

a) Title:

## **An Automated Protocol for Performance Benchmarking a Widefield Fluorescence Microscope**

b) Authors:

Michael Halter<sup>1</sup>, Elianna Bier<sup>2</sup>, Paul C. DeRose<sup>1</sup>, Gregory A. Cooksey<sup>1</sup>, Steven J. Choquette<sup>1</sup>, Anne L. Plant<sup>1</sup>, John T. Elliott<sup>1</sup>

*<sup>1</sup>Biosystems and Biomaterials Division  
Materials Measurement Laboratory  
National Institute of Standards and Technology  
Gaithersburg, MD 20899*

*<sup>2</sup>Augsburg College  
Department of Physics  
2211 Riverside Avenue  
Minneapolis, MN 55454*

c) Running Headline:

Benchmarking the Analytical Performance of an Image Cytometer

d) Corresponding Author:

Michael Halter

Biosystems and Biomaterials Division  
Materials Measurement Laboratory  
National Institute of Standards and Technology  
Gaithersburg, MD 20899  
Phone: 301 975 4195  
Fax: 301 330 3447  
Email: michael.halter@nist.gov

## ABSTRACT

Widefield fluorescence microscopy is a highly used tool for visually assessing biological samples and for quantifying cell responses. Despite its widespread use in high content analysis and other imaging applications, few published methods exist for evaluating and benchmarking the analytical performance of a microscope. Easy-to-use benchmarking methods would facilitate the use of fluorescence imaging as a quantitative analytical tool in research applications, and would aid the determination of instrumental method validation for commercial product development applications. We describe and evaluate an automated method to characterize a fluorescence imaging system's performance by benchmarking the detection threshold, saturation and linear dynamic range to a reference material. The benchmarking procedure is demonstrated using two different materials as the reference material, uranyl-ion-doped glass and Schott 475 GG filter glass. Both are suitable candidate reference materials that are homogeneously fluorescent and highly photostable, and the Schott 475 GG filter glass is currently commercially available. In addition to benchmarking the analytical performance, we also demonstrate that the reference materials provide for accurate day to day intensity calibration.

## Key Terms

cell measurement, image cytometer, high content screening, calibration, quantitative microscopy, instrument qualification, benchmark

## Introduction

Despite the widespread use of optical microscopy to characterize biological cells, there are no widely accepted protocols and reference materials for monitoring day to day instrument performance(1). In addition to research applications, fluorescence microscopy is used in clinical diagnostics, drug development, and development of cell therapy products. Reproducibility of research results will be improved by the benchmarking of microscopy data. Also, in clinical and commercial applications, reference materials and benchmarking methods are needed for instrument validation to assure data quality and facilitate acceptance by regulatory organizations. To address the need for microscopy reference materials, we have developed a simple procedure to benchmark the analytical performance of a widefield fluorescence microscope. Robust reference materials that are photostable and

homogeneously fluorescent allow the performance of an instrument to be monitored over time to assure the repeatability of results.

In this study, we evaluated two candidate reference materials: uranyl-ion-doped glass and Schott 475 GG filter glass. The uranyl-ion-doped glass is a modified composition of an existing National Institute of Standards and Technology (NIST) Standard Reference Material (SRM®) that is certified for its fluorescence emission spectrum(2). The other material, Schott 475 GG filter glass, is commercially available as a long pass filter. The fluorescence property of the Schott 475 GG is an artifact of colloidal particles present in the glass and it is not certified for any fluorescence characteristic, so the emission spectra and intensity can vary from batch to batch(3). These candidate materials were selected because they exhibit stable and homogeneous fluorescence when excited on a widefield fluorescence microscope. They have emission intensities similar to levels frequently associated with fluorophore labeling of cells, which is an advantageous property because the microscope settings for routine imaging of cells can be used for the benchmarking procedure. The glass materials are also easy to use and durable during routine laboratory handling.

In this technical note, we describe a procedure for benchmarking the analytical performance of a widefield fluorescence microscope to a reference material. The benchmarking procedure normalizes for differences in magnification, physical pixel sizes of the detector and bit depths of the image data so that the analytical metrics can be compared over a wide range of instrument configurations. A script written in MicroManager, an open-source microscopy control software(4), has been developed to automate the procedure and return the benchmarked parameters. We evaluated the procedure by making systematic changes to the optical system, including changes to the read rate and gain of the CCD detector and by inserting additional optical components into the optical train. We demonstrate that our single, automated protocol is sensitive to changes in the low light level detection, the light throughput, the signal to noise ratio and the detector saturation of a widefield microscope. In addition to the analytical performance, we also demonstrate that the reference material provides for accurate day to day intensity calibration. The protocol and glass reference materials used in this study provide a method for evaluating the performance of a fluorescence microscope. Tools such as these for assessing instrument performance are essential for measurement assurance of image-derived data, and will enable the reporting of data that are

meaningful, reproducible, and repeatable between laboratories.

## Materials and Methods

### *Fluorescence Reference Glass and Fluorescence Spectroscopy Analysis*

A description of the fabrication of the uranyl-ion-doped glass, specifications for the Schott 475 GG filter glass, and the fluorescence emission spectra of the glasses are included in the Supplemental Information.

### *Image Acquisition*

Images of the glass were collected using an automated fluorescence microscope (Axiovert 200M, Carl Zeiss, Jena, Germany)<sup>1</sup> equipped with a CoolSnap HQ camera (Photometrics, Tucson, AZ) with 12-bit analog to digital conversion. All images were acquired using a blue light emitting diode (LED) excitation source (470 nm, 25 mW; Thorlabs, Newton, NJ), a 470 nm centered excitation filter with a bandpass of 40 nm (full width at half maximum), a dichroic beam splitter centered at 495 nm, a 525 nm centered emission filter with a bandpass of 50 nm (full width at half maximum) (HE38 GFP filter set, Zeiss, Part # 489038-9901-000), and a 10x/0.3NA objective (Zeiss, Part # 420341-9911-000). The pixel to  $\mu\text{m}^2$  ratio,  $A$ , was determined by imaging a spatial calibration slide. To measure the emission intensity of the glass as a function of the excitation intensity, the power input to the LED was varied and the output power of the LED at the front of the objective was monitored with a photodiode (Newport, Irvine, CA; Model 1815-C). The reference material was then placed on the stage and the edge of the fluorescent glass, which provided a sharp line of contrast, was manually brought into focus. Images were then acquired with the glass filling the entire field of view with varying power inputs to the excitation LED. Photostability measurements were performed by continuously exposing the glass to the excitation light and acquiring time lapse images at 30 s intervals.

### *Automated Algorithm for Determining Benchmarked Parameters*

The exposure time that results in saturation of the camera response ( $t_{Sat}$ ), the detection threshold ( $t_{DT}$ ), and the intensity normalization factor ( $r_{glass\_REF}$ ) for calibrating intensities due to day to day fluctuations are the three parameters determined using the reference glass. The

<sup>1</sup> Certain commercial products are identified in order to adequately specify the experimental procedure; this does not imply endorsement or recommendation by NIST.

procedure for identifying  $t_{Sat}$  begins with the analysis of images acquired at a short exposure time, then incrementally increases the exposure time, and finally analyzes the newly acquired images. At each exposure time, two images are acquired to determine the mean intensity ( $I_t$ ) and the pixel variance ( $\sigma_t^2$ ).  $I_t$  is determined by averaging the mean pixel value from each of the two images,

$$I_t = \frac{\sum_{i=1}^{Np} (X1_i + X2_i)}{2 \times Np} \quad (1)$$

where  $I_t$  is the mean intensity corresponding to exposure time,  $t$ ,  $X1_i$  is the individual pixel value for the  $i$ th pixel of the first image,  $X2_i$  is the individual pixel value for the  $i$ th pixel of the second image and  $Np$  is the number of pixels in each image.  $\sigma_t^2$  is calculated by subtracting one image from the other and computing the variance in the pixel intensities, and then dividing by two as described by Equation 2(5),

$$\sigma_t^2 = \frac{\sum_{i=1}^{Np} (X1_i - X2_i)^2}{2 \times Np} \quad (2)$$

where  $\sigma_t^2$  is the image intensity variance corresponding to exposure time,  $t$ . This formula for  $\sigma_t^2$ , which is not the typical method for computing statistical variance, is insensitive to systematic variations between pixels and assumes that the mean for each pixel is constant for both images.

