). The image was analyzed at a single resolution by apportioning the UFOV into groups of every 4 adjacent pixels and performing some statistical operation (unspecified at this point) in order to identify pixels that are different from the background. Pixels that were identified as statistically different were left unaltered, and all other pixels were set to a value of zero. The non-uniform regions were identified by the accumulation of non-zero pixels in certain locations. However, some pixels that were outside of the known uniformity defects were also identified as being statistically distinct. Because the same statistical test was applied at thousands of pixel locations in the image, there was a chance that some pixels would randomly be identified as significantly different, even though they were outside of a true region of non-uniformity.

MCE could have adverse results clinically in that regions identified by the algorithm as statistically distinct may be mistakenly interpreted as defective areas of the detector. This effect is minimized by analyzing the data at different levels of resolution, in which subsequently larger and larger, or smaller and smaller, spatial regions of the image are inspected. Finding regions to be statistically distinct at multiple resolution levels adds confidence that the identification is the result of a systematic rather than random error. Coarser resolutions provide better SNR characteristics by including a larger number of pixels in the parameter estimation, while finer resolutions increase the sensitivity of this method to detecting spatially smaller uniformity defects (32).



**Figure 18 – Statistical Segmentation Demonstrating MCE**

*Left: Segmented pixels in a single resolution parametric map. Right: Figure identifying uniformity defect locations*

One method of multi-resolution spatial decomposition is quadtree analysis (33). Historically, this has been a 2D image processing technique that recursively decomposes or partitions the image into successively smaller quadrants. At each decomposition level, a statistical operator is applied, and, based on its result, each quadrant may itself be subdivided into quadrants. This process is repeated until higher resolution regions may be statistically distinguished from the background. Spatial decomposition has two flavors: 1) an all-inclusive hierarchical analysis, in which the analysis is performed at all resolution levels, and, 2) an *adaptive* decomposition, in which the decomposition may terminate or proceed to higher resolution based on the evaluation of the operator (32). It was discovered that if a small, subtle non-uniform region were contained within the FOV and decomposition began at a coarser pixel size, the region may become averaged into the background and the non-uniformity would go undetected. The all-inclusive decomposition approach was chosen in order to maintain sensitivity in detecting smaller uniformity defects.

The work reported here extends the quadtree concept to three-dimensional volumes of data that might not be square in the cross-section, as is the case of gamma cameras whose UFOVs are rectangular. The predefined time window,  $w$ , (see 5.2.1.1) is used as the constant time resolution to form a two-dimensional image in which each pixel contains the average (or total) over  $w$  time points of the corresponding pixels in the time series, and multi-resolution decomposition is then performed in two spatial directions.

At each volumetric decomposition level, the hypothesis that the regions at time point  $t$  are statistically no different from the spatially corresponding voxels in the baseline set was tested. Several statistical tests were used to evaluate the hypothesis (described later in this chapter).

Based on the result of each individual test, a p-value (see 2.3.3) is derived and then mapped to the corresponding 2D pixels of a parametric image. A predetermined threshold is then applied directly to the parametric map in order to segment regions that had changed to a statistically significant degree. These are inferred to be uniformity defects. As new floods are added to the time series, the fixed window of temporal extent  $w$  advances and the decomposition is repeated.

The multi-resolution, three-dimensional analysis is parameterized into a two-dimensional map because the detector performance for a given day is evaluated in two spatial dimensions. In one technique, a two-dimensional map is constructed by incrementing the value of a pixel by unity for each level of decomposition at which that pixel fell within a region for which the hypothesis was rejected. In other words, the pixel is incremented when it does not belong to a uniform region. For display purposes, the pixel values are then cubed in order to improve the contrast of these regions for purposes of visual interpretation. For techniques in which the associated p-values are calculated, all p-values from each resolution level are summed in a single parametric map of the same  $x, y$  dimensions as the original time series volume. The p-value map is then averaged over the number of resolution levels used, resulting in a parametric map containing the average p-value per pixel. The number of resolution levels may be calculated according to the equation below (33).

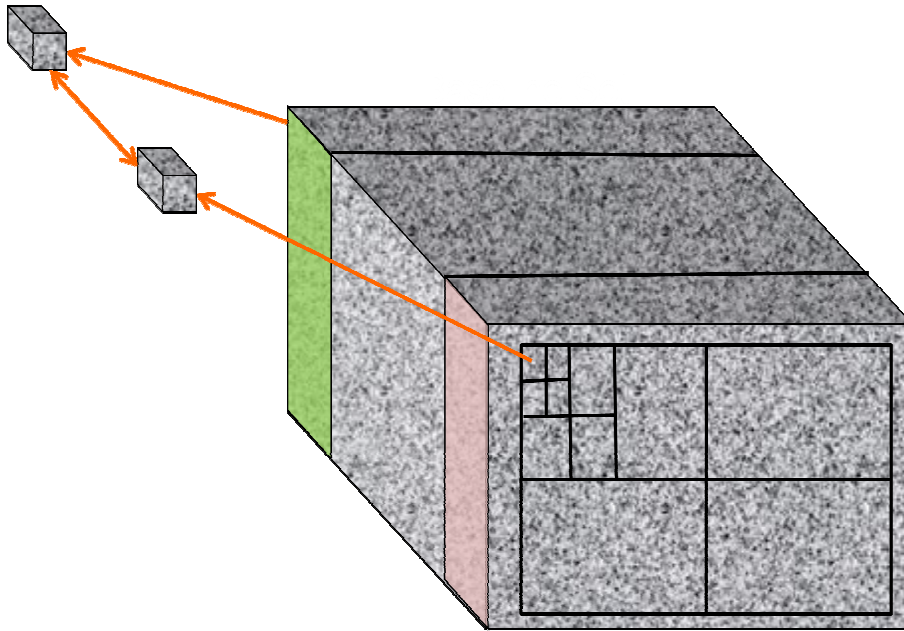
$$\text{Decomposition Levels} = \frac{\ln (\text{dims}_{img(x,y)})}{\ln (2)}$$

**Equation 14 – Spatial Decomposition Levels**

In Equation 14,  $\text{dims}_{img(x,y)}$  are the  $x, y$  dimensions of the original image. The result is that  $x, y$  pixel locations that are in the parametric map identified at multiple resolution levels will converge to lower p-values. A threshold based upon a predetermined level of significance is applied to the final p-value map to produce the final parametric map that depicts the segmented uniformity defects.

To produce the parametric map, the time series flood volume is divided into four large quadrants, or sub-volumes, using the  $(x, y, z)$  indices of the pixel values in the three-dimensional array. The independent pixel samples within each isolated quadrant are tested. Depending upon the statistical test used, either the entire time range (all  $z$  values at each  $x, y$  pixel location) or a select window size retrospective to the current flood is analyzed. Next, the algorithm recursively repeats this analysis for successively smaller sub-volumes. The results of each analytical step are mapped to a 2D parametric image. A new parametric map is produced each time a daily flood is added to the original time series volume. The statistical tests may utilize the stationary, windowed baseline volume of floods or the entire time range of floods present in the volume, including the baseline replicate sets. The advantage

of performing multi-resolution decomposition is that it characterizes the *spatial extent* of non-uniform regions by examining successively larger and larger, or smaller and smaller regions of the UFOV.



**Figure 19 – Graphical Representation of Spatial Decomposition**

*The green portion of the volume represents the baseline set of images, while the rose portion represents the time series sample that includes the most recent daily flood image*

Four distinct algorithms were developed based upon different statistical tests, each testing a different parameter, including total counts, mean counts, and count correlation with time. The first two algorithms use the fixed time window method, and the other two analyze the entire time series of floods in a spatially multi-resolution manner. The tests are evaluated at each resolution level as a function of time, either comparatively with the baseline replicates or over the entire time series. The resulting 2D parametric map of the results indicates spatial changes in pixel values from a baseline time to the current time point. On the following day, the time window advances and the spatial decomposition is repeated.

#### 5.2.1.3 Multi-resolution Minimum Detectable Activity

The first statistical test used was the minimum detectable activity (MDA), which is based on the signal difference between two count distributions (17), the counts from the sample and those from the background or baseline.

$N_{\text{Sample}}$  represents the total counts in each segmented region of the current sample sub-volume, and  $N_{\text{Baseline}}$  is the total counts in the spatially corresponding sub-volume of the baseline replicate set of floods. The variances of independent random processes add in quadrature, therefore (17):

$$\sigma_{\text{Signal}}^2 = \sigma_{\text{Sample}}^2 + \sigma_{\text{Baseline}}^2$$

When a gamma camera detector is functioning properly, there should be no significant difference in counts between the sample sub-volume and the baseline sub-volume at each resolution level (11), and therefore the error terms should be equal (17).

$$\sigma_{\text{Signal}} = \sqrt{2}\sigma_{\text{Baseline}}$$

The critical limit for detecting a change in the detector performance is set to the  $3\sigma$  level (17):

$$L_C = 3\sqrt{2}\sigma_{\text{Baseline}}$$

#### **Equation 15 – MDA Calculations**

If the difference between the two sample counts is greater than this limit, the corresponding 2D location in the parametric map is increased by one. This process is repeated for all resolution levels and the pixel values of the final map are cubed in order to make the uniformity defects appear more prominently above the background. Pixels that are randomly identified by the algorithm at a resolution level due to multiple comparison error are suppressed within the background when pixels identified at more than one resolution level are cubed in the final map.

##### *5.2.1.4 Multi-resolution Two Sample Z and t-tests*

The following statistical techniques test for the difference in means between the baseline and sample image sub-volumes over all resolution levels. Using the same spatial decomposition scheme employed in the MDA algorithm, the means and variances in each decomposed sub-volume are calculated. A test statistic is calculated between the sub-volumes and a p-value is derived. This method allows the final parametric map to be segmented based on statistical significance or the probability that the difference in samples occurred by random chance. Setting an adjustable predetermined p-value limit allows greater sensitivity in detecting regions-of-non-uniformity while rejecting false positives, because the pixel values in the map reflect an actual probability of the



presence of systematic error. The actual statistical test is of the hypothesis that there is no difference in the means between each segmented, current sub-volume of pixels and its baseline counterpart, and thus:

$$H_0: \mu_{sample} - \mu_{baseline} = 0$$

**Equation 16 – Z-test Null Hypothesis**

A test statistic is then generated at each resolution level (15):

$$Z_0 = \frac{\bar{x}_{sample} - \bar{x}_{baseline}}{\sqrt{\frac{S_{sample}^2}{n_{sample}} + \frac{S_{baseline}^2}{n_{baseline}}}}$$

**Equation 17 – Z-test Statistic**

that has a standard normal distribution ( $N(0,1)$ ), (16). This general inference test typically assumes that the variances are known, but in this case, the unknown variance of each sample is estimated by the sample variance. The analysis also assumes a fixed significance level,  $\alpha$ , and rejects the test statistic on the basis of the critical statistic drawn from the standard normal distribution at that significance level (16):

$$Z_0 < -Z_{\alpha/2} \text{ OR } Z_0 > Z_{\alpha/2}$$

This is a two-sample, two-sided Z-test. Although this technique works effectively, it was supplemented with the Student t-test statistic, which does not assume that the variances in each sub-volume are known exactly, but are instead approximated by calculating the sample variance. Because the variances are not assumed to be known exactly, this is a more statistically appropriate test.

A two-sided Student t-test is used to test whether or not the mean pixel values of the sample sub-volume and the baseline sub-volume are equal. The ‘two-sided’ test examines only whether or not the means are different, rather than which is the larger of the two. The number of degrees of freedom (DOF) is the number of pixels in each sub-volume minus one. The calculated test statistic (see 2.3.3) follows a Student-t distribution with  $n_{sample} + n_{baseline} - 2$  degrees of freedom, and is calculated by (16):

$$t = \frac{|\bar{x}_{sample} - \bar{x}_{baseline}|}{\sqrt{\frac{2(s_{sample}^2 + s_{baseline}^2)}{n_{baseline} + n_{sample} - 2}}}$$

**Equation 18 – 2 Sample t-test Statistic**

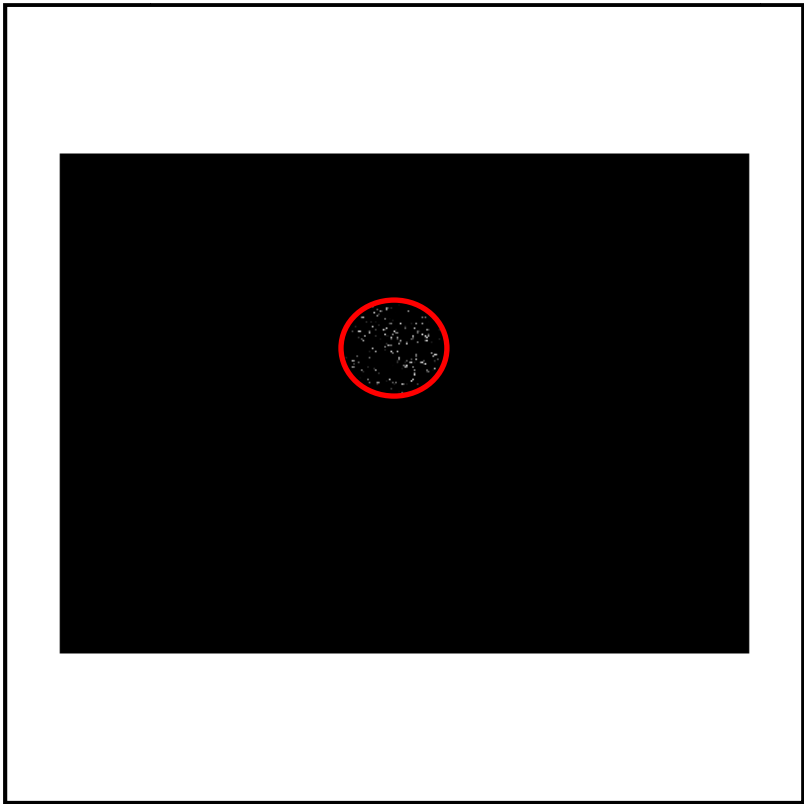
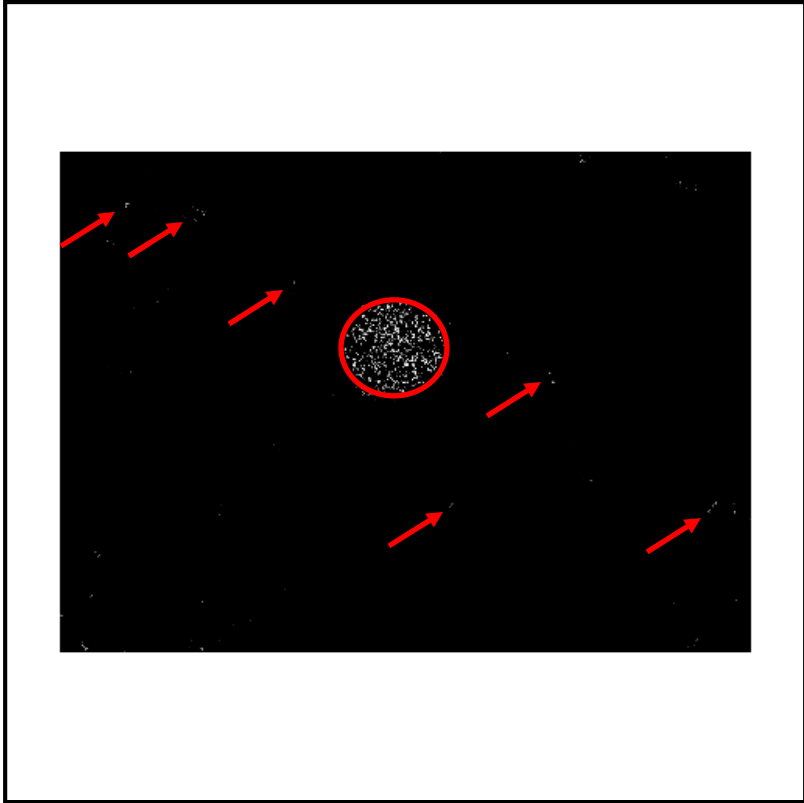
The Student-t distribution, from which the p-values are calculated, is based upon the incomplete gamma function. The p-value is calculated using the following equation:

$$p(t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})} \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}$$

**Equation 19 – Student-t Probability Density Function**

The p-values are summed over all resolution levels into a single parametric map and then an average is calculated over all of the resolution levels used in the time series decomposition. Non-uniformities are segmented in the final 2D map at time t by applying a threshold to the p-values.

Several commonly used levels of significance were investigated as thresholds, including p-values of 0.05 and 0.01. The algorithms were repeated for several different time series phantoms using different threshold levels. By applying larger p-values as thresholds, more pixels outside of the true uniformity defects were identified as being statistically distinct, and thus higher numbers of false positives were observed. For example, the parametric map at day 85 for a generic time series phantom with an 8 cm region of non-uniformity in the center of the CFOV was segmented using a threshold of 0.05 and 0.01 respectively. The difference in the number of false positive pixels is demonstrated in the following images:



**Figure 20 – Parametric Results Using 2 Different Thresholds**

*Top: Segmentation at day 85 using  $p < 0.05$ . Region of non-uniformity outlined in red. False positive regions noted. Bottom: Segmentation at day 85 using  $p < 0.01$ . No false-positive regions observed. Fewer pixels within true region segmented.*

Fewer pixels within the true region of non-uniformity were segmented in the map using the lower threshold value. This demonstrates the tradeoff that exists between the sensitivity of the algorithms to detecting uniformity defects and their ability to suppress false positive regions. By using a threshold of  $p < 0.01$ , it was discovered that the true uniformity defects were segmented in the parametric image without resulting in a large number of false positives at each resolution level. The choice of critical values helped to suppress false positive pixel values from being identified by the spatial decomposition analysis. It was important to minimize false positives present in the parametric images in order to ensure that the gamma camera does not receive service (2) which may result in unnecessary scanner downtime. Results of the algorithms using a threshold of  $p < 0.01$  are demonstrated in sections 5.2.2.2 and 5.2.2.3. Although the sensitivity to detecting uniformity defects is increased using a higher threshold, false positive suppression was determined to be important for reducing the potential for a burden to the clinic, and therefore a threshold of  $p < 0.01$  was used for the Z- and t-test techniques.

#### 5.2.1.5 *Multi-resolution Strength-of-correlation*

In order to test for trends over the entire time series volume, the multi-resolution technique was combined with a method to fit a pixelwise linear model at each x, y pixel location to evaluate the strength-of-correlation between pixel values over time. Unlike the multi-resolution MDA and t-test algorithms, which use the pixel indices at the original resolution to recursively isolate sub-volumes of the time series array and examine these volumes as independent samples, the multi-resolution strength-of-correlation approach recursively combines the x- and y- dimensions into coarser pixel sizes by summing together a subset of adjacent pixels to form a single larger pixel while preserving the original count density. As with the previous spatial decomposition techniques, the larger pixels improve the statistics at the cost of lower spatial resolution. All resolution levels are evaluated and combined into a single 2D parametric map. The entire time dimension, including the baseline replicate set of floods, is analyzed. Beginning with the coarsest resolution, a linear model is fit at each x, y pixel location over all time in the volume, similar to the method suggested by Kalemis (1). From Equation 7 – Derivation of LSR,

$$\beta_0 = \overline{pv_{x,y}} - \beta_1 \bar{t}$$

To test the correlation of pixels at a given resolution with time, the error of  $\overline{pv_{x,y}}$  given  $\bar{t}$  is calculated.

$$s_{\overline{pv_{x,y}t}}^2 = \frac{n-1}{n-2} (s_{\overline{pv_{x,y}}}^2 - \beta_1^2 s_{\bar{t}}^2)$$

The variable n is the number of observations within each fit. Next, the estimates of the uncertainties in the two model parameters are made (3).

$$s_{\beta_1} = \frac{s_{\overline{pv_{x,y}t}}}{[s_{\bar{t}}\sqrt{n-1}]}$$

$$s_{\beta_0} = s_{\overline{pv_{x,y}t}} \sqrt{\frac{1}{n} + \frac{\bar{t}^2}{(n-1)s_{\bar{t}}^2}}$$

The error in the slope parameter is used to calculate a test statistic and evaluate the hypothesis that the slope is equal to zero. This test statistic is then compared to critical values of the t-distribution (3).

$$t = \frac{\beta_1}{s_{\beta_1}}$$

### Equation 20 – Strength of Correlation Test Statistic

As in 5.2.1.4, a probability value is derived by using the Student-t probability distribution function. As with the Z- and t-test multi-resolution techniques, these p-values are summed into a final parametric map, which is then averaged over the number of resolution levels. A final threshold of  $p < 0.01$  is used. Pixels with values above this threshold are smoothed to the same value. If a non-uniformity is present and is a function of time, the map indicates the trend in pixel values at each resolution by evaluating whether or not the slope of the fit is statistically different from zero.

#### 5.2.1.6 Sliding Window t-test (SWTT)

In the final method of time series analysis that was investigated, a smoothing function is applied to the time series volume at each x, y location with respect to time, and the flood image that is calculated from the temporal moving average is convolved with a spatial smoothing kernel, yielding

images of the local mean and variance. A t-statistic image is then calculated from the local mean and variance images and a pixelwise t-test is performed between the sample and a similar image derived directly from the baseline set. This method worked effectively for floods acquired with resolutions of  $512^2$  and higher, but convolving coarser matrices blurred out smaller uniformity defects, reducing its sensitivity when the floods were less than  $512^2$  resolution. Thus, only image matrices greater than  $256^2$  were analyzed using this technique.

As with the other techniques, the 2D floods are sequentially ordered into a time series volume. A moving average filter is applied in time at each x, y pixel location. The size of the window w is the same as the numbers of floods in the baseline replicate set.

$$MA(t) = \frac{img_{1,x,y} + img_{2,x,y} + \dots + img_{t,x,y}}{w}$$

**Equation 21 – Moving Average Filter Process**

The MA window shifts forward in time with every new flood observation. After the time series volume is smoothed, the most recent image in the sequential set of floods is convolved with a 2D spatial smoothing kernel (linear dimension 3-5 pixels for a  $512^2$  flood) that calculates the local mean for every pixel in the image on which the kernel is centered. The result of the convolution is a smoothed image of the local mean values across the field-of-view. Similarly, a local variance image is calculated. This procedure is repeated using a time-averaged image of the baseline replicate set and the resulting images of the local means, and variances for both the sample and baseline are used to calculate a t-statistic image.

$$t = \frac{|\bar{x}_{local\ sample} - \bar{x}_{local\ baseline}|}{\sqrt{\frac{2(s_{local\ sample}^2 + s_{local\ baseline}^2)}{n_{local\ window} + n_{local\ window} - 2}}}$$

**Equation 22 – t-test Statistic for SWTT**

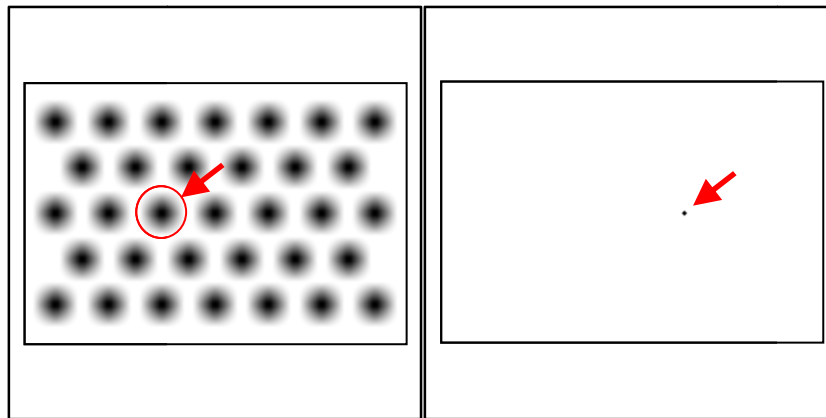
The t-statistic image is then used to calculate a p-value at every x, y spatial pixel location, and the resulting probability map is used to segment any non-uniformity by applying a threshold. Because of the zero-padded region of pixels in the flood image outside of the actual UFOV, the convolution operation causes blurring along the edges of the UFOV. A binary mask of the CFOV is multiplied by the resulting image in order to mask these edges. Therefore, the parametric map produced for this algorithm only depicts the CFOV.

### 5.2.2 Results and Discussion

The time series phantoms presented in 5.1 were processed using the algorithms described above. Each data set was analyzed over its entire volume, resulting in parametric maps for each day that the time window incremented. The results for each phantom at sequential time points are provided in the Appendix in order to demonstrate the detectability of the non-uniformities over time. Descriptions of the phantom characteristics are in Table 4. Clinically, these algorithms are intended to be run once daily in order to calculate the parametric maps from the entire time series volume up to and including the current day's QC flood. The reviewer would then evaluate uniformity performance using the calculated maps, which provide a temporal component to flood monitoring.

Kalemis' pixelwise trend analysis was also applied to these data sets for comparison. The results of each algorithm for all of the phantoms at incremental time points have been provided in the Appendix in order to demonstrate the performance of the methods as the non-uniformities degrade over time.

For synthetic phantoms 13, 15, 17, and 19, a spatial reference for the locations of the non-uniform regions has been noted in the images below.



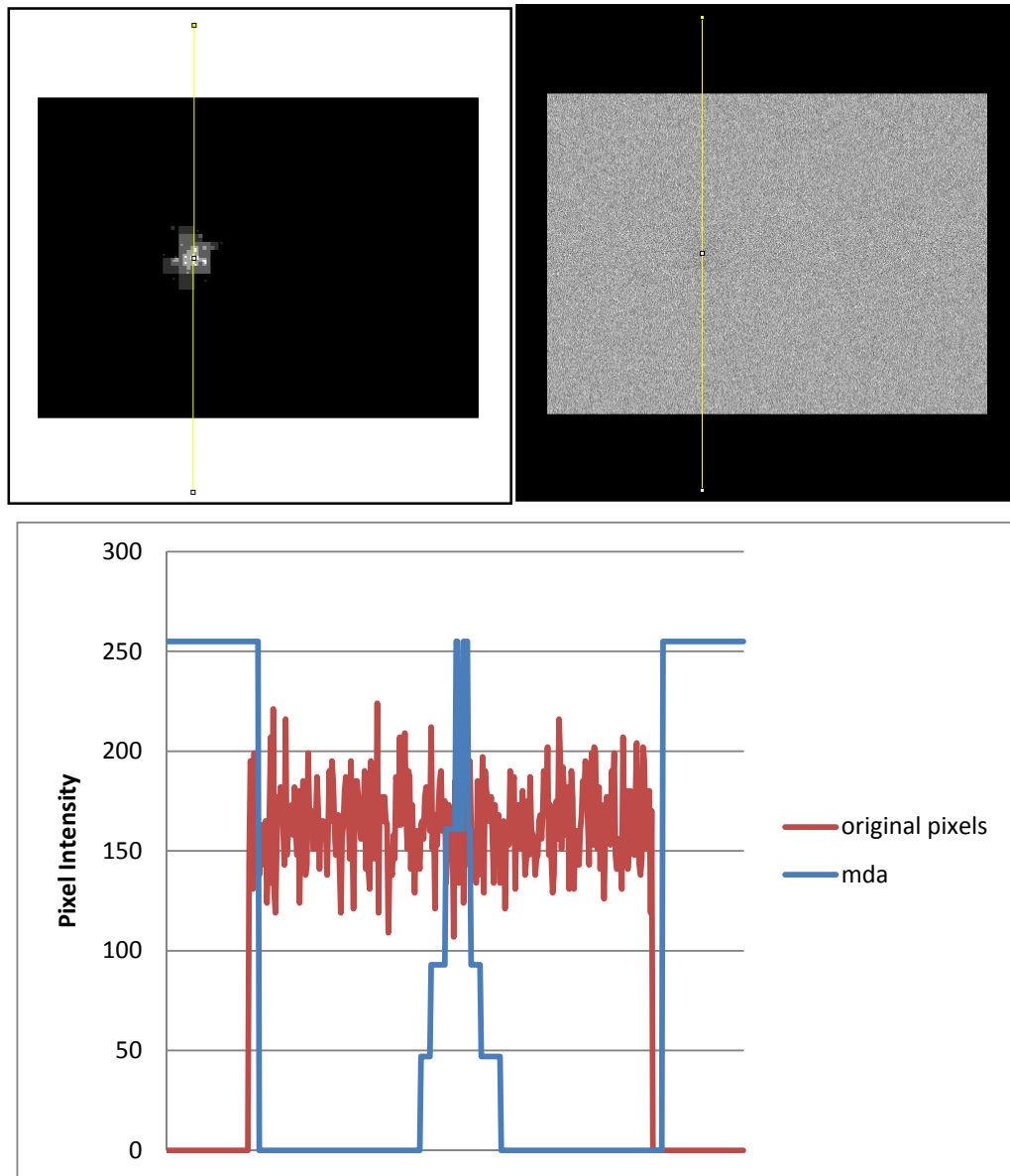
**Figure 21 – Locations of Uniformity Defects in Synthetic Time Series**

*Left: synthetic\_13 and synthetic\_15 (8cm uniformity defect); Right: synthetic\_17 and synthetic\_19 (7mm uniformity defect)*

#### 5.2.2.1 Multi-resolution Minimum Detectable Activity (MDA)

The multi-resolution MDA algorithm segmented the uniformity defects early in the time series for phantoms containing larger defects. This identification of the uniformity defects occurred prior to any single flood image's exceeding the NEMA 5% threshold. Synthetic\_13, which contained an 8 cm defect within the CFOV, exhibited integral and differential NEMA uniformity values below the 5% threshold at time point 40 (4.24% and 2.66%, respectively), and the MDA algorithm segmented the

region well in advance of this time (beginning at day 13 in the time series). The NEMA threshold was crossed repeatedly beginning at approximately day 45 for phantom synthetic\_13. Line profiles were drawn through the non-uniform region in order to demonstrate the contrast gains from the original raw flood image. The region of non-uniformity in the MDA parametric map exhibits marked differentiation from the surrounding background pixels in the UFOV. The profile plot for the original flood at time point 40 showed a noisy distribution of pixel values and the region of non-uniformity were not clearly distinguished from adjacent, uniform pixels.

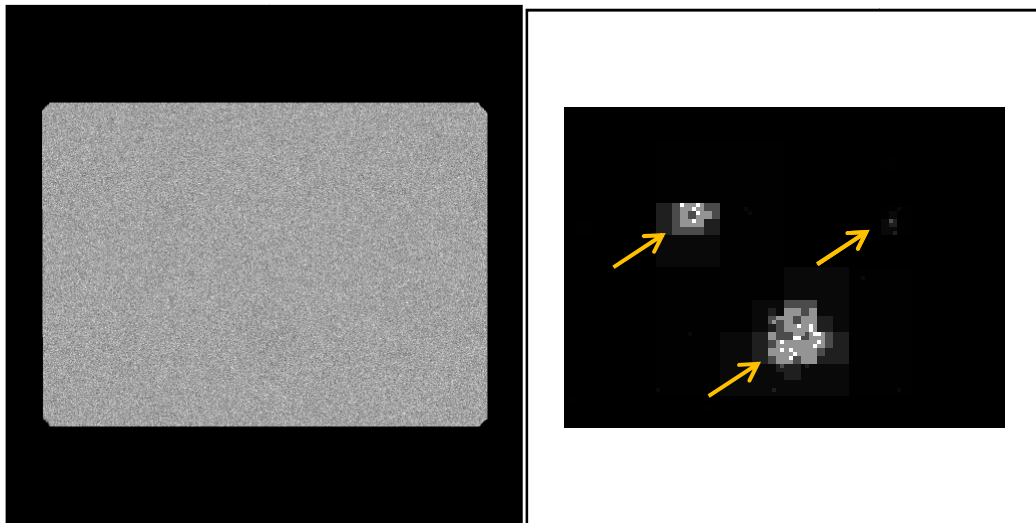


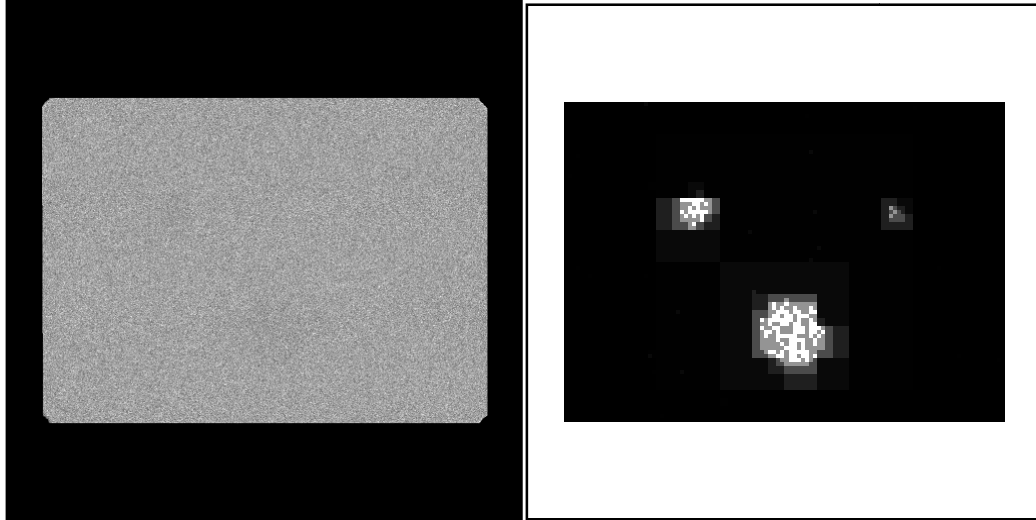
**Figure 22 – Uniformity Defect Contrast Enhancement of Image-space MDA Parametric Map**



*Top Left: Profile through MDA parametric map at observation 40. Top Right: Profile through the original flood image at the location of the uniformity defect at observation 40. Bottom: Plot of superimposed line profiles.*

The MDA algorithm also segmented out each region of non-uniformity in the acquired phantom data set, *ts\_vol\_3disk* (see Figure 18 for location of uniformity defects), in which the rate of degradation was slower than that of *synthetic\_13*. Given 7% attenuation over the 80-day time series, the degradation rate was approximately 0.1% per flood. The larger of the 3 attenuation disks was the first region segmented in image-space (at approximately day 24), and all three regions, including the 2 cm disk size, were segmented by day 50, when the integral and differential uniformity values had risen only to 3.23 and 2.27%, respectively. As time advanced and the regions degraded further, the size of the segmented regions grew in the MDA parametric maps because more pixels contained fewer counts in the regions corresponding to the locations of the disks, due to the increased attenuation of photons. As more attenuation occurred, the diameter of the uniformity defects in the parametric images converged to the actual size of the non-uniformity in the UFOV, and the non-uniformities became easier to discern visually. This is important because it demonstrates that the algorithm more easily segments non-uniformities of greater magnitude, which is a function of the x, y size of the region (i.e., how many pixels are affected by the region in the image), and the rate of degradation (how different the counts are over a region from one day to the next).





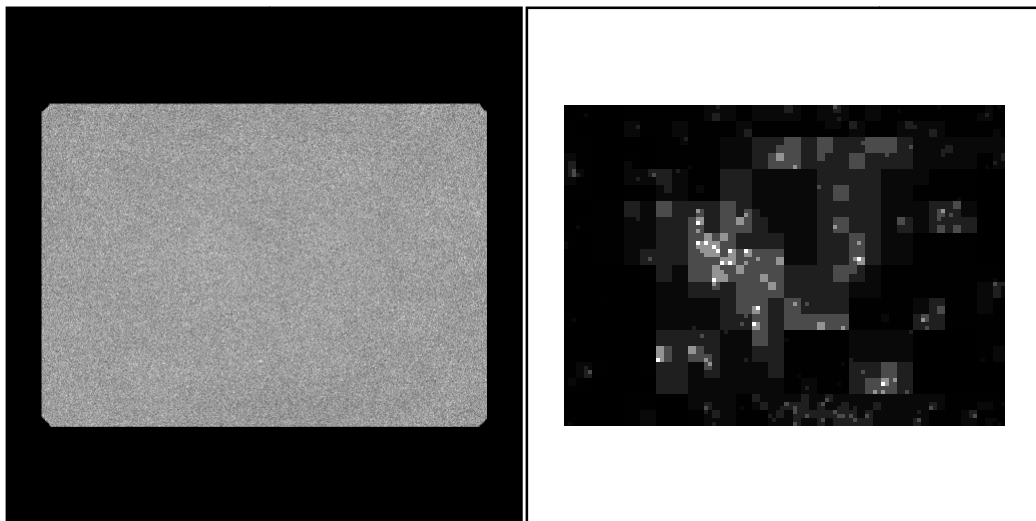
**Figure 23 – Regions of Non-uniformity Segmented via Multi-resolution MDA**

*Top Left: Original flood image, day 50; Top Right: MDA parametric map, day 50  
 Bottom Left: Original flood image, day 70; Bottom Right: MDA parametric map, day 70  
 Phantom: ts\_vol\_3disk*

The multi-resolution MDA algorithm performed well when processing the degradation-free synthetic control data in that no false positive regions were segmented. Moreover, regions outside of the known non-uniformities in phantom ts\_vol\_3disk were not identified by the algorithm, indicating strong false positive performance over the 100-flood time series (including baseline replicates). However, the algorithm did indicate some false positive regions when processing the synthetic\_17 and 19 phantoms, which contained a region of non-uniformity less than 1 cm in diameter. False positive pixels identified outside of the intentionally degraded portions of the UFOV were not sustained in the parametric maps, but decayed rather quickly over the time series. In other words, the algorithm segmented a region found to be statistically different from the baseline, and because this region was outside of a true region of non-uniformity, the segmentation did not persist in time, but faded away as the time window advanced. This demonstrates that by monitoring the results in the parametric maps over time as more maps accumulate, a reviewer would be able to distinguish a true evolving non-uniformity from a random false positive based on whether or not the segmented region persists in the maps. The algorithm also demonstrates some size-dependent sensitivity as the region of non-uniformity was not detected in phantom synthetic\_17 and was only later detected in the time series phantom synthetic\_19, in which the rate of degradation was twice that of synthetic\_17.

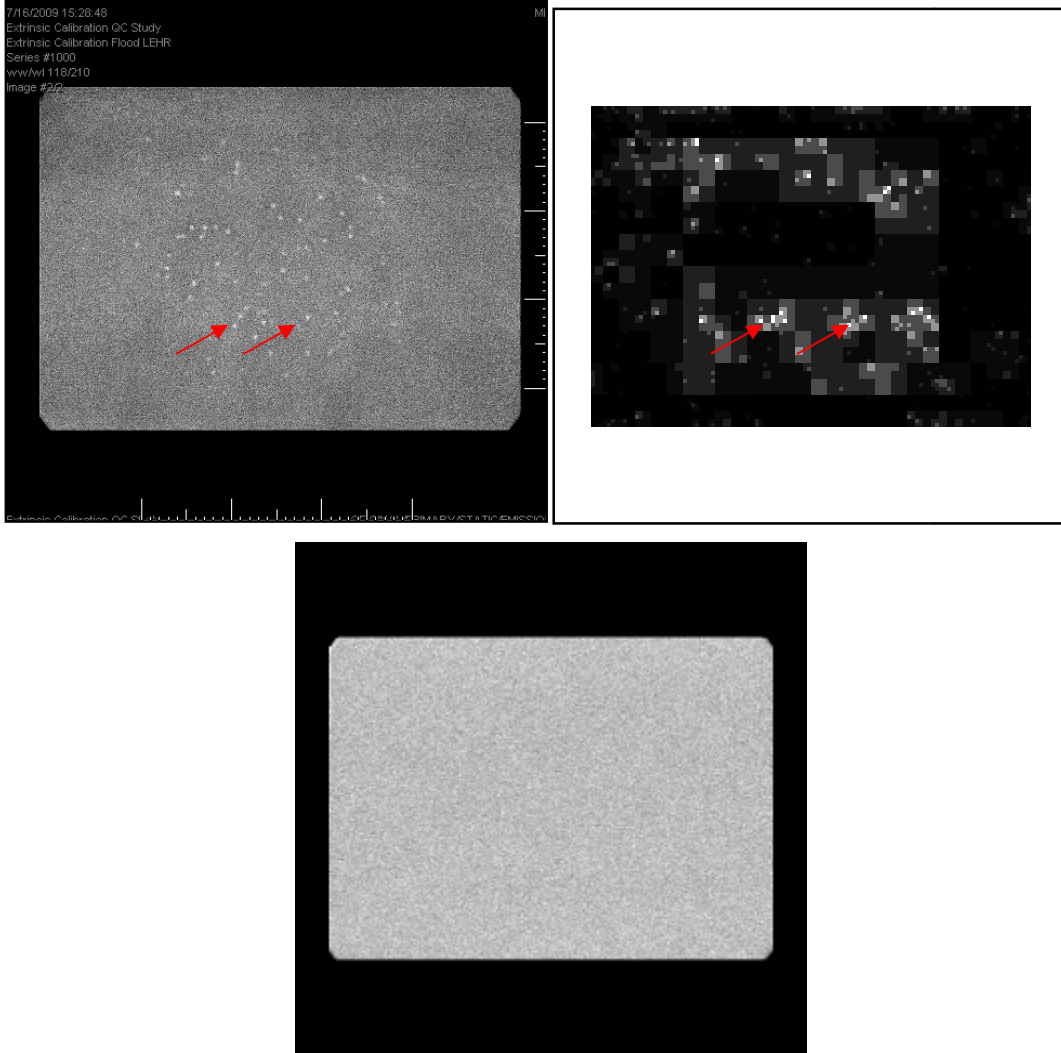
The off-peak phantom data did not contain degradation that was isolated to a specific location within the UFOV of the detector, and yet the algorithm still detected diffuse regions across the UFOV

as being statistically different from the baseline condition. This demonstrates that such image-space techniques are sensitive to degradation even when it is not confined to a specific shape or location within the UFOV. Given that false positives were effectively suppressed in the synthetic\_control phantom and in regions outside of the attenuation disks in phantom ts\_vol\_3disk, the algorithm segmented distinct regions in the off-peak data set, indicating the presence of true systematic error, as opposed to random false positive detection. The diffuse pattern present in the off-peak parametric maps was therefore indicative of the type of non-uniformity occurring in the detector. Thus, the pattern of segmentation in the maps may be an effective diagnostic tool in determining the cause of the non-uniformity present.



**Figure 24 – Multi-resolution MDA Results for Off-peak Time Series, Day 50**  
*Left: Original flood, day 50. Integral uniformity: 4.27%; Right: MDA parametric map demonstrating non-uniform segments throughout UFOV*

The retrospective time series processed using the multi-resolution MDA algorithm presented some interpretive challenges. The time series phantom that contained crystal hydration spots (ts\_vol\_S5) demonstrated regions identified by the algorithm early in time that were sustained throughout the time series, indicating that the uniformity defects were continuously identified after an initial point. These regions may be referenced in the high-count calibration flood acquired at the end of the time series, when the artifact was first visually observed. Several corresponding uniformity defects are noted in Figure 25:



**Figure 25 – Crystal Hydration Uniformity Defects Identified Using Multi-resolution MDA**

*Left: High count calibration flood demonstrating regions of hydration. Right: Parametric map acquired at day 62 in the 99 day time series Bottom: Original flood (Integral NEMA uniformity of corresponding flood: 3.24%)*

While the pixels identified at multiple resolution levels are brighter in appearance and corresponding hydration regions were discerned at a point very early in the time series (over 30 days before the corresponding spots were seen in the high count flood), the quadrants that converged to these regions appeared as larger grayscale squares and contributed noise to the background pixels outside of the non-uniform regions. Furthermore, there were regions identified in the parametric image shown above that do not necessarily correspond to known regions of degradation. Therefore, the algorithm does demonstrate sensitivity, but at the cost of potential false positive identification.

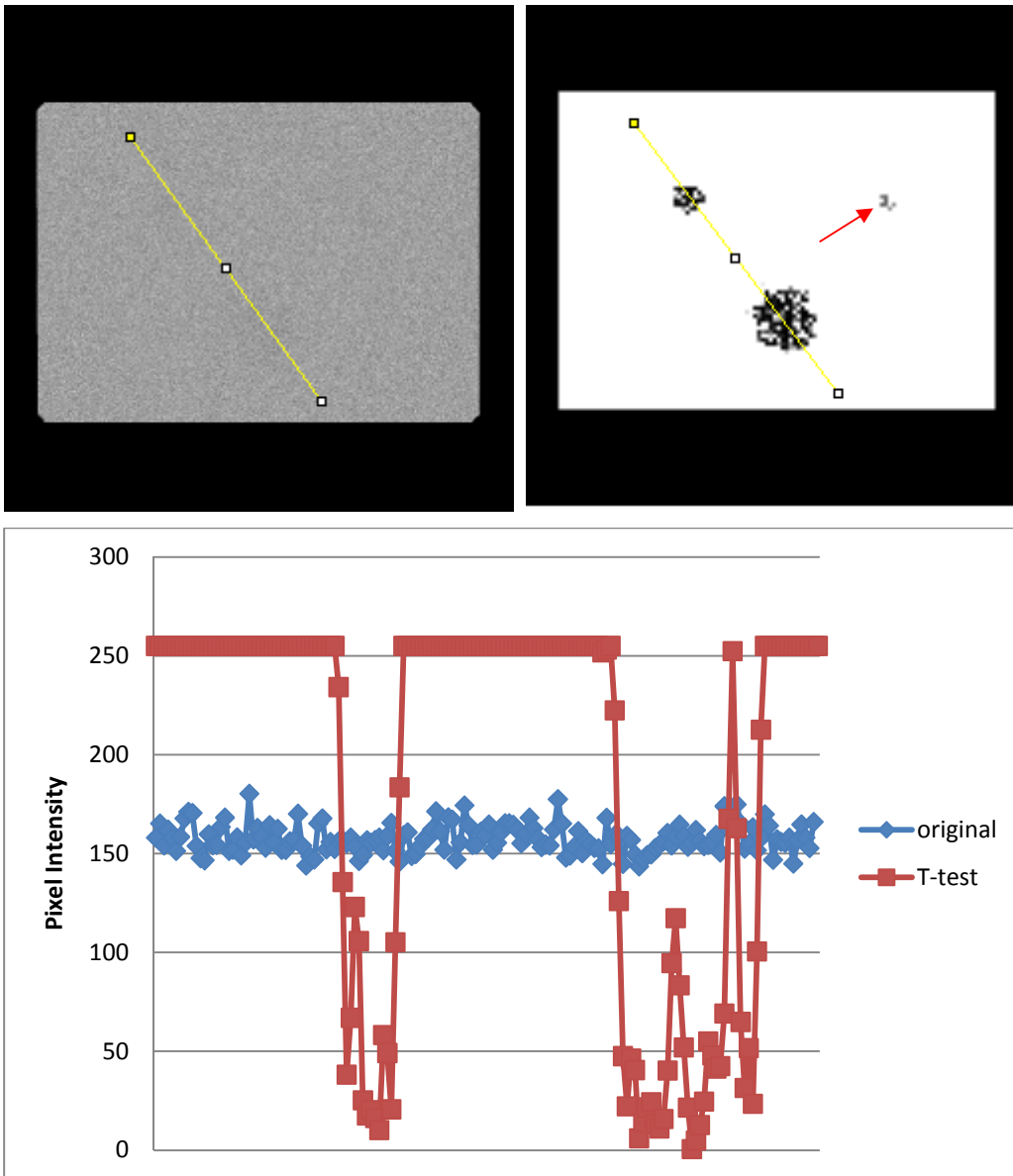
Similar results were observed for the retrospective time series phantom leading up to the failure of a PMT. As the time series advanced, distinct horizontal bands of non-uniform pixels were identified by the algorithm. These bands were sustained throughout the time series after their initial identification and grew more intense as the time series progressed. It is unclear whether this artifact was the result of the PMT's fluctuating prior to failure (the PMT location was in a region of banding) or non-uniformity in the radioactivity concentration in the rod source used to tomographically acquire the flood, causing a count gradient to appear in the parametric map.