$t_{Sat}$  is mathematically defined as the exposure time that corresponds to the maximum pixel variance as a function of the exposure time,

$$t_{Sat} = \underset{t>0}{\operatorname{argmax}}(\sigma_t^2) \quad (3)$$

where  $\sigma_t^2$  is the pixel variance as a function of the exposure time,  $t$ . This definition for  $t_{Sat}$  is used because CCD and CMOS detectors are known to exhibit a sharp decrease in the image variance when the sensor reaches full well capacity(6). This provides a well-defined exposure time for benchmarking the  $t_{Sat}$  of the instrument.

The process for finding  $t_{Sat}$  is broken into two parts: a search for two bounding points and an optimized search for

the maximum variance. The goal of the search for two bounding points is to find an exposure time that is within the linear range of the camera and one that is outside the linear range. Thus the saturation must lie between these two points. In the search for the bounding points, the intensity,  $I_t$ , and variance,  $\sigma_t^2$ , are found at exposure times starting at 5 ms and doubling as long as the exposure time is still within the linear range. To determine whether the current exposure time is within the linear range of the system, linear regression is performed on the  $\sigma_t^2$  versus  $I_t$  for all previous exposure times. This regression line is used to predict the corresponding pixel variance for the current exposure time. If the predicted variance differs from the measured variance by more than 15%, the exposure time is considered outside of the linear range and the bounding points are set as the previous exposure and the current exposure time. A golden section search routine(7) is then used to determine subsequent exposure times. The golden section search routine efficiently finds the maximum of a strictly unimodal function by successively narrowing the range of values inside which the maximum is known to exist. This search routine is used to minimize the cumulative incident light on the reference material. The search for the maximum variance continues until the change in exposure is within 0.001 % of the square of the exposure time. Thus, the threshold for stopping the search for  $t_{Sat}$  increases for longer exposure times.

$t_{DT}$  is defined as the shortest possible exposure time when the SNR of fluorescence emitted from the glass is greater than or equal to 3 and expressed as,

$$\frac{\mu_{glass+bkg}(t) - \mu_{bkg}(t)}{\sqrt{\sigma_{glass+bkg}^2(t) + \sigma_{bkg}^2(t)}} \geq 3 \quad (4)$$

where  $\mu_{glass+bkg}(t)$  is the mean counts detected per  $10 \mu\text{m}^2$  of the fluorescent glass as a function of the exposure time,  $\mu_{bkg}(t)$  is the mean counts detected per  $10 \mu\text{m}^2$  from background as a function of the exposure time,  $\sigma_{glass+bkg}(t)$  is the variance of the counts detected per  $10 \mu\text{m}^2$  of the fluorescent glass as a function of the exposure time, and  $\sigma_{bkg}(t)$  is the variance of the counts detected per  $10 \mu\text{m}^2$  from background as a function of the exposure time. By scaling to intensity per area the parameters are, in principle, insensitive to changes in magnification or pixel size that can occur after inserting a magnifying/demagnifying optic into the light path, changing the binning on the camera, or using a different

camera all together. The area of  $10 \mu\text{m}^2$  was chosen because many image derived measurements, such as total intracellular fluorescence of a reporter molecule inside of cell, are made at this scale since mammalian cells are typically  $20 \mu\text{m}$  to  $30 \mu\text{m}$  in diameter (8). The formulation in Eqn. 4 is analogous to the ‘separation parameter’ used to assess the fluorescence detection performance of a flow cytometer by measuring the instrument’s ability to distinguish two populations of dimly fluorescent beads(9).  $\mu_{glass+bkg}(t)$  and  $\sigma_{glass+bkg}(t)$  are functions estimated from the analysis of images acquired by focusing on the surface of the glass proximal to the objective.  $\mu_{bkg}(t)$  and  $\sigma_{bkg}(t)$  are measured from images acquired after removing the glass from the field of view while leaving the excitation source on so that background fluorescence from sources, such as unintended emission from optical components or other sources of stray light, can be measured. The background intensity and variance is measured at exposure times previously determined to be within the linear range of detection for the reference glass.

The two terms in the numerator of Eqn. 4,  $\mu_{glass+bkg}(t)$  and  $\mu_{bkg}(t)$  are defined as,

$$\begin{aligned} \mu_{glass+bkg}(t) &= r_{glass+bkg} \times t \\ &\times A; \mu_{bkg}(t) \\ &= r_{bkg} \times t \times A \end{aligned} \quad (5,6)$$

where  $r_{glass+bkg}$  and  $r_{bkg}$  are the slopes of the regression lines calculated from the intensity versus exposure time data, and  $A$  is the number of pixels per  $10 \mu\text{m}^2$ . Intuitively,  $\mu_{glass+bkg}(t)$  and  $\mu_{bkg}(t)$  are functions used to estimate the mean number of counts measured for an exposure time,  $t$ , from each  $10 \mu\text{m}^2$  of the fluorescent glass and from background fluorescence measured after removing the fluorescent glass from the microscope stage.

The two terms in the denominator of Eq. 4,  $\sigma_{glass+bkg}^2(t)$  and  $\sigma_{bkg}^2(t)$ , are defined as,

$$\begin{aligned} \sigma_{glass+bkg}^2(t) &= \mu_{glass+bkg}(t) \times m \\ &+ \frac{b}{A}; \sigma_{bkg}^2(t) \\ &= \mu_{bkg}(t) \times m + \frac{b}{A} \end{aligned} \quad (7,8)$$

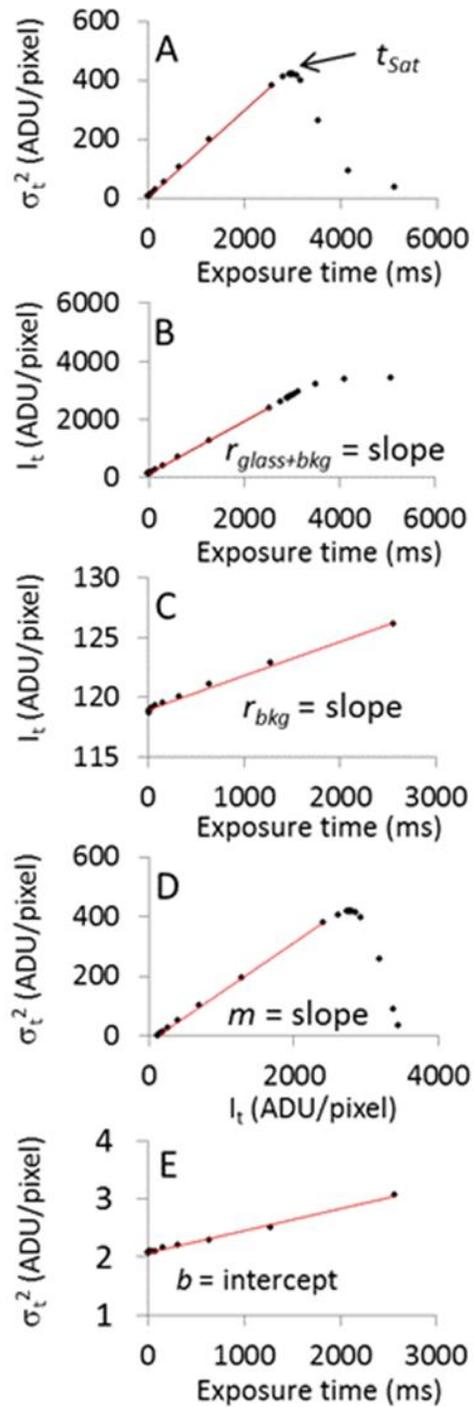
where  $\mu_{\text{glass+bkg}}(t)$  and  $\mu_{\text{bkg}}(t)$  are defined above and  $m$  is the slope of the regression line calculated from the  $\sigma_t^2$  versus  $I_t$  data and is used to estimate the noise present for a given signal level within the linear range of the detector.  $b$  is the signal independent noise associated with the read noise of the camera, which is estimated as the intercept of the  $\sigma_t^2$  versus  $t$ . The read noise originates from the electronic processing on the camera that converts charge carriers on the CCD or CMOS sensor into digital numbers or counts. Dark current, which contributes signal even in the absence of photons incident on the sensor, is not explicitly included in this formulation. In our experience the dark-current measured over exposure times needed to benchmark the instrument was negligible. If dark current were significant, it would contribute additional signal and additional noise. These additional contributions would be accounted for in the  $r_{\text{bkg}}$  measurement and are not expected to change the benchmarking procedure. Substituting the expressions for  $\mu_{\text{glass+bkg}}(t)$ ,  $\mu_{\text{bkg}}(t)$ ,  $\sigma_{\text{glass+bkg}}^2(t)$  and  $\sigma_{\text{bkg}}^2(t)$  in Eqns. 5-8 into Eqn. 4 and solving the resulting quadratic equation for exposure time provides an estimate of the time,  $t_{\text{DT}}$ , required to obtain a signal to noise ratio (SNR) equal to 3 from fluorescence emitted from the glass.