#### 5.2.2.2 *Multi-resolution Two Sample Z and t-tests*

The multi-resolution Z- and t-test algorithms each performed similarly to the multi-resolution MDA technique in terms of their sensitivity to the rate and size of the non-uniformities. One notable distinction between these and the MDA method was an enhanced suppression of false positives outside of the regions of degradation, as well as an improved segmentation of the shape of these regions. This was due to the elimination of the larger quad segments surrounding the converging pixels and was accomplished by averaging over the multiple resolution levels used and then thresholding based on the calculated probability values, rather than cubing the final map for contrast. It had the effect of statistically segmenting out only those pixels identified at multiple resolution levels. The threshold also made intuitive sense, in that it was based upon an actual probability that differences between sub-volumes of pixels at two different time points were the result of random chance. The Z- and t-test results were almost identical in performance for all of the phantoms used, but because the t-test was a more statistically appropriate test, as it does not assume that the variance was known, only its results are demonstrated below.

The results of the algorithm for phantoms synthetic\_13 and synthetic\_15 identified the location of the degraded region beginning at day 14 and 10, respectively, when the respective NEMA integral uniformities were only 3.76 and 4.31%. The uniformity defects were maintained in each of the subsequent parametric maps and became more visibly distinguished as the magnitude of the non-uniformity increased in time.

The algorithm results for the actually acquired phantom ts\_vol\_3disk segmented the 8 cm attenuation disk beginning at day 25 and this region persisted continuously in each of the subsequent parametric maps. The 2 cm region of degradation was segmented beginning at day 67, when the integral NEMA uniformity was a mere 3.87% and the differential uniformity was only 2.83%. The contrast gains for the two larger uniformity defects are illustrated using line profile plots (Figure 26).



**Figure 26 - Uniformity Defect Contrast Enhancement of Image-space t-test Parametric Map**

*Top Left: Profile through original flood at observation 67 (NEMA integral uniformity: 3.85%, differential: 2.83%). Top Right: Profile through the t-test parametric map at observation 67, with the 2 cm disk noted. Bottom: Plot of superimposed line profiles*

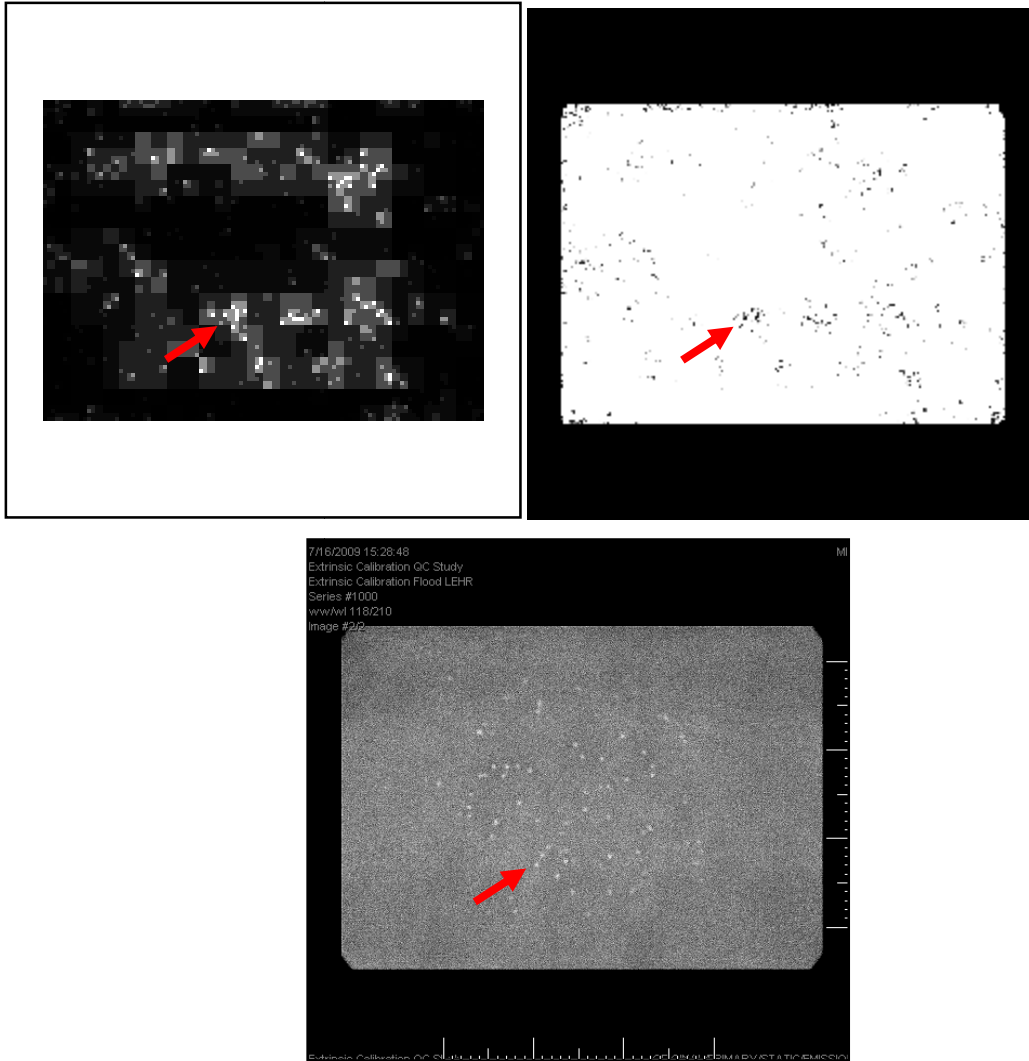
*Phantom: ts\_vol\_3disk*

These contrast gains demonstrated in the acquired flood time series using the hypothetical rate of degradation help prove the hypothesis that temporal image-space segmentation techniques may be used to segment true uniformity defects prior to the NEMA values being above the 5% threshold. The t- and Z- parametric maps did not segment any significant regions outside of the intentional

regions of degradation. The algorithms did not detect the region of degradation in synthetic\_17 and only detected the region in synthetic\_19 consistently within the last 5 days of the time series. This showed insensitivity in the techniques to very slowly progressing, small uniformity defects. It should be noted, however, that none of the original floods in these time series ever resulted in a NEMA value that was greater than 5%.

The multi-resolution t-test algorithm utilized a threshold value of  $p < 0.01$ , resulting in rejection of false positives. This was demonstrated by the rejection of any regional segmentation in the synthetic\_control time series phantom.

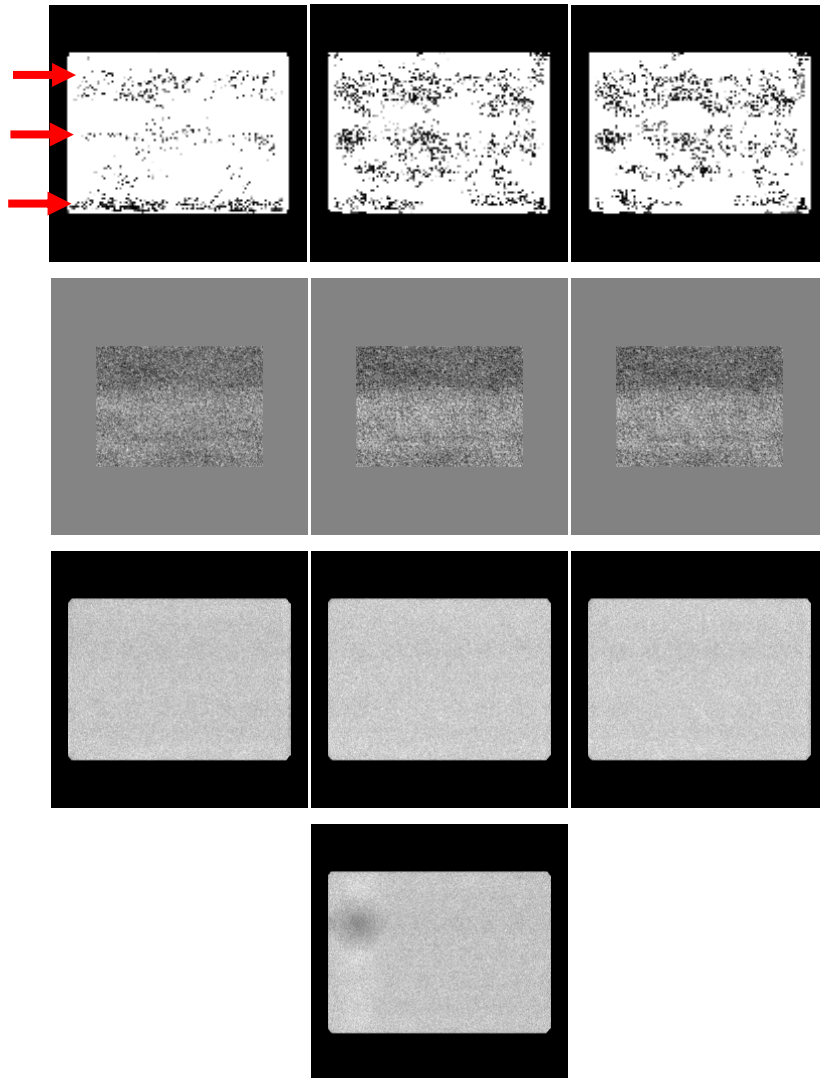
The regions of hydration in retrospective phantom ts\_vol\_S5 were better characterized using the multi-resolution t-test algorithm as opposed to the multi-resolution MDA method because of the suppression of the background quad regions in the former. To demonstrate this, the MDA and t-test parametric results for day 73 in the time series are displayed side by side. In Figure 27, the grayscale quadrants around the most intense pixels in the MDA map have been suppressed in the t-test parametric result, providing an automated threshold. However, when compared to the high count calibration flood acquired some 30 days later, the t-test parametric results still seem to indicate uniformity defects which were not necessarily observed in the 200 million-count flood image. This indicates that the temporal image-space algorithms detected statistically significant changes in pixels from the baseline condition that were not captured in the single temporal resolution 200 million-count flood. What is not clear, however, is the impact of such changes on clinical images, or whether these uniformity defects were the result of attributable causes of degradation or simply the expected fluctuation in detector response over time.



**Figure 27 – Detailed View of t-test Map Results for Hydration Spots**  
*Top Left: Multi-resolution MDA at day 73. Top Right: Multi-resolution t-test at day 73*  
*Bottom: High-count flood for hydration reference*

When the algorithm was applied to the defective PMT series, the horizontal banding across the UFOV was more clearly demonstrated than in the original flood images (Figure 28). There is no way to know whether the results indicate the instability of the PMT or the non-uniform source decay. Kalemis’ single resolution pixelwise slope maps (“s-maps”) demonstrated the banding very effectively and have been provided with the multi-resolution t-test maps in Figure 28. What is important to note is that the regions identified by the multi-resolution algorithm were not manifested as speckled, random noise throughout the parametric map, but were structured in appearance.





**Figure 28 – Multi-resolution and Single Resolution Parametric Maps for Ts\_vol\_T6**

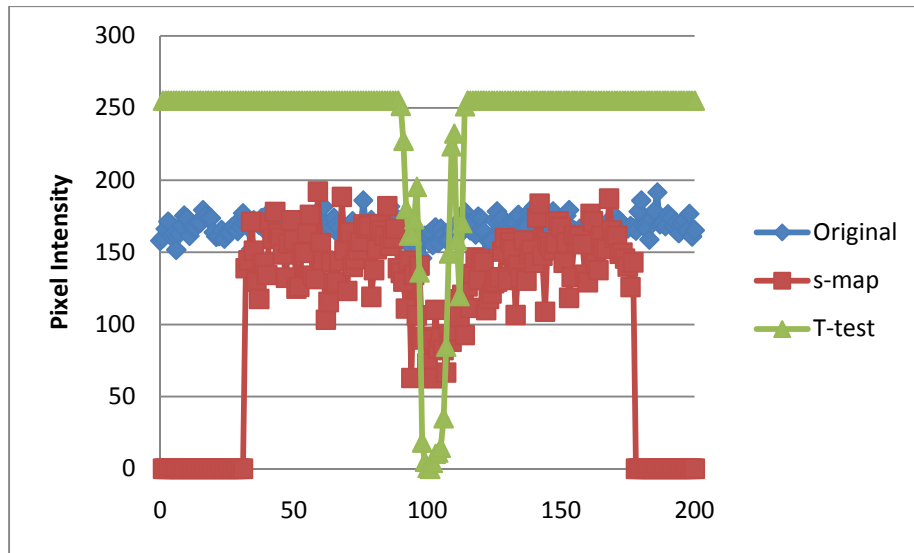
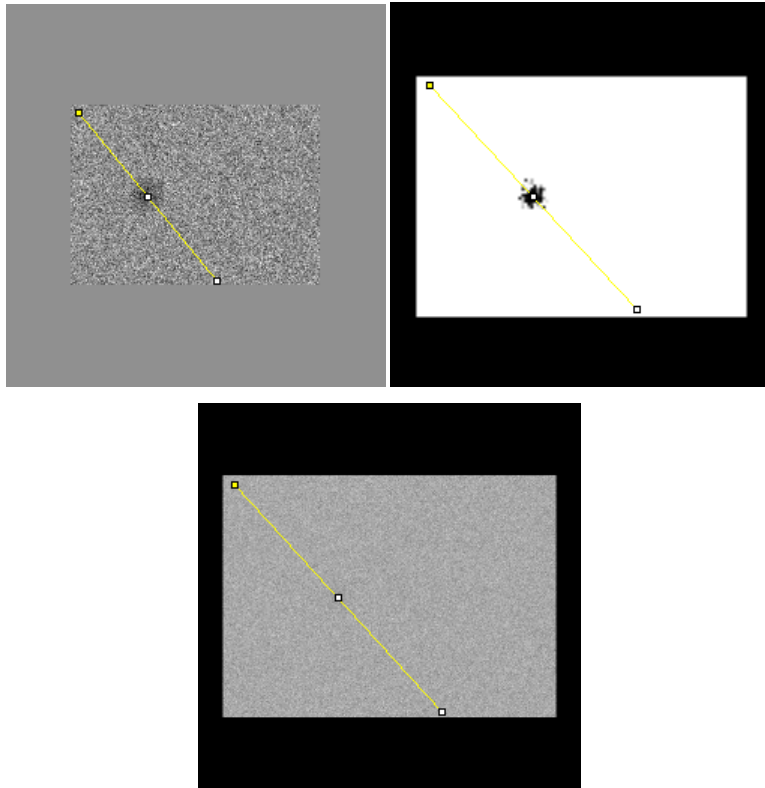
*Top Row: Multi-resolution t-test maps for days 109, 139, and 151. Failure occurred on day 156.*

*Middle Row: Slope Maps for corresponding days (only CFOV shown).*

*Bottom Row: Original images for days 109, 139, and 151. NEMA integral uniformity values were 3.42, 3.43, and 4.18% respectively*

*Bottom: Location of failed PMT. NEMA integral uniformity 19.56%*

Although the single resolution slope maps demonstrated the count gradients in the UFOV, there was a contrast gain obtained in the multi-resolution t-test algorithm. This was demonstrated by plotting a line profile in the parametric results at day 42 in the time series of the phantom synthetic\_13.



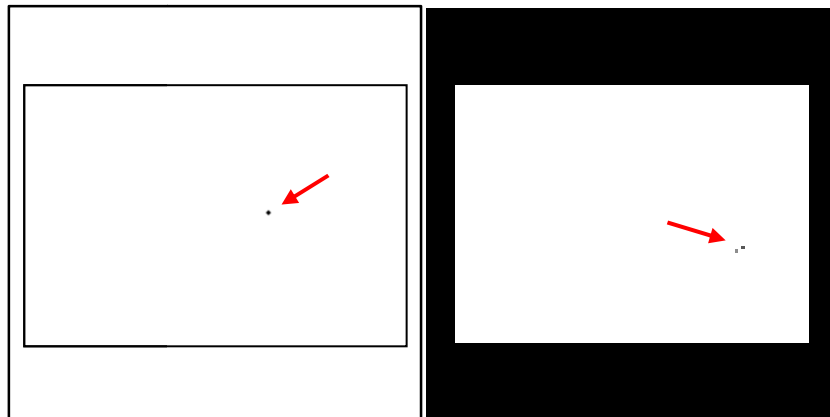
**Figure 29 - Uniformity Defect Contrast Enhancement of Image-space t-test Parametric Map**

*Top Left: Kalemis' single resolution s-map at day 42 (only CFOV shown). Top Right: Multi-resolution t-test result at day 42. Middle Row: Original flood at day 42. Bottom: Line profile plots across uniformity defect*

### 5.2.2.3 Multi-resolution Strength-of-correlation

The results of the multi-resolution strength-of-correlation statistical decomposition demonstrated a slight gain in sensitivity in the time to detection of the uniformity defects in phantoms with more rapidly degrading regions. The degraded region in phantom synthetic\_15 was detected beginning as early as day 8 in the time series, which is 3 days prior to detection in the multi-resolution t-test method. The region of non-uniformity in phantom synthetic\_13, in which the rate of degradation was slower, was not clearly segmented until observation 15, which is similar in performance to the t-test technique. There were no false positive segmentations observed in the 100-flood synthetic control set, and no discernable false positive segmentations outside of the degraded regions were seen in the acquired time series phantom ts\_vol\_3disk. There was an almost identical sensitivity to the size of the non-uniformity as a function of the time to segmentation, in that the 2 cm disk in phantom ts\_vol\_3disk was first observed beginning at day 61 in the time series, which is similar to the results of the multi-resolution t-test.

This technique also demonstrated a similar insensitivity to the very small uniformity defects detected in phantoms synthetic\_17 and synthetic\_19, in which the uniformity defects were never segmented. Furthermore, the multi-resolution strength-of-correlation technique did detect some false positive regions, or regions outside of the intentionally degraded area, for phantom synthetic\_19 due to random error. These regions faded with time.

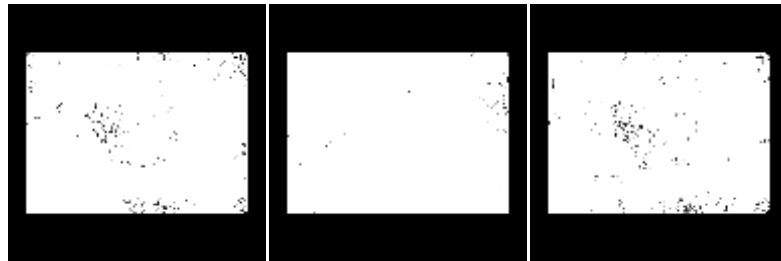


**Figure 30 – False Positive Segmentation**

*Left: Location of uniformity defect in phantoms synthetic\_17 and synthetic\_19. Right: False positive segmented region using multi-resolution strength-of-correlation test*

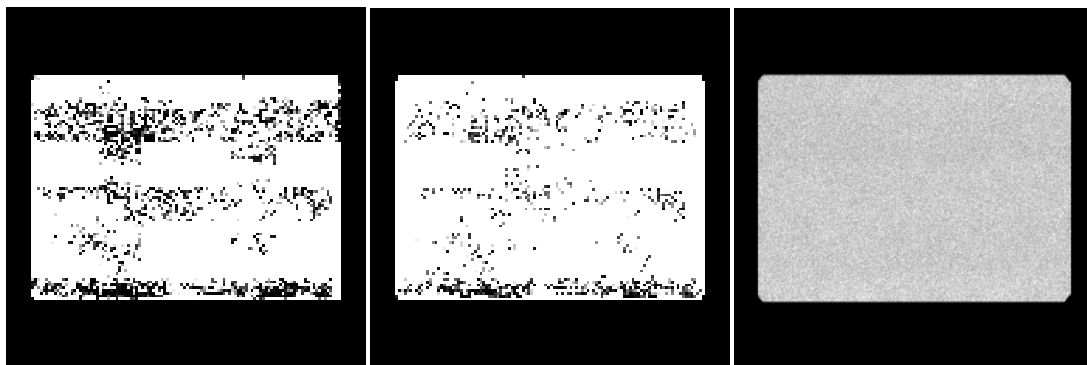
The results for phantom ts\_vol\_off-peak exhibited a similar pattern of diffuse segmented pixels concentrated toward both the center and edges of the UFOV, as previously observed from the multi-resolution t-test algorithm. These regions fade as the energy window shifts toward the photopeak,

and then worsen as the window moves away from the photopeak in time. This demonstrates that the parametric maps adapt to the actual changes of the detector performance. As the uniformity improved when the window was centered on the photopeak, the map results became more uniform; and as the energy window was shifted away from the photopeak, uniformity defects were segmented, indicating the presence of systematic error.



**Figure 31 – Multi-resolution Strength-of-correlation Results for Ts\_vol\_off-peak**  
*Left: Day 19, Integral Uniformity: 3.7%; Middle: Day 30, Integral Uniformity: 3.37; Right: Day 49, Integral Uniformity: 3.8%*

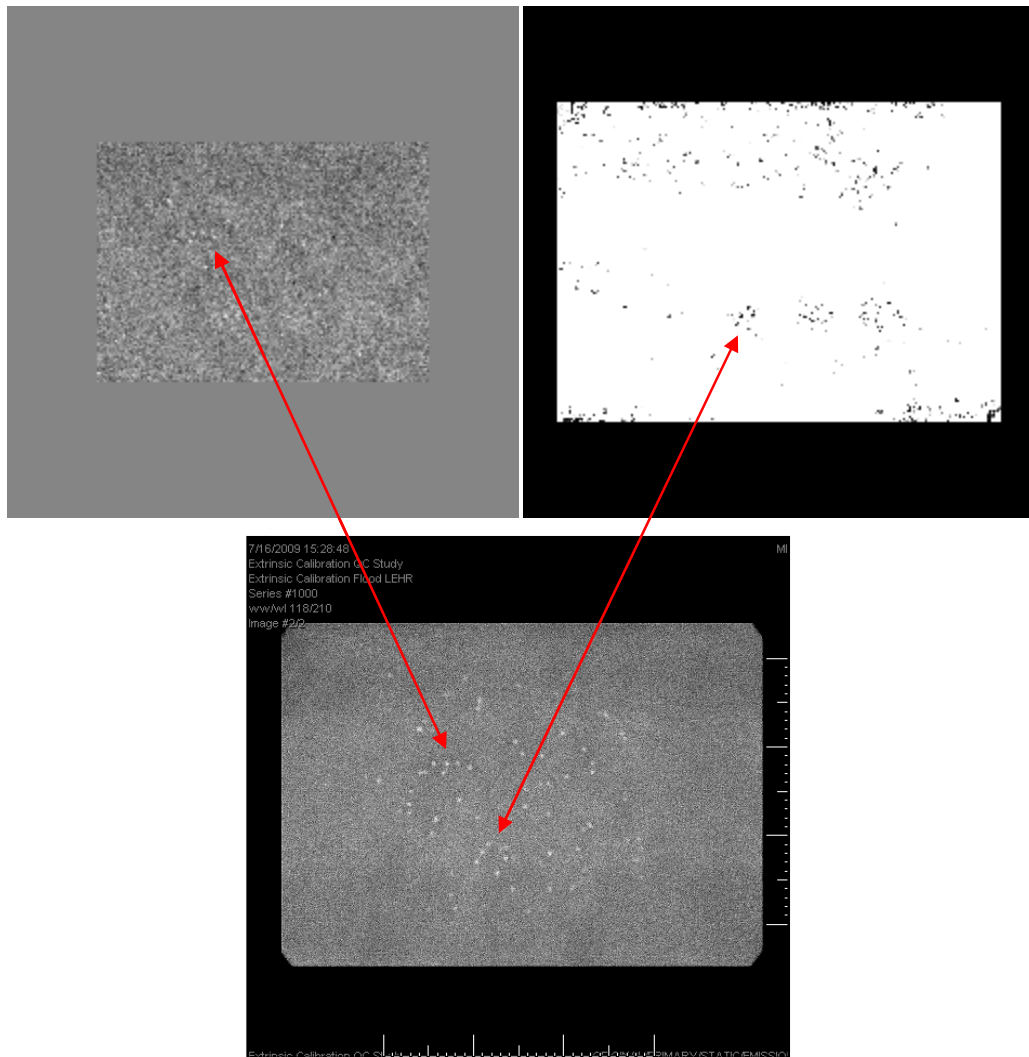
The results for the retrospective phantom ts\_vol\_T6 demonstrated some enhanced sensitivity when the results of the multi-resolution strength-of-correlation are compared to the results using the multi-resolution t-test. The horizontal bands that were observed in the UFOV are better characterized over time using the strength-of-correlation approach, due to the increased number of segmented pixels.



**Figure 32 – Multi-resolution Results for Ts\_vol\_T6, Day 121**  
*Left: Strength-of-correlation; Middle: t-test; Right: Original flood*

The results for the time series in which crystal hydration was observed demonstrated similar sensitivity to the multi-resolution t-test algorithm. Some hydration spots were also difficult to detect

when compared to the high count calibration image, but as with the previous algorithms, several consistent regions exhibited a sustained segmentation with time and seem to correspond to several regions observed in the high SNR flood. Likewise, the algorithm did not segment some regions of hydration noticed in the calibration flood image. For the regions that went undetected, a possible explanation is that several hydration spots appeared later in the time series, but the degradation was not significant enough to make the corresponding slopes of the linear fits statistically different from zero. The Kalemis trend analysis demonstrated the hydration spots consistent with the high-count calibration flood. However, the background pixels within the UFOV contribute a significant amount of noise and obscure the locations of the artifact.

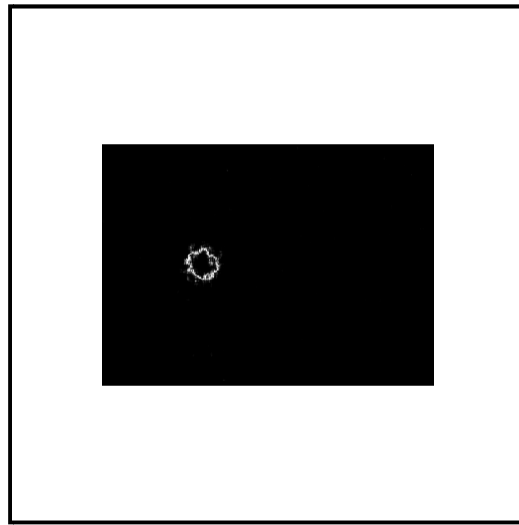


**Figure 33 – Strength-of-correlation Results for Hydration Time Series**

*Top Left: Kalemis S-map 10 days prior to artifact discovery (only CFOV); Top Right: Multi-resolution Strength-of-correlation Results 10 days prior to artifact discovery; Bottom: High SNR image depicting location of hydration spots.*

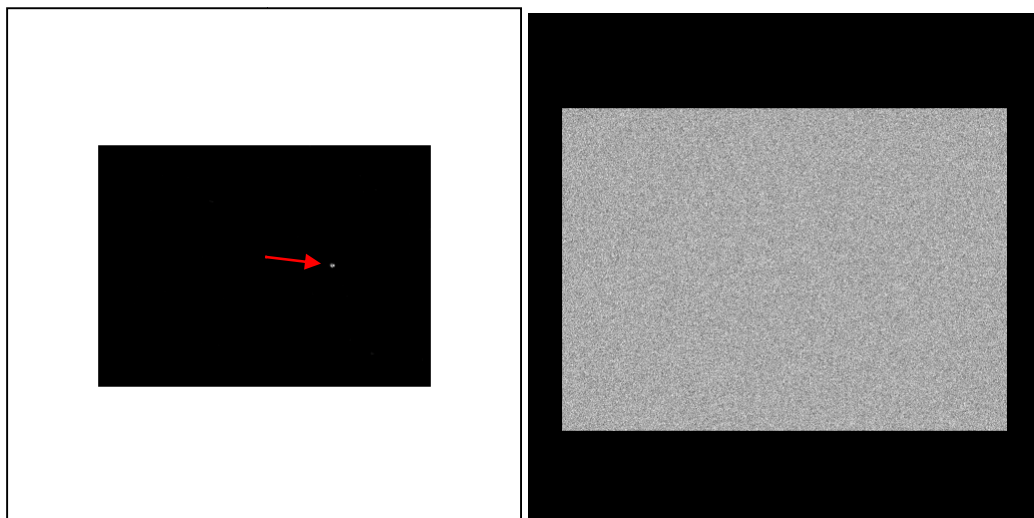
#### 5.2.2.4 Sliding Window t-test

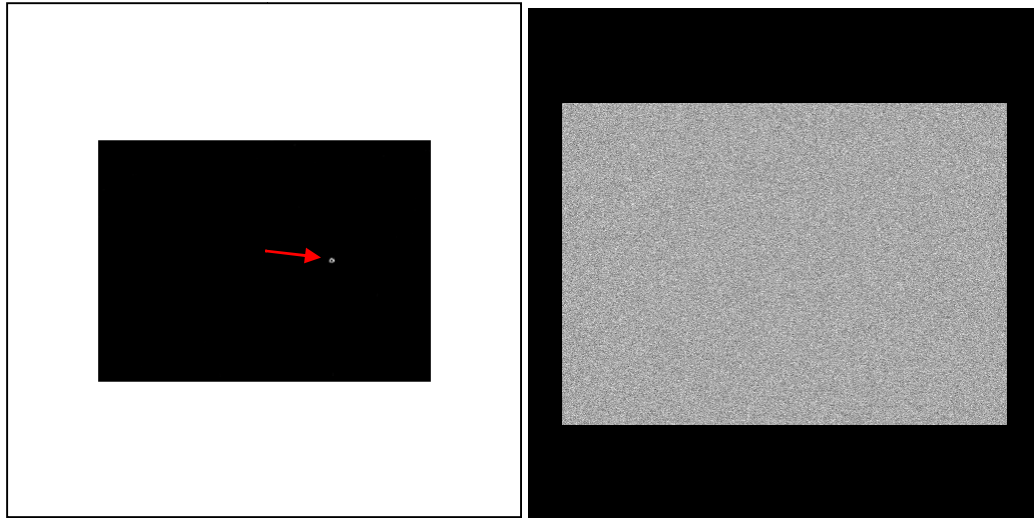
The results for the sliding window (convolution) t-test algorithm demonstrated a marked increase in sensitivity to the smaller uniformity defects. The regions were segmented early in both synthetic\_17 and synthetic\_19 time series phantoms and were sustained throughout time. As the regions degraded the diameter of the segmented region widened, similar to other algorithms. The algorithm results also demonstrated an as-of-yet unexplained phenomenon in that only the outer edge of the uniformity defect region is detected. This was observed for several different spatial window sizes:



**Figure 34 – Sliding Window t-test Results for Synthetic\_13, Day 81**

*Note that the outer ring of the uniformity defect was segmented. The diameter of the uniformity defect is 8 cm*





**Figure 35 – Synthetic\_19 Sliding Window t-test Results and Original Floods, Days 48 and 84**

*The diameter of the uniformity defect is approximately 7 mm. The ring cannot be fully seen due to the small size of the uniformity defect (only CFOV processed in map)*

The parametric maps calculated using Kalemis' technique also demonstrated superior sensitivity in distinguishing smaller uniformity defects from the background when compared to the previous multi-resolution techniques, but contrast was enhanced when using the sliding window t-test algorithm because all pixels outside of the region of non-uniformity took on a value of zero on the parametric image.

The limitations of the sliding window t-test include an increased number of random regions segmented outside of the region of true non-uniformity. The control time series expressed a few segmented regions that faded quickly in time. In addition, the algorithm did not segment any regions for the off-peak time series, indicating that it is more sensitive to regions of degradation that worsen linearly with time and were confined to particular spatial locations. The algorithm was also limited to floods acquired in larger matrix sizes (i.e., smaller pixel sizes), due to the size of the spatial window. The algorithm was insensitive to floods acquired with smaller matrix sizes (coarser pixel sizes) because the spatial window blurs out the distinct uniformity defects.

The algorithm results presented above demonstrate that regional non-uniformities can be segmented in order to spatially identify degraded portions of the detector prior to a violation of the 5% NEMA threshold. The multi-resolution MDA algorithm detected uniformity defects early on in most of the time series, but the larger grayscale quadrants characteristic of the algorithm made manual thresholding necessary to help identify uniformity defects. The multi-resolution t-test offered an improved method of automatically thresholding the result using the calculated statistical p-value.

The multi-resolution strength-of-correlation method yielded similar results to the multi-resolution t-test in terms of sensitivity to the size and rate of degradation and offered a means of characterizing degradation over the entire time series, as opposed to an advancing time window. However, these multi-resolution methods still exhibited insensitivity to very small uniformity defects that had little impact on the calculated NEMA uniformity values. The sliding window t-test algorithm along with Kalemis' single resolution trending approach demonstrated an improved sensitivity to smaller regions of degradation over the multi-resolution decomposition techniques.

The table below shows the numbers of days that elapsed until the region(s) of non-uniformity were segmented in the parametric maps by each algorithm for several of the time series phantoms used.

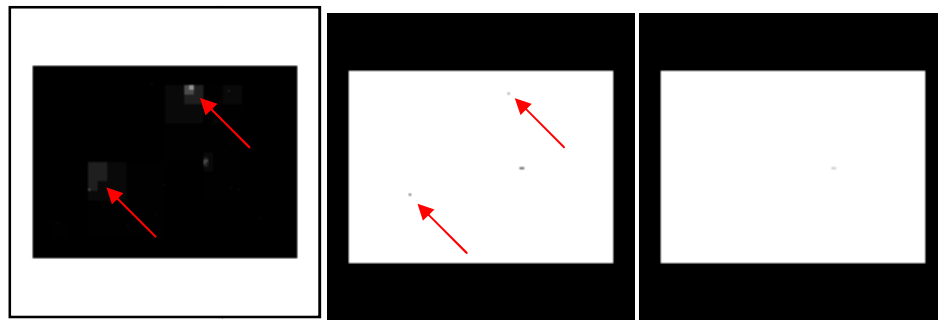


multi-resolution MDA		
phantom	time to detection (days)	integral uniformity (%CFOV)
synthetic_13	13	3.25
synthetic_15	10	2.65
synthetic_17	Not detected	-
synthetic_19	96	2.34
ts_vol_3disk - 8 cm disk	25	2.39
ts_vol_3disk - 4 cm disk	33	2.86
ts_vol_3disk - 2 cm disk	50	3.23
ts_vol_offpeak	18	3.6
multi-resolution t-test		
phantom	time to detection (days)	integral uniformity (%CFOV)
synthetic_13	14	3.76
synthetic_15	10	2.65
synthetic_17	Not detected	-
synthetic_19	95	2.79
ts_vol_3disk - 8 cm disk	26	2.39
ts_vol_3disk - 4 cm disk	32	2.86
ts_vol_3disk - 2 cm disk	67	3.85
ts_vol_offpeak	18	3.6
multi-resolution strength-of-correlation		
phantom	time to detection (days)	integral uniformity (%CFOV)
synthetic_13	14	3.76
synthetic_15	9	3.44
synthetic_17	Not detected	-
synthetic_19	Not detected	-
ts_vol_3disk - 8 cm disk	25	2.39
ts_vol_3disk - 4 cm disk	32	2.86
ts_vol_3disk - 2 cm disk	62	3.59
ts_vol_offpeak	16	3.57
sliding window t-test		
phantom	time to detection (days)	integral uniformity (%CFOV)
synthetic_13	19	2.82
synthetic_15	13	3.76
synthetic_17	33	2.39
synthetic_19	19	2.38
ts_vol_3disk - 8 cm disk	33	2.86
ts_vol_3disk - 4 cm disk	41	3.32
ts_vol_3disk - 2 cm disk	56	3.04
ts_vol_offpeak	Not detected	-

**Table 5 – Summary of the Algorithm Performance on Phantoms**

In the table above, the time to detection was considered the first day on which the non-uniformities were segmented in the maps and sustained throughout the subsequent parametric maps. The entire phantom non-uniformities were segmented by at least one of the algorithms employed prior to a single flood image resulting in a NEMA uniformity value that exceeded 5%. Of note was

the observation that each of the multi-resolution techniques demonstrated similar sensitivity in terms of the number of days taken to begin segmenting the non-uniformities. The exception to this was the time taken to begin segmentation of the 2 cm disk. In that case, the multi-resolution MDA algorithm demonstrated superior sensitivity to the other two multi-resolution techniques. The number and location of false positives was similar between the multi-resolution t-test and MDA algorithms, and better false positive performance was observed for the multi-resolution strength-of-correlation algorithm. For example, consider the parametric maps produced for each algorithm at day 46 for the phantom synthetic\_19 in Figure 36.



**Figure 36 – Multi-resolution Parametric Results for Phantom Synthetic\_19, Day 46**

*Left: MDA (false positives noted); Middle: t-test; Right: strength-of-correlation*

In general, the multi-resolution MDA maps showed more false positives than the other multi-resolution algorithms. This result may be because a threshold based on the level of significance is not applied directly to the MDA parametric maps, as in the t-test and strength-of-correlation algorithms. Instead, the maps are cubed in order to create contrast. This may result in regions outside of the uniformity defect still visible in the map, whereas when using the thresholding approach all values above a certain level of significance are effectively eliminated.

The sliding window t-test method showed the most sensitivity to small, slowly degrading regions and did not seem to be sensitive to the size of the non-uniformity. This is because the spatial window function slides over each pixel in the flood image rather than isolating discrete groups of pixels, effectively centering itself over each pixel and testing the local mean and variance. Because this algorithm smooths the pixel values in time rather than examining discrete voxels in three dimensions as the multi-resolution algorithms do, some statistical power seems to be lost given that the time to detection is somewhat slower using this method and the off-peak non-uniformities were never segmented. This justifies the use of multiple techniques clinically. The multi-resolution t-test and strength-of-correlation algorithms, with improved false positive suppression over the MDA

technique, demonstrated improved performance over the sliding window t-test in some scenarios, while the sliding window t-test showed improved performance in others.

### **5.3 Specific Aim 3 – Statistical Process Control Techniques**

*Apply statistical process control to conventional methods of quantitatively monitoring uniformity in order to characterize how these metrics change over time in response to the mimicked degradation processes.*

#### **5.3.1 Materials and Methods**

The results in 5.2 indicate that the image-space methods of statistical segmentation provide an improvement in specificity over global figures-of-merit, such as NEMA or CRSD uniformity values, because they identify the spatial locations and extent of uniformity defects. However, conventional measures of uniformity, such as the NEMA uniformity index, provide a widely accepted global figure-of-merit, and thus the natural reference against which to compare alternative techniques.

Statistical process control is widely used in manufacturing to ensure the quality and consistency of products. Fundamental to statistical process control is the characterization of what is called an “in-control” process. This characterization includes the calculation of statistical parameters (such as the sample mean and variance) of certain quantitative attributes of a product at a time when they are deemed to be acceptable in quality and free from any defect that may be the result of a systematic, non-random cause (16). Time Series observations of these attributes are plotted in control charts, which are graphical tools used to determine if a quality process is in a state of statistical control. An out-of-control process is one in which the observations of the control parameters continuously fall outside of the established control limits (16). The in-control parameters are used to calculate control limits against which future observations of the process are graphically compared in a control chart. Observed defects that result from systematic causes may introduce a sustained shift in the process observations in a time series. These process shifts represent changes in the product characteristics that are outside of the expected statistical fluctuation and signal a problem with the process that should be addressed in order to maintain quality. Control plots provide a sensitive means of determining trends in time series data by providing real-time signals for observations made outside of the previously established limits. Methods of using control charts for quantitative quality control have been used extensively for bone density scanners (34).

To cast this project in statistical process control terms, a baseline replicate set of acceptably uniform floods serves as the in-control portion of the daily flood processes. The ‘product attributes’

that were evaluated using statistical control methods were the global uniformity figures-of-merit. The sample mean and variance of the sets of the baseline replicate NEMA uniformity (or CRSD) values were calculated (this is known as a phase I action). These parameters were then used to monitor the performance characteristics of subsequently observed uniformity values (a phase II action (16)). The control limits for the synthetic data were derived from the large sample synthetic\_control time series (100 synthetic floods). The control limits for the retrospective phantom data were based on the first 20 floods in the time series because the NEMA values in this period were all under the 5% threshold. While it would have provided more statistical power had more floods been utilized for the control limit characterization, twenty consecutive floods represented an adequate number of observations for a phase I action (16) while minimizing the time span over which detector performance could have changed. Furthermore, twenty consecutive floods of 10-million counts each are consistent with the current practice of acquiring a 200-million count flood for calibration purposes. For the acquired flood data (attenuation disk and off-peak time series), the phase I portion was observed over 40 consecutively acquired uniform replicates, in order to provide a case in which greater statistical power was used.

Several different types of control charts were constructed and applied to the time series phantom data and were used to monitor the points in time at which systematic causes of degradation occurred, indicating that the uniformity process was out-of-control. Control plots signal both large and small shifts in the expected values of a time series and alert the reviewer of systematic changes in the uniformity performance.

#### *5.3.1.1 Shewhart Control Chart for Individual Observations*

Shewhart control charts are tools for monitoring a process mean and variance and are the most common type of control chart used for monitoring normally distributed data (15). These charts are ideal for detecting large shifts in the mean and/or variance of a time series of observations. In most applications of Shewhart charts, one observation of a process attribute in time is derived from the mean of several repeated measurements at that particular observation point. For example, if measuring the weight of widgets on a conveyor belt, the observation at time  $t$  will be the average weight of 5 different widgets pulled off the belt at time  $t$ . However, because the calculated uniformity observations are relatively infrequent for daily uniformity floods (4-5 per week, with 1 uniformity value per day), a chart that can be used to monitor individual measurements rather than individual sub-samples is necessary. There is a version of the Shewhart control method for single-measurement processes, in which each observation in time on the control chart is a result of a single measurement

(15). In order to derive the Shewhart single-observation control limits, the moving range is calculated for each of the phase I observations

$$MR_i = |x_i - x_{i-1}|$$

**Equation 23 – Shewhart Moving Range Calculation**

where  $i$  denotes the point in time (observation number) among the baseline measures and  $x_i$  and  $x_{i-1}$  are the current and previous points in time NEMA values in the set of baseline observations (16). The average of these moving range values is calculated over all of the baseline measurements. Next, the control limits for the chart are calculated using the following equations (16).

$$UCL = \bar{x}_{Baseline} + 3 \frac{\overline{MR}_{Baseline}}{1.128}$$

$$Centerline = \bar{x}_{Baseline}$$

$$LCL = \bar{x}_{Baseline} - 3 \frac{\overline{MR}_{Baseline}}{1.128}$$

**Equation 24 – Shewhart Control Limit Calculations**

$\overline{MR}$  is the average moving range over the baseline replicate set of floods and  $\bar{x}_{Baseline}$  is the average NEMA (or CRSD) value over the baseline set of images. The factor of 1.128 in Equation 20 is derived in the literature (16) for single-measurement time series. The upper and lower control limits essentially reflect the  $3\sigma$  limit for the phase I mean. Obviously, the lower control limit may not be clinically relevant, as there is no limitation on how low a value would be desirable for the global figures-of-merit. However, a minimum value is expected based upon the randomness of the count data, therefore if things were operating “too good” that might indicate a problem.

**5.3.1.2 CUSUM Control Chart**

Cumulative summation (CUSUM) control charts are an effective means of detecting subtle shifts in a process mean and variance and have been shown to operate well with Poisson distributed data (15). The test statistic used to construct these charts is the continuous summation of the residuals about an expected process mean over time. This form of statistical control is useful because it amplifies signals in the time series data by summing the differences from the baseline condition. If there are no significant shifts in the mean value of a process, then the CUSUM statistic will always

have a value essentially of zero, as there is no expected significant difference between the current observation and the phase I mean value.

A standardized test statistic was calculated for each phantom time series global uniformity by subtracting the phase I mean from the current uniformity value in the series and dividing by the phase I standard deviation (16):

$$y_i = \frac{x_i - \bar{x}_{Baseline}}{S_{Baseline}}$$

The standardized two-sided CUSUM values at time point  $i$  may be calculated using the following equations (16):

$$C_i^+ = \text{MAX}[0, y_i - k + C_{i-1}^+]$$

$$C_i^- = \text{MAX}[0, -k - y_i + C_{i-1}^-]$$

**Equation 25 – CUSUM Statistic Calculations**

where  $k$  is a reference value, usually taken to be one-half of one standard deviation (16), and  $C_{i-1}$  is the previous CUSUM statistic in the time series. The MAX [0,\*] notation indicates that the CUSUM value is actually the larger value of 0 and the calculated statistic. The upper CUSUM reflects a positive shift in the process mean, while the negative CUSUM reflects a reduction in the process mean. The initial CUSUM value ( $C_0^+$  and  $C_0^-$ ) is 0. The smaller the value of  $k$  selected, the smaller the shift in the process mean that will be detected. The reference value  $k$  is normally chosen in order to give good Average Run Length (ARL) values. The ARL is the number of consecutive observations made before randomly indicating an out-of-control signal by chance. A tradeoff with ARL exists in the selection of  $k$ , in that larger values reduce the number of false positives shown to be outside of the control limits, but smaller values of  $k$  improve the sensitivity to detecting shifts in the process mean. A value of  $k=1$  was selected because it demonstrated good ARL performance for the uniform time series control data. The CUSUM statistics were plotted as a function of time in the control chart, and if any values exceeded the decision value,  $H$ , an *out-of-control* shift was inferred, especially if the signal was sustained, or continuously above the decision value. According to the literature, for the value of  $k$  chosen, a reasonable decision interval for normally distributed data is  $H=3$  (16). (One might argue for clinical practice that a QC measurement should be repeated immediately if it fails this test, and a true positive declared when  $H > 3$ ).

### 5.3.1.3 EWMA Control Chart

As with the CUSUM method of statistical control, exponentially weighted moving average (EWMA) control charts are ideal for monitoring relatively small shifts in process means and are very effective tools for single-measurement time series measurements (16), such as NEMA and CRSD uniformity values. EWMA is distinct from the CUSUM in that it is actually a predictor of the process mean at a time point one day in advance of the process, and thus it is often used in time series analysis and forecasting applications (16). Another benefit of this method of statistical control is that it is insensitive to the assumption of normality, meaning that data do not necessarily have to be normally distributed for changes in process parameters to be detected using this chart. Both normally distributed data as well as data from other probability distributions may be plotted in CUSUM and EWMA control charts (16). The Shewhart control chart is sensitive to detecting large process shifts and works best for normally distributed data. Shewhart charts are recommended by Montgomery (16) to be used concomitantly with control charts designed to detect small process shifts, such as EWMA and CUSUM charts.

For an independent observation  $x_i$ , the exponentially weighted moving average (EWMA) at the point of observation is defined as

$$z_i = \omega x_i + (1 - \omega)z_{i-1}$$

**Equation 26 – EWMA Statistic**

The EWMA is effectively a weighted average of all past and current observations and is insensitive to normality (16). The variable omega is the weighting coefficient that controls the relative influence of the current measurement,  $x_i$ , and the previous EWMA value,  $z_{i-1}$ . The initial value for  $z_0$  in this case is the observed sample mean of the phase I observations. For a derivation of the exponential weighting coefficient, the reader may refer to Montgomery (16). The upper and lower control limits for this type of chart are as follows:

$$UCL = \bar{x} + L\sigma_s \sqrt{\frac{\omega}{(2 - \omega)} [1 - (1 - \omega)^{2i}]}$$

$$\text{Center line} = \bar{x}$$

$$LCL = \bar{x} - L\sigma_s \sqrt{\frac{\omega}{(2 - \omega)} [1 - (1 - \omega)^{2i}]}$$

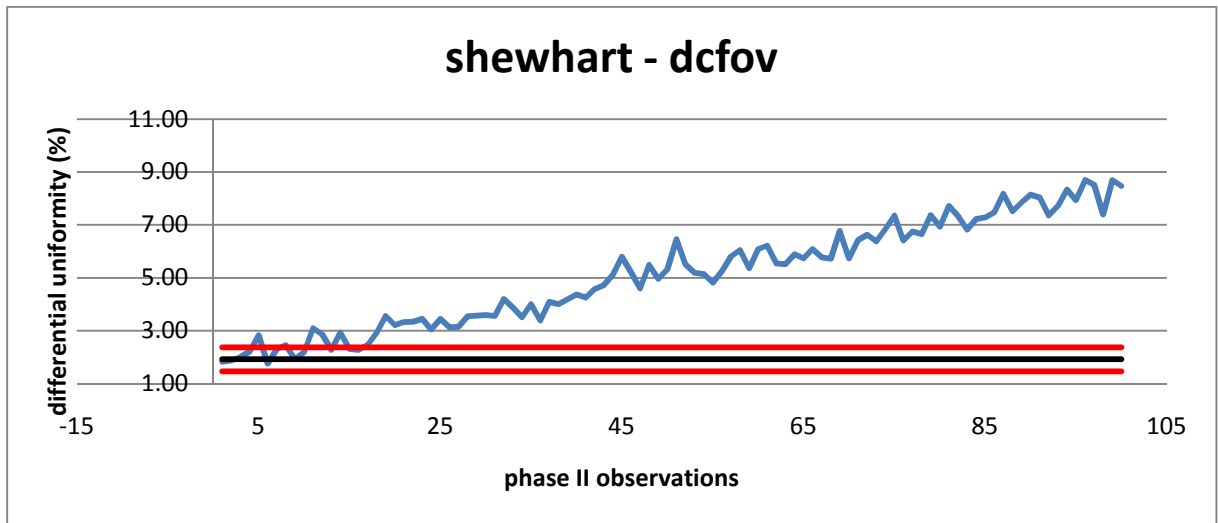
**Equation 27 – EWMA Control Limit Calculations**

The centerline for the chart ( $\bar{x}$ ) is the mean value of the preliminary data, or phase I observations. The value L is essentially the width of the control limits, or rather the threshold for number of standard deviations from the mean. According to the literature (15), there are commonly selected values for  $\omega$  and L, and for this application they were chosen to be  $\omega=0.1$  and  $L=3$  (15). These values have been found by Montgomery to work well in most applications and demonstrate good ARL performance, meaning that false positives are rarely signaled in a time series run of observations (15).

### 5.3.2 Results and Discussion

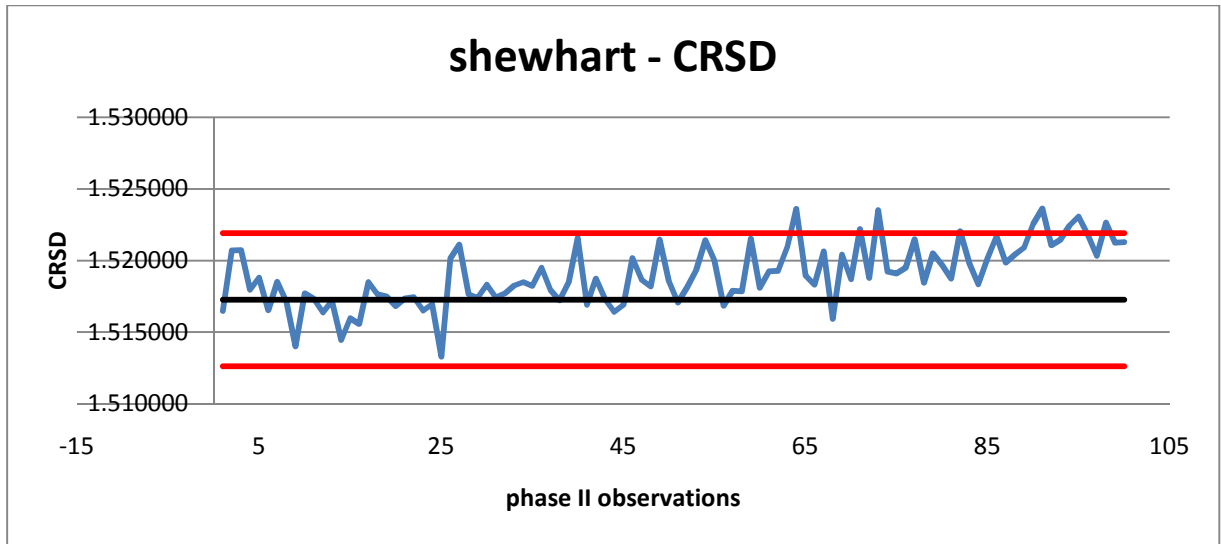
NEMA uniformity and CRSD values were calculated for the CFOV of each image in the time series phantoms. The results were plotted as a function of time using the statistical control chart methods described in 5.3.1 (see Appendix for charts). The time series data for the uniform control floods were plotted within the calculated control limits and there were no false positives noted in the uniform time series data. This indicates that appropriate limits were established.

The CRSD control charts did not exhibit the same sensitivity to changes in uniformity performance as the NEMA calculations. An out-of-control signal was noted when the value crossed the upper control limits and remained above this set limit. This demonstrated a trend or change in the process parameters. The time series plots for phantom synthetic\_15, which had the highest rate of degradation, showed that both the integral and differential uniformity (Figure 37a.) signaled an out-of-control shift weeks before control limits for the CRSD plot (Figure 37b.) were violated.



a.





b.

**Figure 37 – Shewhart Control Charts for Phantom Synthetic\_15**

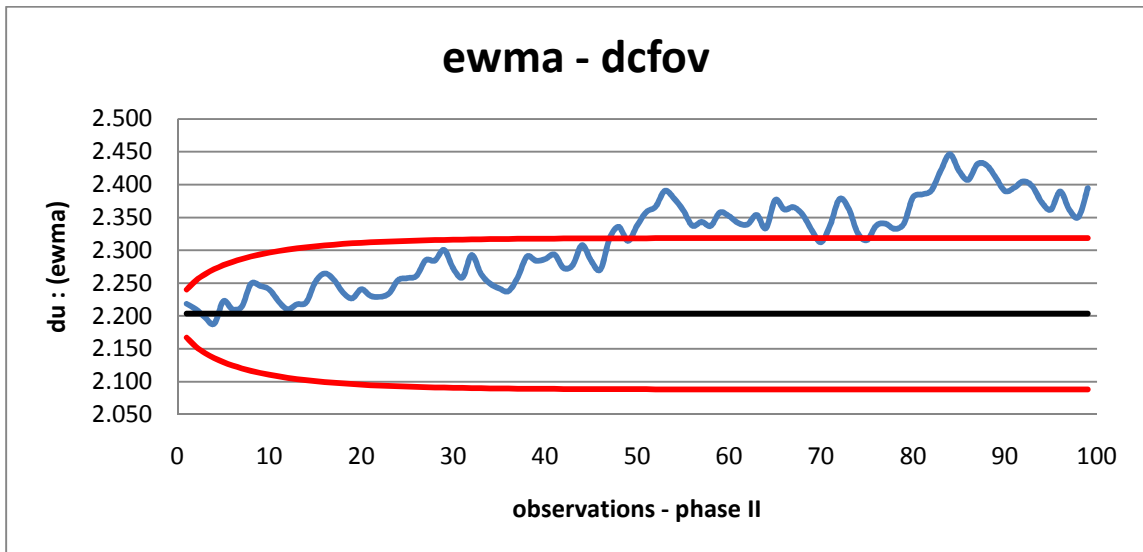
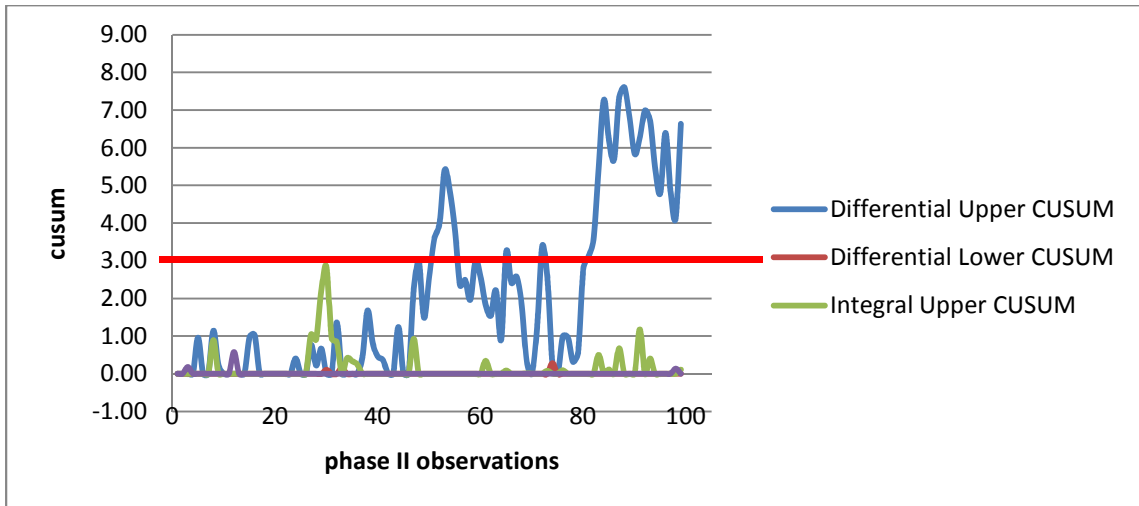
*Top (a): Differential uniformity signals an early trend outside of the control limits (red lines).*

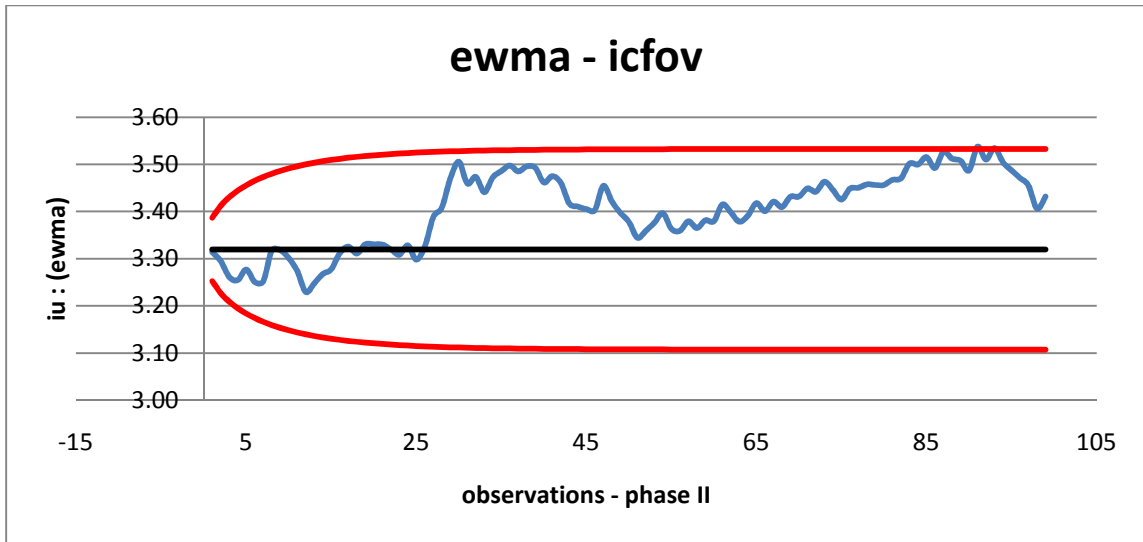
*Bottom (b): CRSD uniformity is within the set limits for most of the time series.*

This observation is explained by the fact that the CRSD is calculated over all of the pixels present in the UFOV. Any subtle differences in pixel values can be masked by the entire distribution of pixels. The NEMA calculation, on the other hand, is essentially a range calculation between the maximum and minimum values within the FOV. The result is that while NEMA may give rise to more random fluctuations in a time series plot of values (as demonstrated by Young, et al, (15)), it does seem to be more sensitive in identifying sustained parametric shifts. Moreover, CRSD has not gained wide use in clinical settings. The remaining results focus on the NEMA uniformity index.

The integral uniformity value was more sensitive than the differential uniformity to detecting trends in most of the time series phantoms. For the 8 cm diameter non-uniform region in phantom synthetic\_13, the Shewhart control charts for the integral uniformity signaled a shift approximately 20 days prior to shift detection by the differential uniformity method. The off-peak phantom data yielded similar results, in that the differential uniformity Shewhart chart never signaled an out-of-control process, while the integral uniformity Shewhart chart demonstrated multiple violations over the course of the time series as the energy window shifted toward and then away from the photopeak. The EWMA and CUSUM control charts exhibited similar differences between the integral and differential uniformity values. The only exceptions to this observation were the results for the crystal

hydration time series, ts\_vol\_S5 (Figure 38). The differential uniformity index signaled a process shift in both the CUSUM and EWMA control charts, while the integral uniformity never did.





**Figure 38 – CUSUM and EWMA Control Charts for Crystal Hydration Time Series**

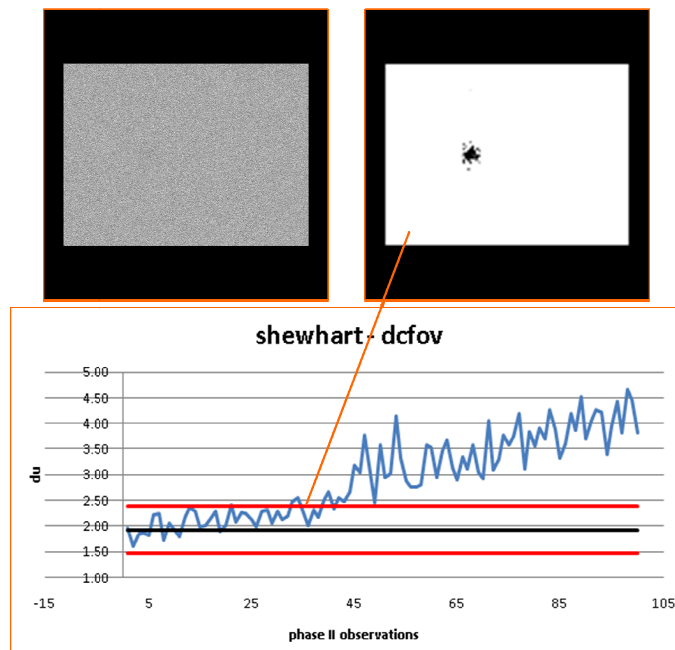
*Phantom: ts\_vol\_S5*

The chart data for ts\_vol\_S5 also demonstrated the similarities in sensitivity between the EWMA and CUSUM methods. Both methods signaled a violation at approximately day 50 in the time series, but the CUSUM statistic eventually fell back under the threshold of three. While the CUSUM later exceeded this value again, the EWMA demonstrated a sustained shift above the upper control limit. Both the CUSUM and EWMA control charts showed improved sensitivity over the Shewhart method. For example, in the crystal hydration time series images, the Shewhart control limits were never violated for either the integral or the differential uniformity values as opposed to the violations noted in Figure 38 for the EWMA and CUSUM charts.

The control plot results for the time series phantom ts\_vol\_T6, in which a defective PMT was observed in the last image of the image set, presented an interpretive challenge. The control limits for the integral uniformity in all three plotting methods showed violation within the first few days of the time series. The Shewhart and EWMA chart indicated an eventual return back within control limits, but the CUSUM chart never stabilized. Because the control limits for the time series were based on only 20 floods acquired over successive days, the parameters may not have been accurately characterized in the control limit calculation. However, this same approach was used for the hydration time series data and the performance of the charts suggests that the control limits were appropriate because no discernable false positives were present early in the time series. Furthermore, the detector used to acquire the floods for ts\_vol\_T6 had several service events over the course of the time series run, indicating that image quality issues had been noticed by the imaging technologist who acquired the QC floods. The performance of the detector was improved in order to address these

deficiencies. Therefore, it is unclear whether the initial control plot violations were due to a systematic defect of the gamma camera detector or this was expected random error for which less restrictive control limits need to be established.

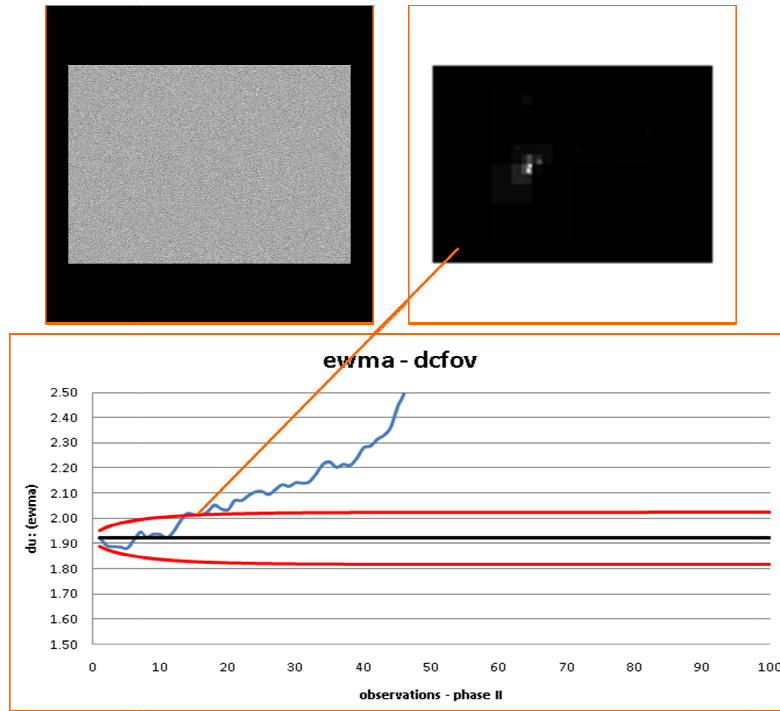
The multi-resolution decomposition algorithms complemented the sensitivity to uniformity changes exhibited in the control charts. In cases where the size of the region of non-uniformity was large, as in synthetic\_13 and synthetic\_15, the control plots for the NEMA values effectively detected the sustained shift in performance early on in the time series. The figures below depict the parametric images calculated around the time points at which the trends in uniformity values exceeded the control limits illustrating that the specificity provided by the maps in terms of visually locating the uniformity defects complemented the sensitivity of the time series control charts. Again, the shift in the NEMA integral uniformity values was more sensitive, and the corresponding parametric map at the time point of the control violation depicted the region of non-uniformity in a nascent stage of progression.



**Figure 39 – Multi-resolution Strength-of-correlation at Point of Shewhart Differential Uniformity Control Violation**

*Original flood shown to the left of the parametric result*

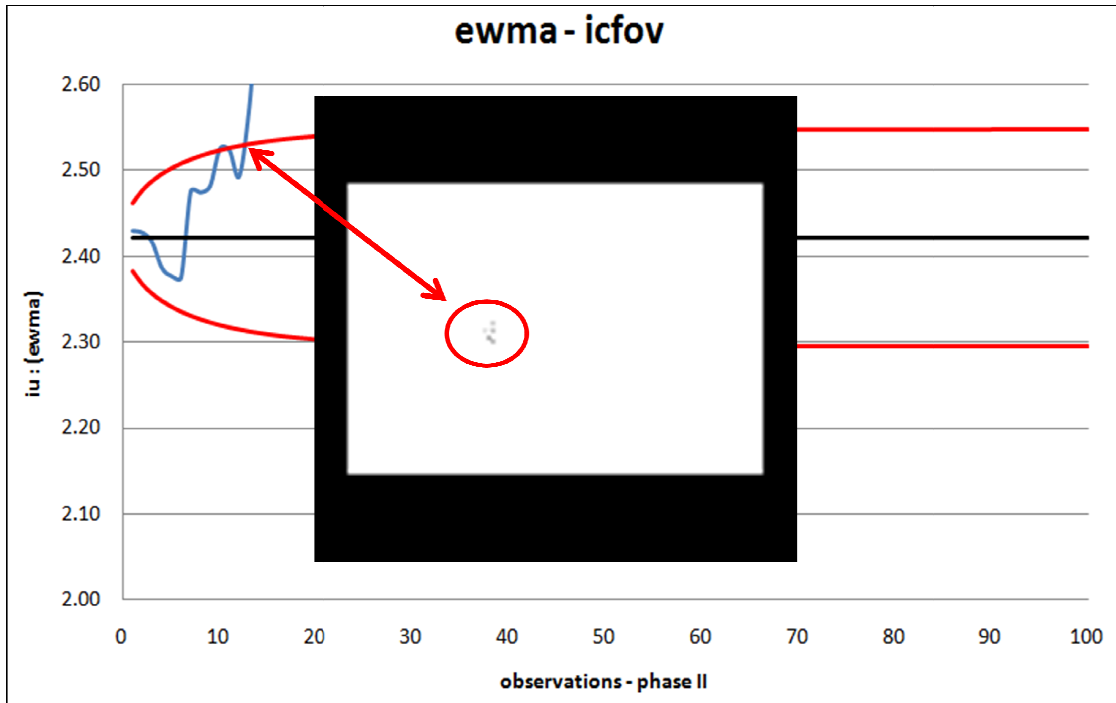
*Phantom: synthetic\_13*



**Figure 40 - Multi-resolution MDA at Point of EWMA Differential Uniformity Control Violation**

*Original flood shown to the left of the parametric result*

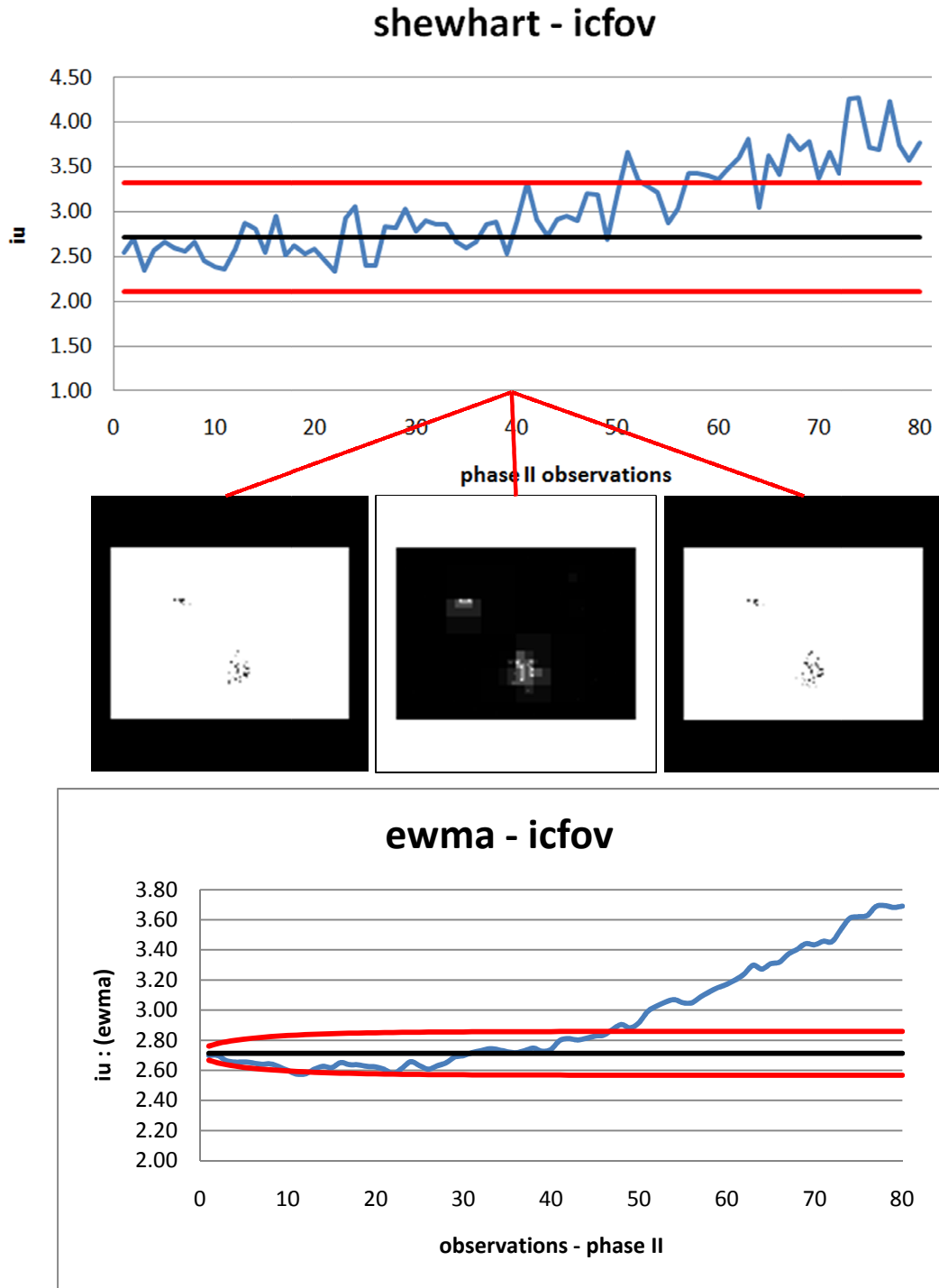
*Phantom: synthetic\_13*



**Figure 41 – Multi-resolution Strength-of-correlation Result at the Point of EWMA Integral Uniformity Control Chart Violation**

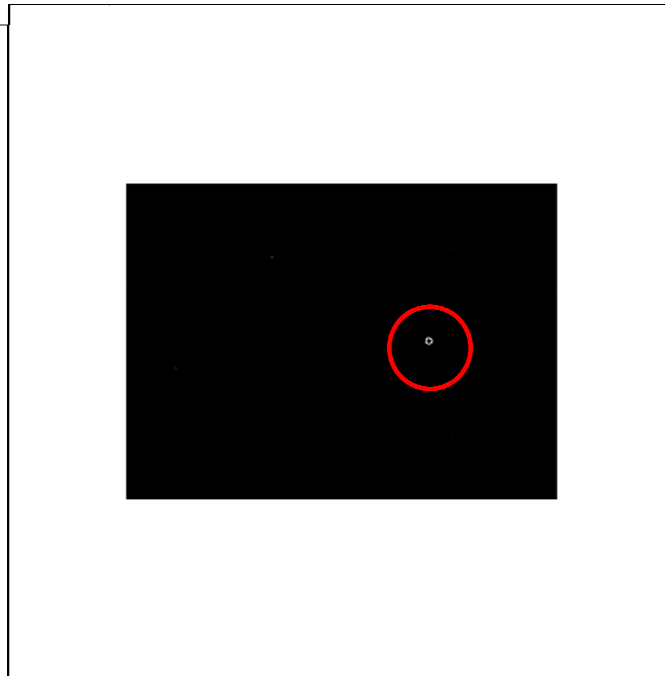
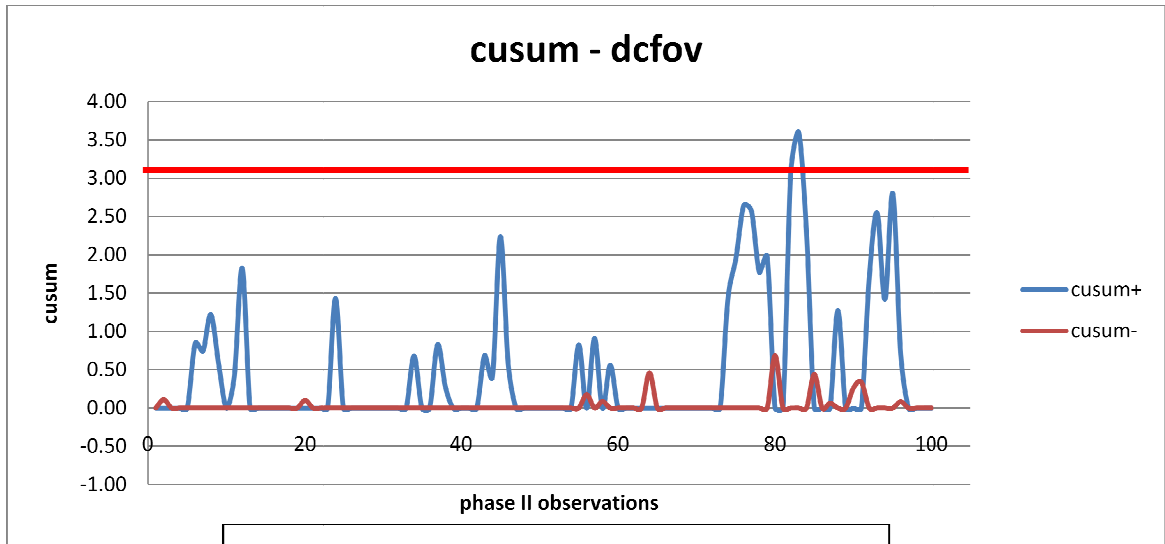
*Phantom: synthetic\_13*

The time points of control limit violations in the attenuation disk time series were similar for the three different statistical process control techniques. The multi-resolution algorithms segmented the larger two disk sizes well in advance of the control limit violation, demonstrating a case when the image-space techniques offered greater sensitivity than the global uniformity trend analysis. The rate of degradation in this time series was low (0.1%) compared to the synthetic phantom data and the NEMA uniformity values were less sensitive to detecting the subtle, localized count losses in the FOV. Linear, slower rates of degradation were segmented the multi-resolution algorithms before trends were signaled in the global control charts, while larger, rapidly changing regions of degradation signaled trends in the charts prior to being segmented in the parametric maps. Furthermore, small uniformity defects such as those present in synthetic phantoms 17 and 19 were segmented in the time series images using the sliding window t-test approach, but never resulted in a sustained shift in the values above a control limit. The CUSUM and EWMA results for synthetic\_19 contained points toward the end of the time series run that crossed the control limit, but in each case, the trend returned to being within control. However, the region was segmented in the parametric result well in advance of this.



**Figure 42 – Parametric Results for Attenuation Disk Images at Day 40**

*Top: Shewhart chart for integral uniformity. Middle: multi-resolution t-test, MDA, and Strength-of-correlation results at observation 40. Bottom: EWMA control chart*



**Figure 43 – Parametric Result for Synthetic\_19, Day 40**

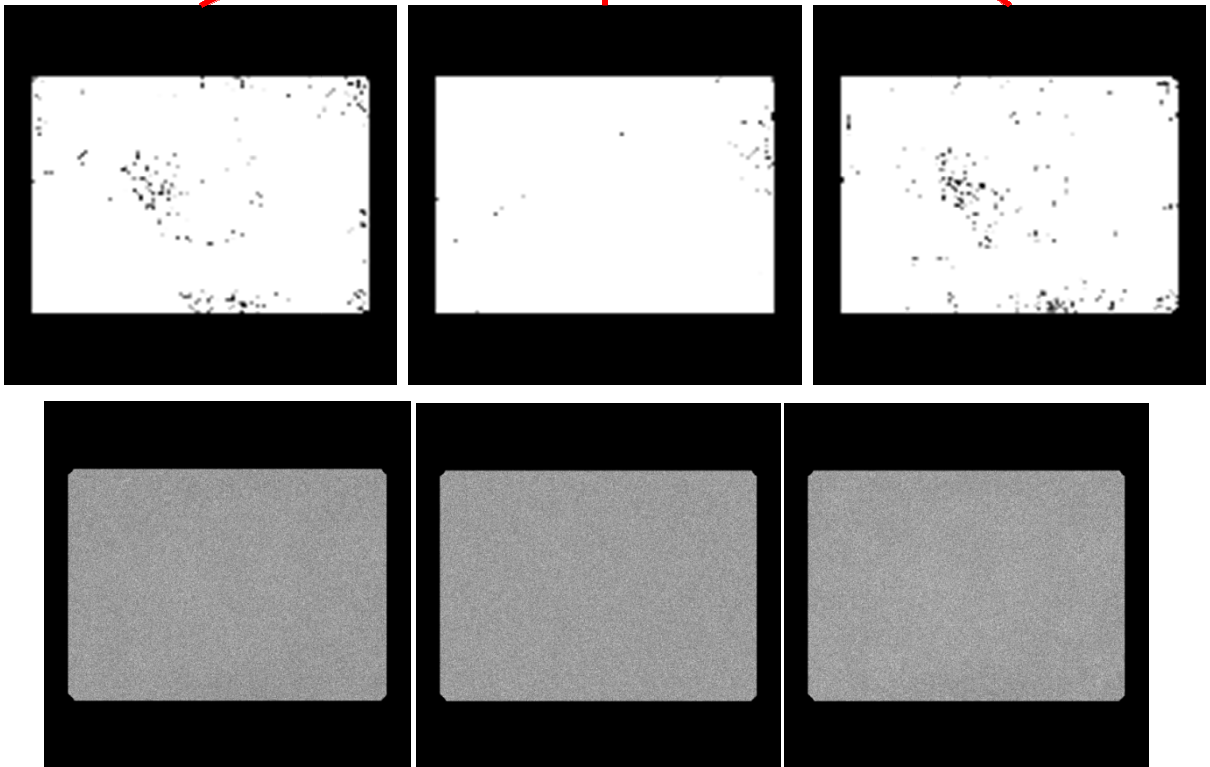
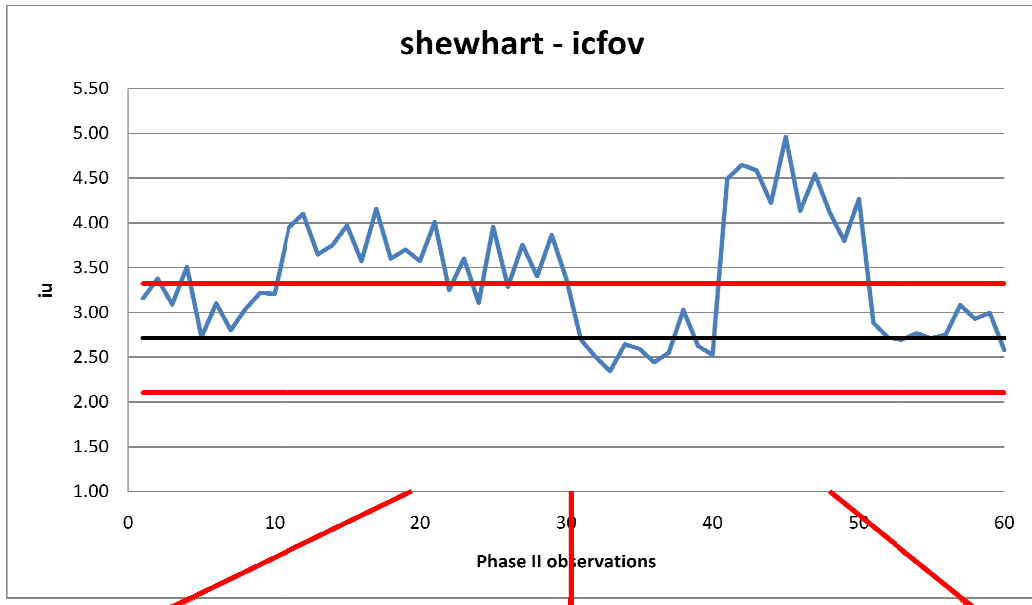
*Violation of CUSUM differential uniformity control chart after day 80. Non-uniformity segmented well in advance.*

The EWMA and CUSUM differential uniformity value control charts for the differential uniformity values for the crystal hydration time series signaled a change in uniformity performance well in advance of the hydration spots' being discerned in the multi-resolution parametric results. While no confirmation is available that this was the result of the formation of hydration spots in the CFOV, it does demonstrate the utility of using the global trend charts concomitantly with the parametric maps. If the global uniformity results begin trending beyond the established control



limits, the parametric maps help to confirm the presence of isolated uniformity defects. Likewise, if segmented regions observed in the parametric maps correspond to the initiation of a trend in the global figures-of-merit, there is a more sound statistical case that a significant change in performance has occurred and that a service action should be initiated.

The control charts effectively signaled trends in the non-linear degradation process demonstrated by the off-peak time series phantom (Figure 44). As the acquisition energy windows fluctuated toward and away from the  $^{57}\text{Co}$  photopeak, corresponding shifts outside of the established control limits were noted. This corresponded with the diffuse segmentation pattern that resulted in the parametric maps for days in which the shift was signaled in the control charts.



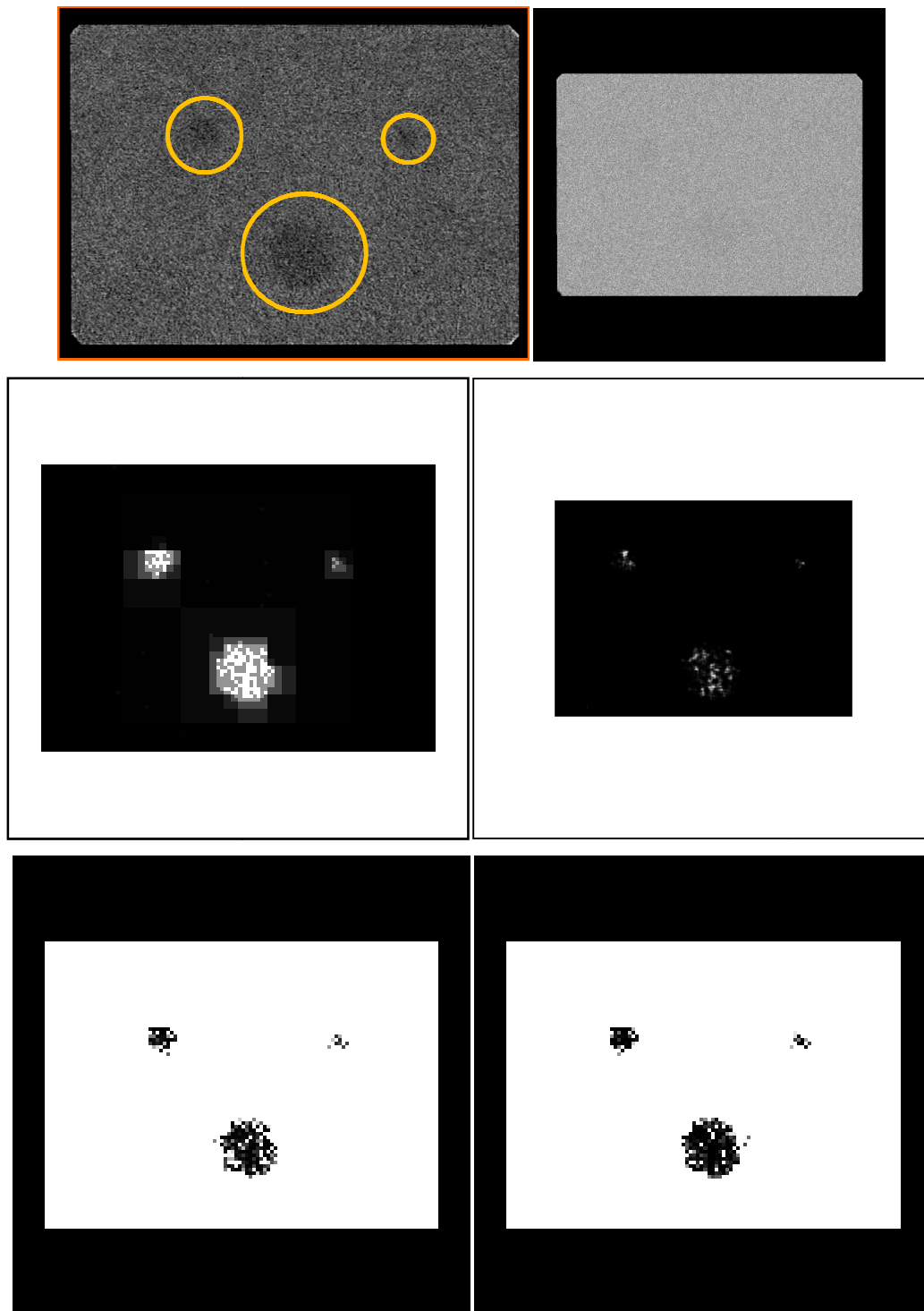
**Figure 44 – Comparison of Multi-resolution Maps to Shewhart Control Chart for Off-peak Time Series**

*Top: Shewhart Control Chart; Middle: Strength-of-correlation maps for days 19, 30, and 48, respectively; Bottom: Original flood images for corresponding days*

## 6. CONCLUSIONS

This project was a proof-of-principle study that implemented several novel methods of using registered time series floods in order to enhance spatial regions of performance degradation within the detector field-of-view for improving detectability. The results demonstrated a sensitive means of monitoring gamma camera uniformity by improving the detectability of progressive uniformity defects. Clinically, this is expected to minimize scanner downtime by identifying trends in uniformity performance at a point in time before they are manifested in patient images.

The hypothesis stated a specific, clinically relevant rate of degradation against which to evaluate the algorithms. The time series attenuation disk phantom (ts\_vol\_3disk) represented a degradation rate of less than 0.1%. The highest value for the NEMA integral uniformity was 4.27%, which is below the 5% threshold. The addition of attenuating material to the FOV was a linear process, occurring at evenly spaced time intervals, and the NEMA uniformity values were expected to eventually exceed 5%. The multi-resolution parametric maps at observation 70 in the time series (10 days prior to the last observation) showed that the 2, 4, and 8 cm attenuation disk regions were all segmented from the background pixels (Figure 45). The NEMA integral uniformity value for this day was 3.37%. Thus, given this rate of degradation over regions of diameter 2, 4, and 8 cm, multi-resolution algorithms recognized the non-uniformities before there was a single flood with a NEMA value in excess of 5%.



**Figure 45 – Parametric Results for Attenuation Disk Phantom – Day 70**

*Top Left: Disk location reference; Top Right: Original daily flood at day 70. Middle Left: Multi-resolution MDA map at day 70; Middle Right: Sliding Window t-test map at day 70; Bottom Left: Multi-resolution t-test map at day 70; Bottom Right: Multi-resolution strength-of-correlation map at day 70.*

In addition to the algorithms' segmentation of the non-uniformities in the FOV prior to the NEMA uniformity value being above 5%, the time series control plots identified the trend in the uniformity values before this threshold was crossed. All three charts signaled a control limit violation approximately 30 days prior to the end of the time series. This implies that a distinct trend in the data was signaled prior to the 5% threshold violation. Furthermore, the segmentation algorithms were able to detect the two larger regions of degradation in the image before a trend was indicated in the NEMA uniformity values. While this was the case for slowly progressing non-uniformities, larger, rapidly degrading regions were quickly detected by the control chart mechanisms, even before the spatial extent of the non-uniformity was fully realized in the parametric maps. Overall, the phantoms synthetic\_17, synthetic\_19, ts\_vol\_S5, and ts\_vol\_3disk illustrated the scenarios in which the time series segmentation tools were most useful: smaller regions of degradation that slowly progress in magnitude as a function of time. In these cases, the NEMA threshold was never violated and there is relatively little shift in the trending performance of the global figures-of-merit. These methods provide the user with the spatial location and size of the region of non-uniformity based on a difference from the baseline value. The effectiveness of the control charts at detecting rapid degradation and the sensitivity of the multi-resolution techniques at segmenting subtle degradation indicates that concomitant use of these tools would provide optimal sensitivity in monitoring flood uniformity. There are time series plots of the NEMA uniformity values over time for each of the phantoms provided in 9.1, along with a summary table of the performance of the different monitoring techniques. For the global figure-of-merit control chart techniques, the point in time at which each chart began to show multiple points plotted outside of the control limits is marked with a corresponding colored line. For the image-space techniques, the point in time at which pixels are segmented in the parametric maps at the location of the true defect for at least 5 continuous days is marked with a corresponding colored line. These plots demonstrate the relative performance of each technique. In summary, the control chart techniques provide greater sensitivity to the size and rate of degradation than the image-space techniques. The EWMA and CUSUM charts consistently outperform the Shewhart charts in terms of sensitivity. The charts provided a truly autonomous method of monitoring for defects that is not dependent upon visual interpretation. However, the image-space algorithms provide more specificity than the control chart results do because they provide the reviewer with the exact location and relative size of the defects. For this reason it is recommended that the control charts be used to automatically monitor the time series uniformity performance, and when signals are triggered by observations that fall beyond the control limits, the parametric maps from the multiresolution t-test and strength-of-correlation test be used to diagnose the defects.