The above method for determining the benchmarking parameters assumes that  $I_t$  increases proportionally with  $t$  within the linear dynamic range, and that  $\sigma_t^2$  also increases proportionally with  $t$  within this range. The second assumption is generally true because the generation of photoelectrons by the CCD results in 'countable events' and the fluctuations in pixel intensities follow Poisson statistics(5). The image acquisition protocol, image analysis, linear regression operations, and all other computations reported in this study were implemented within MicroManager(4), an open source platform for microscope hardware control that is itself implemented as a plugin within ImageJ (<http://rsbweb.nih.gov/ij/>). The program script for obtaining the benchmarking parameters is contained in the Supplemental Information.

## Results

### *Determination of the Saturation Exposure Time, $t_{\text{Sat}}$ , and the Detection Threshold, $t_{\text{DT}}$*

The data shown in Figure 1 were collected and analyzed using the MicroManager script (Supplemental Information). The five plots shown in Figure 1 are used to derive two analytical performance parameters,  $t_{\text{Sat}}$  and  $t_{\text{DT}}$ . A plot of the image variance,  $\sigma_t^2$ , versus exposure time,  $t$ , is shown in Figure 1A.  $t_{\text{Sat}}$  is determined when  $\sigma_t^2$  versus exposure time becomes non-linear, indicating detector



**Figure 1** (A) Plot of the image variance,  $\sigma_t^2$ , reported as counts in ADUs per pixel from the Schott 775 GG filter glass versus the camera exposure time. (B) Plot of the measured intensity reported as counts in ADUs per pixel from the Schott 775 GG filter glass versus the camera exposure time. (C) Plot of the measured intensity of a blank sample (the fluorescent glass was removed from the microscope stage) versus the camera exposure time to obtain the background signal. (D) Plot of the image variance,  $\sigma_t^2$ , reported as counts in ADUs per pixel from the Schott 775 GG filter glass versus measured intensity reported as counts in ADUs per pixel. (E) Plot of the image variance,  $\sigma_t^2$ , reported as counts in ADUs per pixel from the background fluorescence versus the camera exposure time.

saturation(6). In the representative example shown,  $t_{Sat} = 2.97$  sec, the exposure time when  $\sigma_t^2$  is maximum.

Figure 1A indicates that the linearity ends at approximately 3 seconds whereas the  $I_t$  versus exposure time in Figure 1B appears to be linear for more than 3.5 seconds. This is typical of our observations on CCD detectors with our benchmarking procedure. Our interpretation of this is that as the full well capacity of the CCD is approached, charge is exchanged between neighboring pixels and the intensity values of neighboring pixels become correlated. Thus, the  $\sigma_t^2$  calculated according to Eqn. 2 is reduced but the  $I_t$  calculated according to Eqn. 1 is not. An additional benefit of using the  $\sigma_t^2$  versus exposure time plot to determine the saturation is that it provides a well-defined maximum.

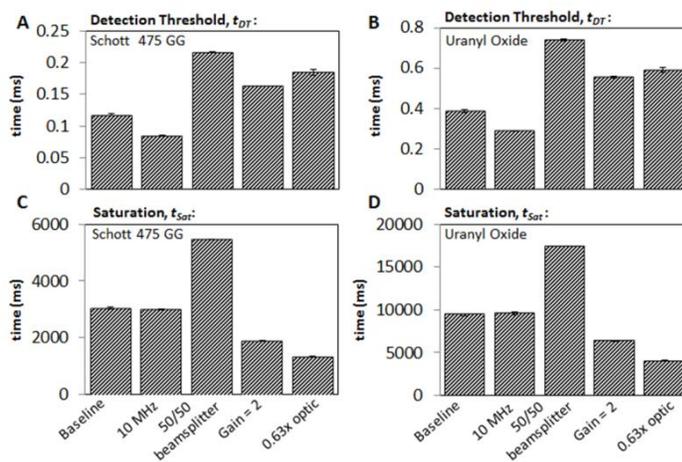
For the determination of  $t_{DT}$ , each of the terms in Eqn. 4 is estimated using the data shown in Figures 1B-1E.  $r_{glass+bkg}$  and  $r_{bkg}$  are determined from linear regression on the  $I_t$  versus exposure time data plotted in Figures 1B and 1C and  $\mu_{glass+bkg}(t)$  and  $\mu_{bkg}(t)$  can be determined using Eqns. 5 and 6 with knowledge of the number of pixels per  $10 \mu m^2$ . The ratio of  $\sigma_t^2$  to  $I_t$ ,  $m$ , is the slope of the regression line calculated from the data in Figure 1D. The signal independent noise,  $b$ , is determined from the intercept of the  $\sigma_t^2$  versus exposure time data from the background intensities plotted in Figure 1E. With  $m$  and  $b$ ,  $\sigma_{glass+bkg}^2(t)$  and  $\sigma_{bkg}^2(t)$  can be determined using Eqns. 7 and 8.

For the data shown in Figure 1, the analysis resulted in an estimated  $t_{DT} = 1.17 \times 10^{-4}$  sec. This estimate is a function of both the efficiency of the imaging system to convert emitted photons into useful signal and the noise floor of the detector. Limitations on the exposure time that may arise from mechanical or electronic camera shutters or the frame read time are not accounted for in the estimate for  $t_{DT}$ .  $t_{Sat}$  and  $t_{DT}$  can be used to establish the linear dynamic range of an imaging system. For the example shown in Figure 1, the system has a linear dynamic range =  $t_{Sat} / t_{DT} \approx 25,500 \approx 10^{4.4}$  (based on the mean signal detected over  $10 \mu m^2$  area of the fluorescent glass).

### Monitoring the Saturation Exposure Time, $t_{Sat}$ , and Detection Threshold, $t_{DT}$ , in Response to Variations in the Fluorescence Imaging System

We made several systematic differences in the microscope configuration to examine the effect on  $t_{Sat}$  and  $t_{DT}$ . Figure 2 contains plots of  $t_{Sat}$  and  $t_{DT}$  in response to differences in the read-out rate of the CCD, the gain setting of the CCD, and the presence or absence of a beam

splitter or a reducing optic in the light path. The benchmarking strategy provided a sensitive method for assessing the performance of an imaging system in response to these changes. For example, when the read-out rate of the detector was changed from 20 MHz to 10 MHz,  $t_{DT}$  was reduced by a factor of  $\sqrt{2}$  due to the reduced read noise at the 10 MHz read-out rate. No significant change was observed in  $t_{Sat}$ , as expected. The addition of the 50/50 beam splitter into the collection path approximately doubled both  $t_{DT}$  and  $t_{Sat}$ , as expected. The 50/50 beam splitter transmittance was measured as 52.7% using the transmitted light source of the microscope with the emission filter in the optical train, inserting/removing the beam splitter, and measuring the reported counts on the CCD detector. Increasing the gain setting from 1 to 2 resulted in an increase of  $t_{DT}$  by approximately 30 %, likely because of additional signal amplification noise at the detector. And  $t_{Sat}$  decreased, likely because the maximum bit depth was exceeded before the full well capacity of the CCD was reached. With a 0.63x demagnifying optic inserted into the light path, a longer time was required for  $t_{DT}$ , likely because of the photon loss due to the additional optic.  $t_{Sat}$  was shorter because the demagnification optic focuses the projected beam and increases the photon flux on the detector.



**Figure 3** (A and B) The detection threshold,  $t_{DT}$ , measured from the Schott 475 GG glass and the uranyl-ion-doped glass, respectively, to changes in the microscope configurations indicated on the x-axis. (C and D) The saturation,  $t_{Sat}$ , measured from the Schott 475 GG glass and the uranyl-ion-doped glass, respectively. Error bars represent the 95% confidence interval for n=3 independent measurements.

At the outset, it was not clear whether either material was sufficiently bright and homogeneous to serve as an effective benchmarking material. The response of  $t_{DT}$  and  $t_{Sat}$  to the systemic perturbations was significant and highly similar between the Schott 475 GG and uranyl-ion-doped glass reference materials, suggesting that both can be used

indicate changes in the performance of the fluorescence imaging system. For the repeat measurements shown in Figure 2, the glass was removed from the stage and the procedure was repeated from the beginning. In our experience, both glasses were sufficiently homogeneous in fluorescence that placement on the stage was not critical.

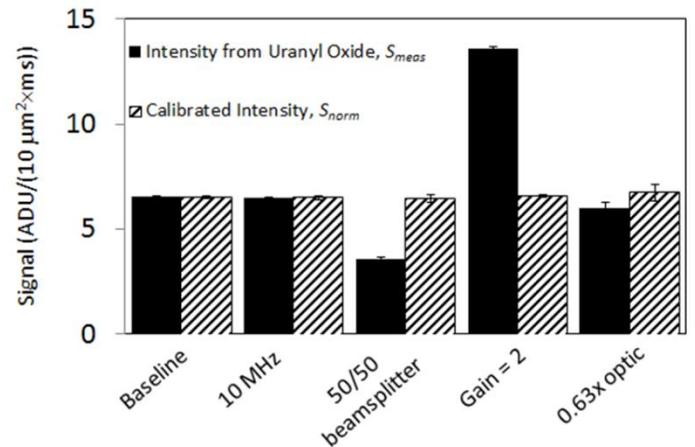
#### Using the Reference Glass to Normalize Signal Intensities

For day to day comparisons the intensities measured from the reference glass can be used to normalize intensity data using the following mathematical procedure,

$$S_{norm} = \frac{r_{glass} \times A}{r_{glass\_REF} \times A_{REF}} \times S_{meas} \quad (9)$$

where  $r_{glass} = r_{glass+bkg} - r_{bkg}$ , which is the counts  $/(10 \mu m^2 \times s)$  detected from the reference glass measured under the experimental conditions.  $r_{glass\_REF}$  is the counts  $/(10 \mu m^2 \times s)$  detected from the reference glass measured at the same time and under the same conditions as the reference signal that you want to normalize to.  $S_{meas}$  and  $S_{norm}$  are the measured and normalized signal from the experiment, respectively, and  $A$  and  $A_{REF}$  are the number of pixels per  $10 \mu m^2$  measured for the experimental and reference conditions.

To demonstrate the calibration of intensities, the Schott 475 GG glass was used as a reference material to normalize intensities measured from the uranyl-ion-doped glass. This analysis was performed on the same microscope after changing the read-out rate and gain setting of the CCD, and the in the presence or absence of a beam splitter or a reducing optic in the light path. For the data shown in Figure 3,  $r_{glass\_REF}$  was measured from Schott 475 GG glass with the microscope at baseline conditions (see Figure S2) and  $r_{glass}$  was measured from Schott 475 GG glass after changing the configuration of the imaging system as indicated on the x-axis of Figure 3.  $S_{meas}$  is the mean rate at which counts were detected for  $10 \mu m^2$  of the uranyl-ion-doped glass, which can be considered a surrogate for a fluorophore used for cellular labeling, such as green fluorescent protein or fluorescein. Changing the read rate of the CCD from 20 MHz at baseline to 10 MHz resulted in no significant change in  $S_{meas}$  and the normalization also resulted in comparable intensities within 0.4%. Placing a 50/50 beamsplitter in the optical path reduced  $S_{meas}$  to 54% of the baseline intensity and increasing the gain from 1x to 2x increased  $S_{meas}$  to 208% of the baseline intensity. After normalization,  $S_{norm}$  was comparable to the baseline conditions to within 1.2% in both cases. When a 0.63x



**Figure 3.** The slope of the regression line of intensity versus exposure time ( $r_{glass}$ ) measured from uranyl-ion-doped glass (solid black bars) is shown for each for the imaging conditions. The same data are shown after normalizing with intensity versus exposure time data measured from Schott 475 GG glass (hatched bars). Error bars represent the 95% confidence intervals (n=3).

reducing lens was placed in the optical path,  $S_{meas}$  was reduced to 91.5% of the baseline intensity, likely caused by the reduction of the light throughput of the system with the additional optic. After normalization,  $S_{norm}$  was comparable to within 3.5%. This example shows that the reference materials can be used to normalize intensities from day to day when the emission spectra of the measured fluorescence emitter and the reference material are different.

#### Using the Reference Glass to Monitor Day to Day Variability and Deriving Instrument Specifications

The benchmarking procedure was executed in the morning and in the evening for three consecutive days to monitor the performance of a system and begin to generate acceptable operating specifications for the instrument. In this way, the glass reference materials were used as check standards to characterize the stability of fluorescence measurements on the microscope using control charts(10). In Figure 4, data for monitoring the performance of the imaging system using the benchmarking procedure is plotted. For each run, the mean performance metric +/- the 95% confidence interval is plotted (n=3). The horizontal lines indicate the mean value (solid line) and +/-  $3\sigma$  values (dashed line) computed for all of the triplicate measurements over the 6 runs (n=18). The  $3\sigma$ /mean values obtained using the Schott 475 GG glass for  $t_{Sat}$ ,  $t_{DT}$  and  $r_{glass}$  were 5.5%, 4.9%, and 4.5%, respectively, and, the  $3\sigma$ /mean values obtained using the uranyl-ion-doped glass for  $t_{Sat}$ ,  $t_{DT}$  and  $r_{glass}$  were 13.6%, 12.7%, and 11.6%, respectively. The larger variability in the uranyl-ion-doped glass measurements is likely due to this piece being

manually cut and polished (Figure S1C). To confirm this possibility, images were acquired from regions across the entire uranyl-ion-doped glass piece and the entire Schott 475 GG filter glass piece. Defects in the uranyl-ion-doped glass were observed as well as systematic differences in the fluorescent intensity (data not shown). To minimize the effect of the non-uniformities observed in the uranyl-ion-doped glass, images were acquired from approximately the same location over the three days. Schott 475 GG filter glass did not exhibit defects or systematic differences in fluorescence intensity across the test piece. Ideally, the uranyl-ion-doped glass would be manufactured under tightly controlled conditions and exhibit uniformity similar to the Schott 475 GG filter glass.