The tools developed and presented above provide methods of alerting the reviewer to subtle changes in uniformity performance by including a temporal component to both the images and global figures-of-merit. They offer a predictive aspect to monitoring uniformity in that they make assumptions about expected behavior based on previous observations. When significant changes in either uniformity indices or pixel values occur, the tools provide a method of signaling the reviewer of the change, either in the form of a control limit violation or the specific location of the degraded region in a parametric map. The segmentation techniques enhance the contrast of statistical non-uniformities and provide an automated method of isolating such regions from uniform pixels.

This thesis has presented several novel, alternative methods of visually and quantitatively monitoring flood uniformity by exploiting the temporal nature of gamma camera detector degradation. These methods were tested using several different types of time series data sets in order to demonstrate the feasibility of the algorithms. Clinically, these algorithms should be run daily for every detector after new QC flood acquisitions are acquired. The reviewer would inspect the parametric maps produced by each algorithm as well as review the control chart of the calculated NEMA uniformity values. As was discussed in section 5.2, the multi-resolution techniques demonstrated similar sensitivity, but given the number of false positives identified by the MDA technique, as well as the somewhat arbitrary manner in which contrast is enhanced in the parametric maps, the most suitable multi-resolution techniques to implement clinically are the t-test, strength-of-correlation, and sliding-window t-test algorithms. In addition to this, the NEMA uniformity time series control charts demonstrate sensitivity to detecting shifts in uniformity performance, especially for larger uniformity defects, and using them concomitantly with the image-space algorithms would provide the reviewer with more certainty when determining the state of performance of the clinical system.

## 7. FUTURE WORK

This project was a proof-of-principle study to demonstrate the utility of improving the sensitivity in uniformity monitoring by adding an automated, temporal component to QC floods as well as statistical process control. Several potential project threads would further lay the foundation for its clinical implementation.

### 7.1 Reviewer-based Studies

Several aspects of this work should be verified in a reviewer-based study. First, multiple reviewers should verify the time to detection of each of the algorithms for the different phantom types used. While the automated segmentation algorithms essentially produce a high-contrast map of the non-uniformity locations, which is easily interpreted, the reviewer must still visually verify the presence and location of the segmented regions and the point at which they are first observed and compare this to the results in original raw flood images. Because this still permits some subjectivity in the review process, a reviewer-based study would adequately characterize the performance of the algorithms. Secondly, the threshold levels established ( $p < 0.01$ ) for the multi-resolution techniques should be validated in a reviewer-based study in order to determine the optimality of the different levels of significance. The levels chosen for this study were kept low in order to minimize false positive segmentations in the parametric maps, however, clinically there may be some benefit to using larger threshold values in terms of gaining sensitivity to detecting non-uniformities earlier in the time series. This can be tested using a reviewer-based approach. Furthermore, it was discovered during the course of this investigation that a potential method of viewing the parametric images produced by the algorithms is in a cine loop. In order to discover if there are inherent sensitivity gains in reviewing QC in this manner, in terms of detecting any uniformity defects, a reviewer-based study would be needed.

### 7.2 Algorithm Optimization

To maximize the sensitivity to detecting steeper rates of degradation at an earlier time point, multi-resolution decomposition in the temporal dimension should be developed and applied to the time-windowed multi-resolution algorithms. While using a larger time bin improves the SNR of the pixels as a function of time, it also has the potential for averaging out rapid changes as floods accumulate. Decomposing the temporal dimension using successively smaller time windows should improve the sensitivity of the algorithm. While this may increase the rate of false positives detected,

it may be beneficial to improve sensitivity to dramatic shifts in the pixel samples. Weighting the time windows differently according to their temporal size such that larger time windows are weighted more heavily than smaller windows should help to minimize the false positive rate while gaining sensitivity to detecting more abrupt changes in the pixel sub-volumes.

The correct registration of the FOVs in the time series is critical in order to ensure that same pixels are being compared between two time points in the multi-resolution techniques. The data analyzed in this study was inherently registered, but shifts in the location of the UFOV within the digital image matrix can occur, and therefore the programs need to account for these potential shifts in order to make sure that the images are registered throughout the time series.

Another project thread would be to use the temporal multi-resolution decomposition maps to derive a function that will help characterize the relationship between the size of the region of non-uniformity, rate or magnitude of degradation, and time.

A running log of the false positive pixels identified by each of the segmentation algorithms may also provide an effective global figure-of-merit that can be used in a similar manner as the NEMA uniformity indices. Calculating the difference between the expected number of false positive pixels and the actual number present in the parametric maps may provide a sensitive method of monitoring flood uniformity. The reviewer would then be able to use the visual specificity provided by the maps along with the statistical probability provided by the false positive rate.

### **7.3 Extensive Retrospective Analysis**

The retrospective data sets processed in this study demonstrated that distinct trends were detected using both the multi-resolution algorithms as well as the time series statistical process control plots. However, the results were somewhat ambiguous and it is not well understood if many of the segmented regions in the parametric maps were true systematic non-uniformities or simply expected changes in the pixels from the baseline condition. It is also not well understood whether the baseline replicate sets used in the retrospective time series data adequately characterized the expected uniformity behavior for these systems. By processing the retrospective daily floods that did not have any noted performance defects, as well as processing more flood images from gamma cameras with both crystal hydration and PMT defects, the differences in expected behavior and systematic detector error can be assessed using the algorithms developed in this project. In this way, it can be determined whether or not common clinical defects such as PMT failure can, in fact, be predicted.



## **7.4 Study of Clinical Implications of Uniformity Defects**

In terms of the global figures-of-merit, it would be valuable to investigate the direct relationship between the calculated uniformity in a single 10-million count flood, and the clinical significance of that uniformity in terms of patient image quality. This would give the 5% threshold, which the algorithm results in this study were compared against, some clinical context in terms of how NEMA uniformity translates to patient image quality. To do this, the binary PMT degradation map (using all PMT locations to adequately sample the UFOV) should be multiplied by the individual projection images of a SPECT data set, and the images tomographically reconstructed. The same degradation map could then be multiplied into a daily flood image and the NEMA values calculated. This would provide a direct correlation of NEMA uniformity values to patient image quality. The magnitude of the degraded regions could be varied in order to determine the percent uniformity at which patient image quality is adversely affected. The degradation maps (see section 5.1.1.1) could also be multiplied directly into a planar patient acquisition, such as a bone scan, and the effects demonstrated using, for example, a contrast-to-noise calculation.

## **7.5 Software Integration**

A final goal is to integrate the algorithm programs presented in the thesis into the clinical software environment at UT MD Anderson Cancer Center. These tools can be used to help characterize actual clinical modes of non-uniformity, such as PMT failure or crystal hydration. These processes are not well understood at this point, and by no means definitively predictable, but including a temporal component in order to improve the ability to detect them could provide useful information regarding their spatial and temporal characteristics.

The groundwork laid in this thesis has identified novel methods of providing a functional, temporal aspect to clinical quality control images and has demonstrated the feasibility of these modifications to uniformity QC review. The future work detailed above will provide several benchmarks for the continuation of research on this topic.

## 8. BIBLIOGRAPHY

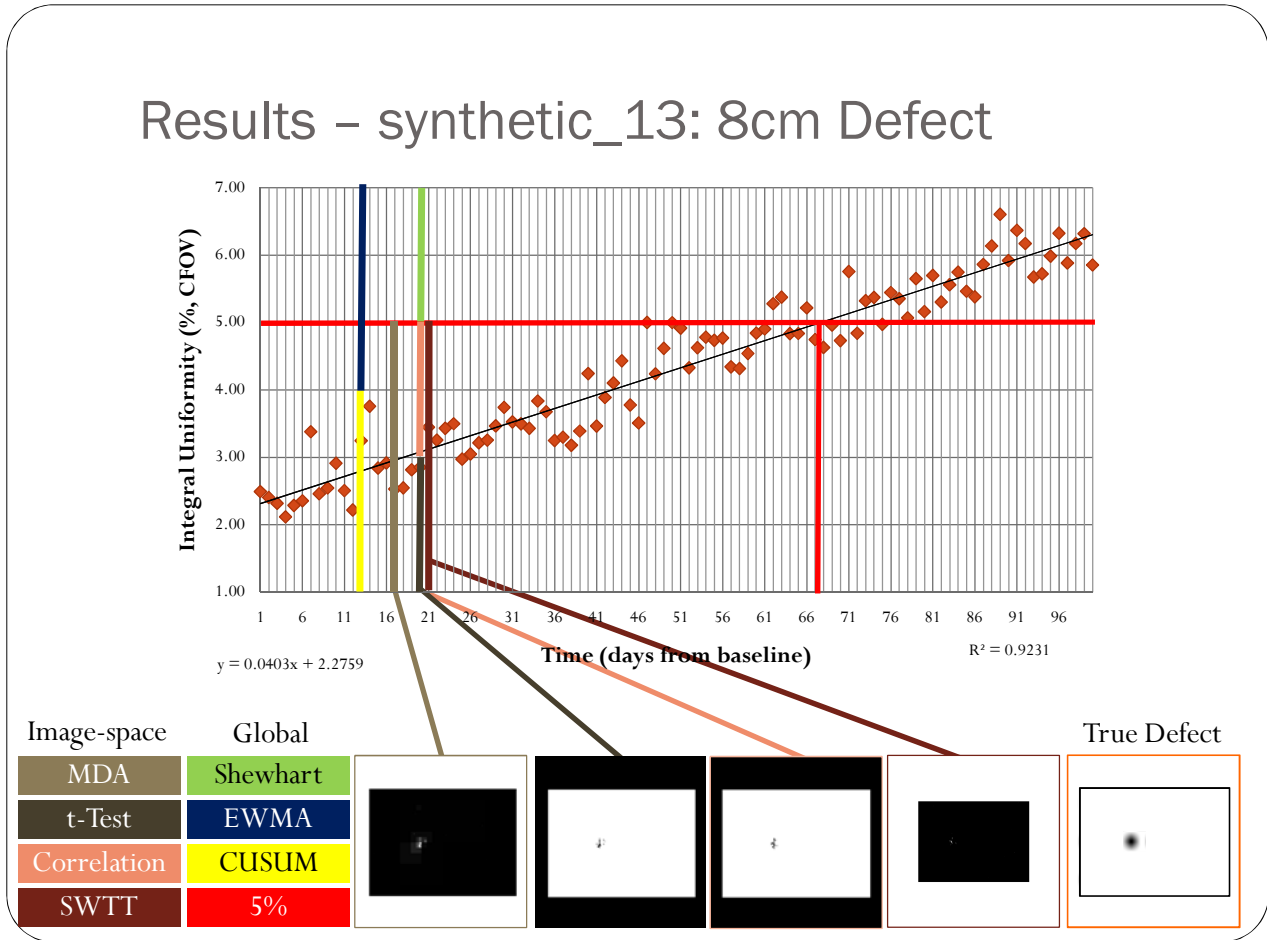
1. Kalemis, A., D. Bailey, M. Flower, S. Lord, and R. Ott. 2004. Statistical pixelwise inference models for planar data analysis: an application to gamma-camera uniformity monitoring. *Phys Med Biol* 49:3047-3066.
2. Hines, H., R. Kayayan, J. Colsher, D. Hashimoto, R. Schubert, J. Fernando, V. Simcic, P. Vernon, and R. Sinclair. 1999. Recommendations for implementing SPECT instrumentation quality control. Nuclear Medicine Section--National Electrical Manufacturers Association (NEMA). *Eur J Nucl Med* 26:527-532.
3. Cherry, S., J. Sorenson, and M. Phelps. 2003. *Physics in Nuclear Medicine*. Saunders, Philadelphia, PA.
4. Peller, P., V. Ho, and M. Kransdorf. 1993. Extrasosseous Tc-99m MDP uptake: a pathophysiologic approach. *Radiographics* 13:715-734.
5. Johns, H., and J. Cunningham. 1983. *The Physics of Radiology*. Charles C Thomas, Springfield, Ill.
6. O'Connor, M. K. 1996. Instrument- and computer-related problems and artifacts in nuclear medicine. *Semin Nucl Med* 26:256-277.
7. Murphy, P. 1987. Acceptance testing and quality control of gamma cameras, including SPECT. *J Nucl Med* 28:1221-1227.
8. Halama, J., and R. Henkin. 1987. Quality assurance in SPECT imaging. *Appl Radiol* 16:41-42, 44, 46 passim.
9. Rogers, W., N. Clinthorne, B. Harkness, K. Koral, and J. Keyes Jr. 1982. Field-flood requirements for emission computed tomography with an Anger camera. *J Nucl Med* 23:162-168.
10. O'Connor, M., and C. Vermeersch. 1991. Critical examination of the uniformity requirements for single-photon emission computed tomography. *Med Phys* 18:190-197.
11. Halama, J. R., Madsen, M.T. 1992. Is your gamma camera-SPECT system installed and operating optimally? *Applied Radiology*:35-41.
12. Uribe, J., H. Li, H. Baghaei, M. Aykac, Y. Wang, Y. Liu, and W. Wong. 2001. Effect of photomultiplier gain-drift and radiation exposure on 2D-map decoding of detector arrays used in positron emission tomography. In *Nuclear Science Symposium, 2001 IEEE*, San Diego, CA. 1960 - 1964
13. Keszthelyi-Landori, S. 1986. NaI(Tl) camera crystals: imaging capabilities of hydrated regions on the crystal surface. *Radiology* 158:823-826.

14. Zanzonico, P. 2008. Routine quality control of clinical nuclear medicine instrumentation: a brief review. *J Nucl Med* 49:1114-1131.
15. Young, K., K. Kouris, M. Awdeh, and H. Abdel-Dayem. 1990. Reproducibility and action levels for gamma camera uniformity. *Nucl Med Commun* 11:95-101.
16. Montgomery, D. C. 2009. *Introduction to Statistical Quality Control*. Wiley, Hoboken, NJ.
17. Knoll, G. F. 2000. *Radiation Detection and Measurement*. Wiley, Hoboken, NJ.
18. Bushberg, J., J. Seibert, E. Leidholdt, and J. Boone. 2002. *The Essential Physics of Medical Imaging*. Lippincott Williams and Williams, Philadelphia.
19. Raff, U., V. Spitzer, and W. Hendee. 1984. Practicality of NEMA performance specification measurements for user-based acceptance testing and routine quality assurance. *J Nucl Med* 25:679-687.
20. NEMA. 2001. *NEMA NU 1-2001: Performance Measurements of Scintillation Cameras*. National Electrical Manufacturers Association, Rosslyn, VA, USA.
21. Cox, N., and B. Diffey. 1976. Letter: A numerical index of gamma-camera uniformity. *Br J Radiol* 49:734-735.
22. Waddington, W., G. Clarke, K. Barnes, G. Gillen, A. Elliott, and M. Short. 1995. A reappraisal of current methods for the assessment of planar gamma camera performance. *Nucl Med Commun* 16:186-195.
23. Hughes, A., and P. Sharp. 1987. Factors affecting gamma-camera uniformity. *Phys. Med. Biol.* 33:259-269.
24. Chapra, S., and R. Canale. 2002. *Numerical Methods for Engineers*. McGraw Hill, Boston, MA.
25. Press W.H., T. S. A., Vetterling W.T., Flannery B.P. 2002. *Numerical Recipes in C++*. Cambridge University Press, Cambridge.
26. Kalemis, A., D. Binnie, D. Bailey, M. Flower, and R. J. Ott. 2003. Scaling images using their background ratio. An application in statistical comparisons of images. *Phys Med Biol* 48:1539-1549.
27. Gonzalez, R., and R. Woods. 2008. *Digital Image Processing*. Pearson Prentice Hall, Upper Saddle River, NJ.
28. Tenhunen, M., J. Pyykkonen, M. Tenhunen-Eskelinen, K. Jaatinen, and J. Kuikka. 1996. Components of the flood-field uniformity index in a gamma camera. *Phys Med Biol* 41:1209-1216.
29. Rosenthal, M., J. Cullom, W. Hawkins, S. Moore, B. M. Tsui, and M. Yester. 1995. Quantitative SPECT imaging: a review and recommendations by the Focus Committee of the Society of Nuclear Medicine Computer and Instrumentation Council. *J Nucl Med* 36:1489-1513.

30. Box, G., and M. Muller. 1958. A note on the generation of random normal deviates. *Ann Math Statist* 29:610-611.
31. Shleien, B., L. Slaback Jr., and B. Birky, editors. 1998. *Handbook of Health Physics and Radiological Health*. Williams and Wilkins, Baltimore, MD.
32. Burt, P., T. Hong, and A. Rosenfeld. 1981. Segmentation and Estimation of Image Region Properties Through Cooperative Hierarchical Computation. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-11:802-809.
33. Wendt III, R. E. 1982. Restoration and Functional Analysis of Nuclear Medicine Images. In *Electrical Engineering*. Rice University, Houston, TX. 88.
34. Pearson, D., and S. Cawte. 1997. Long-term quality control of DXA: a comparison of Shewhart rules and Cusum charts. *Osteoporos Int* 7:338-343.

## 9. APPENDIX

### 9.1 Summary of Results



# Results – synthetic\_15: 8cm Defect, Rtx2

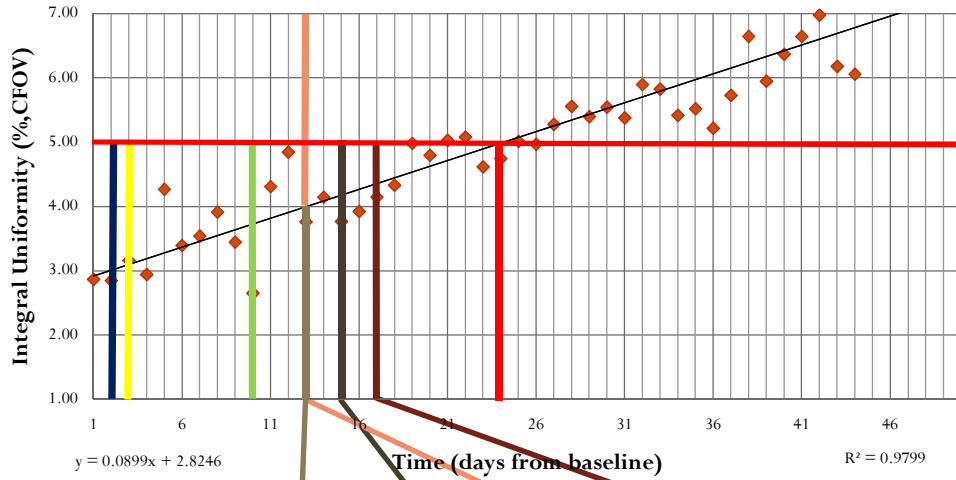


Image-space	Global	True Defect			
MDA	Shewhart				
t-Test	EWMA				
Correlation	CUSUM				
SWTT	5%				

# Results – ts\_vol\_3disk (Attenuation Disks)

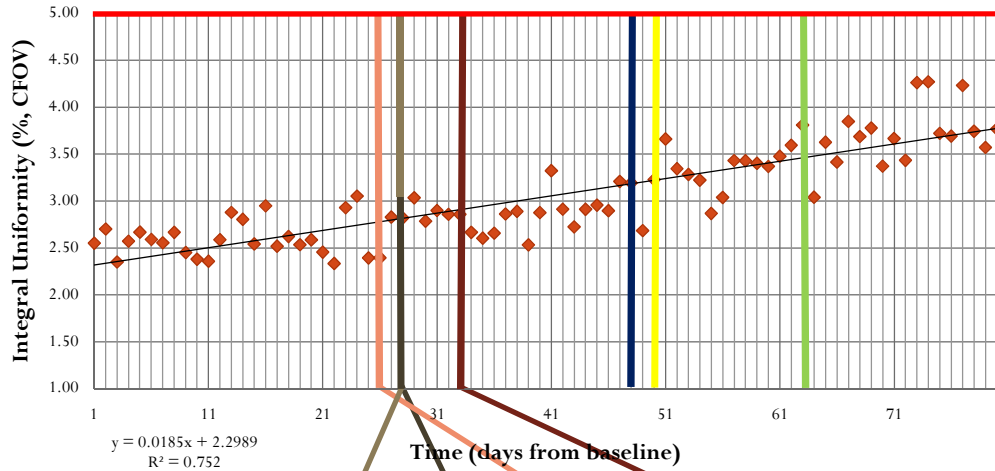
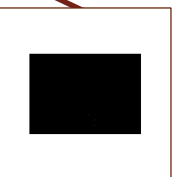
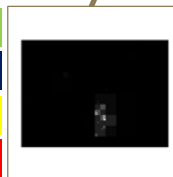


Image-space

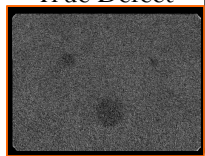
MDA
t-Test
Correlation
SWTT

Global

Shewhart
EWMA
CUSUM
5%



True Defect



# Results – ts\_vol\_3disk (Attenuation Disks)

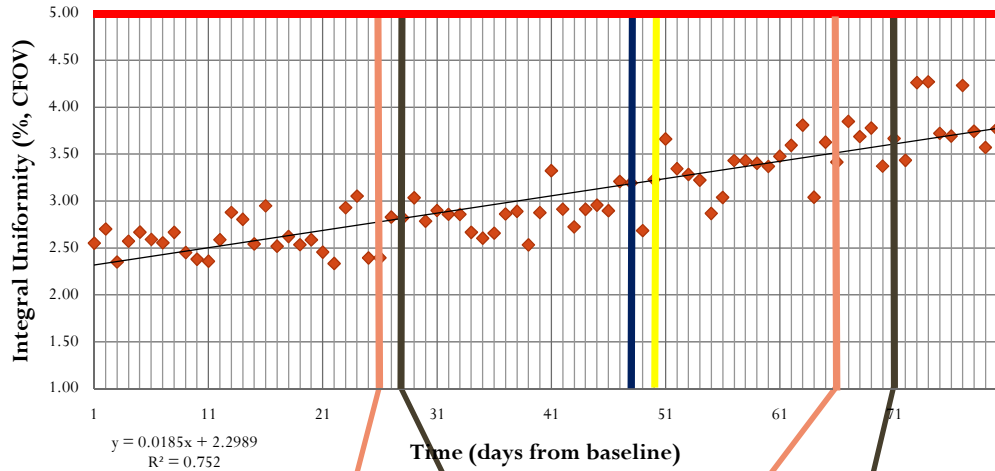


Image-space

t-Test

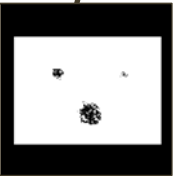
Correlation

Global

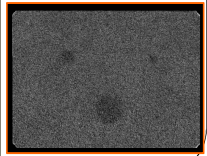
EWMA

CUSUM

5%



True Defect





## Results – ts\_vol\_offpeak (E-window shifts)

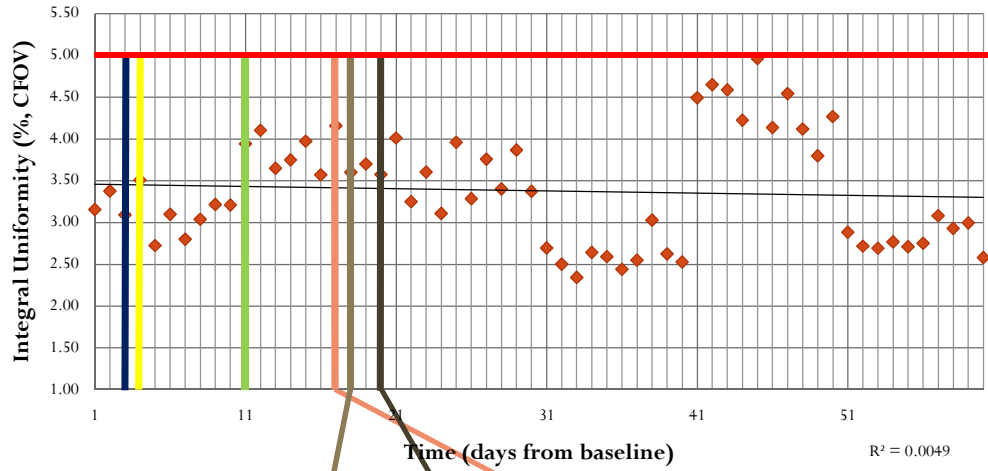


Image-space

MDA

t-Test

Correlation

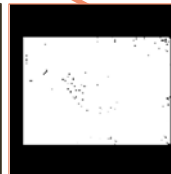
Global

Shewhart

EWMA

CUSUM

5%



# Results – ts\_vol\_S5 (Crystal Hydration)

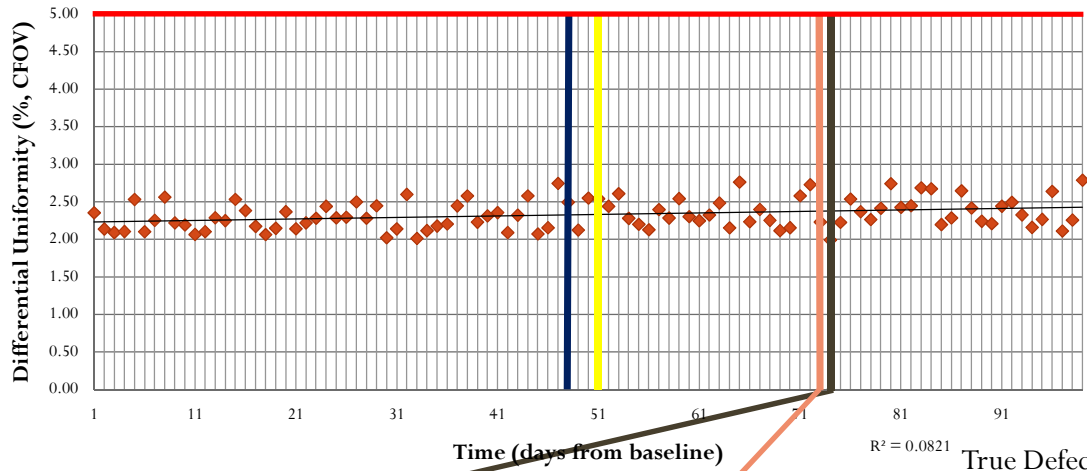


Image-space

Global

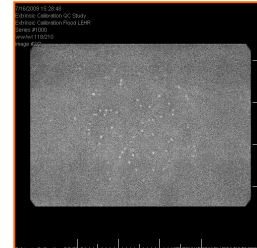
t-Test

EWMA

Correlation

CUSUM

5%



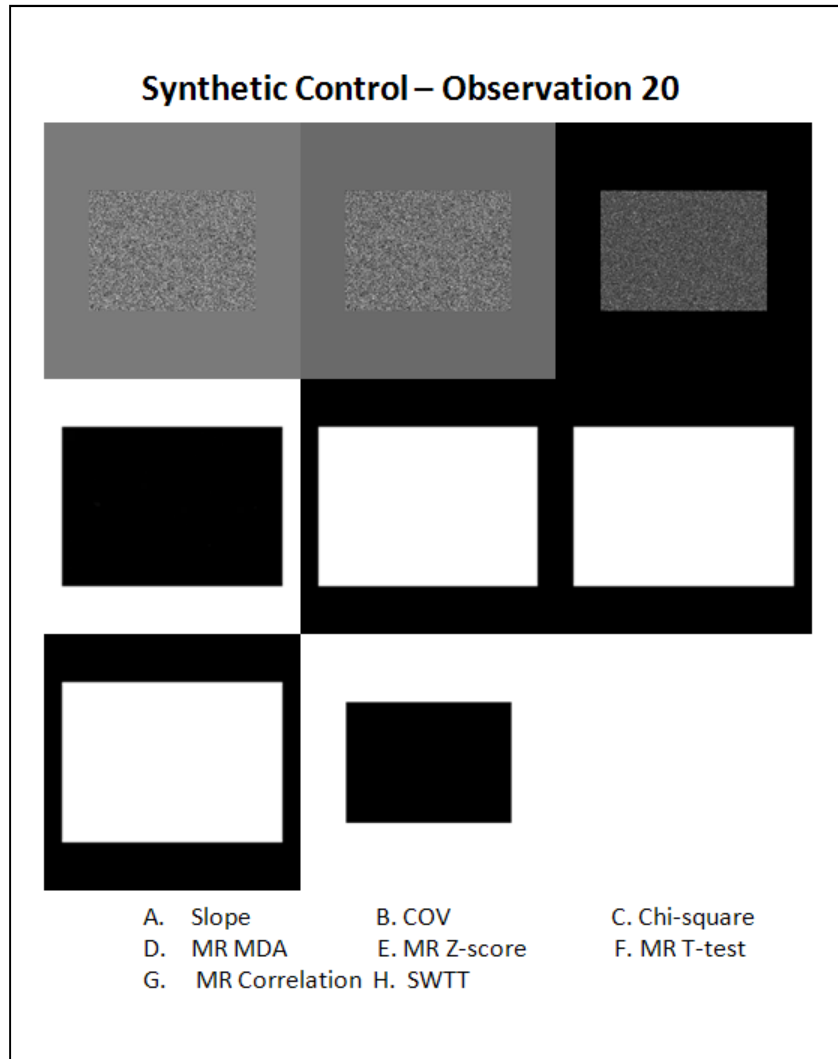
Method	Pros	Cons
Multiresolution MDA	Sensitive; provides diagnostic information	Prone to false-positives due to quad regions; somewhat qualitative to interpret
Multiresolution t-Test	Improved false-positive performance; sensitive; provides diagnostic information	Limited when detecting very small defects; while quantitative, still requires visual interpretation
Multiresolution SoC	Slightly more sensitive than t-Test; good false-positive performance	Limited when detecting very small defects; while quantitative, still requires visual interpretation
SWTT	Ability to detect very small, slowly progressive defects	Larger matrix required; insensitive to larger, less defined defects (off-peak data)

Shewhart Control Charts	Autonomous detection (no interpretation needed); sensitive to non-linear degradation	Less sensitive than EWMA or CUSUM to subtle shifts;
CUSUM Control Charts	Autonomous detection (no interpretation needed); more sensitive than image-space methods; sensitive to non-linear degradation	Limited when detecting very small defects
EWMA Control Charts	Autonomous detection (no interpretation needed); more sensitive than image-space methods; sensitive to non-linear degradation	Limited when detecting very small defects

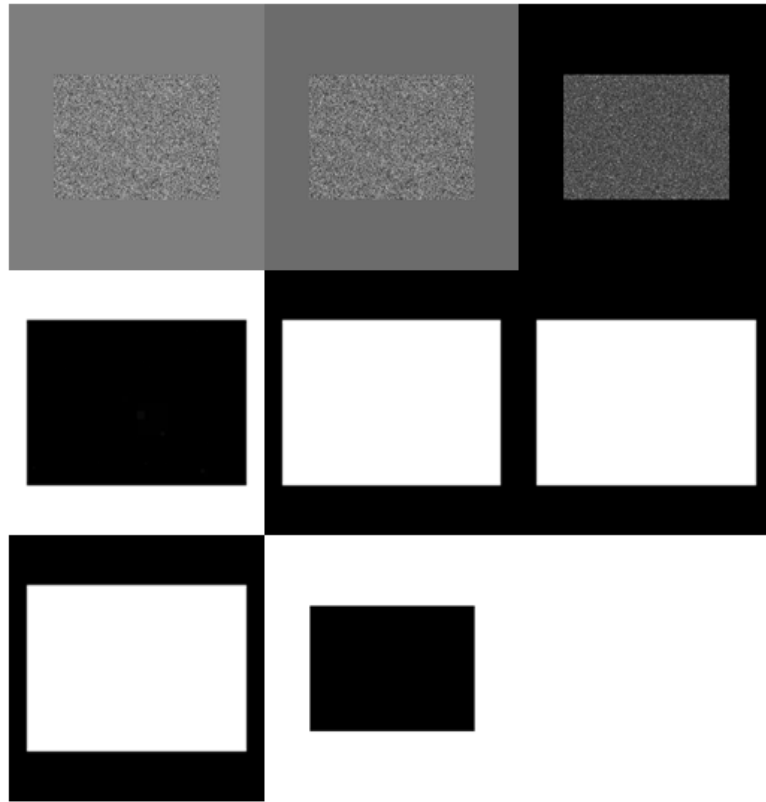
**Table 6 – Summary of Observations**

9.2

Results at Various Time Points from Image-space Algorithms

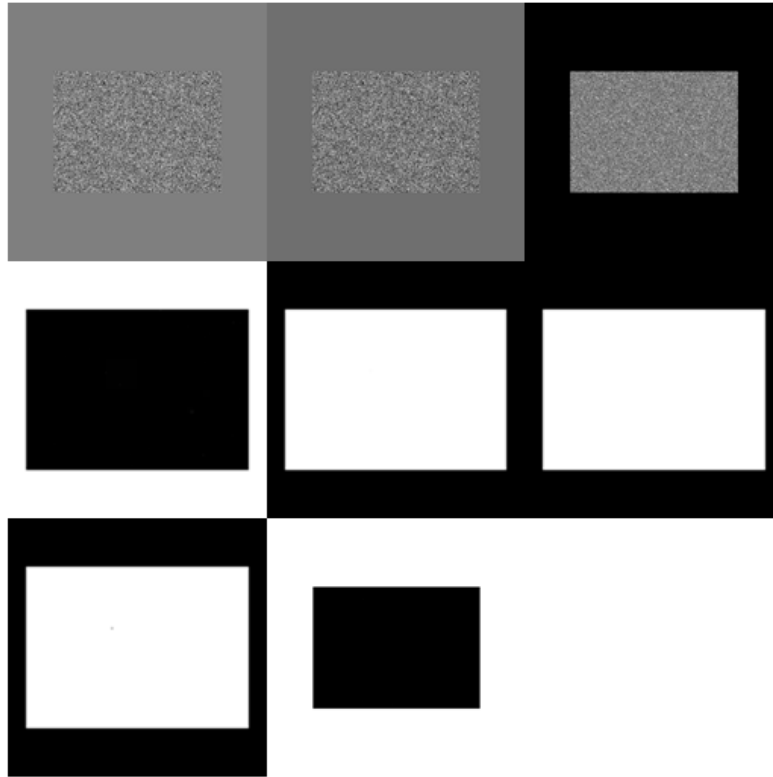


### Synthetic Control – Observation 40



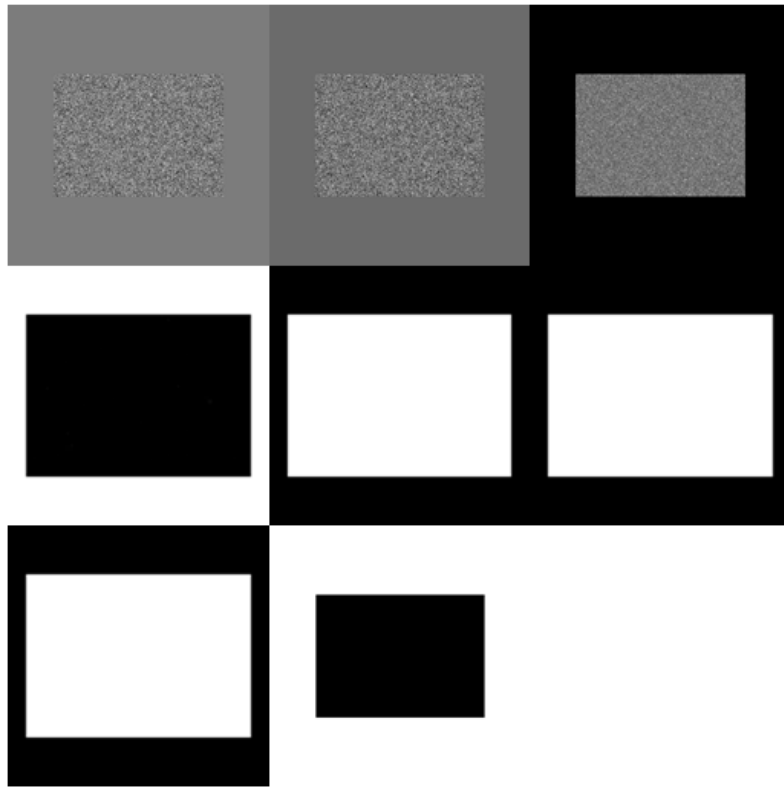
- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MR T-test
- G. MR Correlation
- H. SWTT

### Synthetic Control – Observation 60



- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MR T-test
- G. MR Correlation
- H. SWTT

### Synthetic Control – Observation 80



- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MRT-test
- G. MR Correlation
- H. SWTT

The image displays a diagnostic tool interface. On the left, a large black-bordered box labeled "Original" contains a grayscale noisy image. To its right is a grid of seven smaller images under the heading "Synthetic\_13-Observation 20". The top row of the grid shows three different noisy images. The middle row shows three images, each with a small white crosshair in the center. The bottom row shows a single large black image. Below the grid is a legend with seven items: A. Slope, B. COV, C. Chi-square, D. MR MDA, E. MR Z-score, F. MR T-test, and G. MR Correlation H. SWTT. An orange arrow points from the legend area towards a large empty white box on the right side of the interface.

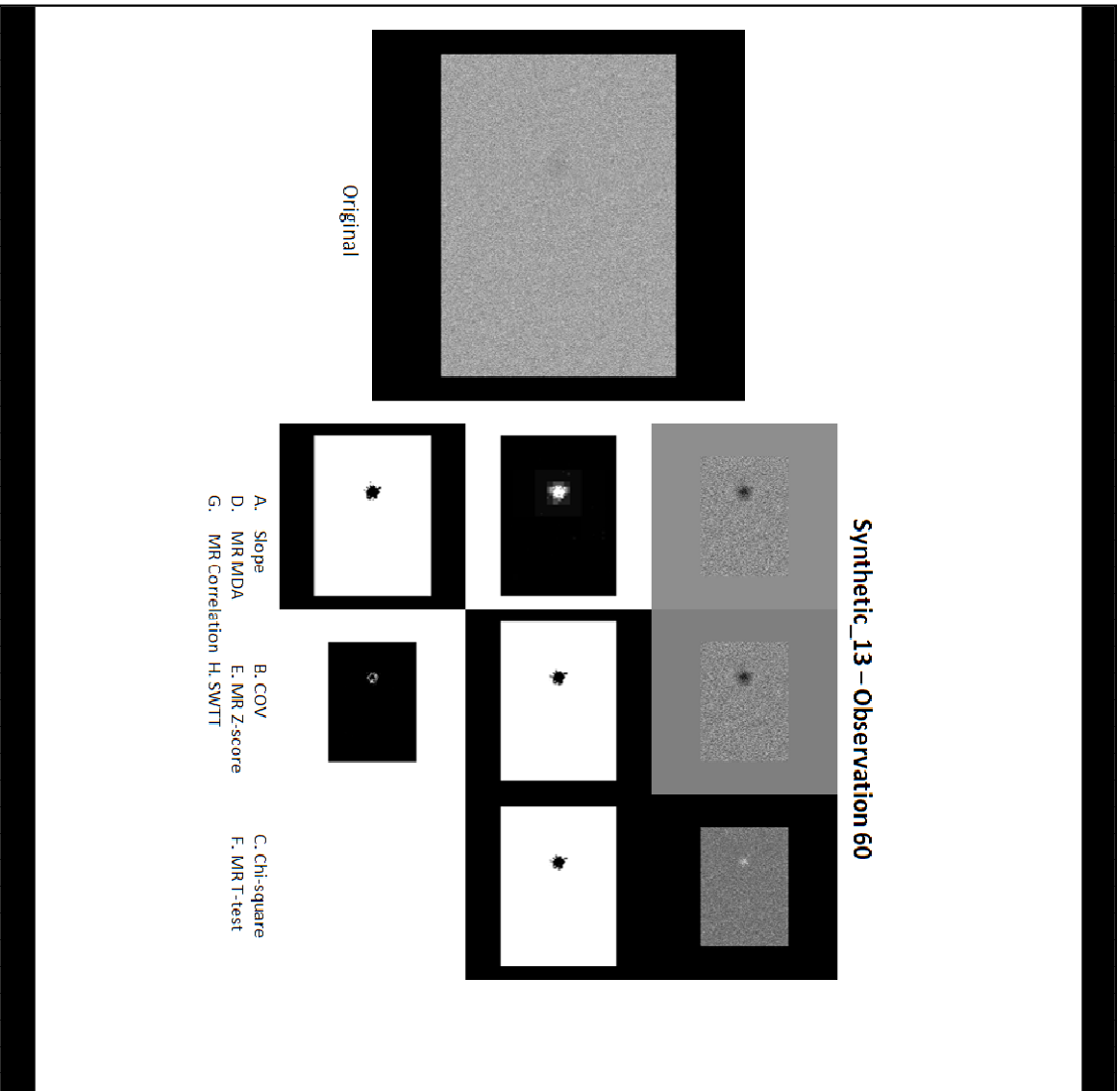


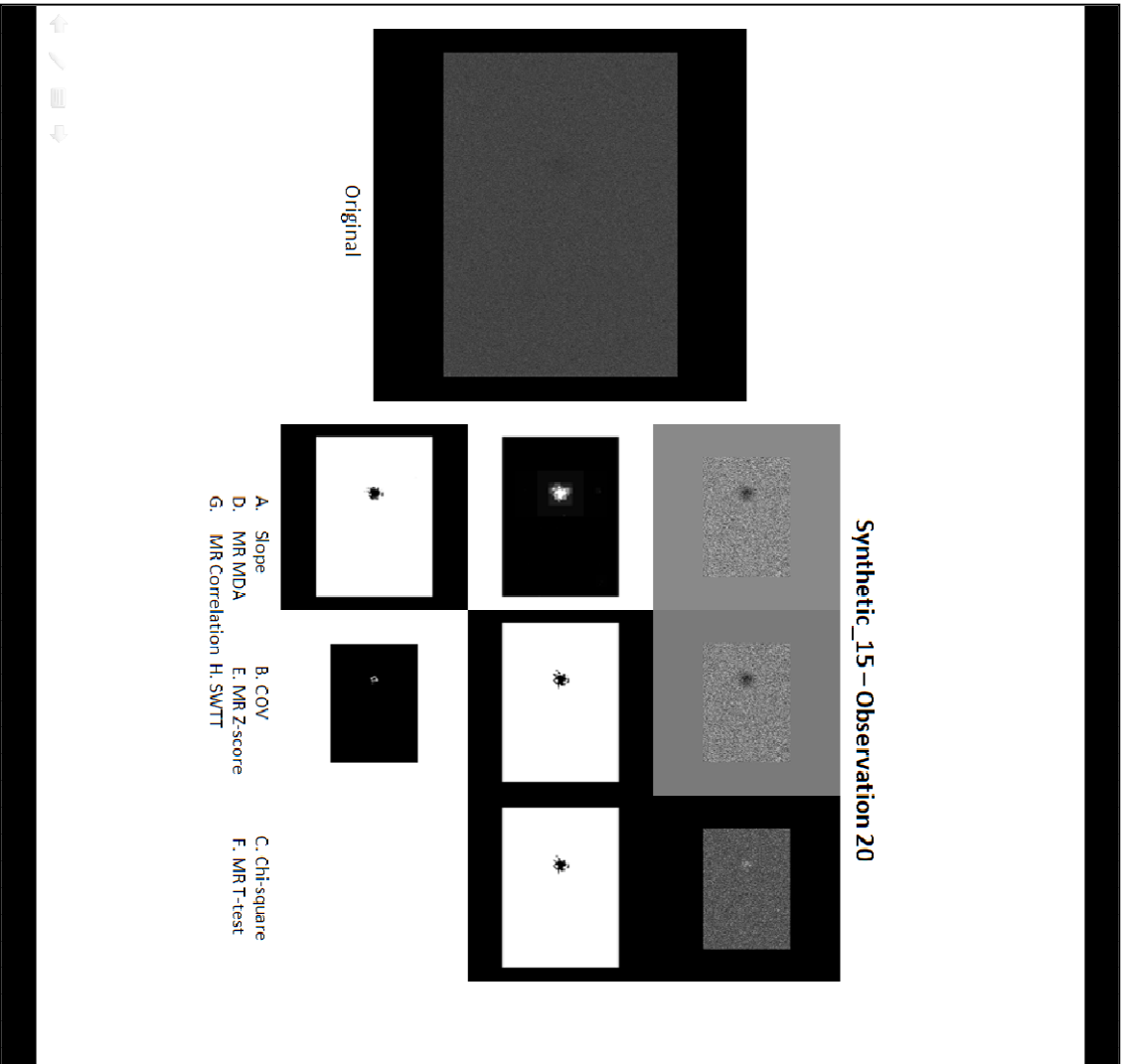
Original

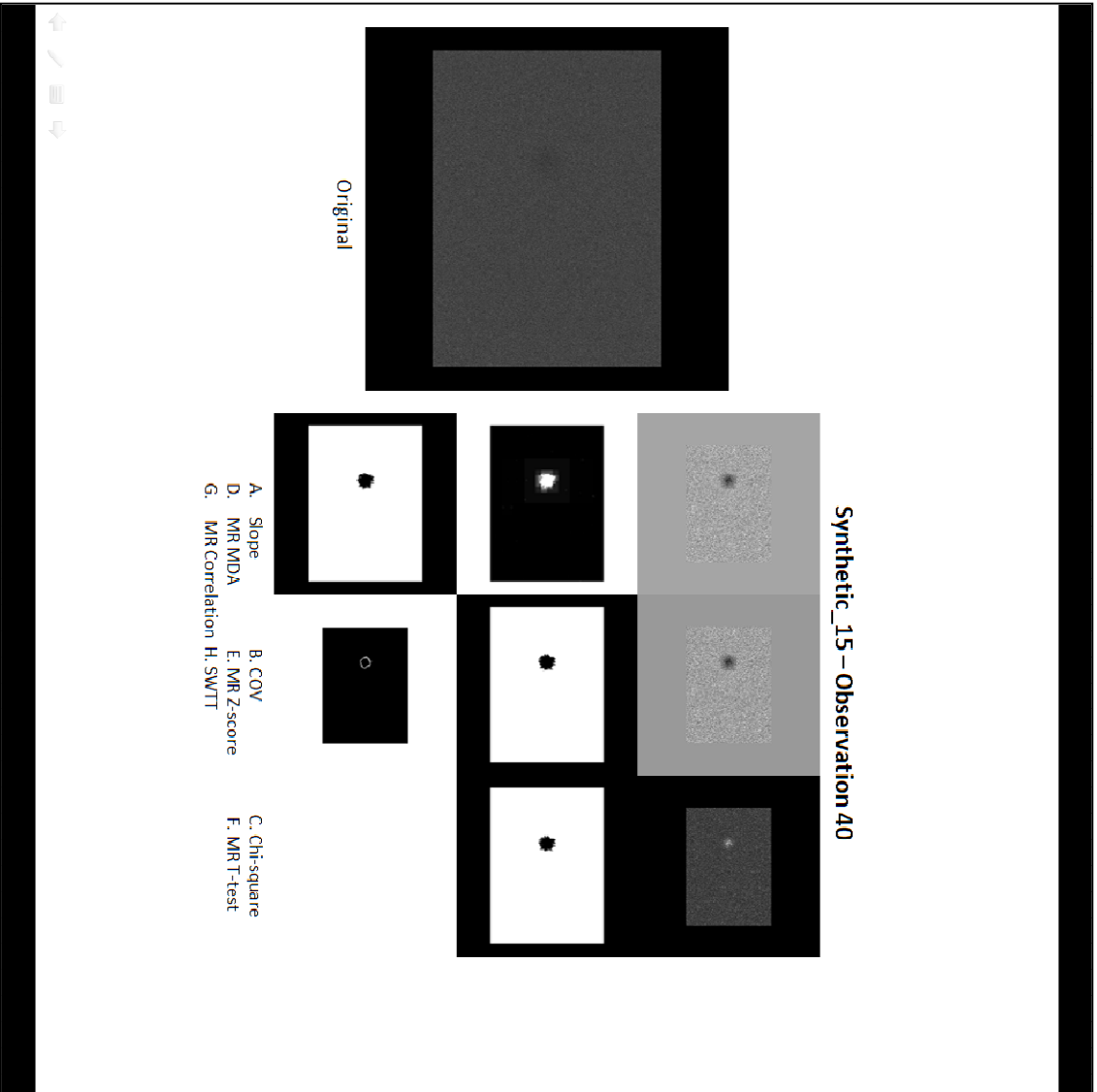
Synthetic\_13 - Observation 40

- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MR T-test
- G. MR Correlation H. SWTT