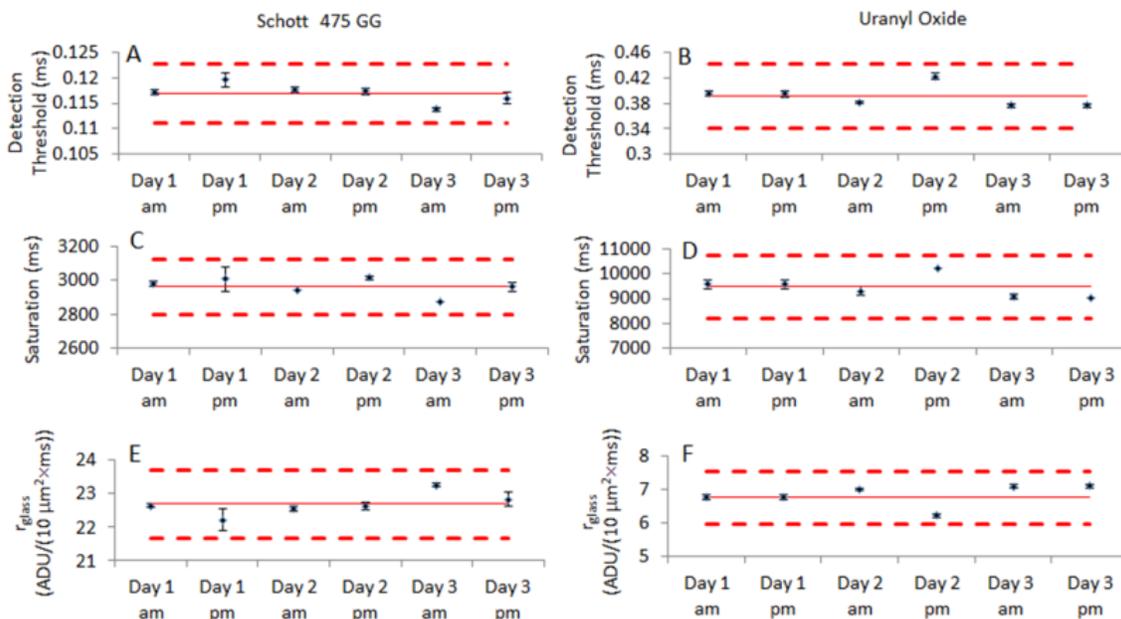
The specifications derived from the analysis in Figure 4 could be used as acceptance criteria for the performance of an instrument to ensure that an imaging system is functioning as expected. When the benchmarked parameter deviates from the mean by a fraction greater than indicated by the  $3\sigma$ /mean value shown in Figure 4, it suggests that the imaging system is operating outside of a normal range.

When either of the reference materials is used to monitor the performance of an imaging system, it is critical that the fluorescence is stable. To directly examine the photostability in the context of microscope benchmarking,

we performed the automated procedure on exactly the same area nine times in a row. No photobleaching could be detected for either glass (Supplementary Information). We also examined the stability of the fluorescence during storage. Over a ten month period during which the glass was stored at room temperature on the benchtop or in the dark, no change in the fluorescence intensity from either glass could be detected (Supplementary Information).

## Discussion

Benchmarking the analytical performance of a microscope is critical for assuring that image data is comparable and that derived measurements are reproducible and repeatable. We describe and evaluate a procedure using a photostable and homogeneously fluorescent material to benchmark the detection threshold and saturation of a widefield fluorescence microscope. From the detection threshold and saturation, the linear dynamic range of the instrument can also be derived. These three properties should be reported to support claims of intensity differences, particularly in experiments where cellular fluorescence varies over a wide range. The procedure has been designed to be convenient to use and sensitive to changes in the performance of the fluorescence microscope. In the experiments performed here, the automated procedure typically required less than five



**Figure 4.** Three parameters used to benchmark the performance of a fluorescence imaging system were measured at six different times: the detection threshold,  $t_{DT}$ , for (A) Schott 475 GG and (B) uranyl-ion-doped glass; the saturation exposure time,  $t_{Sat}$ , for (C) Schott 475 GG and (D) uranyl-ion-doped glass; and the rate at which counts were detected for each  $10 \mu\text{m}^2$  ( $r_{\text{glass}}$ ) of (E) Schott 475 GG glass and (F) uranyl-ion-doped glass. Error bars represent the 95% confidence intervals ( $n=3$ ). The means (solid red line) and  $\pm 3$  standard deviations (dashed red line) computed over the aggregate data sets are also shown ( $n=18$ ).

minutes to complete from start to finish.

In this study, comparisons were made on the same instrument using the same illumination source and the same objective lens. Comparisons between different imaging systems cannot be made using this approach when the spectral properties of the illumination source, the optical path, and the detector response are different. This is because the incident light flux on the fluorescent material and its spectral characteristics will affect the fluorescence emission. The detected intensity will also depend on the relative spectral intensity of the emitted light and the spectral characteristics of the collection optics, filters, and camera. To apply this approach for intensity calibration or performance comparisons between instruments with different spectral properties, several additional measurements would need to be made. The excitation irradiance ( $\text{W}/\text{m}^2$ ) at the focal plane would need to be measured and normalized along with the spectral properties of the excitation source, all optical filters and components in the optical train, and the detector. These are challenging calibrations that are an active area of research in the fluorescence spectroscopy field(11,12).

Materials that are photostable and homogeneously fluorescent are an essential component of the benchmarking procedure. The two materials described here, uranyl-ion-doped glass and Schott 475 GG filter glass, are extremely photostable, especially compared to organic materials (2,13). The emission spectra of the material are an important consideration, and ideally, materials covering the entire useful spectral range would be available. The materials described here cover an emission wavelength region from approximately 500 nm to 600 nm, which is in the range of several popular fluorophores including green fluorescent protein (GFP), Alexa 488, and fluorescein. For practical purposes, the materials should also be easy to manufacture, easy to use, and durable when handled. Glasses, phosphors, and polymers are all likely candidate materials. Fluorescent glass reference materials were the focus of this study because they have all of the aforementioned properties and they have been used successfully as standards for spectral correction in Raman and fluorescence spectroscopies (14,15). As a consequence of being homogeneously fluorescent, these materials can in principle also provide for accurate flatfield correction. Though, the procedure described here does not require flatfield correction to evaluate the system performance and field non-uniformity is not expected to have a significant influence on  $t_{DT}$  and  $r_{glass}$ , but is expected to change  $t_{Sat}$  because brighter parts of the field will saturate faster than dimmer parts (Supplemental Information).

Our analysis is not the first to use homogeneously fluorescent materials, such as slides, to characterize properties of a fluorescence microscope (13,16-18), though to our knowledge, a protocol has not been described for using fluorescent slides for performance benchmarking. In addition to a slide format, fluorescent beads and standard light sources have been considered as reference materials for use on a fluorescence microscope. Though, both of these have drawbacks as reference materials. Two disadvantages of a bead format are that they require image segmentation when they are used on a microscope and they do not immediately provide flatfield information. Nonetheless, practical methods have been described for using fluorescent beads to evaluate the performance of a fluorescence microscope (19,20). LED light sources have also been used to calibrate intensities and evaluate the light collection performance of microscopes (21-23). Two limitations of the LED sources are that they are not as simple to use or manufacture as the glass slide and they do not provide a method to monitor potential changes in the excitation source or excitation optics. By implementing this automated procedure in combination with a photostable and reusable fluorescent slide, we hope to enable laboratories to routinely evaluate the performance of their instruments.

While this study is focused on widefield imaging at low magnification, it appears that our method could be used to evaluate the performance of cameras and imaging systems for single molecule localization and detection. Two recent studies evaluated EMCCDs and sCMOS cameras used for these applications (24,25). The study by Saurabh et al. directly compared the localization accuracy between different cameras using dimly fluorescent beads. It would not be possible for our method to achieve this, but our automated method is likely to provide similar data for evaluating the low light detection sensitivity of a camera because the models used for the analysis of camera performance in these two studies and ours are fundamentally similar. In order to apply our approach to low light detection relevant to single molecule studies, reference materials that are more dimly fluorescent will likely be needed.

Using reference materials to benchmark instrument performance is critical for measurement assurance. While benchmarking has not been applied rigorously to fluorescence microscopy in the past, we hope that these materials and method will facilitate the transition of fluorescence microscopy to a more quantitative measurement tool and result in the reporting of data that are more repeatable from day to day and from lab to lab.

## Acknowledgements

The authors would like to thank Heather Moosher and Bryan Goering for supporting the development of the Micromanager script to automate the benchmarking procedure and Marco vanGemeren at the Mineral Optics Laboratory for making the 170  $\mu\text{m}$  thick glass slides.

## References

1. Stack RF, Bayles CJ, Girard AM, Martin K, Opansky C, Schulz K, Cole RW. Quality assurance testing for modern optical imaging systems. *Microsc Microanal* 2011;17:598-606.

2. Derosé PC, Smith MV, Mielenz KD, Blackburn DH, Kramer GW. Characterization of standard reference material 2941, uranyl-ion-doped glass, spectral correction standard for fluorescence. *Journal of Luminescence* 2008;128:257-266.

3. Resch-Genger U, Hoffmann K, Nietfeld W, Engel A, Neukammer J, Nitschke R, Ebert B, Macdonald R. How to improve quality assurance in fluorometry: Fluorescence-inherent sources of error and suited fluorescence standards. *Journal of Fluorescence* 2005;15:337-362.

4. Edelstein A, Amodaj N, Hoover K, Vale R, Stuurman N. Computer control of microscopes using microManager. *Curr Protoc Mol Biol* 2010;14:20.1-20.17.

5. Sluder G, Wolf DE. Digital microscopy. 3rd edn. Amsterdam: Elsevier Academic Press. 2007.

6. Janesick JR, Klaasen KP, Elliott T. Charge-Coupled-Device Charge-Collection Efficiency and the Photon-Transfer Technique. *Optical Engineering* 1987;26:972-980.

7. Kiefer J. Sequential Minimax Search for A Maximum. *Proceedings of the American Mathematical Society* 1953;4:502-506.

8. Halter M, Elliott JT, Hubbard JB, Tona A, Plant AL. Cell volume distributions reveal cell growth rates and division times. *J Theor Biol* 2009;257:124-130.

9. Wood JC. Fundamental flow cytometer properties governing sensitivity and resolution. *Cytometry* 1998;33:260-266.

10. *NIST/SEMATECH e-Handbook of Statistical Methods*,

<http://www.itl.nist.gov/div898/handbook/pmc/section3/pm c31.htm>. 2012.

11. Resch-Genger U, Bremser W, Pfeifer D et al. State-of-the art comparability of corrected emission spectra. 1. Spectral correction with physical transfer standards and spectral fluorescence standards by expert laboratories. *Anal Chem* 2012;84:3889-3898.

12. Resch-Genger U, Bremser W, Pfeifer D et al. State-of-the art comparability of corrected emission spectra. 2. Field laboratory assessment of calibration performance using spectral fluorescence standards. *Anal Chem* 2012;84:3899-3907.

13. Zwier JM, Van Rooij GJ, Hofstraat JW, Brakenhoff GJ. Image calibration in fluorescence microscopy. *J Microsc* 2004;216:15-24.

14. Choquette SJ, Etz ES, Hurst WS, Blackburn DH, Leigh SD. Relative intensity correction of Raman spectrometers: NIST SRMs 2241 through 2243 for 785 nm, 532 nm, and 488 nm/514.5 nm excitation. *Appl Spectrosc* 2007;61:117-129.

15. DeRose PC, Early EA, Kramer GW. Qualification of a fluorescence spectrometer for measuring true fluorescence spectra. *Rev Sci Instrum* 2007;78:033107.

16. Model MA, Burkhardt JK. A standard for calibration and shading correction of a fluorescence microscope. *Cytometry* 2001;44:309-316.

17. Kaplan DS, Picciolo GL. Characterization of instrumentation and calibrators for quantitative microfluorometry for immunofluorescence tests. *J Clin Microbiol* 1989;27:442-447.

18. Turney SG, Culican SM, Lichtman JW. A quantitative fluorescence-imaging technique for studying acetylcholine receptor turnover at neuromuscular junctions in living animals. *J Neurosci Methods* 1996;64:199-208.

19. Lockett SJ, Jacobson K, Herman B. Quantitative precision of an automated, fluorescence-based image cytometer. *Anal Quant Cytol Histol* 1992;14:187-202.

20. Murray JM. Evaluating the performance of fluorescence microscopes. *J Microsc* 1998;191:128-134.

21. Cho EH, Lockett SJ. Calibration and standardization of the emission light path of confocal microscopes. *J Microsc* 2006;223:15-25.

22. Young IT. Absolute fluorescence calibration. *Proc of SPIE* 2006;6859:151-158.

23. Young IT. Photonic calibration for fluorescence microscopy. Proc of SPIE 2008;6088:0U1-0U9.

24. Huang ZL, Zhu H, Long F, Ma H, Qin L, Liu Y, Ding J, Zhang Z, Luo Q, Zeng S. Localization-based super-resolution microscopy with an sCMOS camera. Opt Express 2011;19:19156-19168.

25. Long F, Zeng S, Huang ZL. Localization-based super-resolution microscopy with an sCMOS camera part II: experimental methodology for comparing sCMOS with EMCCD cameras. Opt Express 2012;20:17741-17759.

## Supplemental Information

### CONTENTS

1. *Fabrication and Specifications for the Uranyl-ion-doped Glass and the Schott 475 GG Glass (pg. 2)*
2. *Characterization of Fluorescence Emission Intensity, Spectra and Photostability of the Fluorescent Glasses (pg. 3)*
3. *Characterization of the Loss of Fluorescence Intensity while Performing the Benchmarking Protocol and During Storage at Room Temperature (pg. 6)*
4. *Considerations on the Effect of a Non-uniform Field on the Benchmarked Parameters (pg. 8)*

### 1. *Fabrication and Specifications for the Uranyl-ion-doped Glass and the Schott 475 GG Glass*

The uranyl-ion-doped glass was melted at 1300 °C using a base glass composition with mass fractions of B<sub>2</sub>O<sub>3</sub> (67.0%), Li<sub>2</sub>O (10.4%), MgO (9.9%), SiO<sub>2</sub> (4.15%), and a dopant mass fraction of U<sub>3</sub>O<sub>8</sub> (0.25%). The melt was cast into a pre-heated aluminum cuvette-form mold 1.25 cm X 1.25 cm X 4.5 cm and then quickly transferred to an annealing oven where the block was equilibrated and held at 625 °C. The annealing oven was then turned off, and the glass was allowed to cool to room temperature. The glass cuvette was buried in sand in an oven for further annealing. The oven was ramped from room temperature to 625 °C at 75 °C/hr, held for an additional 8 hrs, and then ramped down from 625 °C to 380 °C at 4 °C/hr. The temperature was held at 380 °C for an hour, then the oven was turned off, and it was allowed to cool to room temperature over approximately 6 hrs.

The Schott 475 GG glass was purchased as a 1 inch diameter round filter (Edmund Optics, Barrington, NJ; Part # 46-075; Lot # 035480). Both glasses were cut, mounted to microscope slides with optical glue (Crystalbond™, Aremco, Valley Cottage, NY), and polished to approximately 170 μm thick (Mineral Optics Laboratory, Wilder, VT). During imaging, the slide was placed on the microscope with the fluorescent glass proximal to the objective lens.