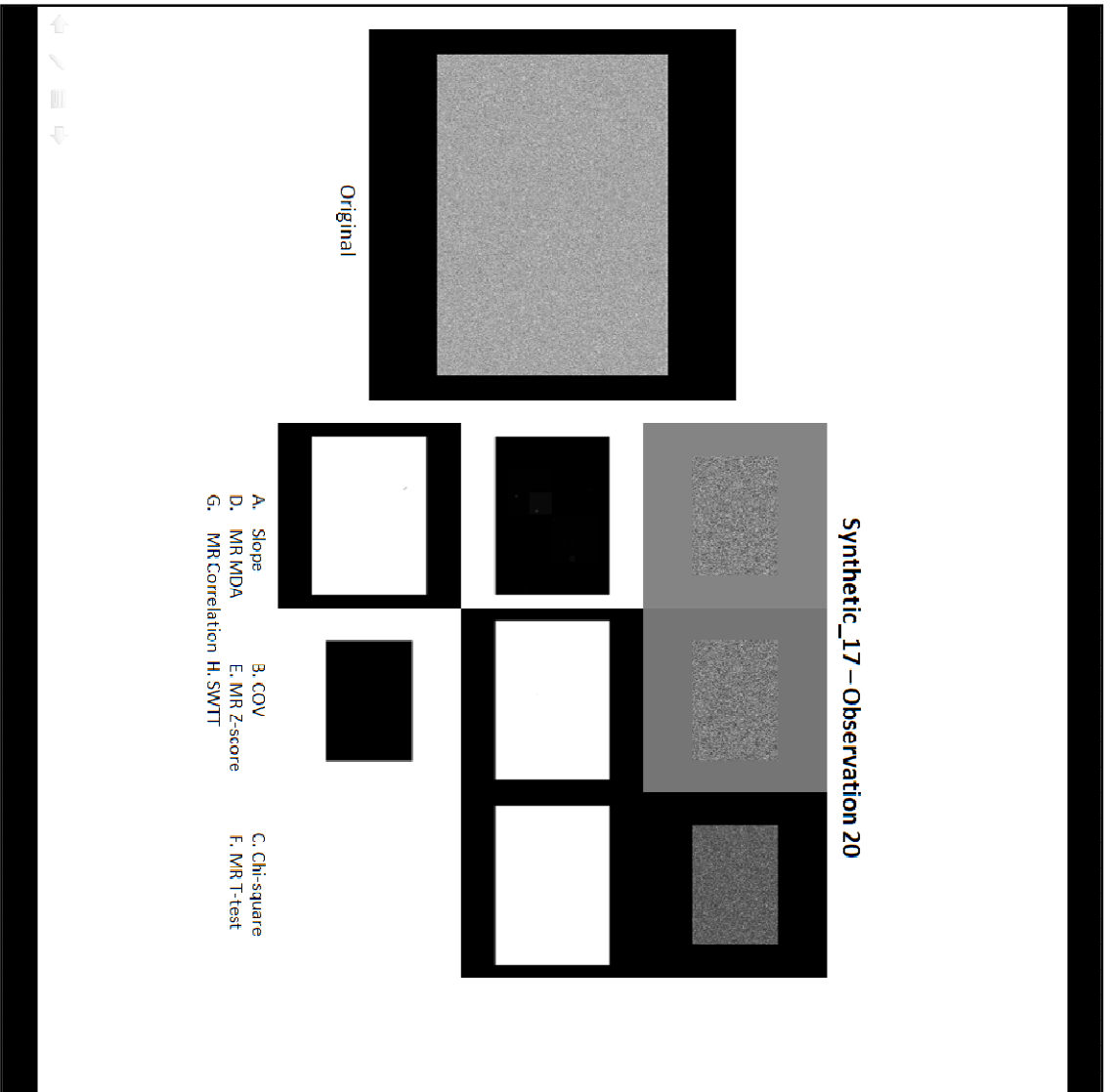
The interface displays a large 'Original' image of a star on a gray background. To its right is a grid of 'Synthetic\_13 - Observation 40' images, which are smaller versions of the original. Below the grid is a legend with seven items (A-G) corresponding to different image processing methods. An orange arrow points from the legend to a larger, high-contrast image of the star on a black background, which is the result of one of the methods.







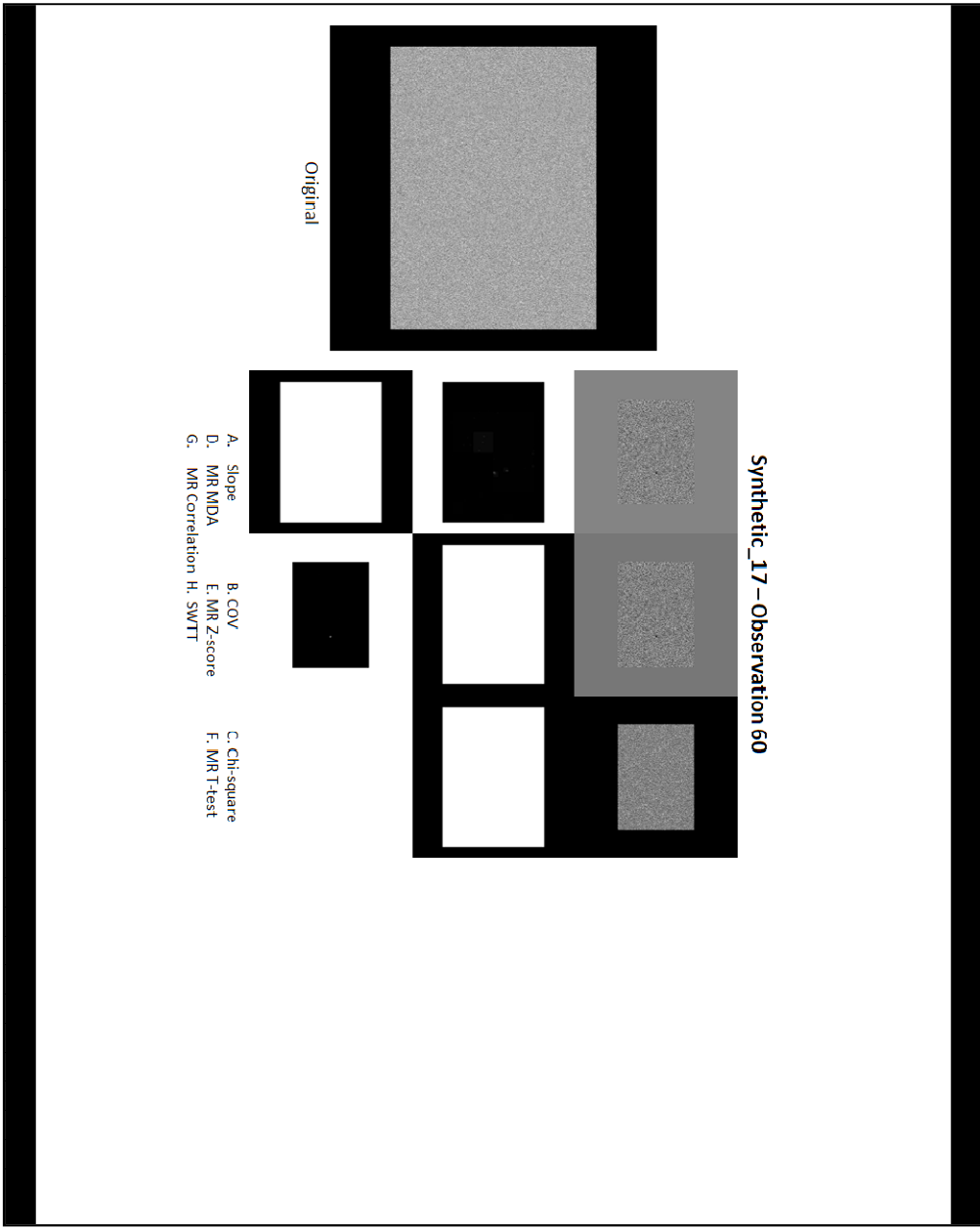
- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MR T-test
- G. MR Correlation
- H. SW/TI

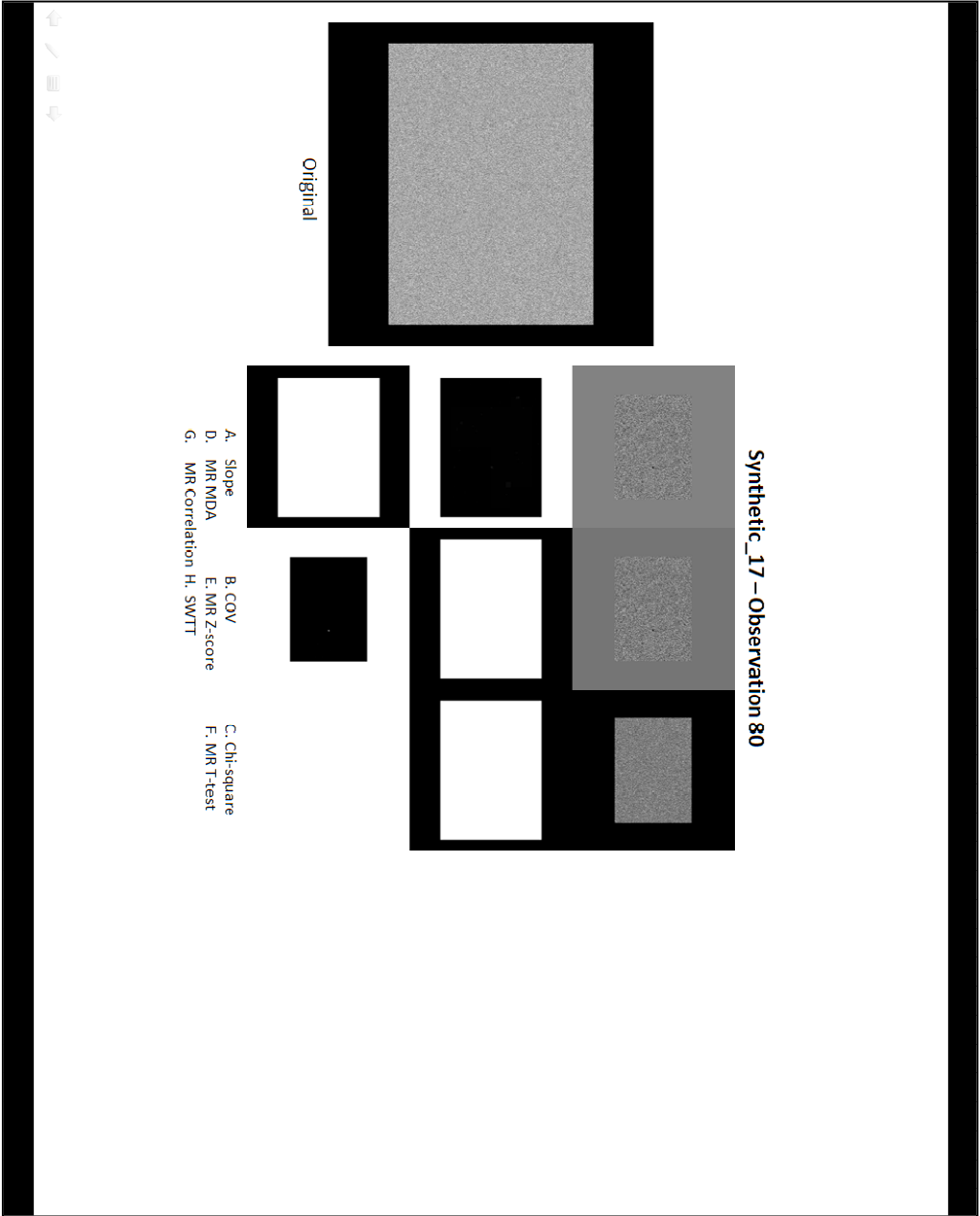


Original

Synthetic\_17 - Observation 40

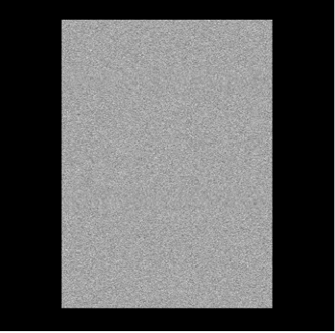
- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MR T-test
- G. MR Correlation H. SWTI



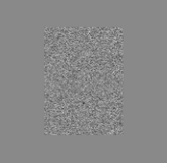


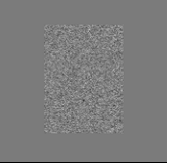


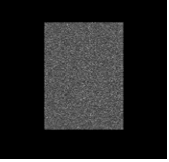
**Synthetic\_19 – Observation 20**





Original







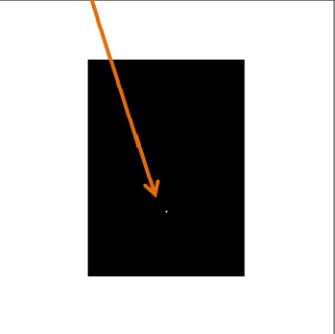













- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MR T-test
- G. MR Correlation H. SWTT



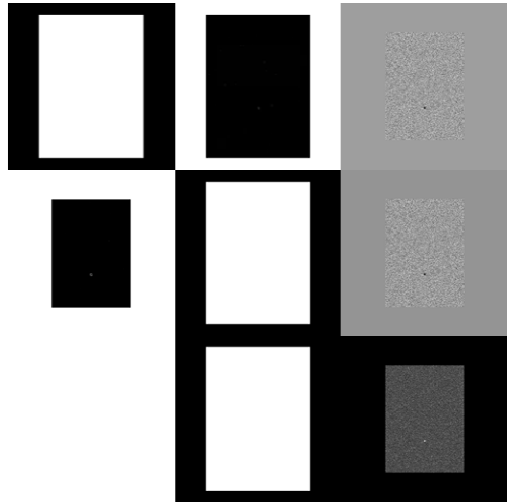
124

**Synthetic\_19 – Observation 40**

**Original**

- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MR T-test
- G. MR Correlation H. SWTT

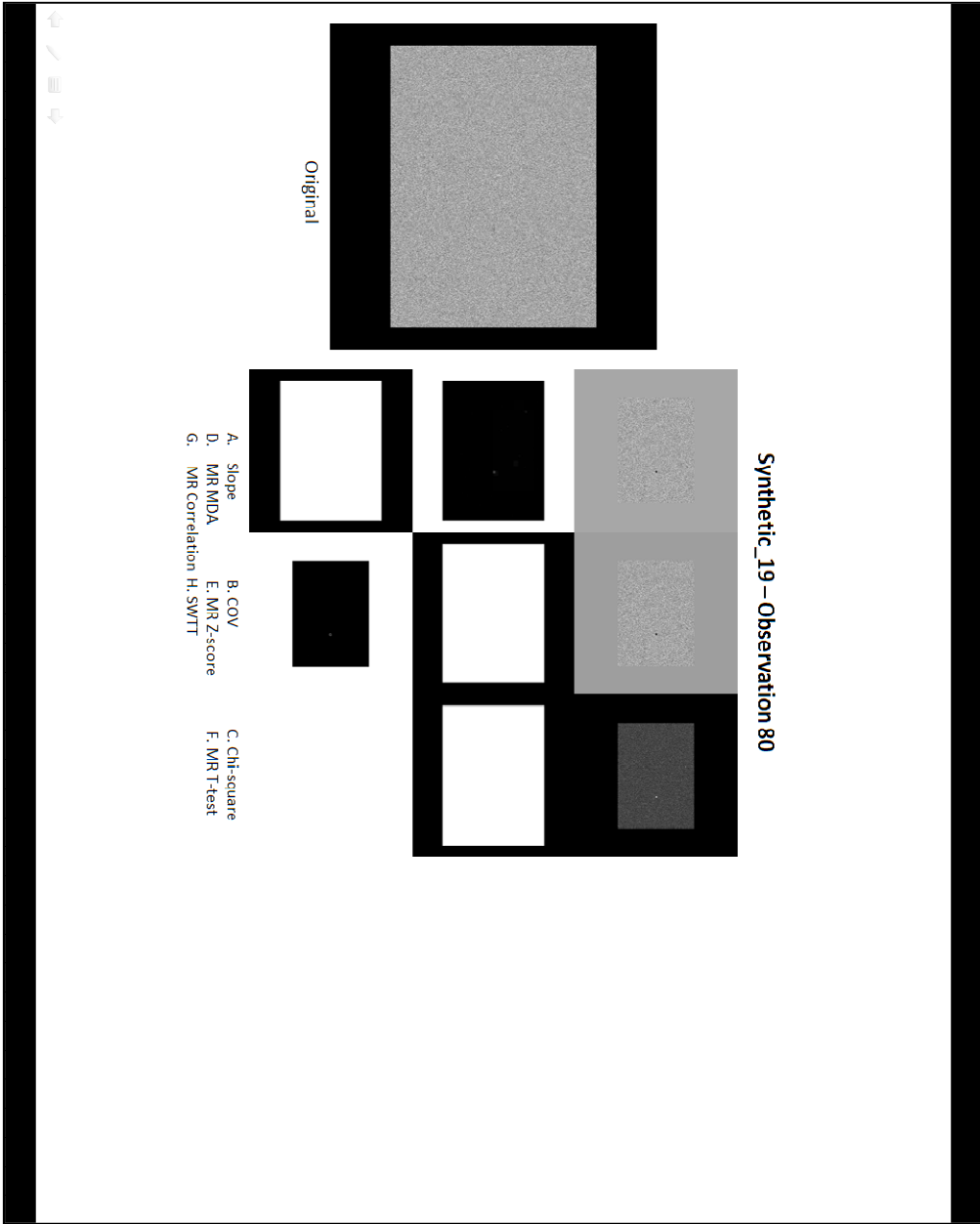
Synthetic\_19 – Observation 60

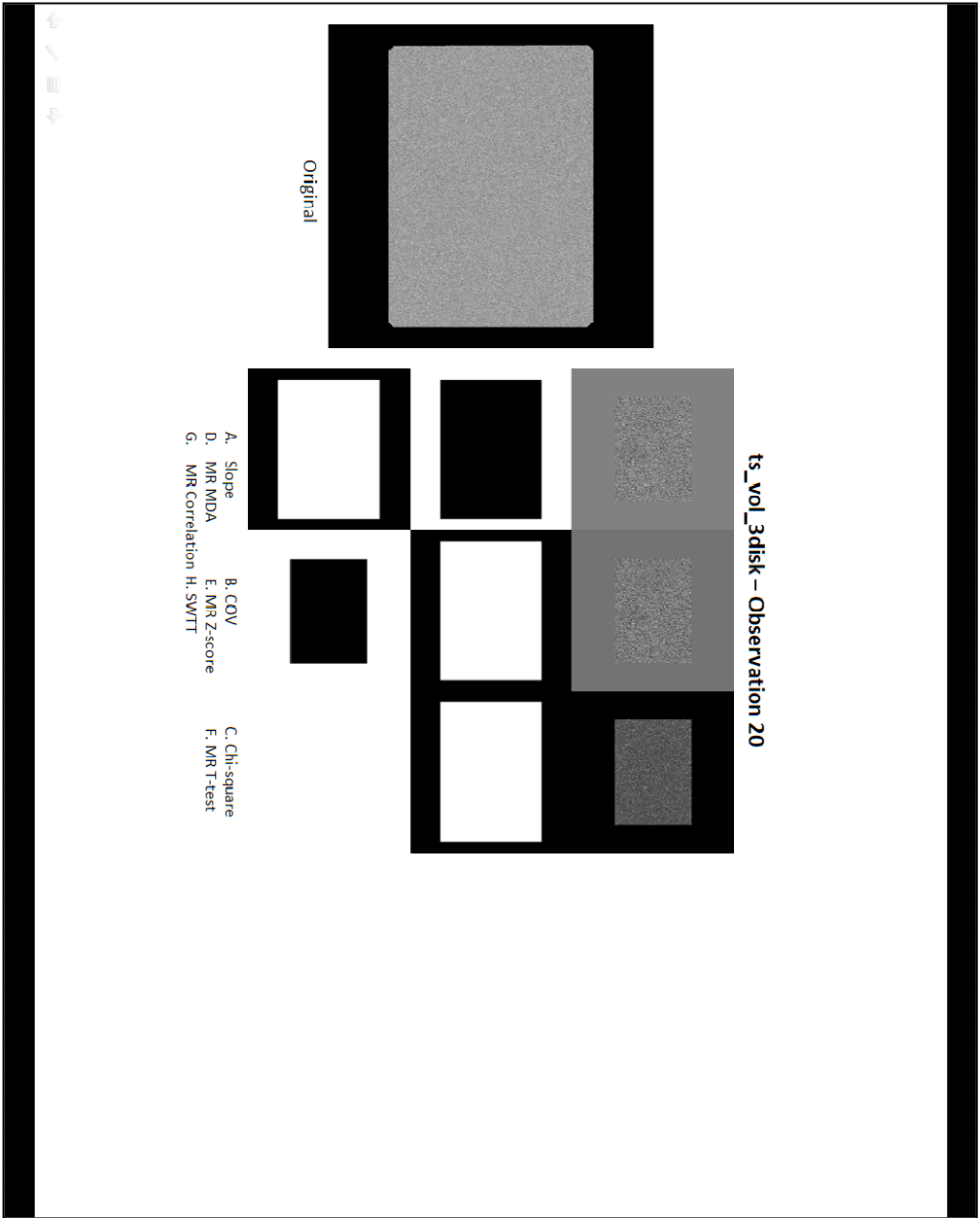


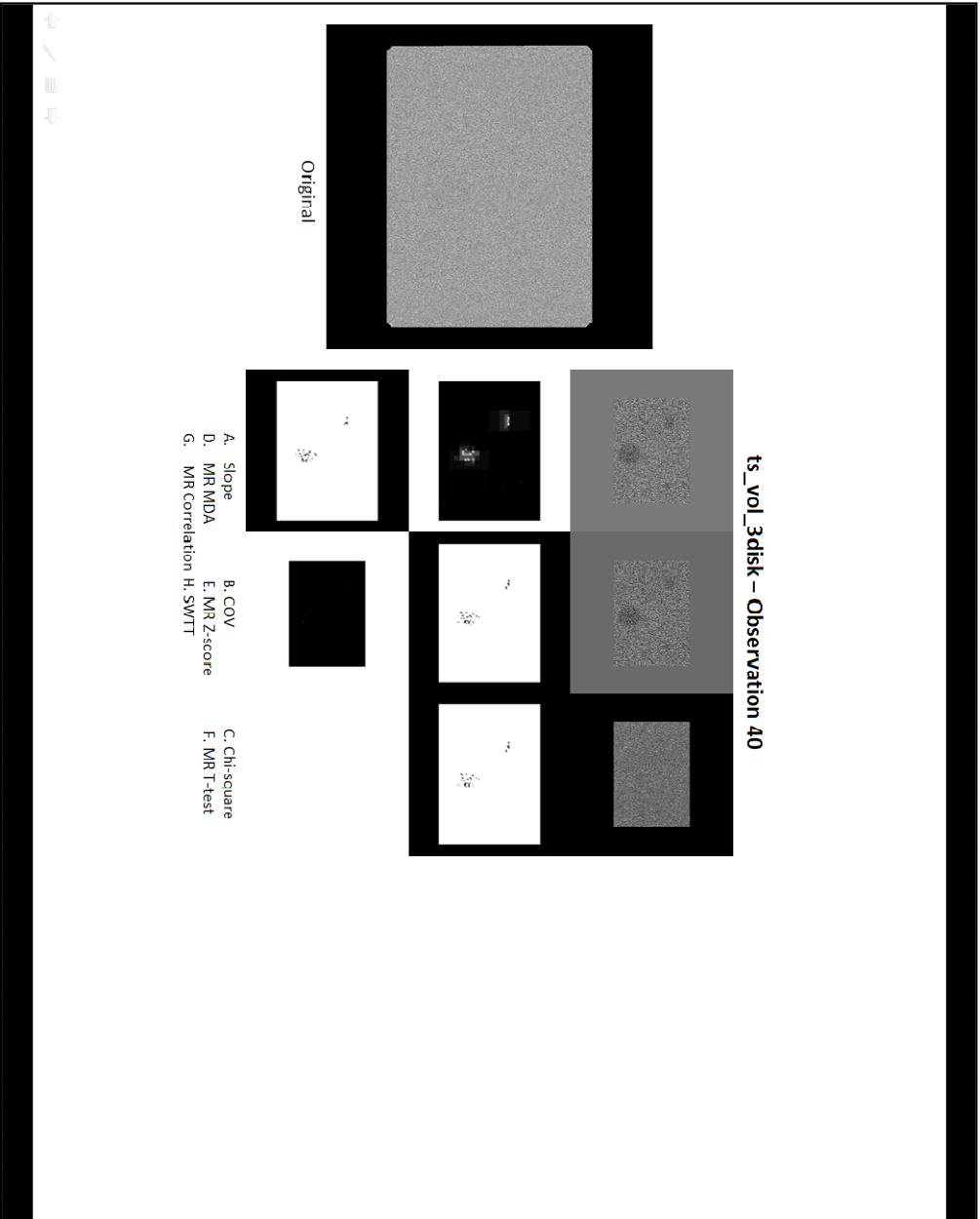
- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MR T-test
- G. MR Correlation
- H. SWTT