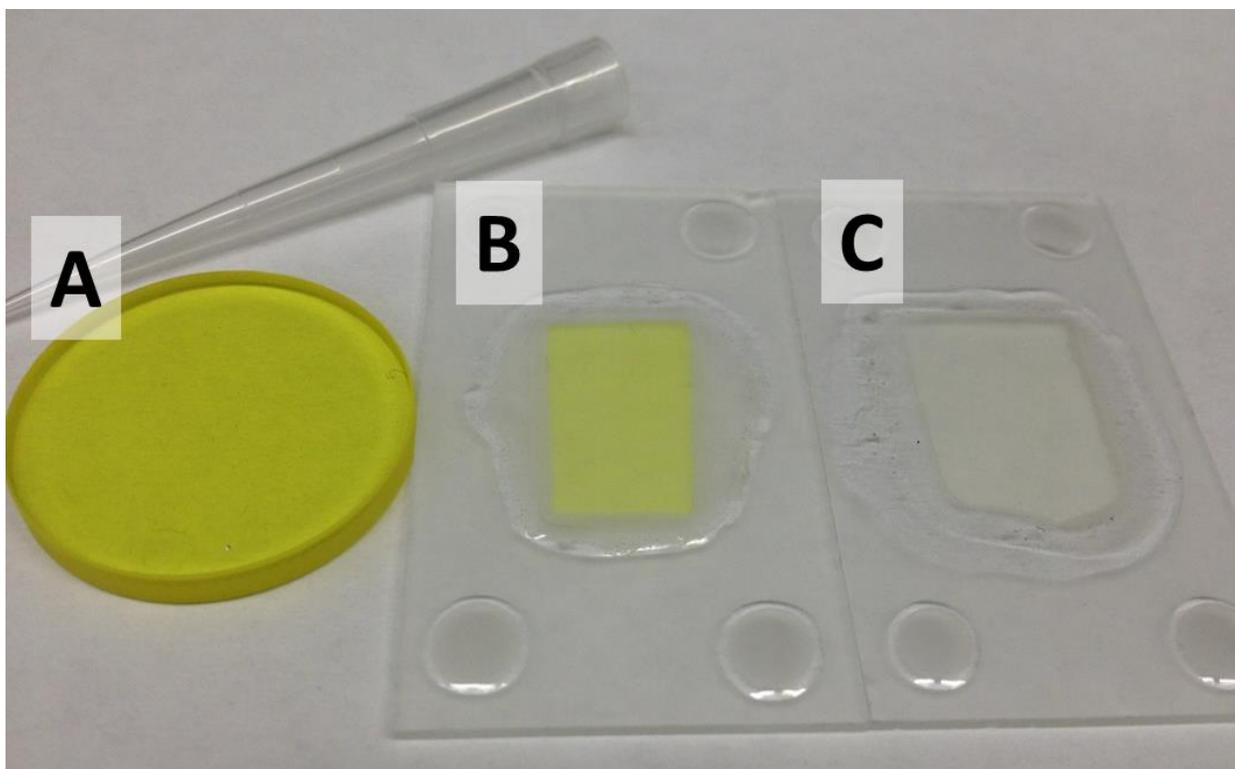
## 2. Characterization of Fluorescence Emission Intensity, Spectra and Photostability of the Fluorescent Glasses

All fluorescence spectra were taken on a SPEX Fluorolog 3 (Jobin Yvon, Edison, NJ) spectrofluorometer using a continuous 450 W Xe lamp as an excitation source, double monochromators for excitation and emission wavelength selection, and a photomultiplier tube for detection of emission. A “reference” photodiode just before the sample was used to measure fluctuations of the relative excitation intensity in time and wavelength, and correct the emission signal for these fluctuations. The wavelength accuracy was determined to be  $\pm 0.2$  nm for both emission and excitation using atomic lamps. The spectra were corrected for the relative spectral responsivity of the emission detection system to give an instrument independent spectral shape. This correction was done using a calibrated light source with a calibrated diffuse reflector traceable to the NIST realization of the International System of Units (SI)(1-4). All fluorescence measurements were taken at 24 °C using an L-shaped optical geometry with the emission being detected at 90° relative to the excitation beam. The sample was positioned at an angle of approximately 35° relative to the excitation beam using a front face geometry, i.e., the emission was collected from the same face of the sample to which the excitation beam was incident.

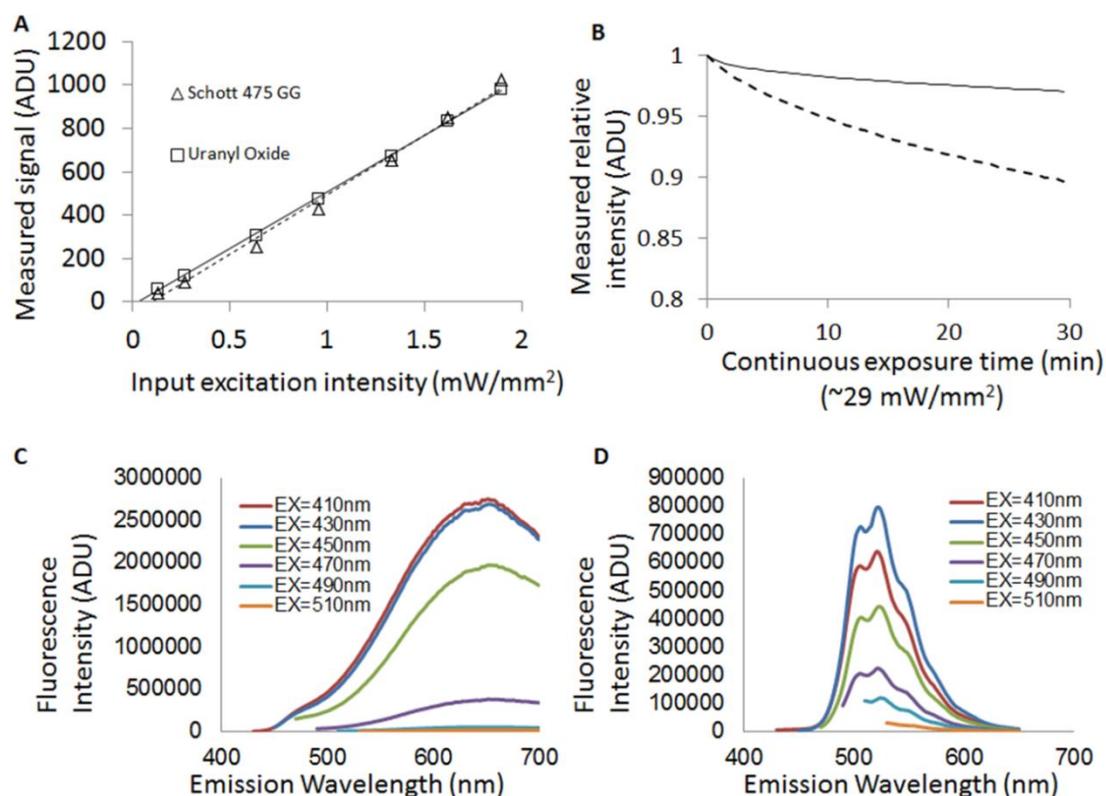
The scanning range for emission spectra was from 430 nm to 700 nm using excitation and emission bandwidths of 5.5 nm. Emission spectra were taken at six excitation wavelengths from 410 nm to 510 nm using a 20 nm increment. A more detailed description of the qualification of the fluorescence spectrometer, related uncertainties, and the determination of spectral correction factors is given elsewhere(4).

Three properties of the glasses were examined to assess their suitability as reference materials: (1) the stability of the emitted fluorescence intensity during continuous excitation, (2) the sensitivity of the fluorescence emission intensity to changes in excitation intensity, and (3) the sensitivity of the fluorescence emission spectra to changes in the excitation wavelength. For these measurements, approximately 170  $\mu\text{m}$  thick glass slides of each material shown in Figure S1 were analyzed. Both Schott 475 GG and the uranyl-ion-doped glass provided an approximately linear response in emission intensity using a 525 nm bandpass filter with increasing excitation intensities centered at  $\approx 470$  nm (Figure S2A). The excitation intensities examined were in the range of incident fluxes typical for widefield fluorescent microscopes in our laboratory, and the linear response indicates that the measured emitted intensity could be used to indicate and normalize for changes in illumination intensity. The photostability of the emitted fluorescence intensity during continuous excitation is plotted in Figure S2B. A small loss in intensity, approximately 10% for Schott 475 GG and 3% for uranyl-ion-doped glass, can be observed over 30 minutes of continuous exposure to a high intensity LED focused with a 10x/0.3 NA objective (29 W/mm<sup>2</sup>, total illumination area  $\approx 0.6$  mm<sup>2</sup>). For the benchmarking data reported in this study, the total exposure time required was always less than 1 minute indicating that photobleaching during routine use of the glass is negligible.

The fluorescence emission spectra are plotted for the Schott 475 GG and the uranyl-ion-doped glass in Figures 2C and 2D, respectively. For each glass, spectra were collected using six excitation wavelength regions with maxima ranging from 410 nm to 510 nm. The emission intensity is strongly dependent on the excitation wavelength for both glasses, especially for the Schott 475 GG. For example, the uranyl-ion-doped glass emission at 525 nm is reduced by approximately 2-fold when the excitation wavelength changes from 450 nm to 470 nm. The same change in the excitation wavelength causes a 6-fold reduction in the emission intensity at 525 nm for the Schott 475 GG glass. The strong dependence of the fluorescence emission on the excitation wavelength suggests that the benchmarked parameters will be highly sensitive to the spectral characteristics of the excitation source and the excitation filters.