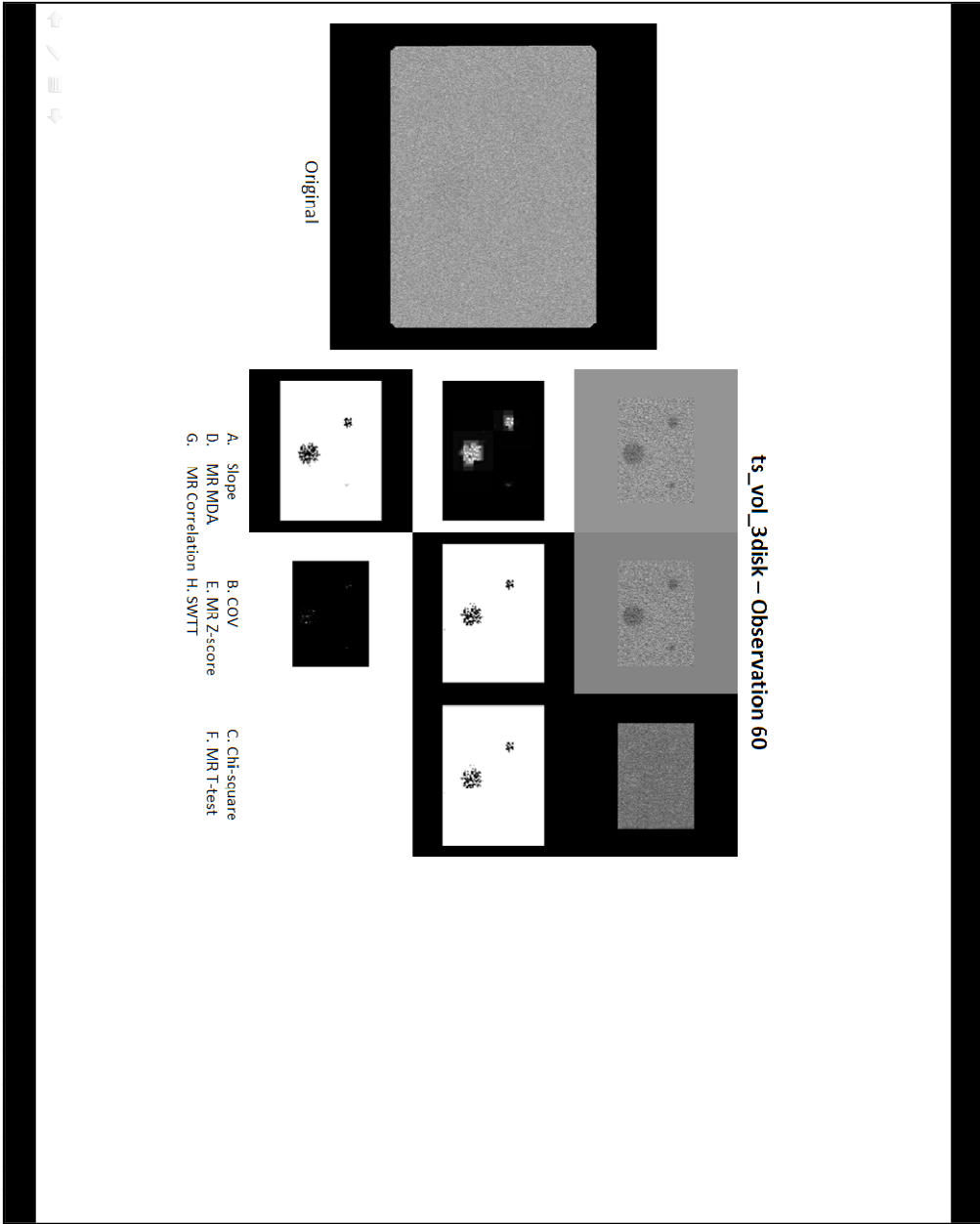
Original

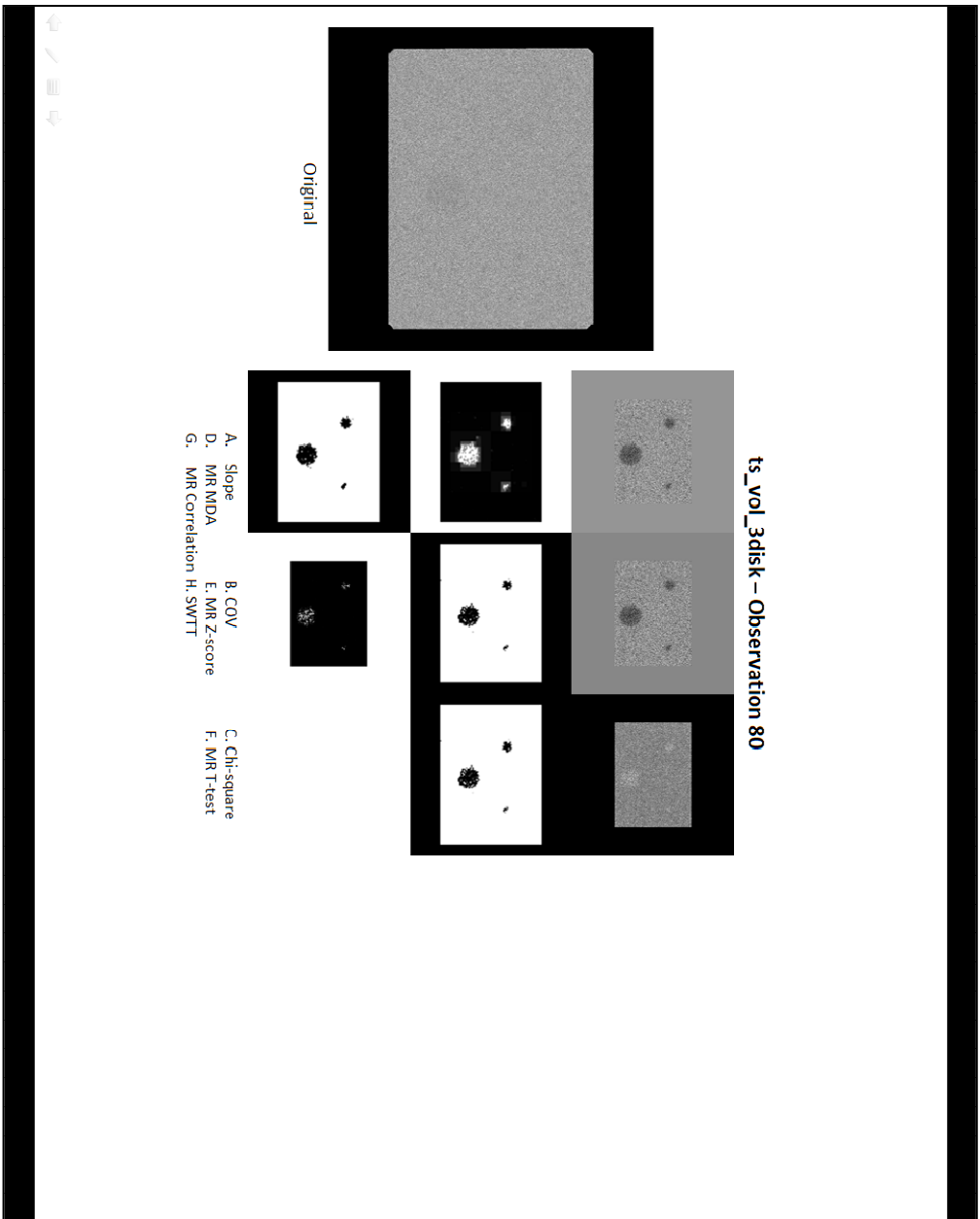






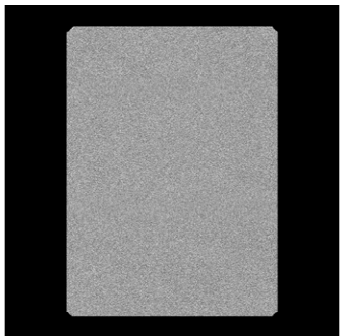




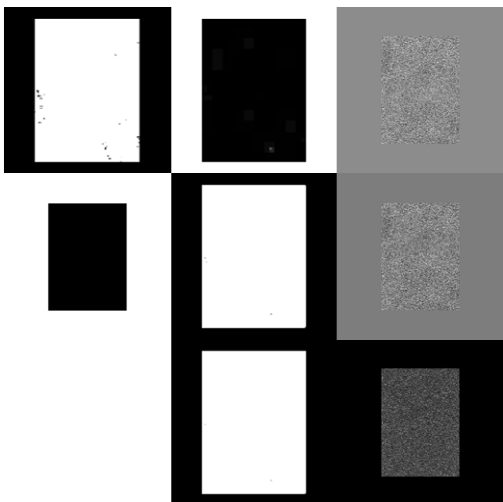




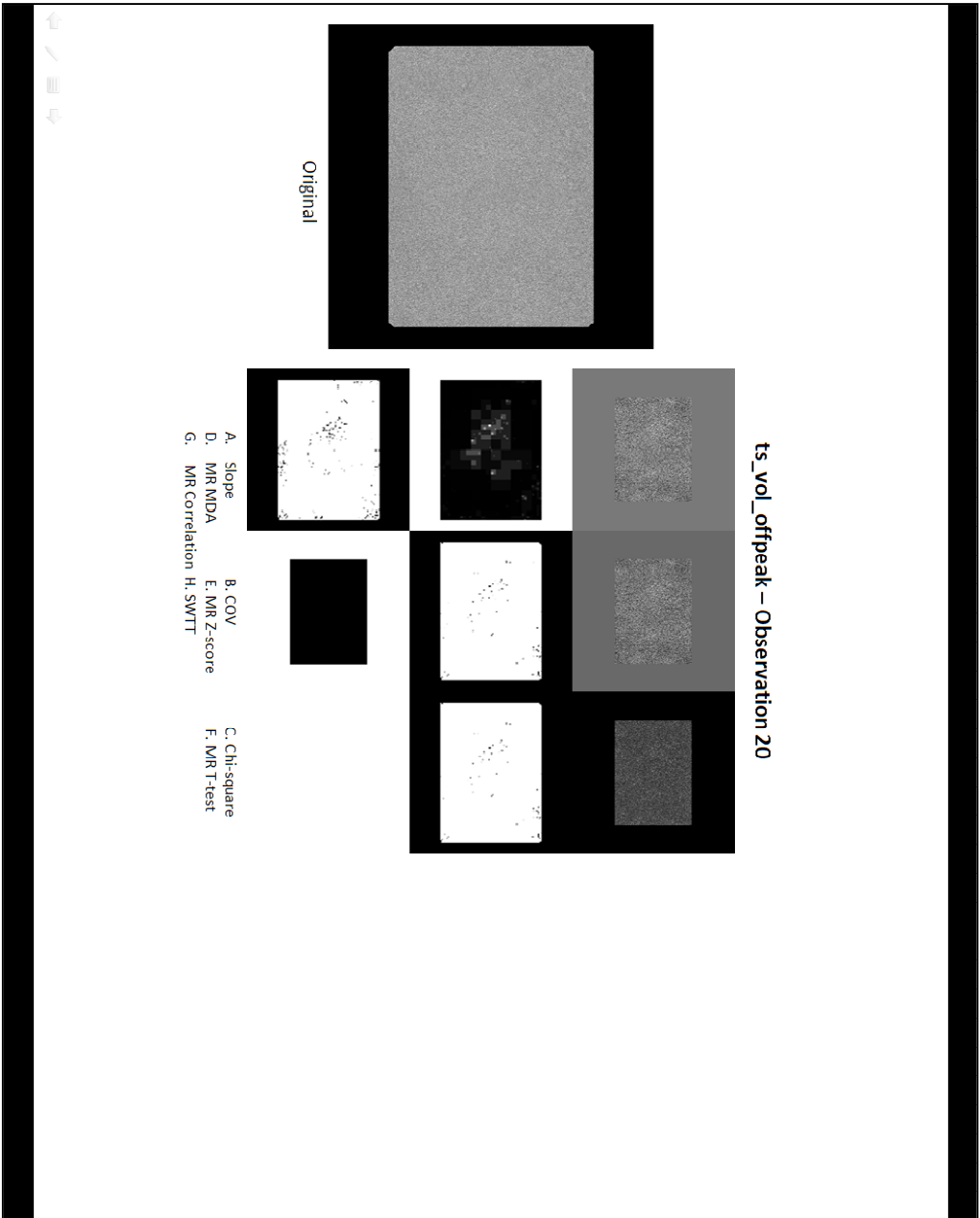
ts\_vol\_offpeak - Observation 10

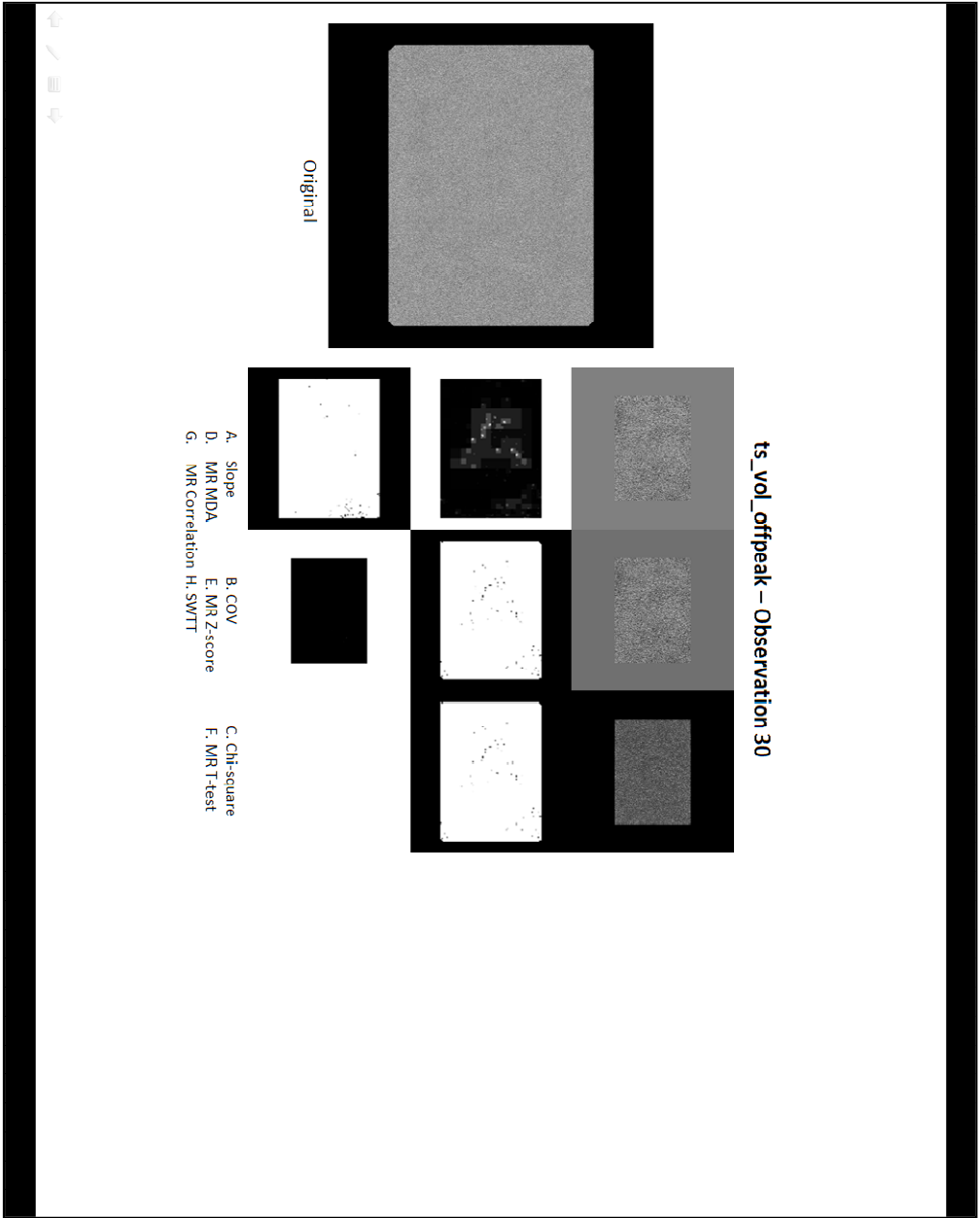


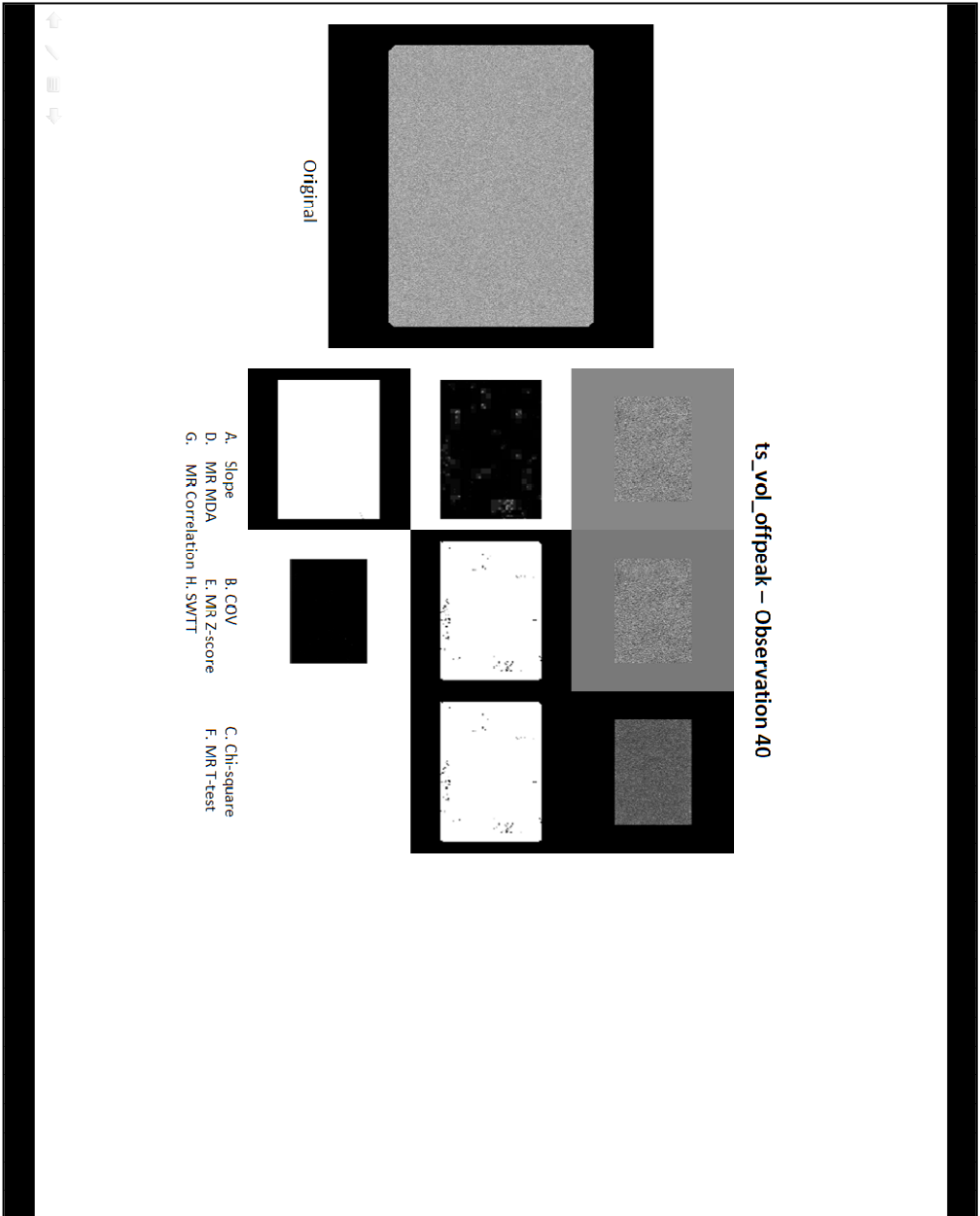
Original

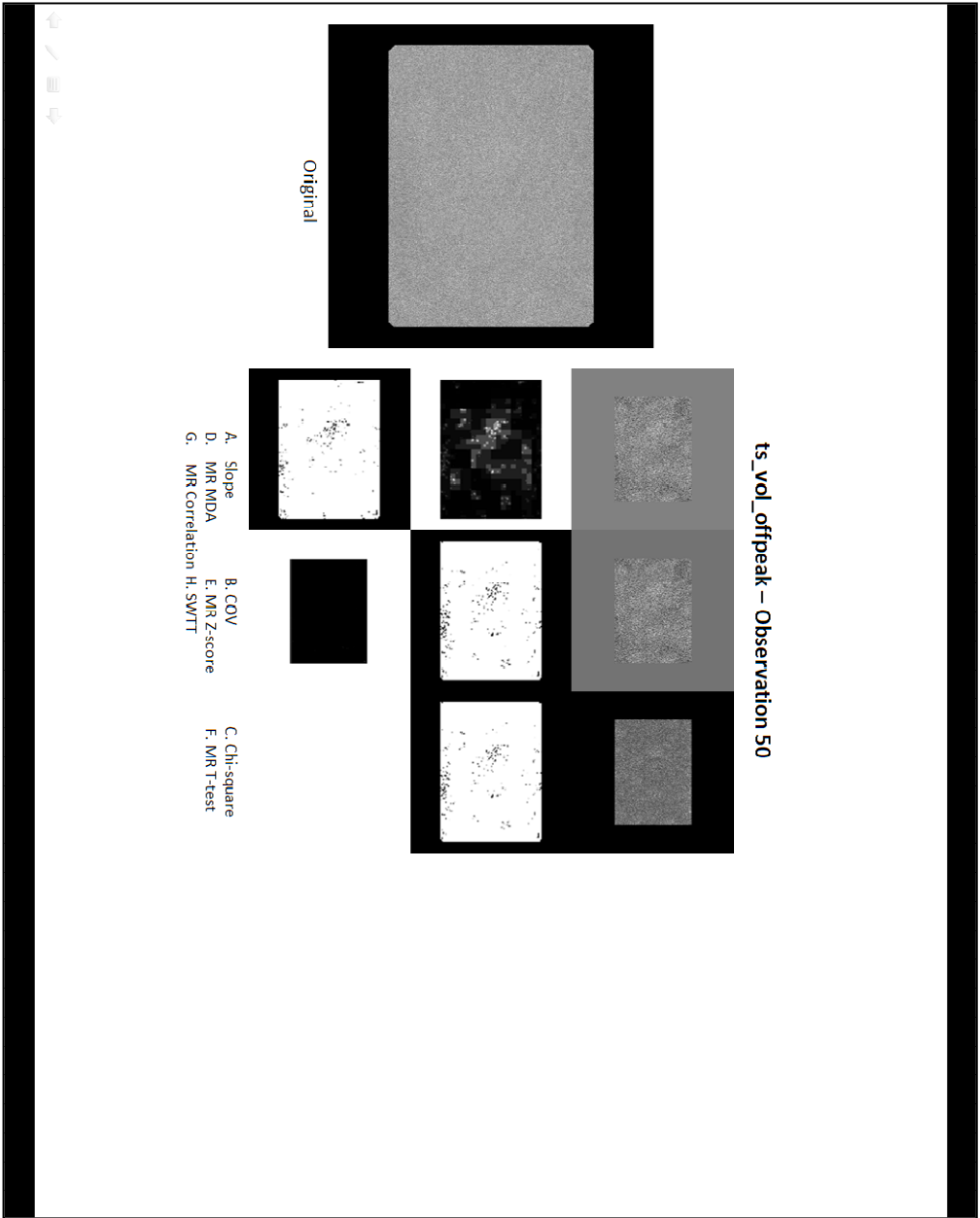


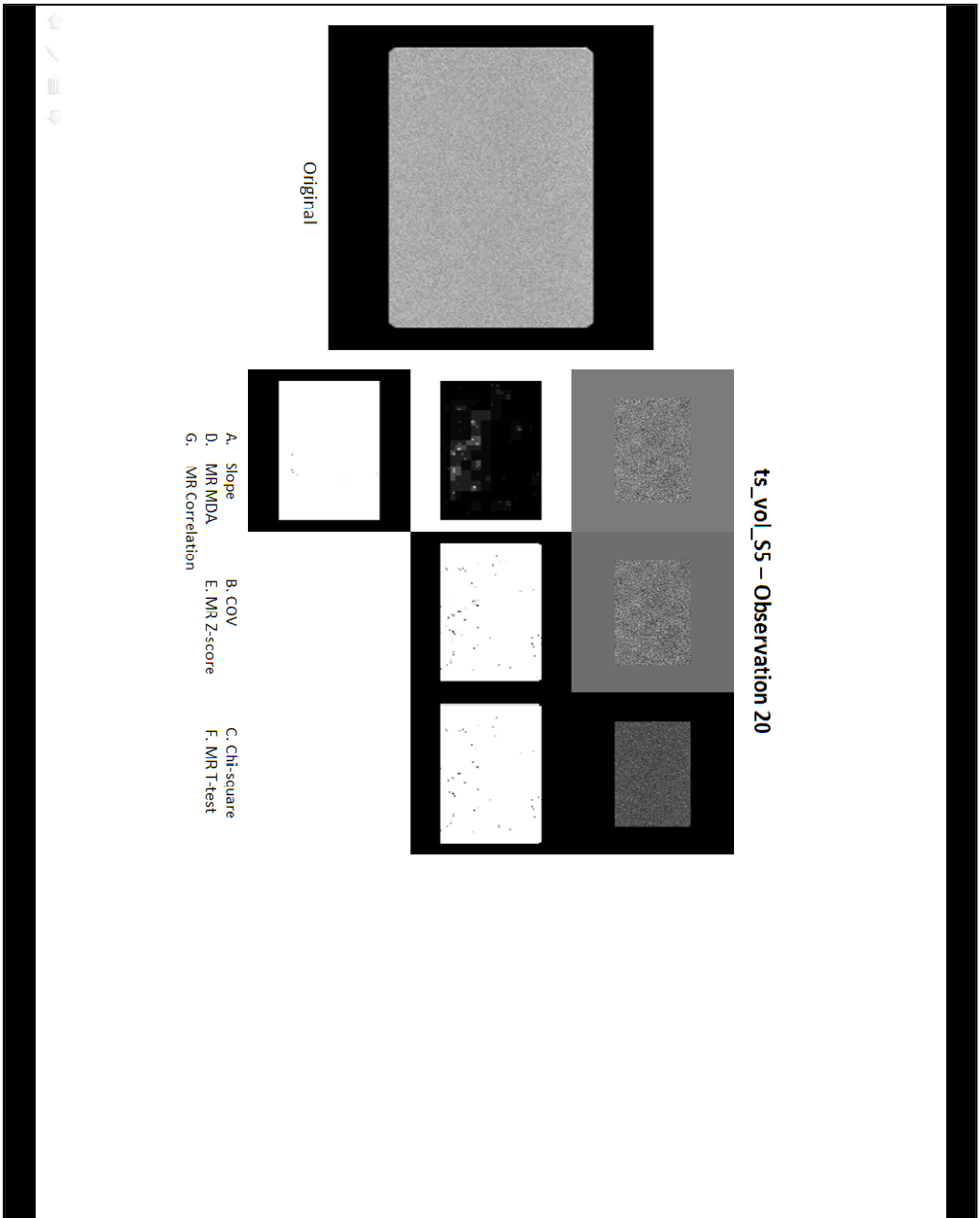
- A. Slope
- B. COV
- C. Chi-square
- D. MR MDA
- E. MR Z-score
- F. MR T-test
- G. MR Correlation H. SWTT

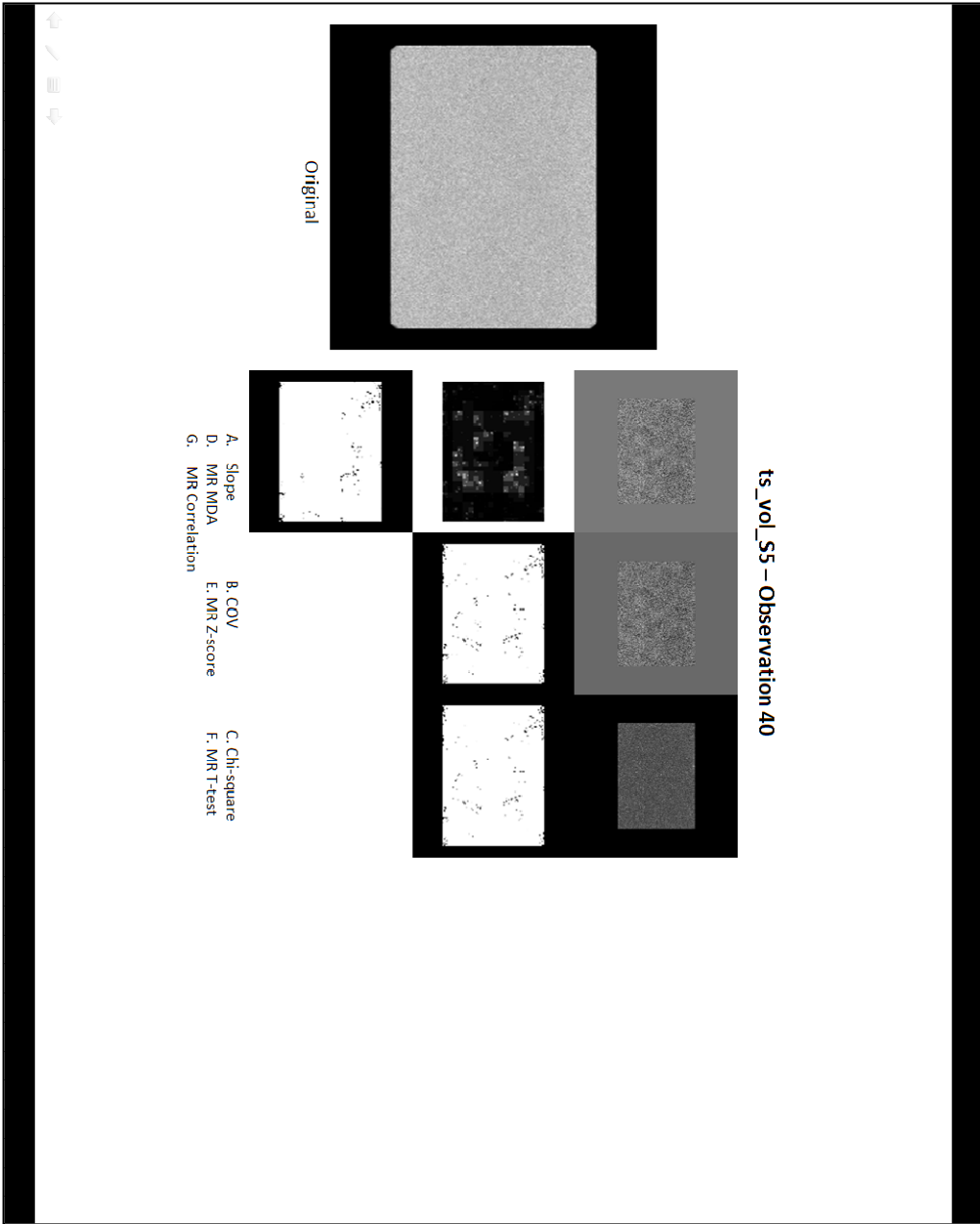


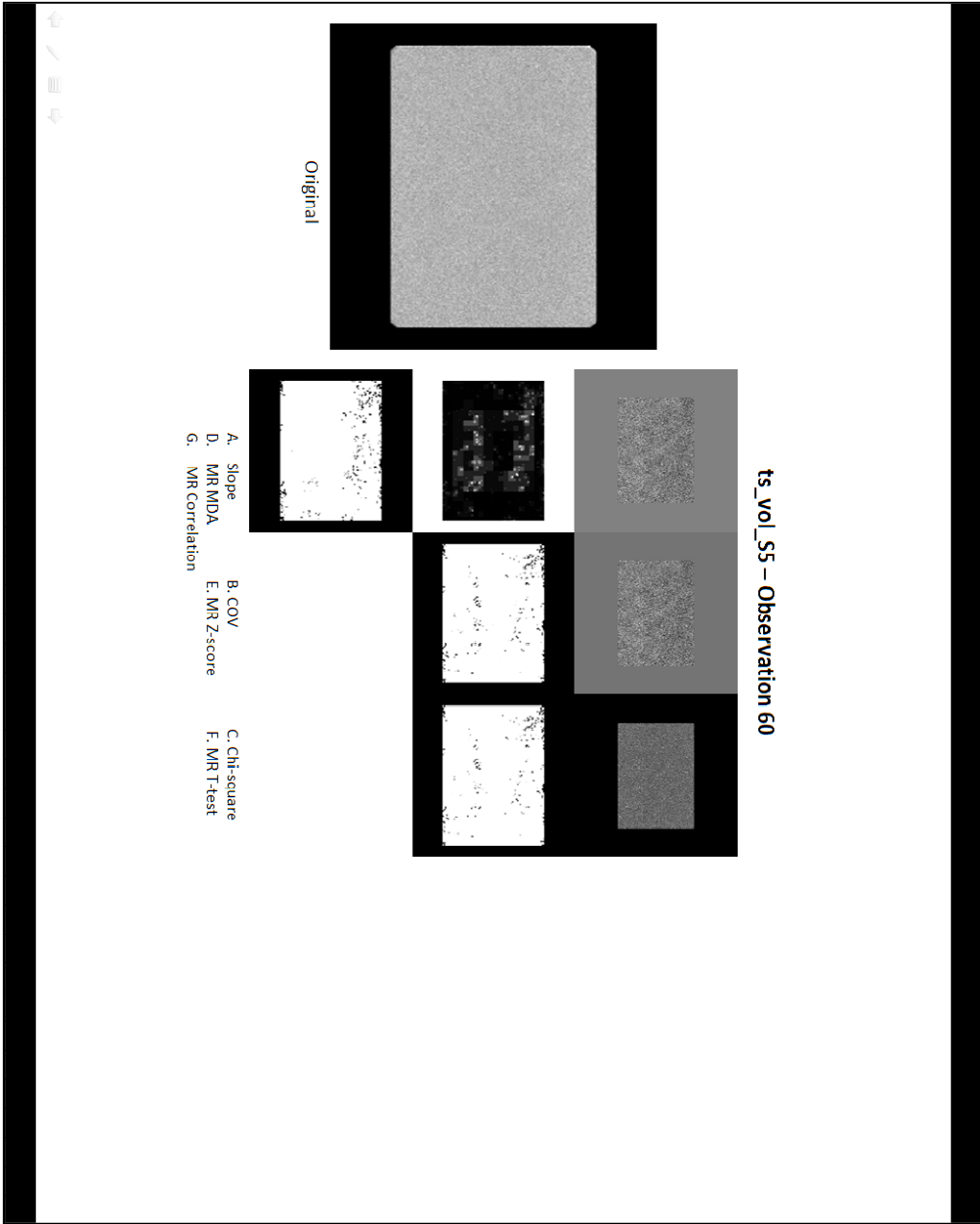




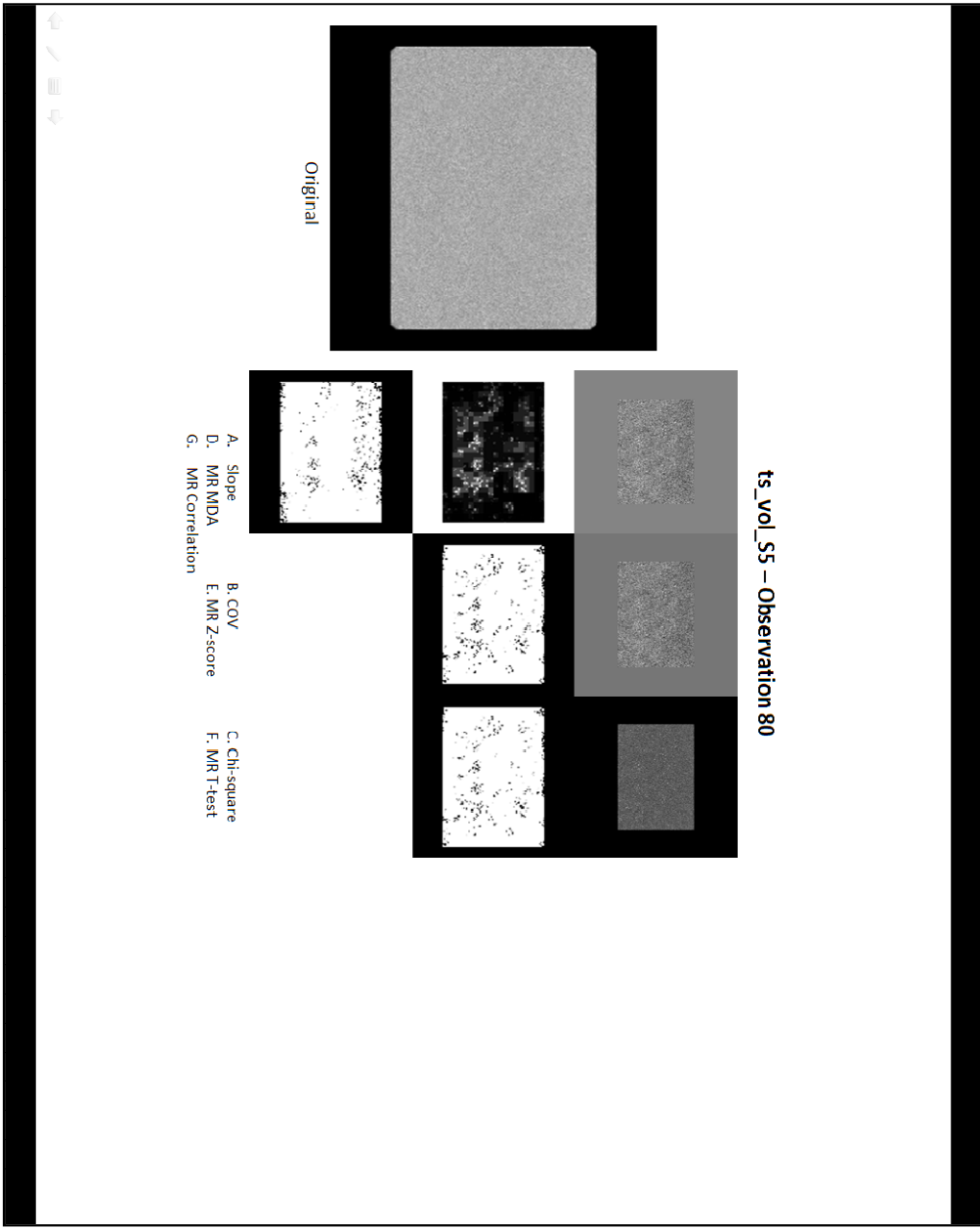




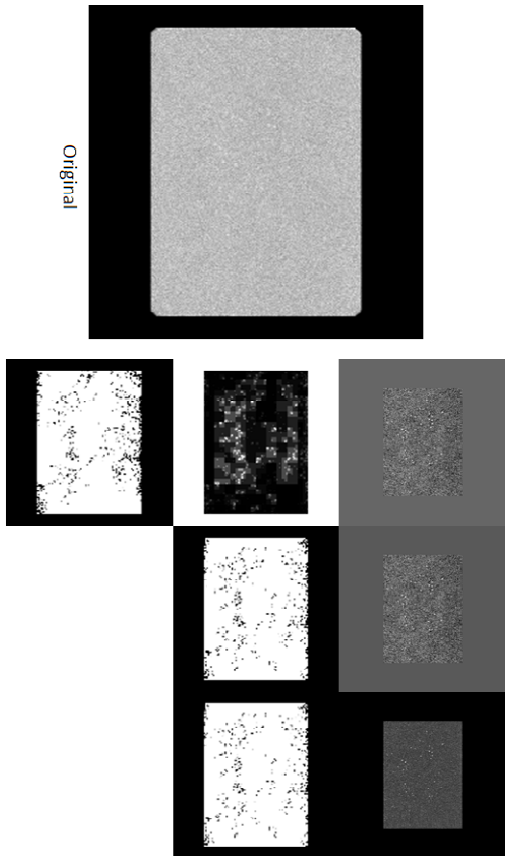




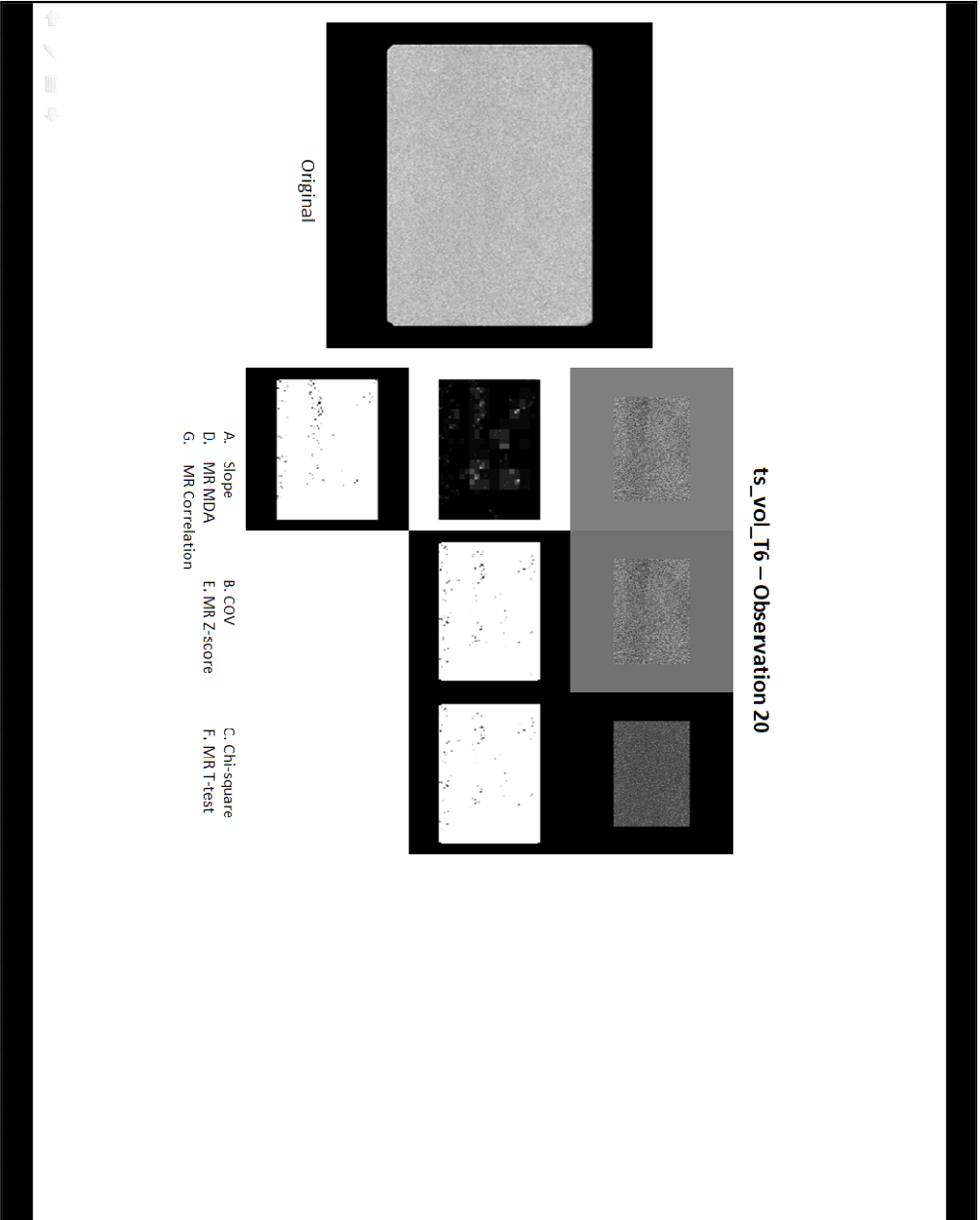




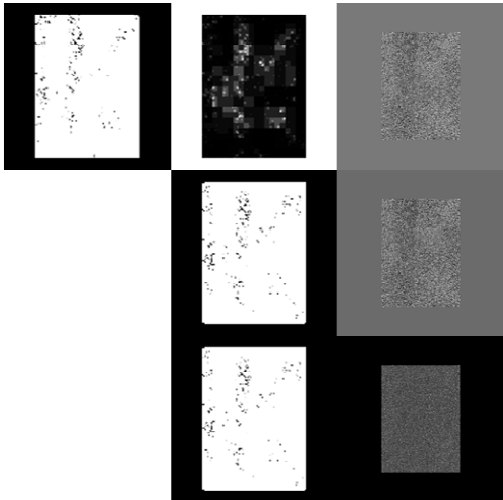
ts\_vol\_55 - Observation 97



Original

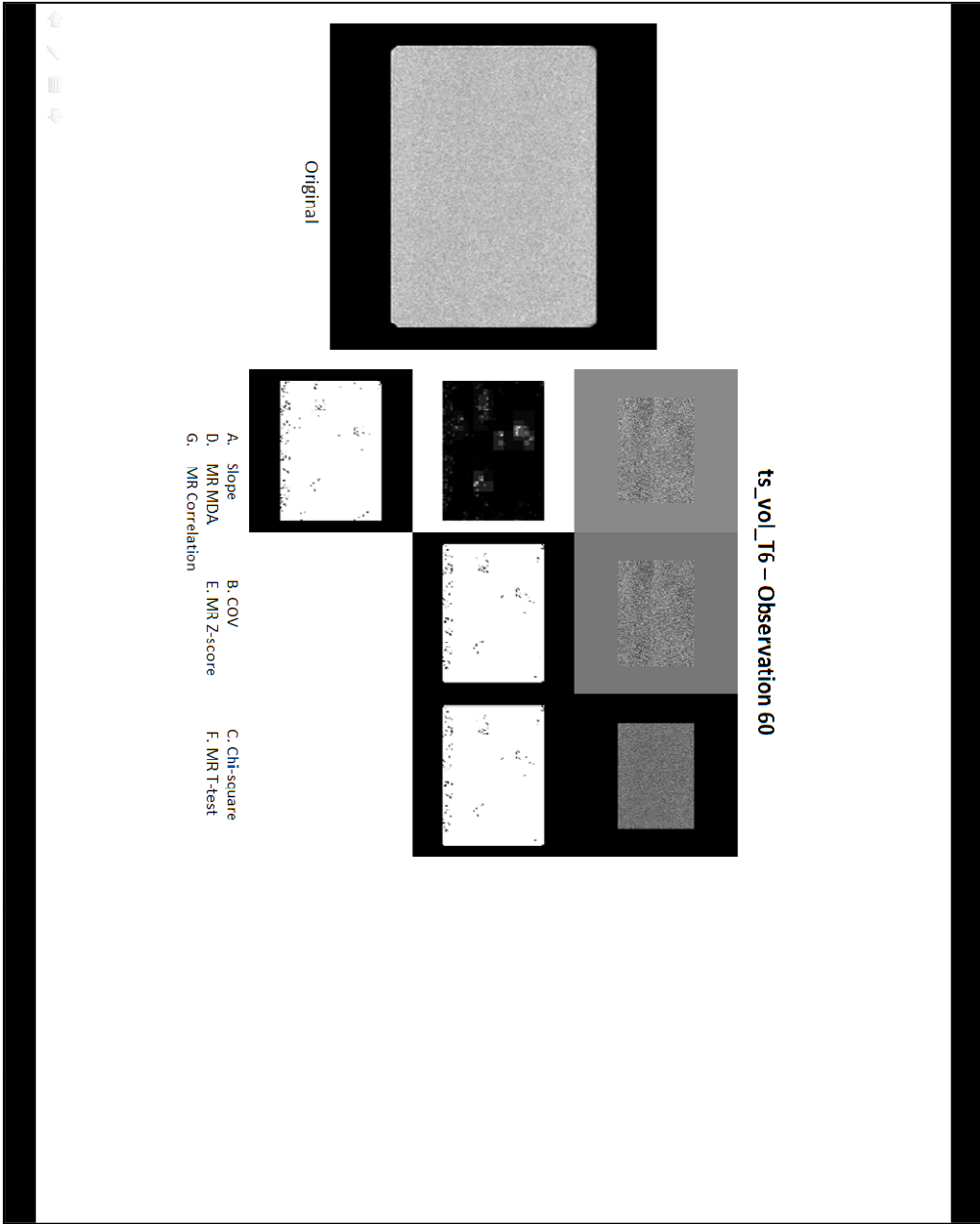


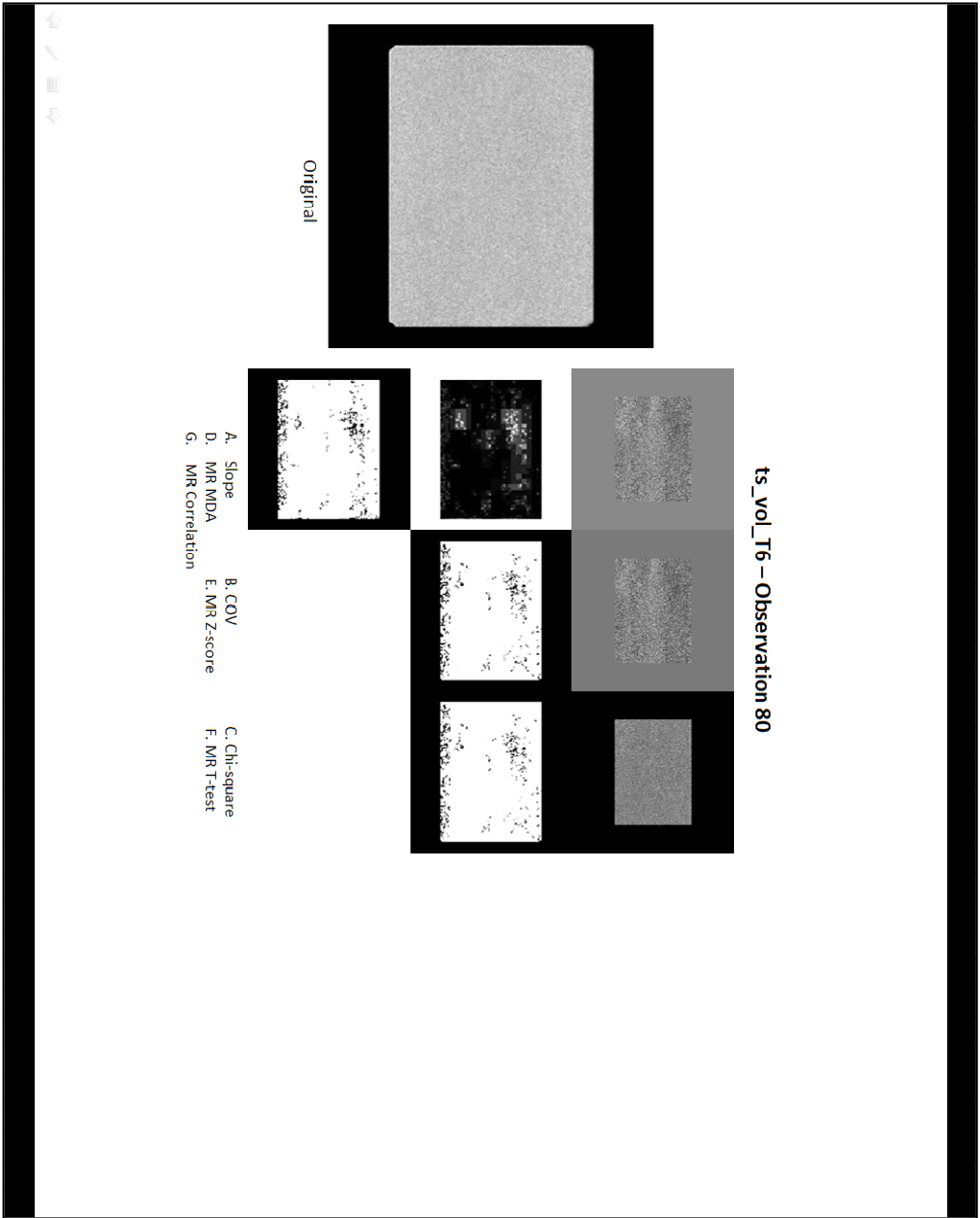
ts\_vol\_T6 - Observation 40

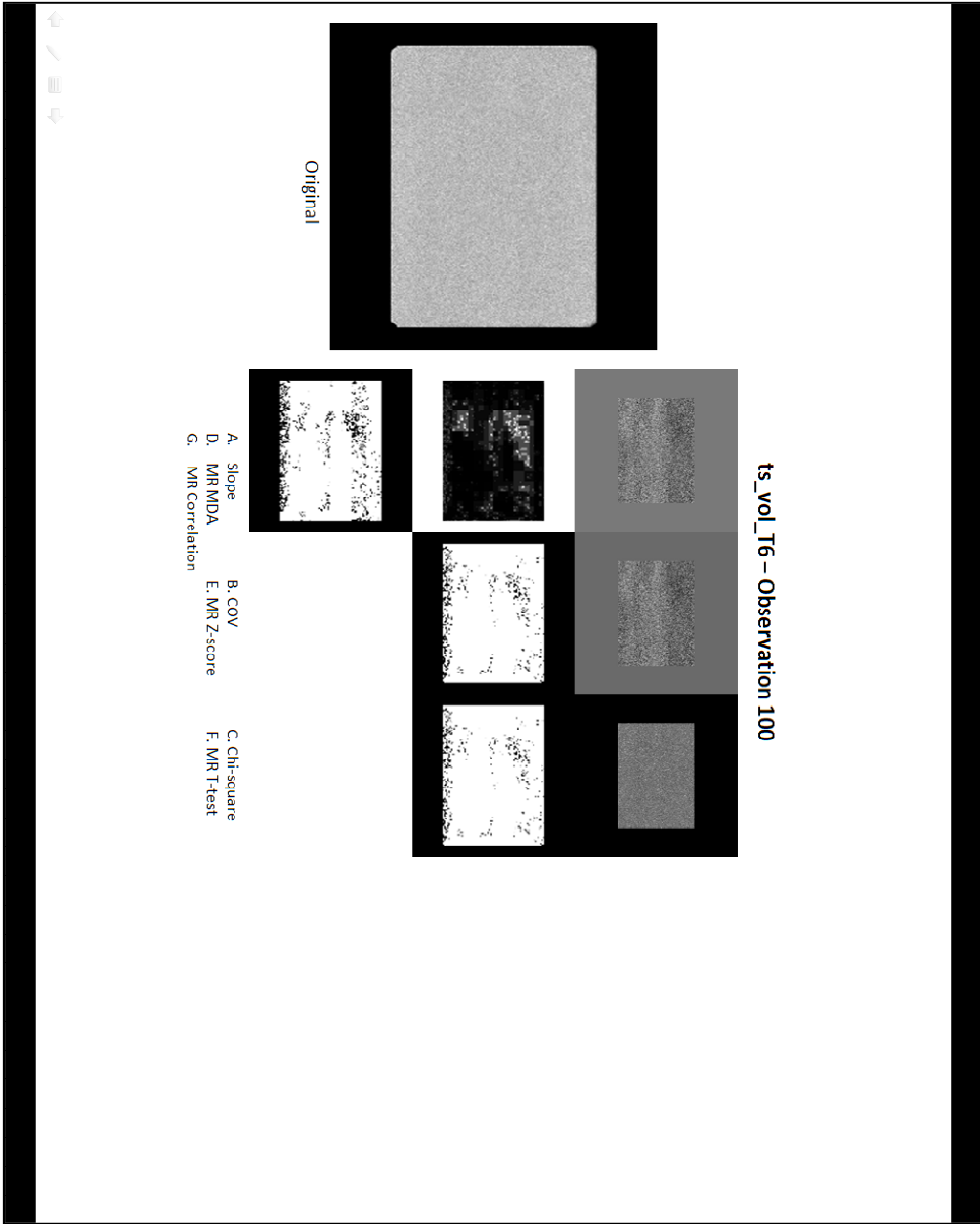


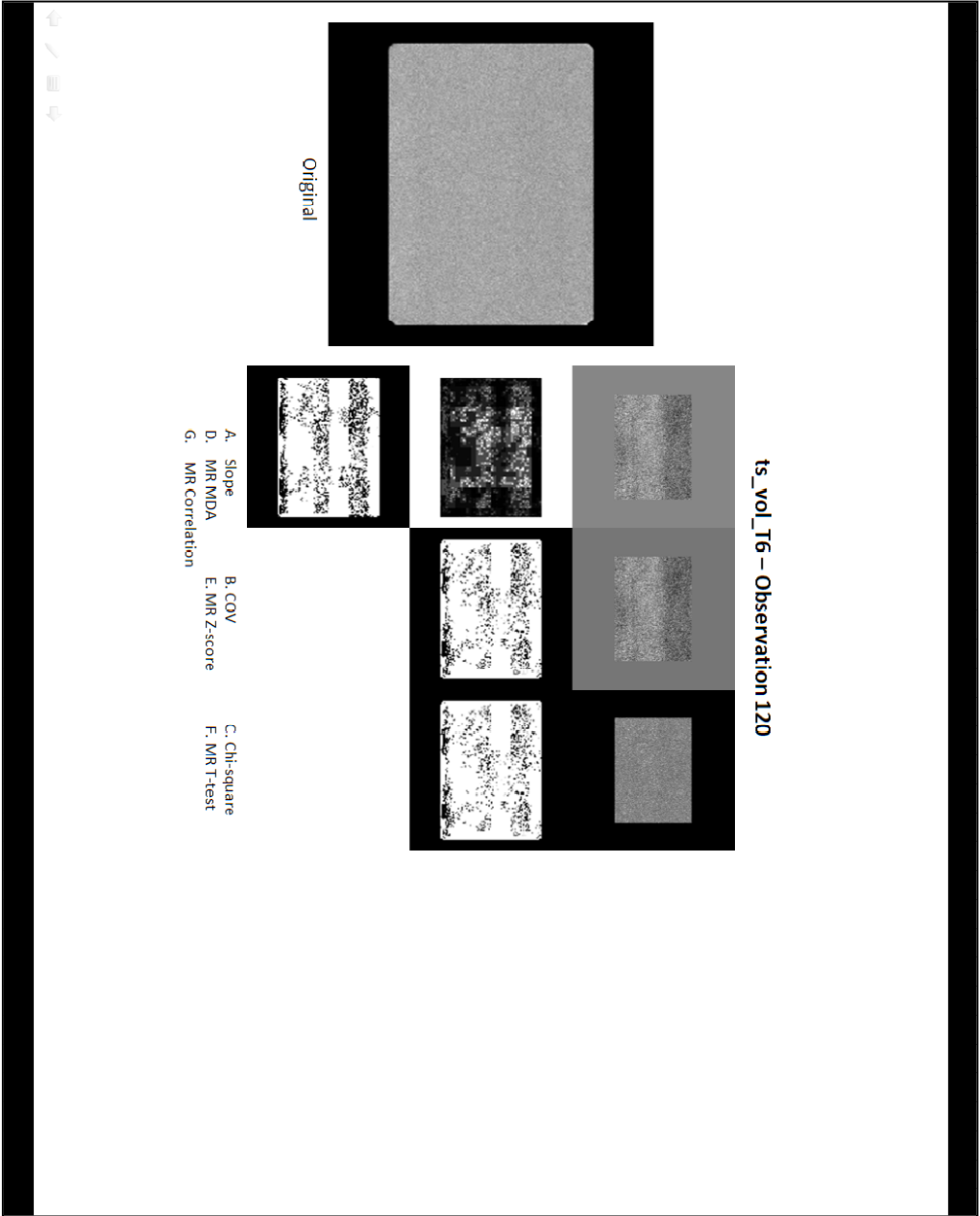
Original



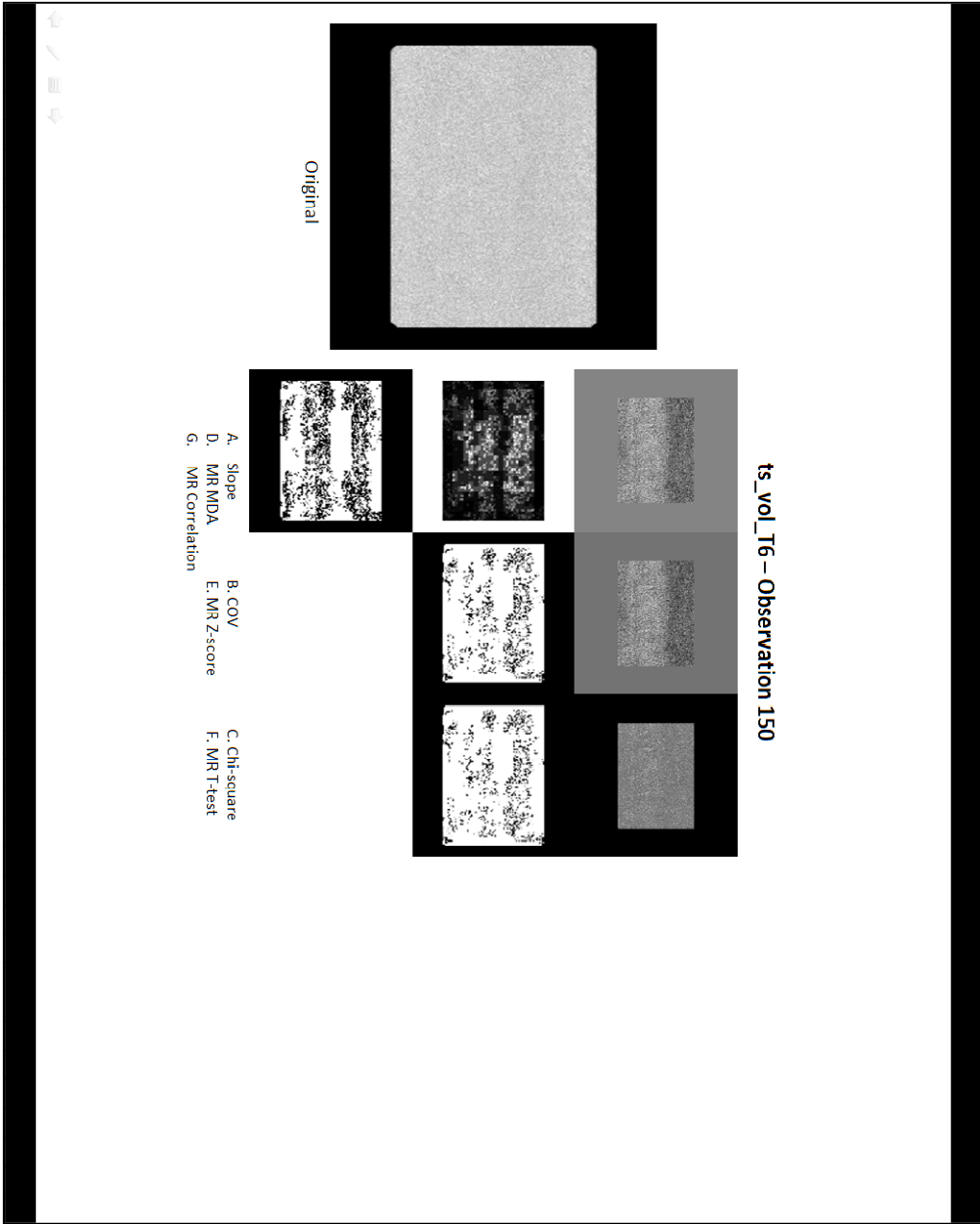






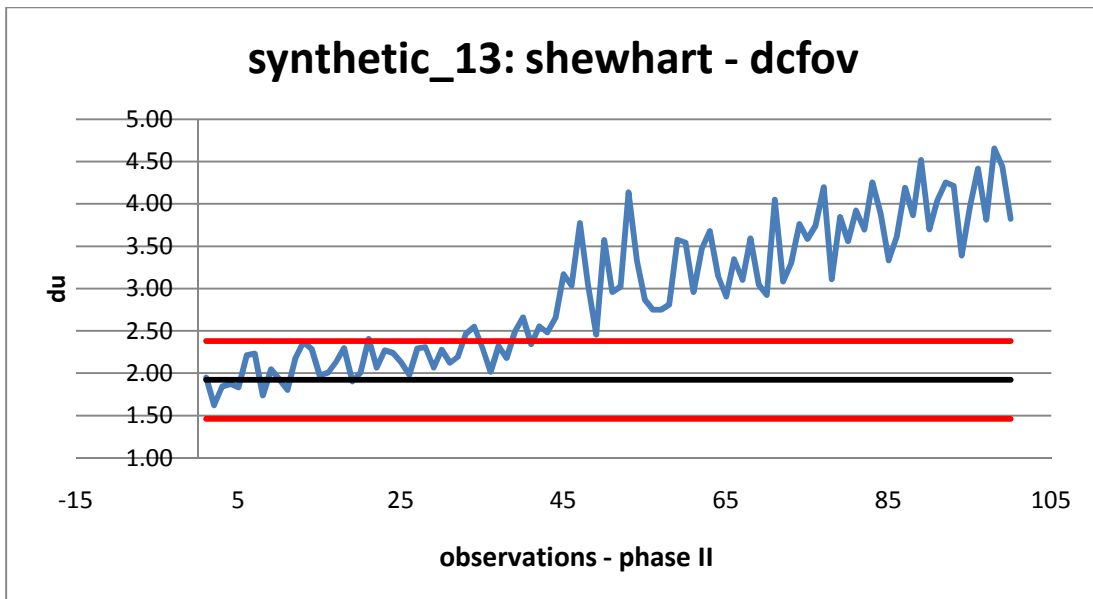
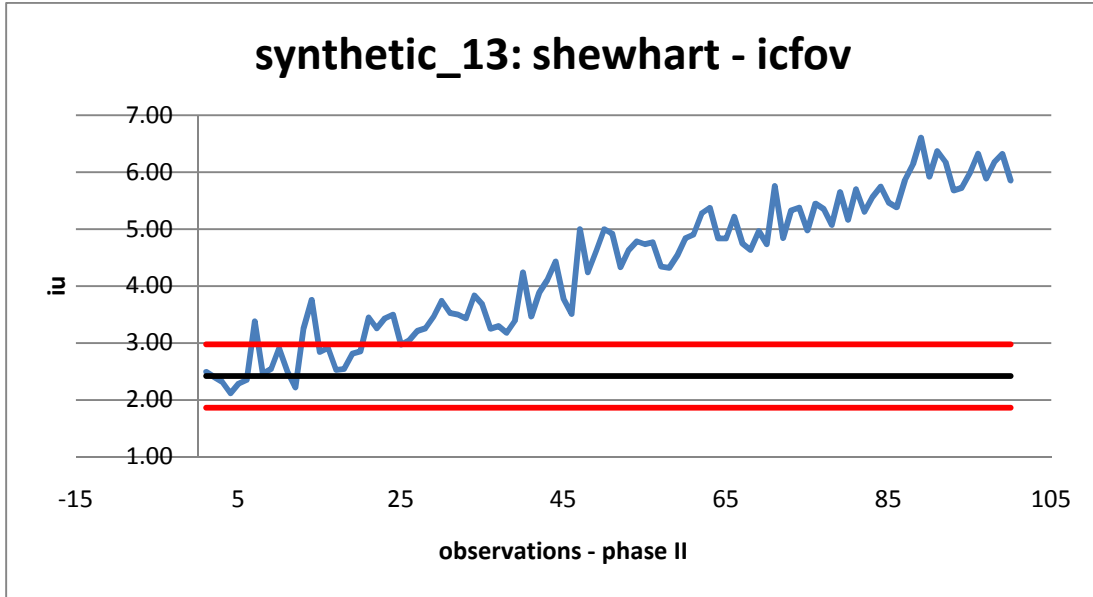


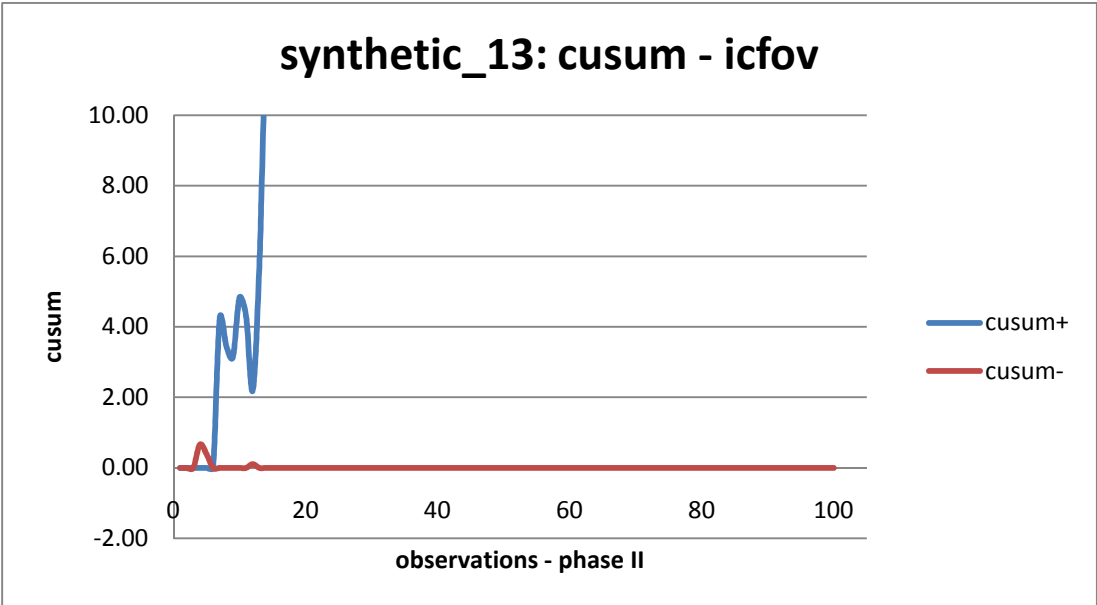
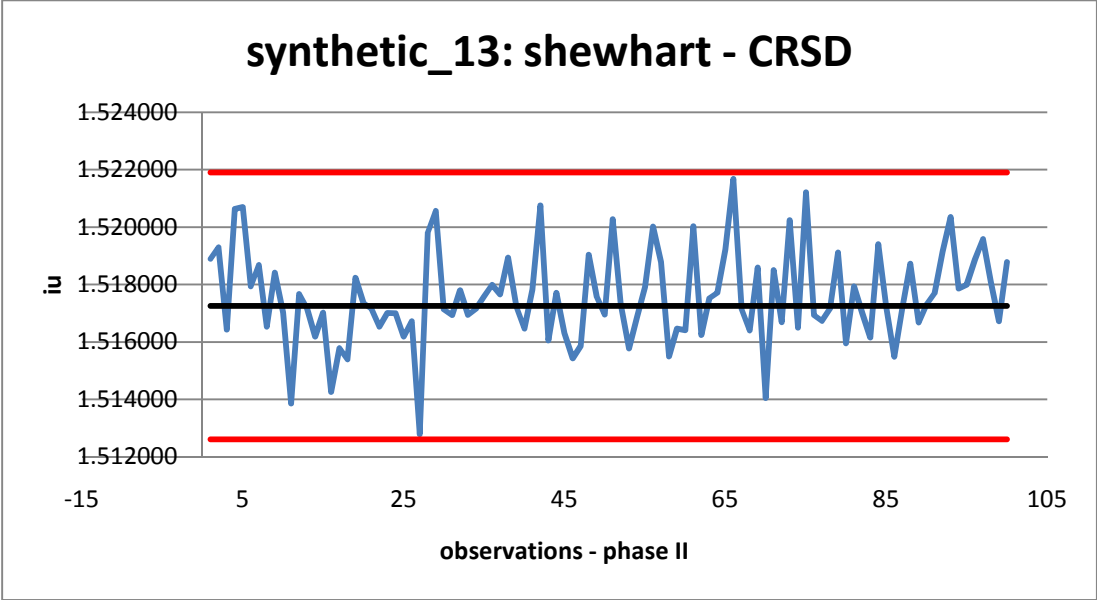


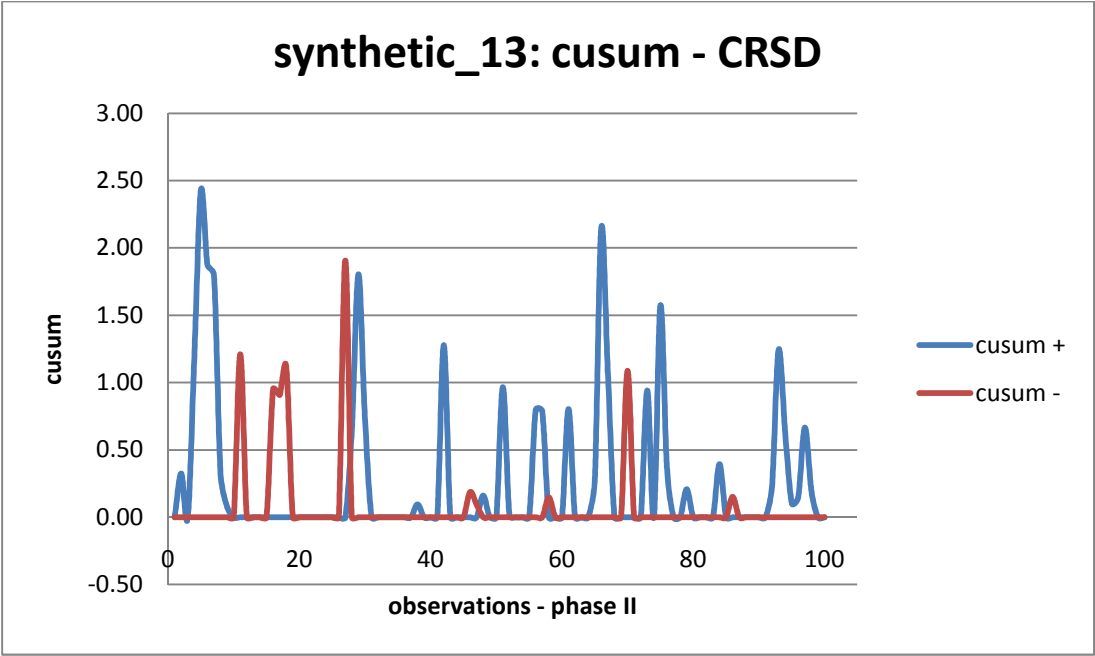
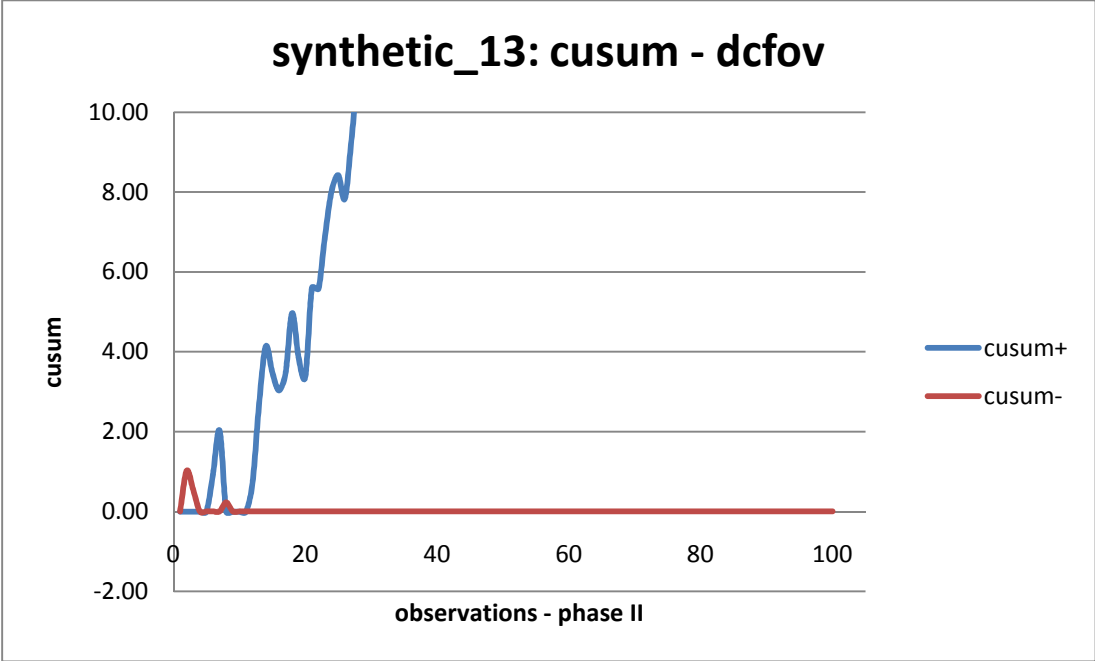


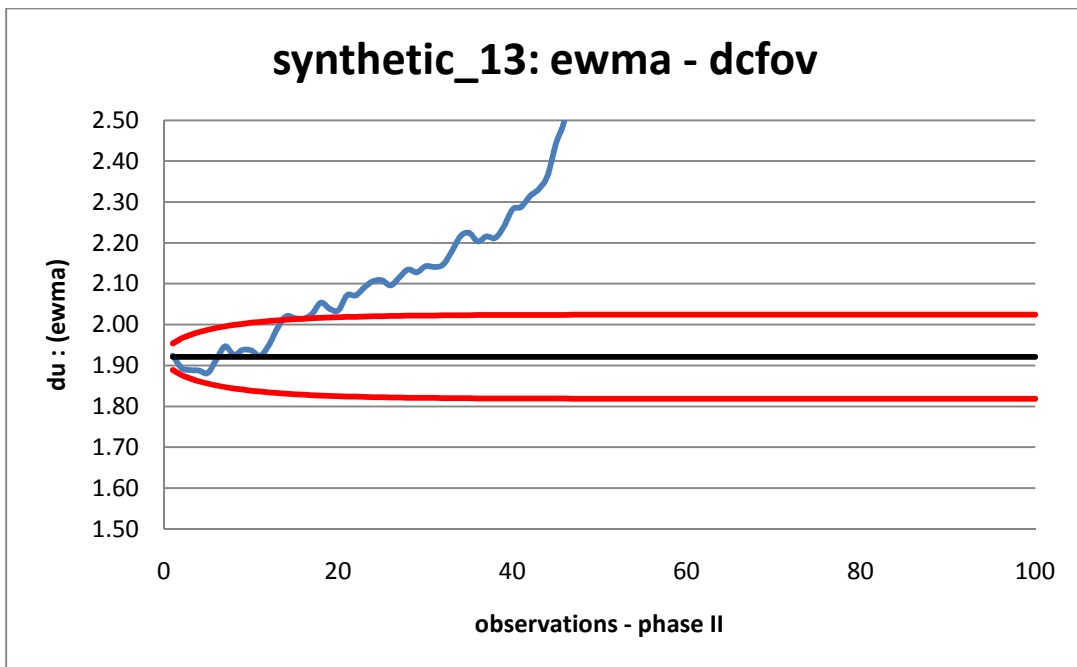
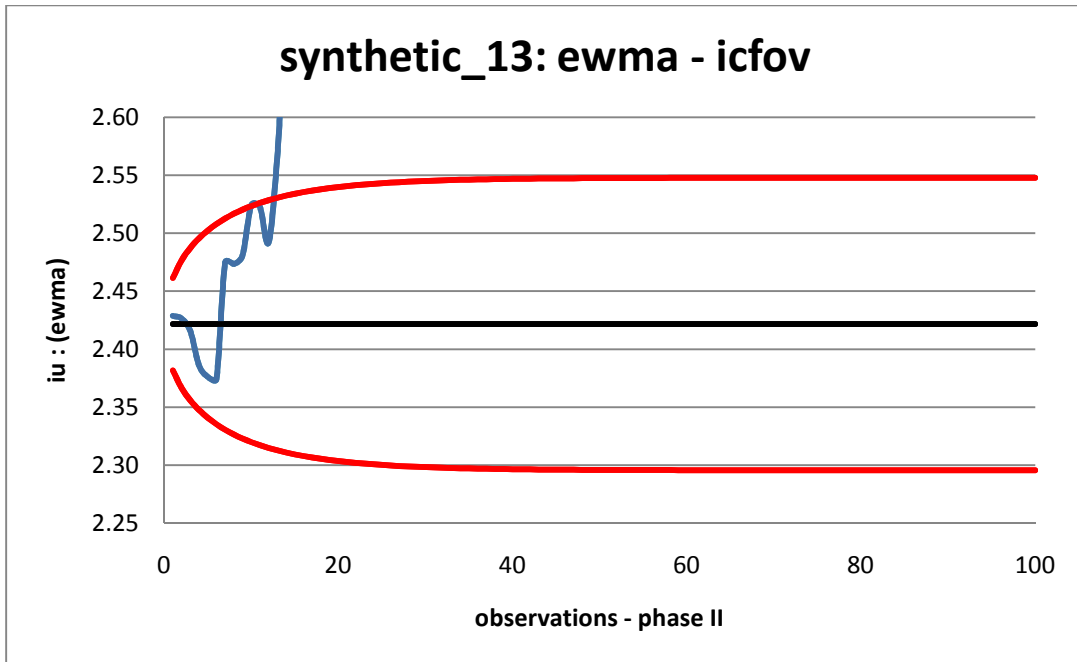
### 9.3

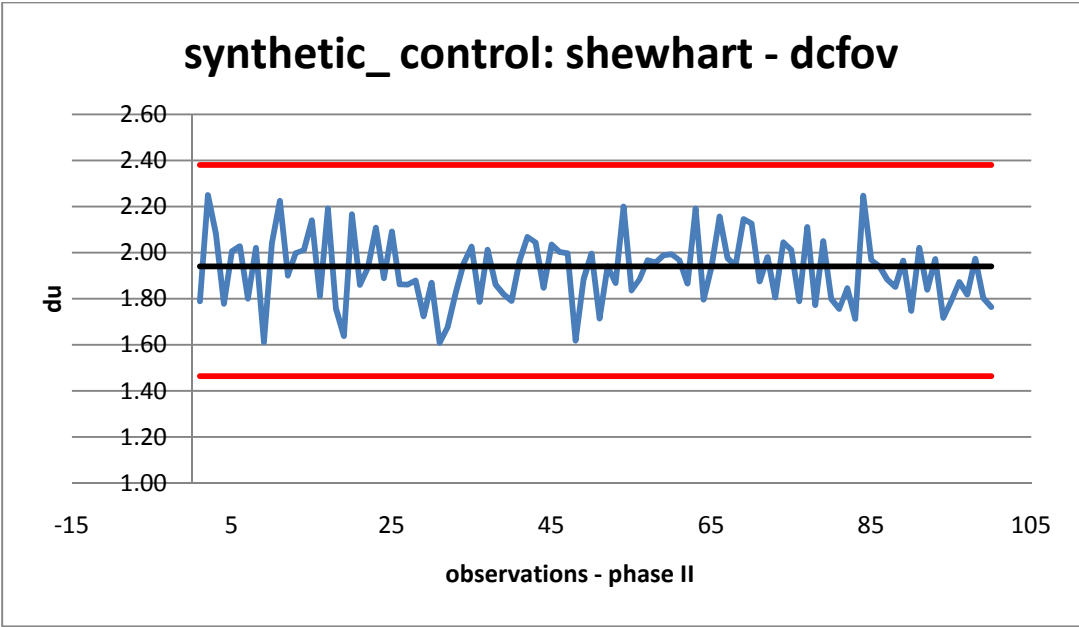
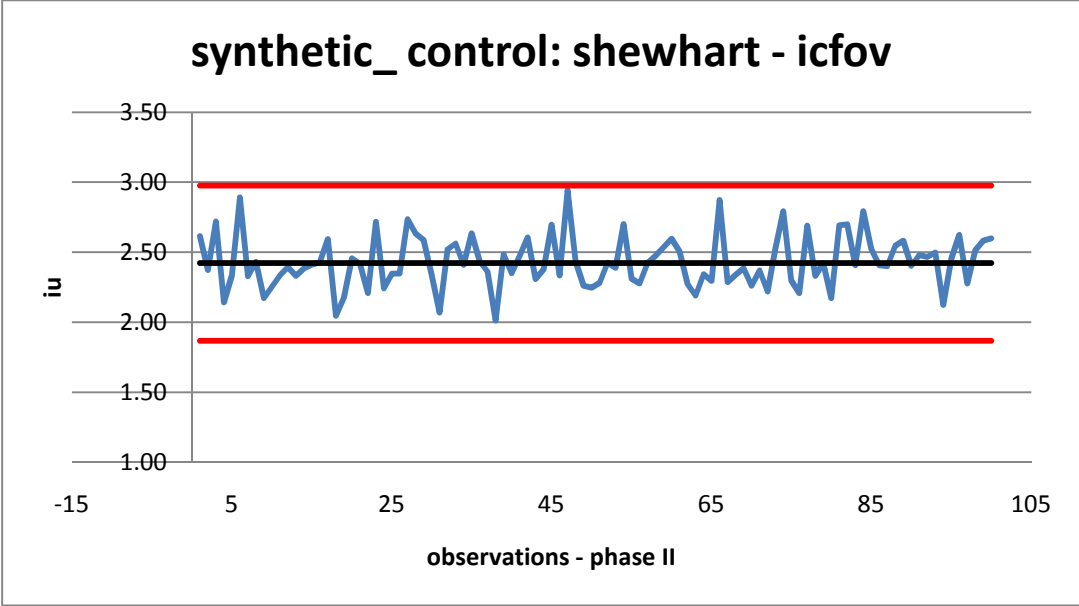
### NEMA Uniformity Trends Using Statistical Process Control

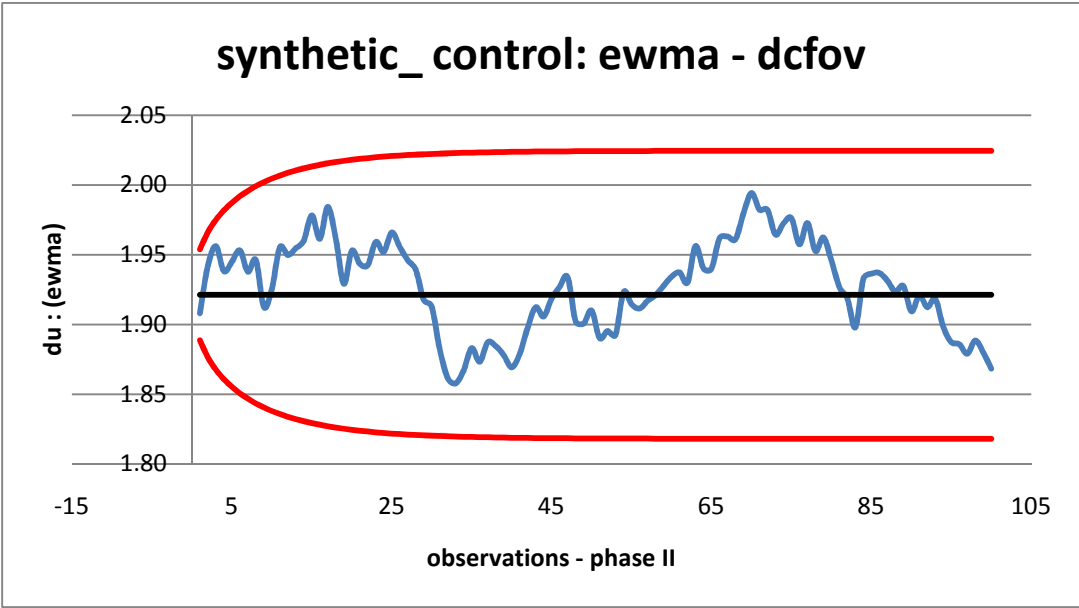
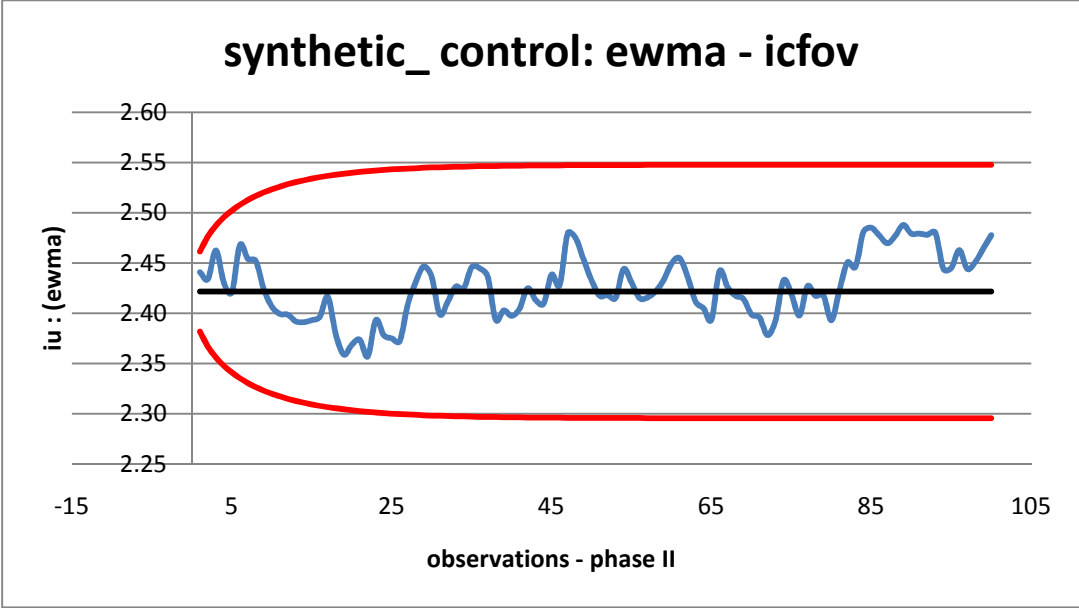


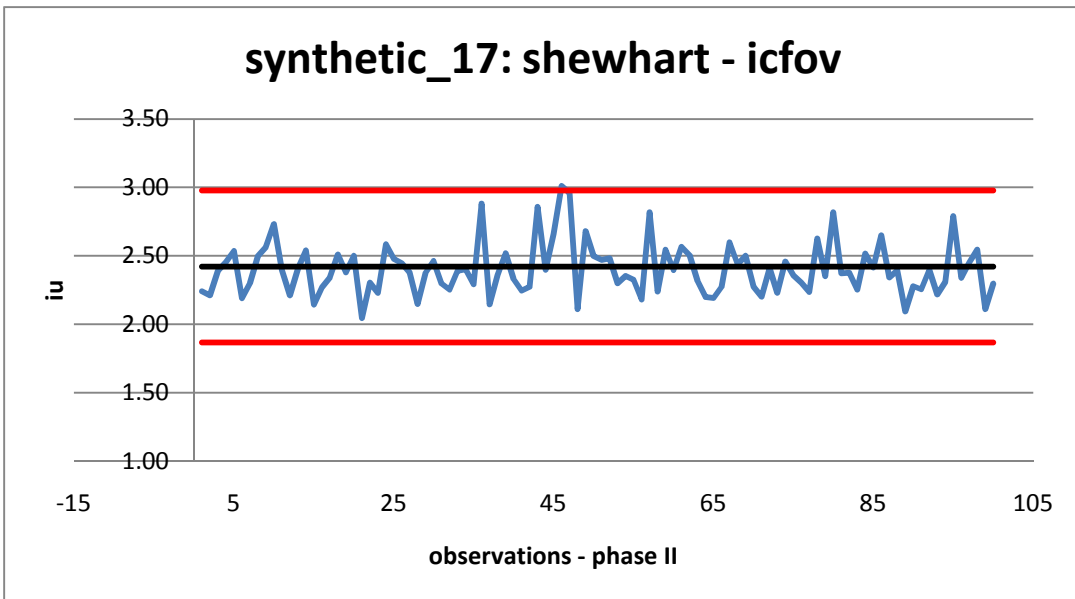
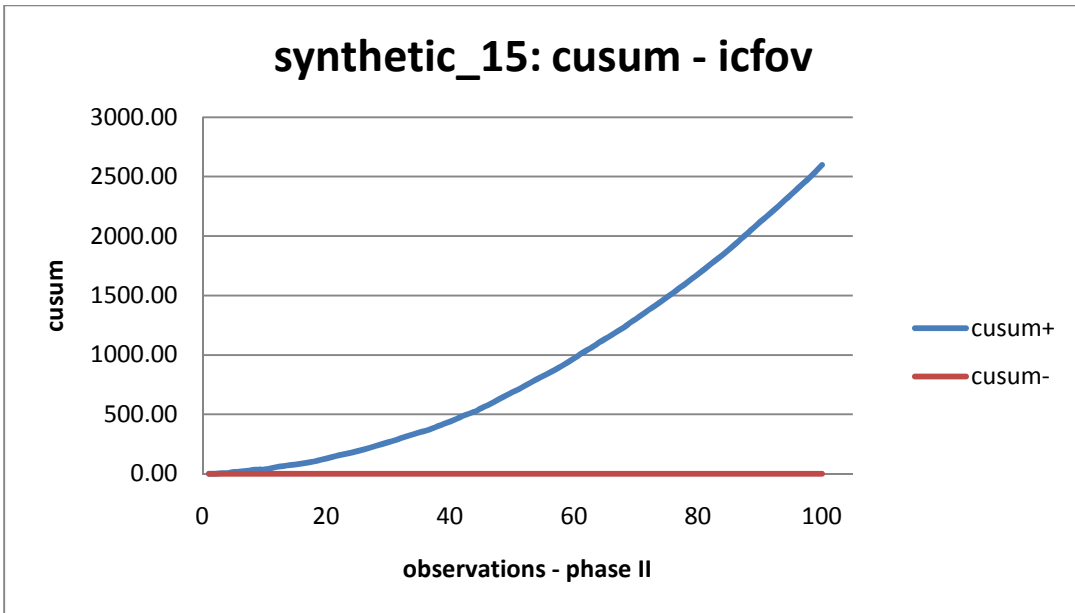




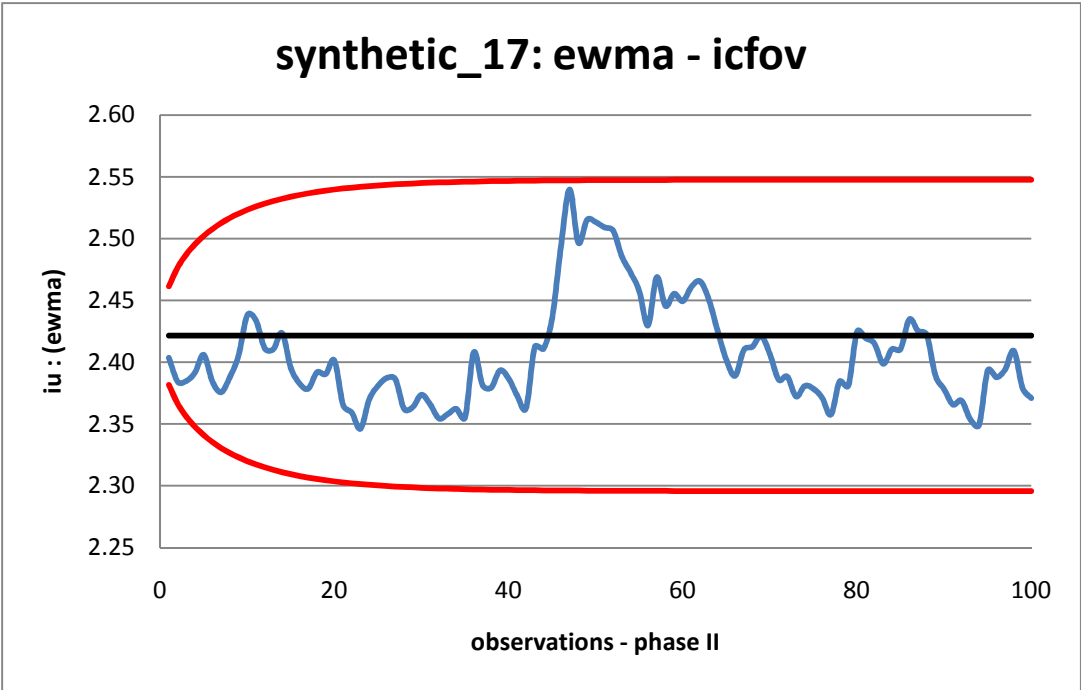
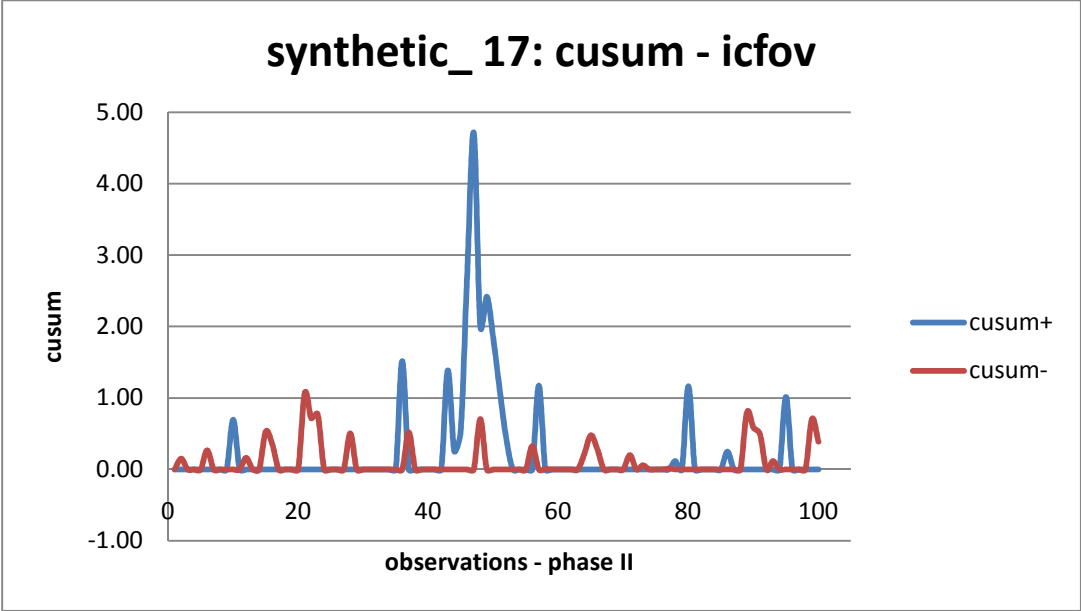




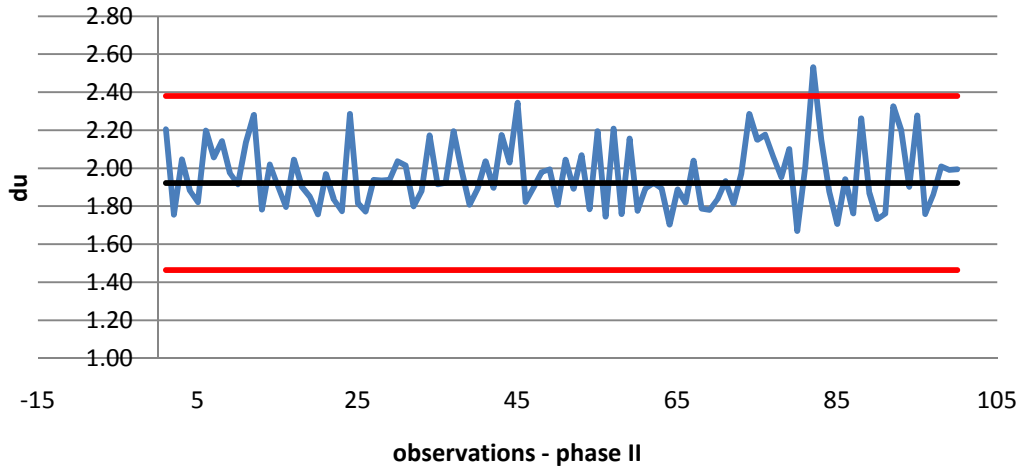




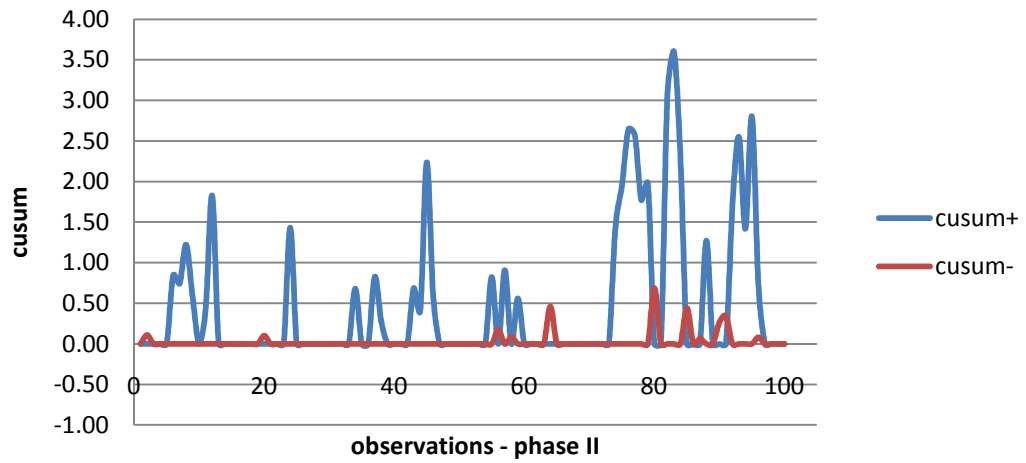


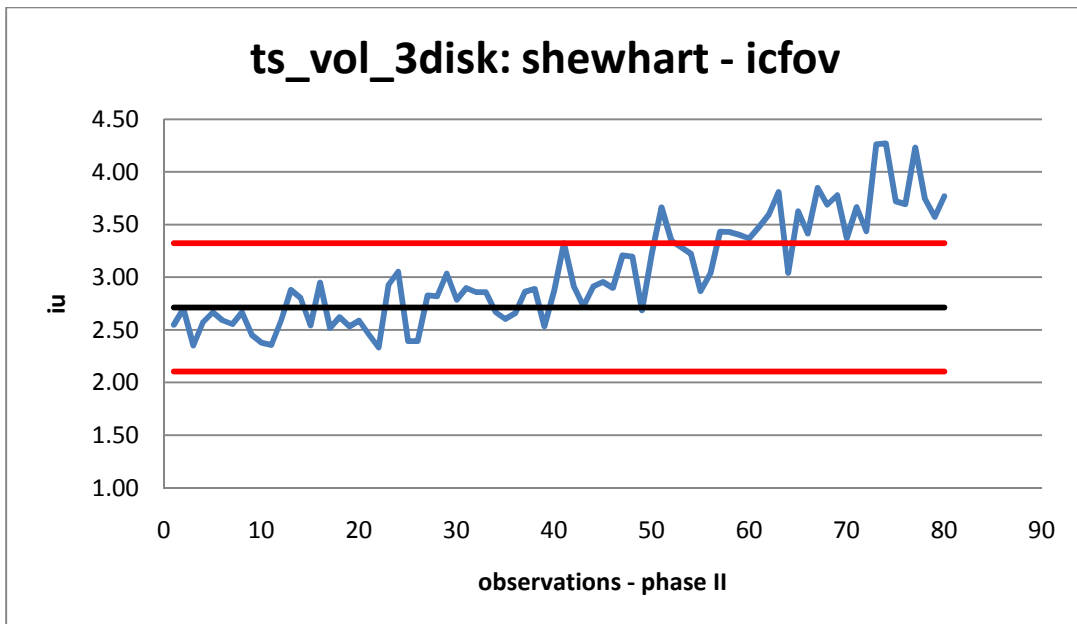
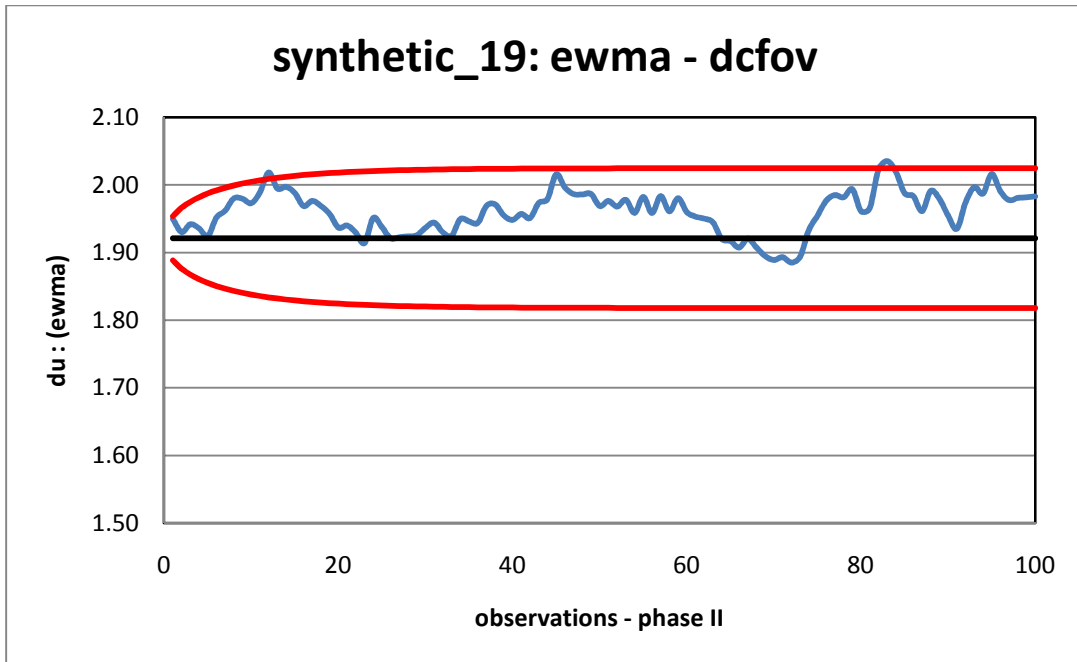


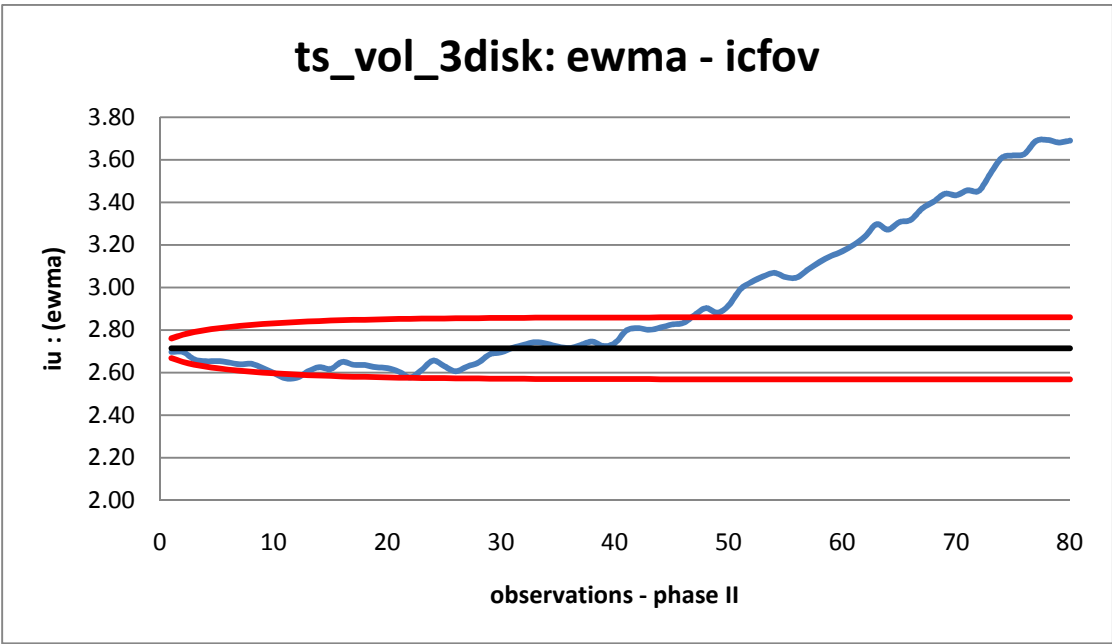
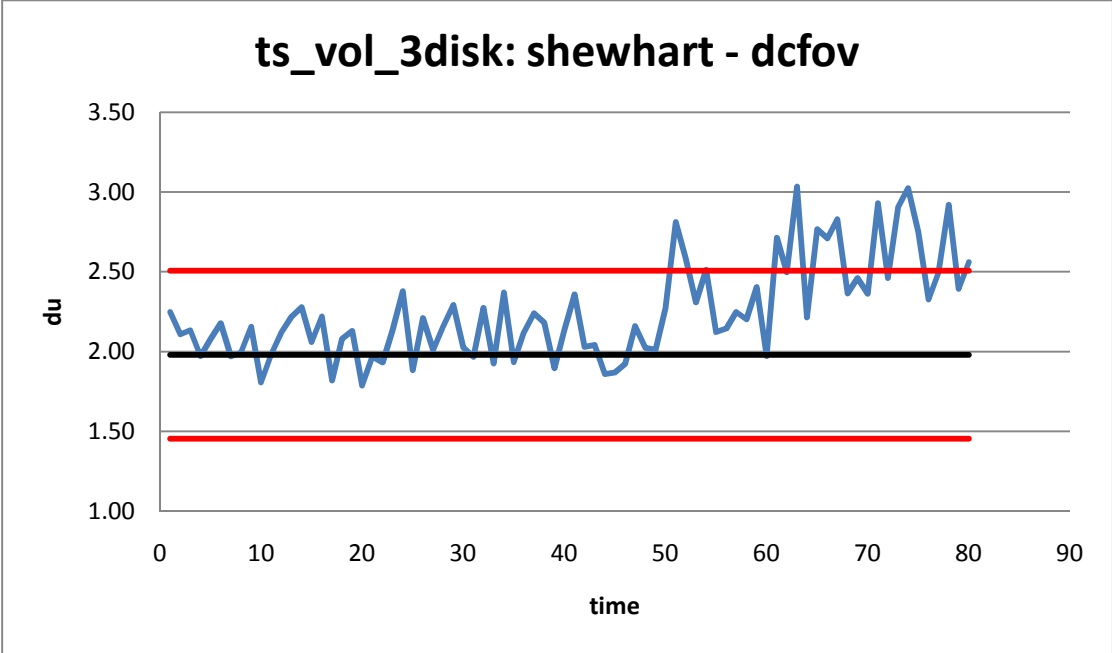
synthetic\_19: shewhart - dcfov

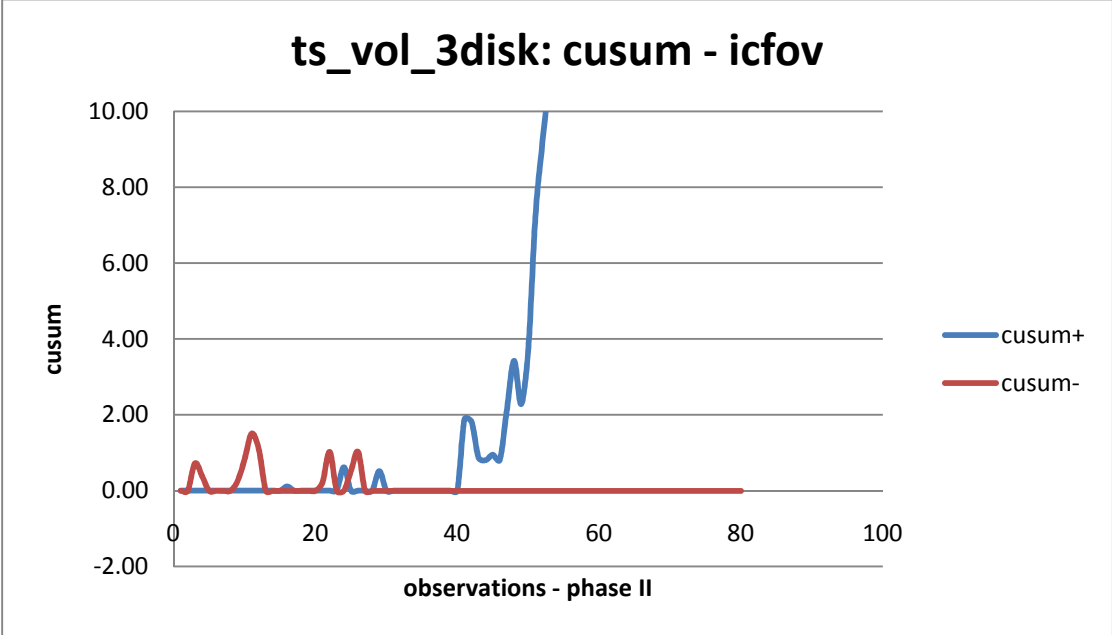
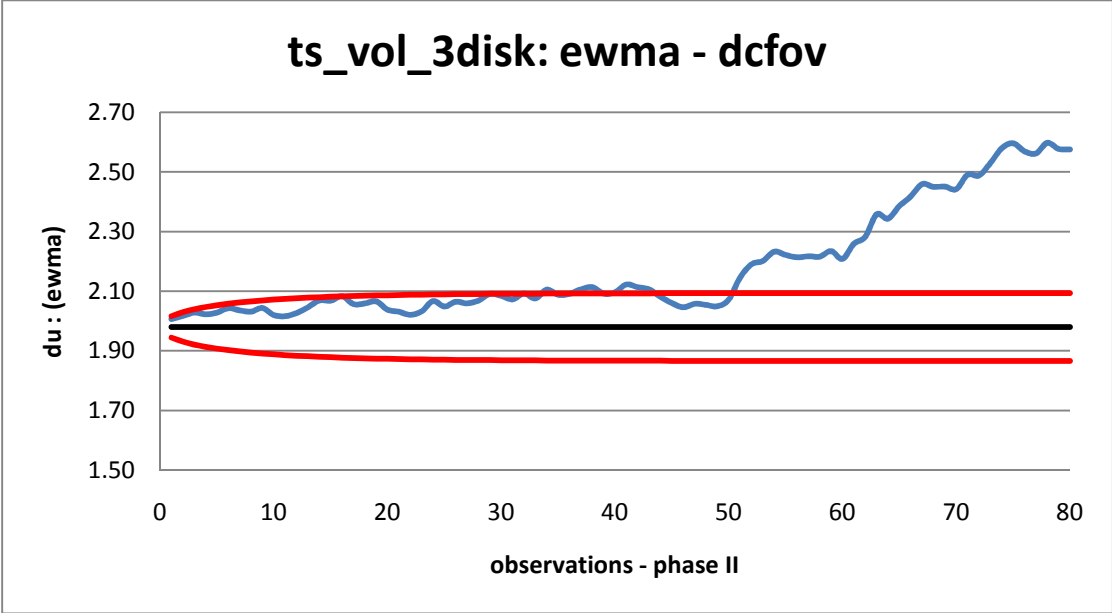


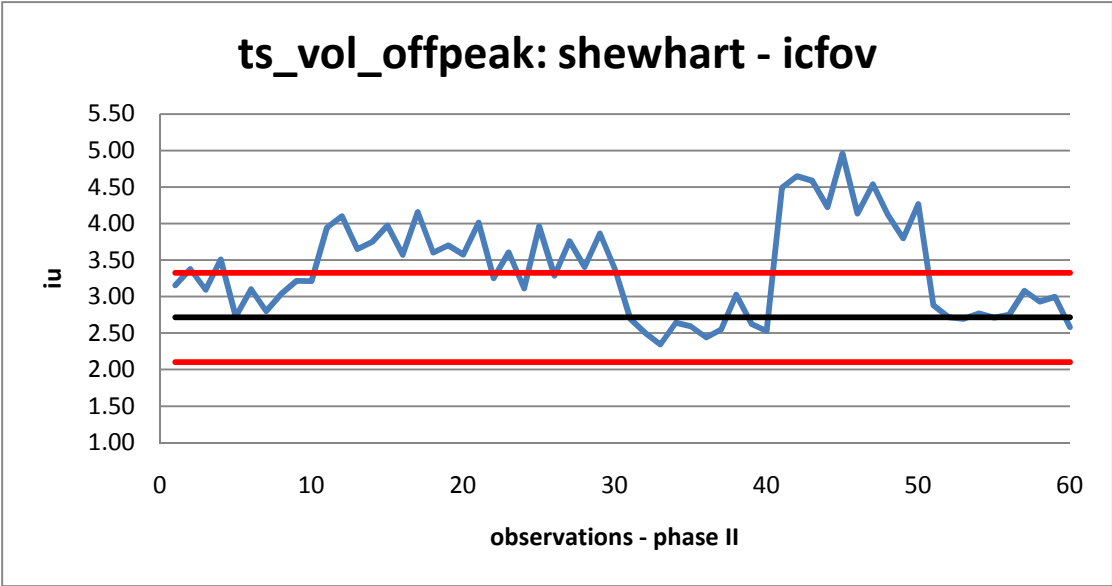
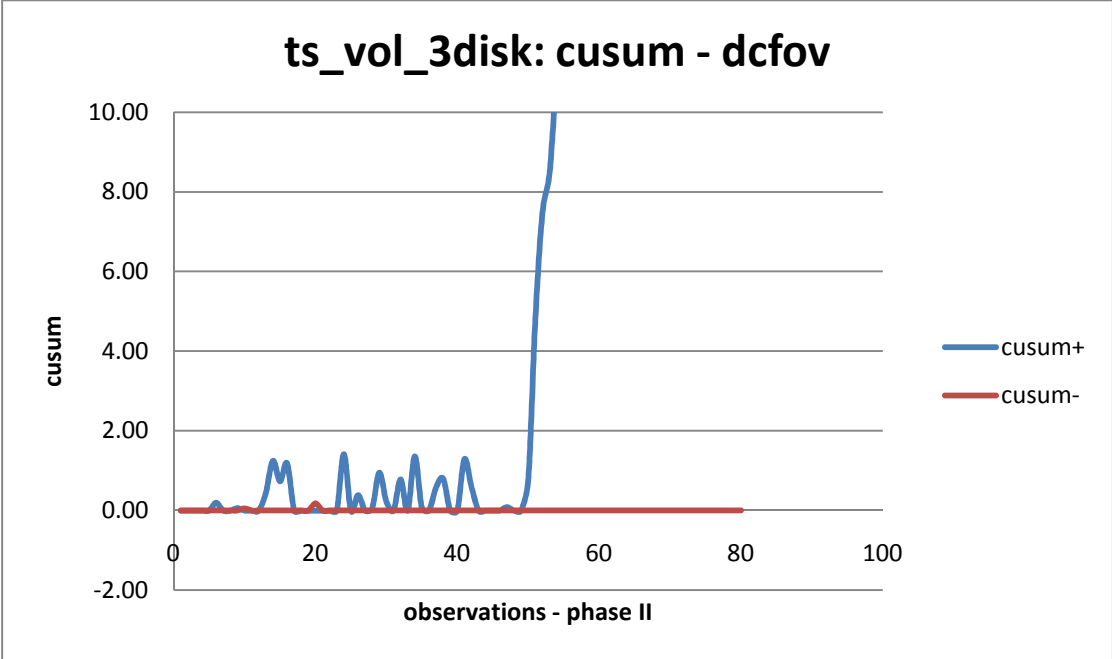
synthetic\_19: cusum - dcfov

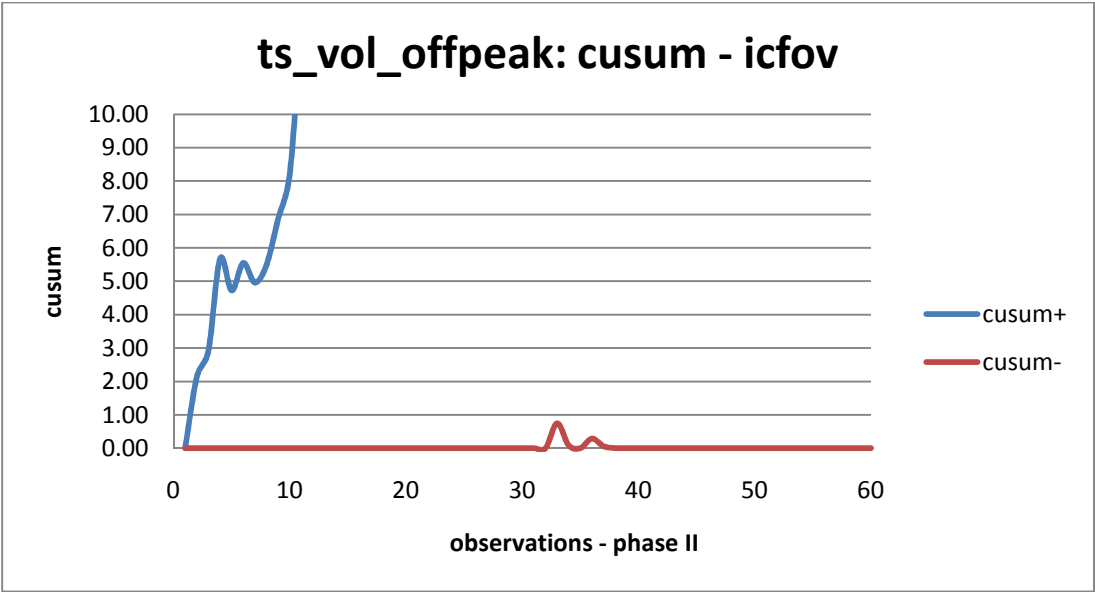
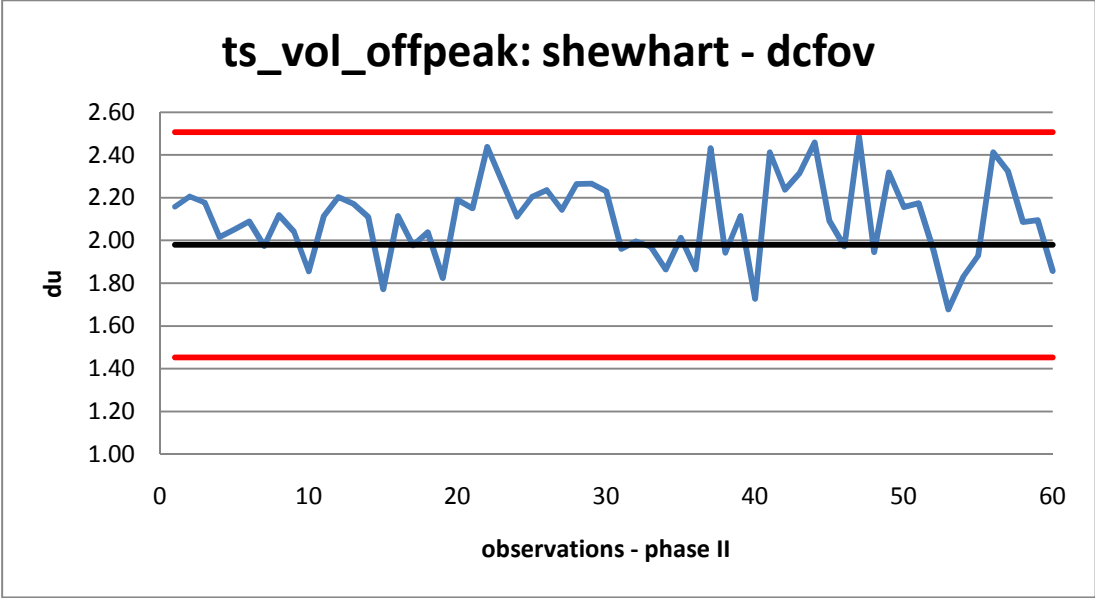


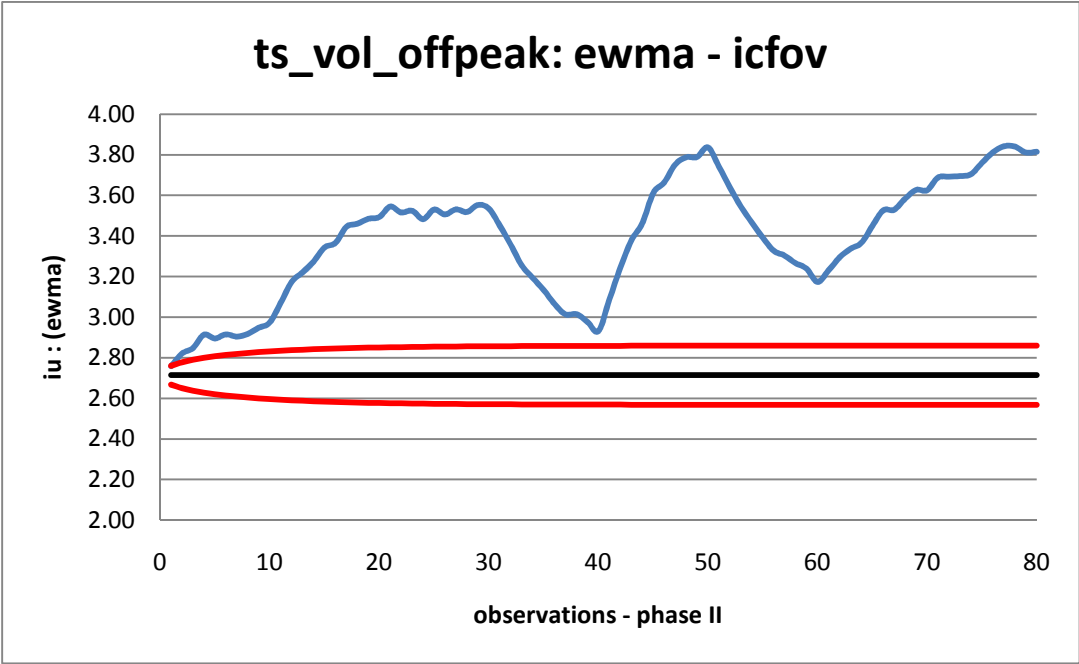
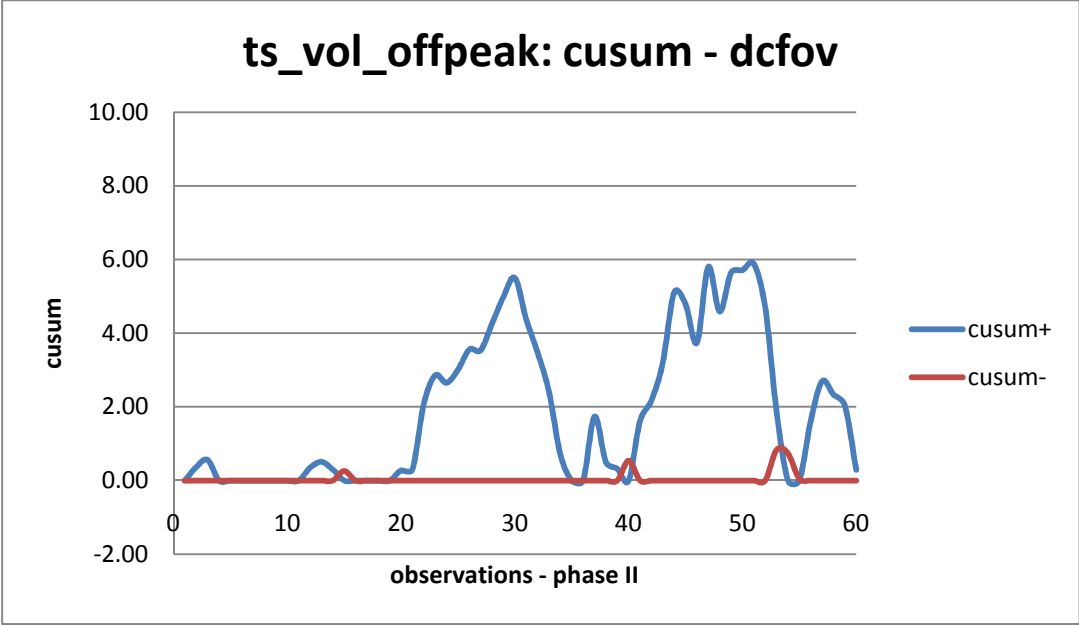




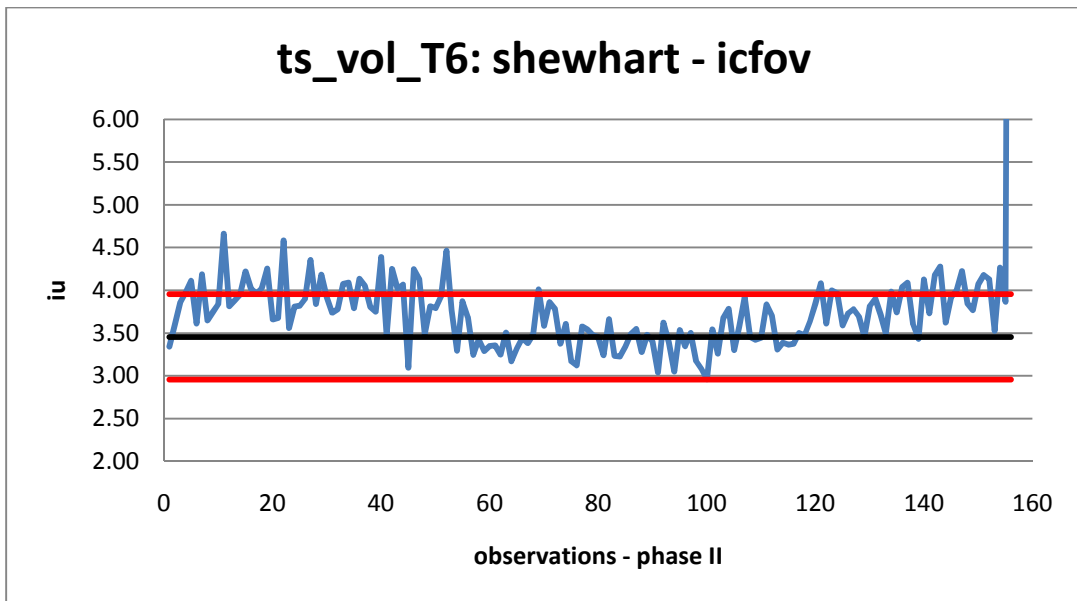
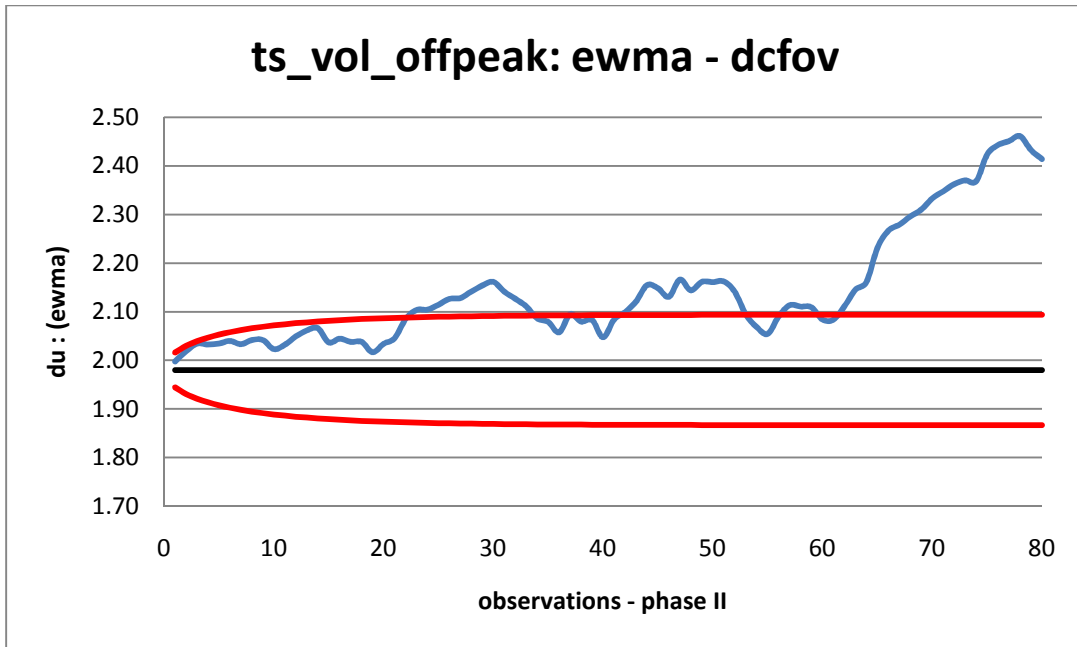


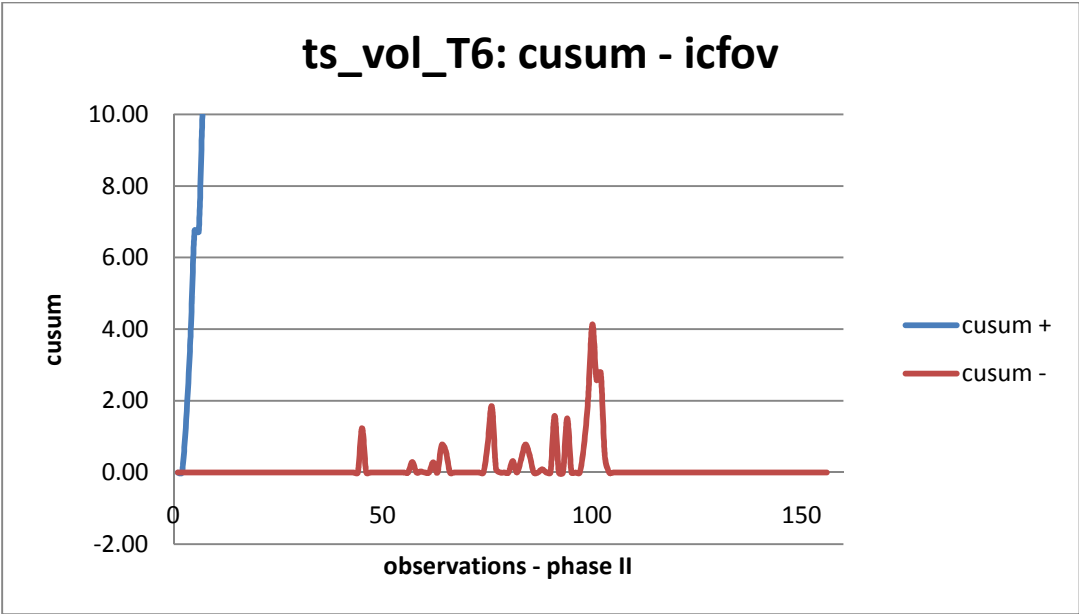
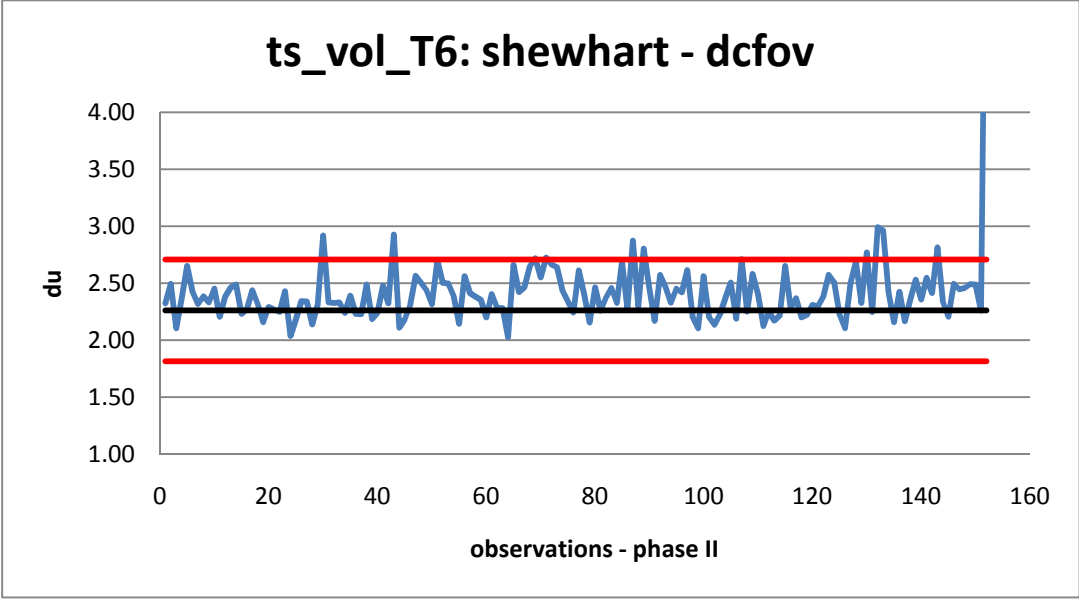


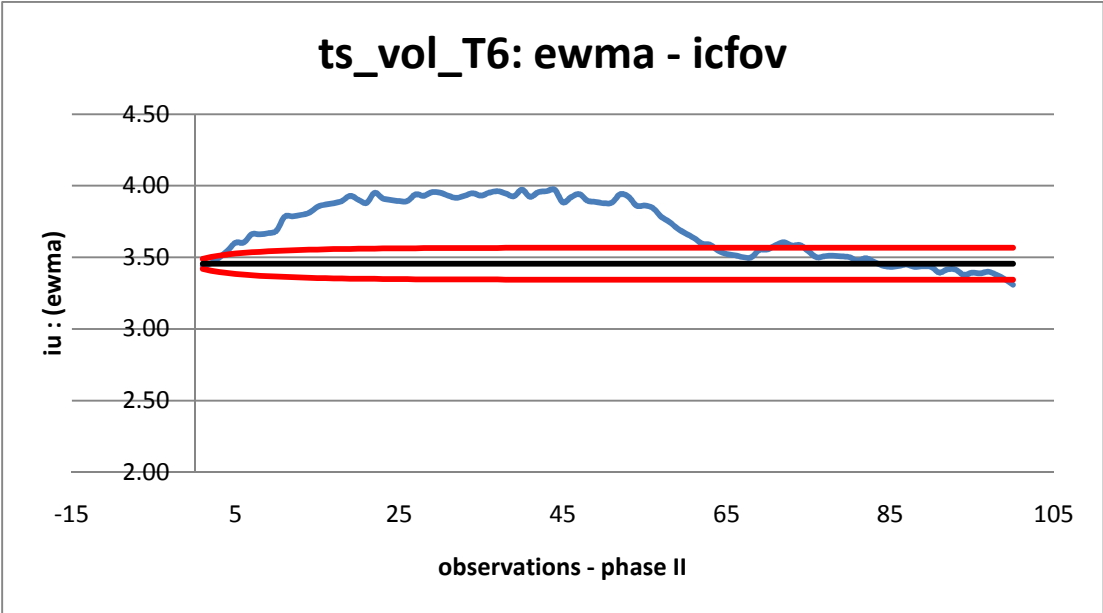
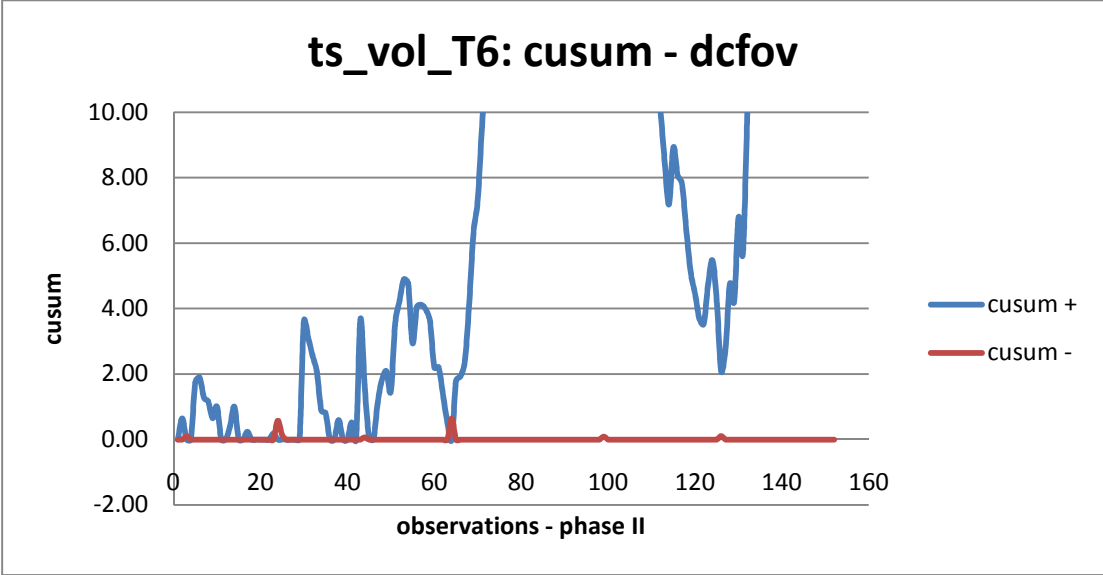


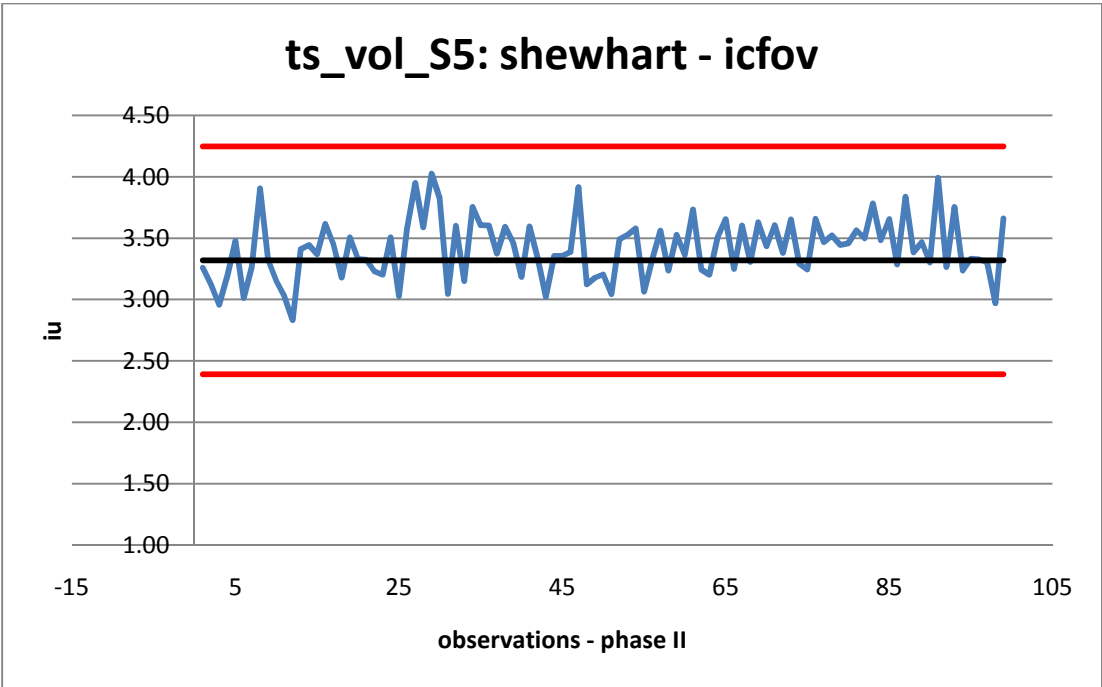
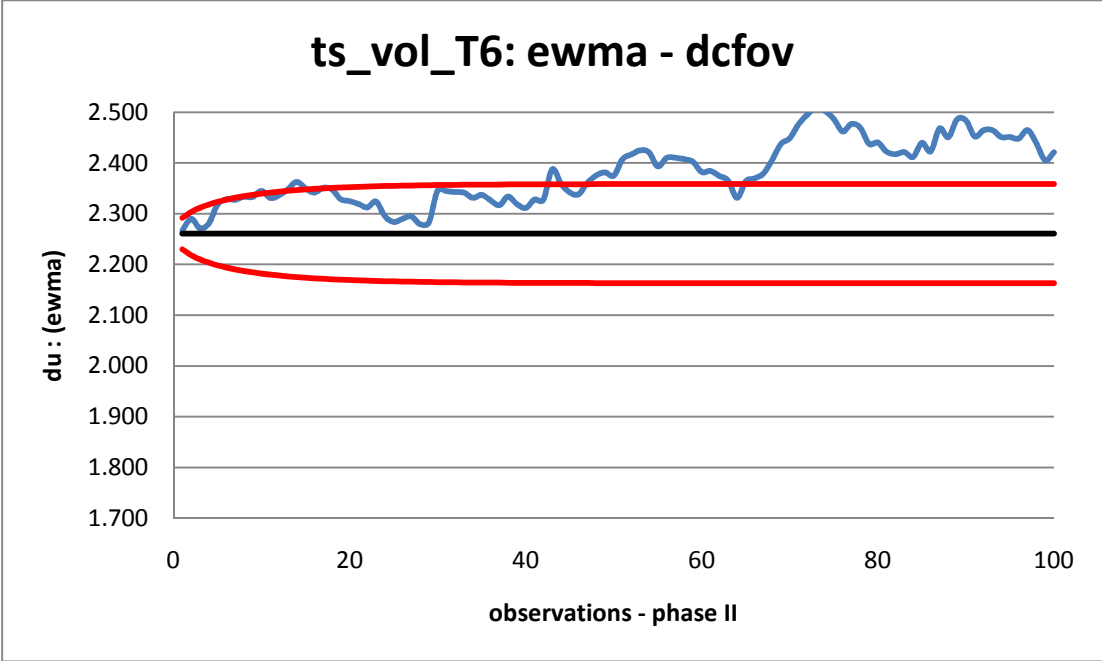


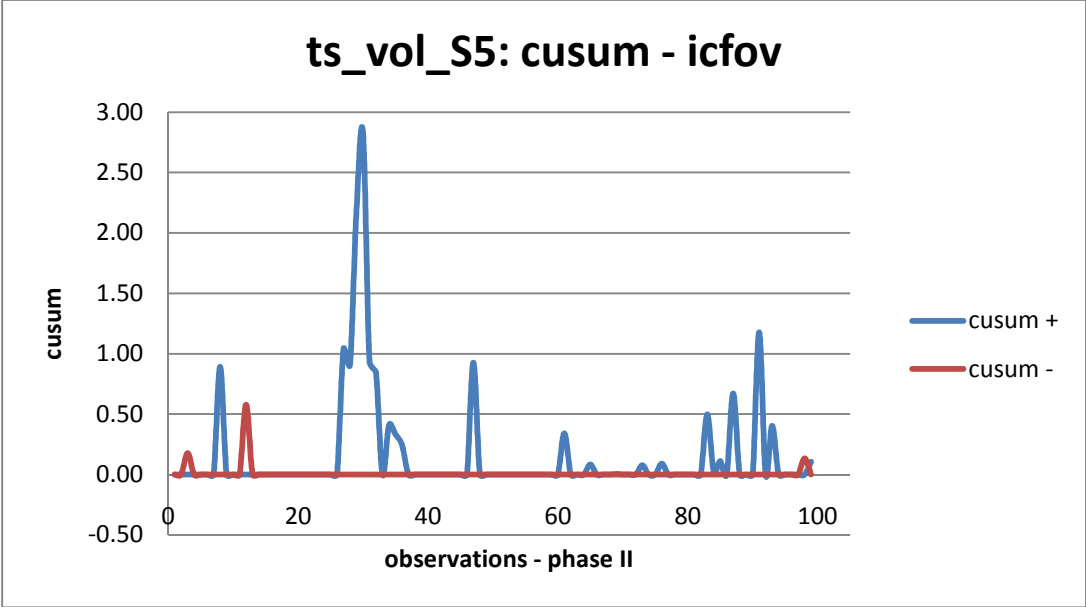
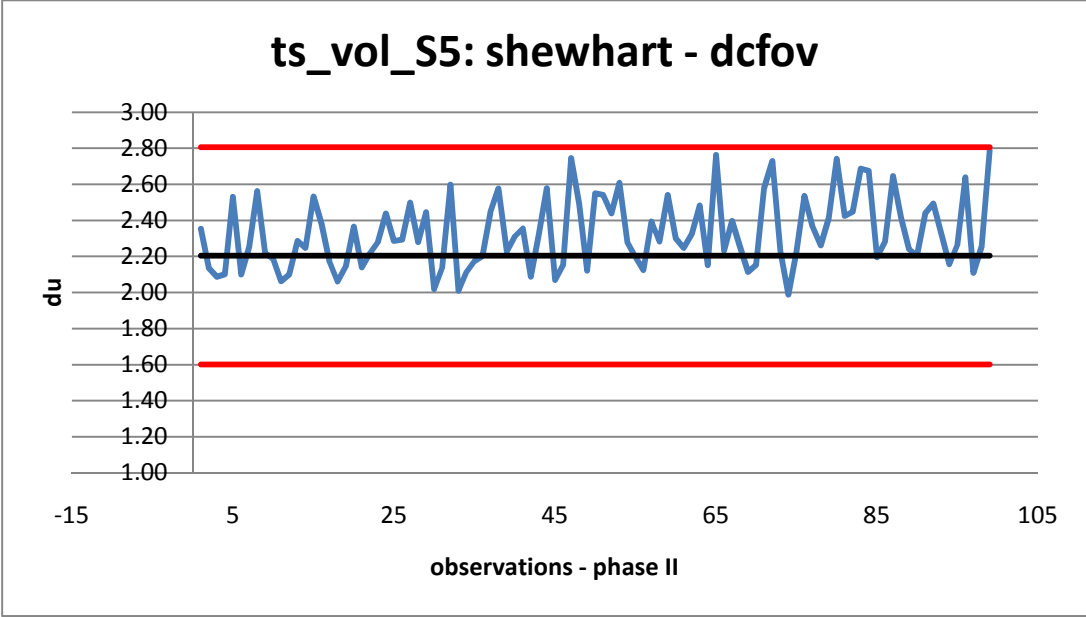


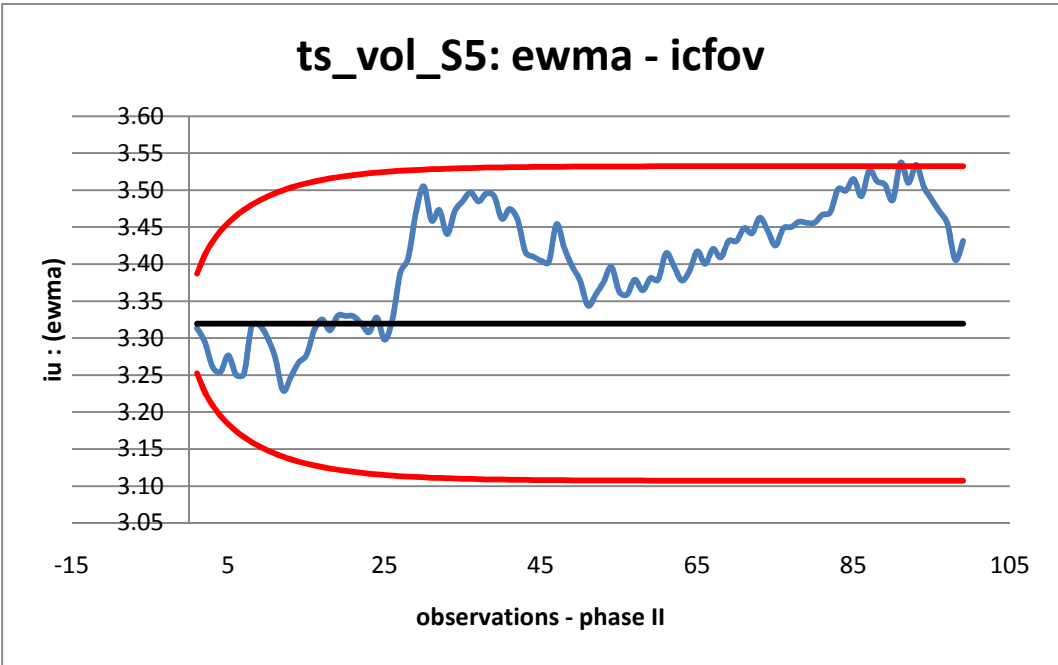
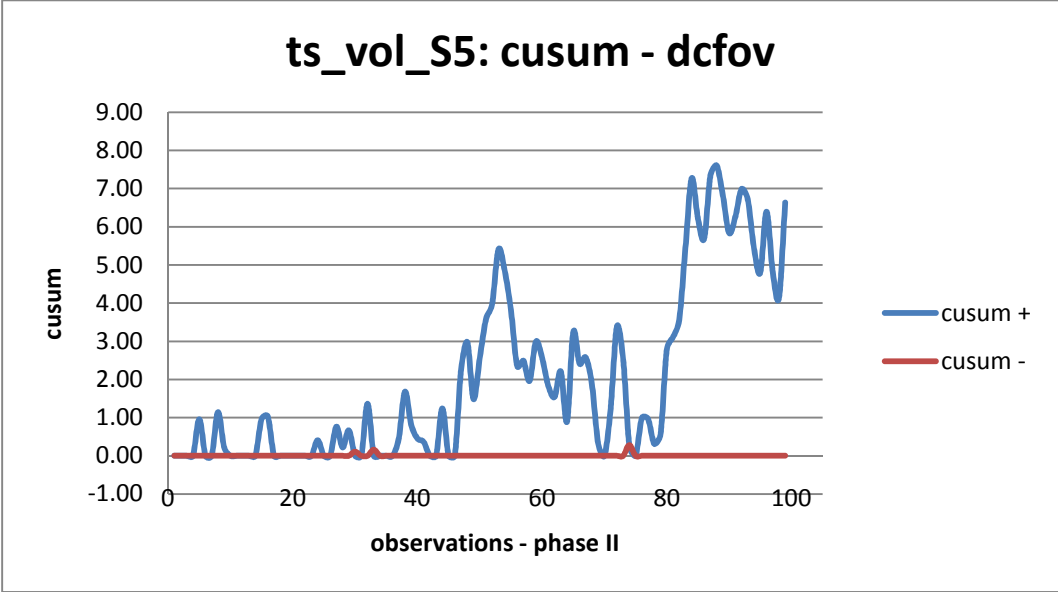


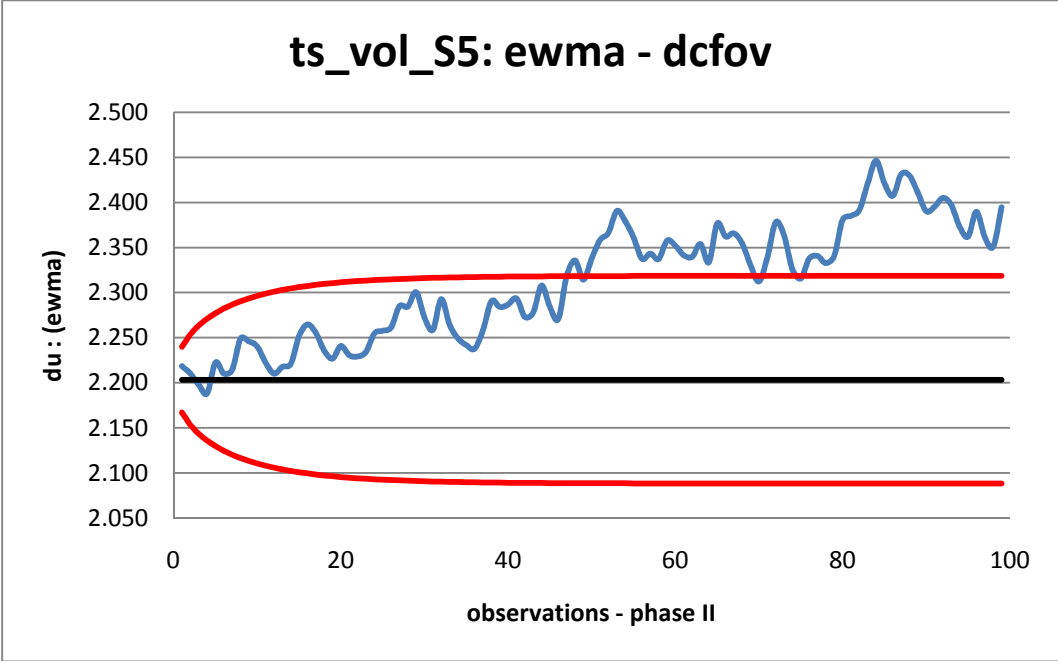












## **Vita**

Bradley Kyle Lofton was born in Houston, Texas on November 28, 1979, the son of Becky Lynn Lofton and Larry Keith Lofton. He grew up in Midland, TX before graduating from Robert E. Lee High School. He began his undergraduate work at Texas A&M University in 1998, receiving a B.S in biomedical engineering in December 2002. After graduating, he worked as a clinical engineer at the Methodist Hospital in Houston, TX before joining MD Anderson's Department of Imaging Physics as a Physics Technologist, where he worked for almost 5 years. In August 2008, he entered The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences.

Permanent address:  
3607 Kennedy Dr.  
Pearland, TX 77584