**Figure S1.** (A) Schott 475 GG glass as purchased as a 1 inch diameter round filter. (B) Cut and polished piece of the Schott 475 GG filter glass mounted to a microscope slide. (C) Cut and polished piece of the uranyl-ion-doped glass mounted to a microscope slide. Both B and C were approximately 170  $\mu\text{m}$  thick.



**Figure S2.** (A) Plot of the excitation intensity versus the measured signal (mean counts, in analog to digital units (ADUs), per pixel). Solid lines are regression lines computed for the Schott 475 GG ( $R^2 = 0.992$ ) and the uranyl-ion-doped glass ( $R^2 = 0.9991$ ) (B) Plot of the measured relative intensity for the Schott 475 GG filter glass (dashed line) and uranyl-ion-doped glass (solid line) versus the time during continuous exposure time to the excitation source. (C and D) Emission spectra for the Schott 475 GG filter glass and uranyl-ion-doped glass, respectively. Emission spectra are plotted for excitation wavelengths 410 nm, 430 nm, 450 nm, 470 nm, 490 nm, and 510 nm as indicated by the colors shown in the legend.

#### References

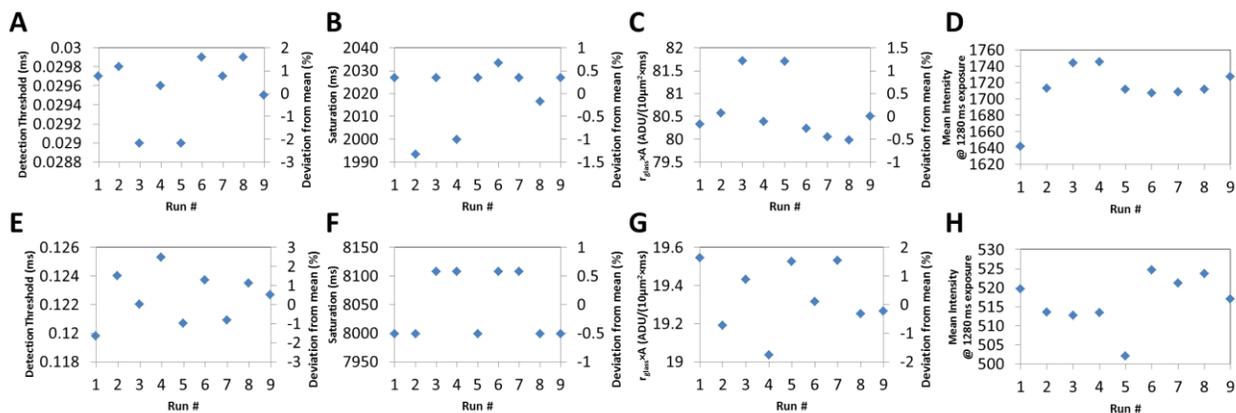
1. Barnes PY, Early EA, Parr AC. Spec. Publ. 250-48: Spectral reflectance. Gaithersburg, MD: National Institute of Standards and Technology. 1998.
2. Walker JH, Saunders AT, Hattenburg AT. Spec. Publ. 250-1: Spectral radiance calibrations. Gaithersburg, MD: National Bureau of Standards. 1987.
3. Yoon HW, Gibson CE. Spec. Publ. 250-89: Spectral Irradiance Calibrations. Gaithersburg, MD: National Institute of Standards and Technology. 2011.
4. DeRose PC, Early EA, Kramer GW. Qualification of a fluorescence spectrometer for measuring true fluorescence spectra. Rev Sci Instrum 2007;78:033107.

### 3. Characterization of the Loss of Fluorescence Intensity while Performing the Benchmarking Protocol and During Storage at Room Temperature

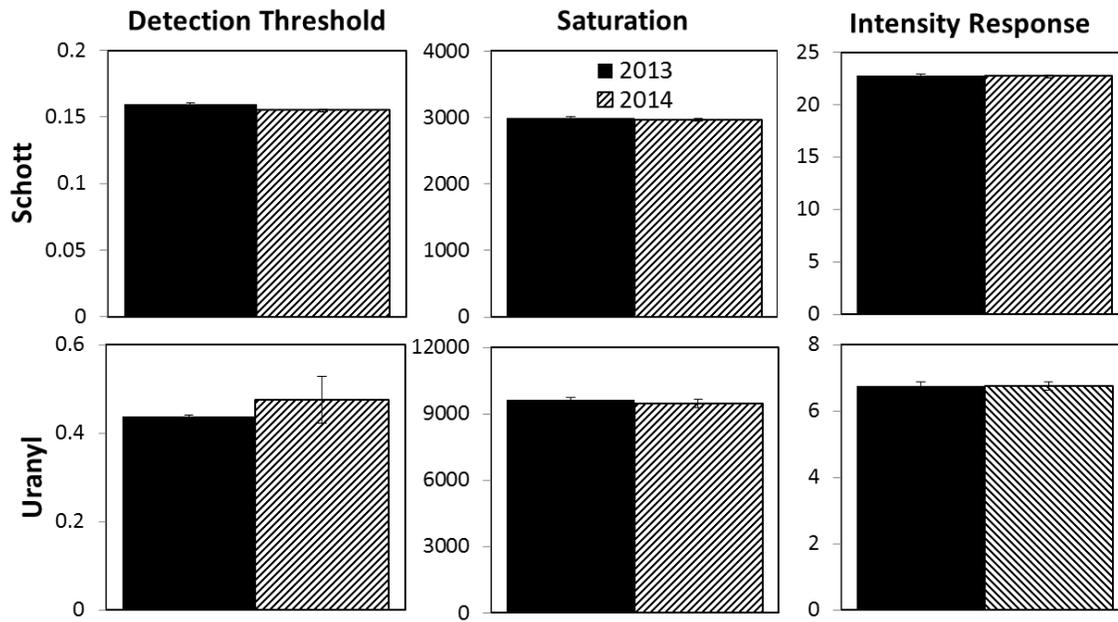
To examine the photostability of the Schott 475 GG filter glass and uranyl-ion-doped glass in the context of the benchmarking procedure, we ran the automated procedure on exactly the same area nine times in a row. During the experiment, the benchmarking protocol was performed as described in the Methods with the exception that during the background intensity measurements a shutter was placed in front of the camera and the glass was left in place on the stage. In Figure S3, the detection threshold ( $t_{DT}$ ), saturation exposure time ( $t_{SAT}$ ), the rate at which counts were detected from each  $10 \mu\text{m}^2$  of the glass ( $r_{glass}$ ), and the mean intensity across the field using a 1280 ms exposure time are plotted. No evidence for photobleaching could be detected for either glass during the nine runs for either glass.

To evaluate the photostability of the Schott 475 GG filter glass and uranyl-ion-doped glass during storage, we compared the fluorescence intensity of each glass over a span of 11 months (August 2013 to June 2014). During this time the two glasses were stored at room temperature in a drawer for most of the time (approximately 10 months) or were on a laboratory bench top where they were occasionally exposed to fluorescent room lights. The approach we took was to first benchmark the microscope with Schott glass, then to characterize the fluorescence of the uranyl-ion-doped glass.

For the measurements taken June 2014, the LED intensity on the microscope (see Methods for description of microscope) was adjusted until the parameters were approximately the same for the Schott 475 GG glass as they were in August 2013. The benchmarking program was then run using the uranyl-ion-doped glass without any additional changes to the LED source. The parameters (detection threshold, saturation, and intensity calibration ( $r_{glass}$ )) from the uranyl-ion-doped glass taken in 2014 were highly similar to the parameters from 2013. The data we collected is shown in Figure S4. The similarity of the benchmarked parameters between August 2013 and June 2014 strongly suggests that both glasses were highly stable when stored at room temperature in a drawer. Although an alternative explanation is that both glasses experienced identical degradation during the 11 months. We consider this unlikely since the two glasses are chemically quite different from one another.



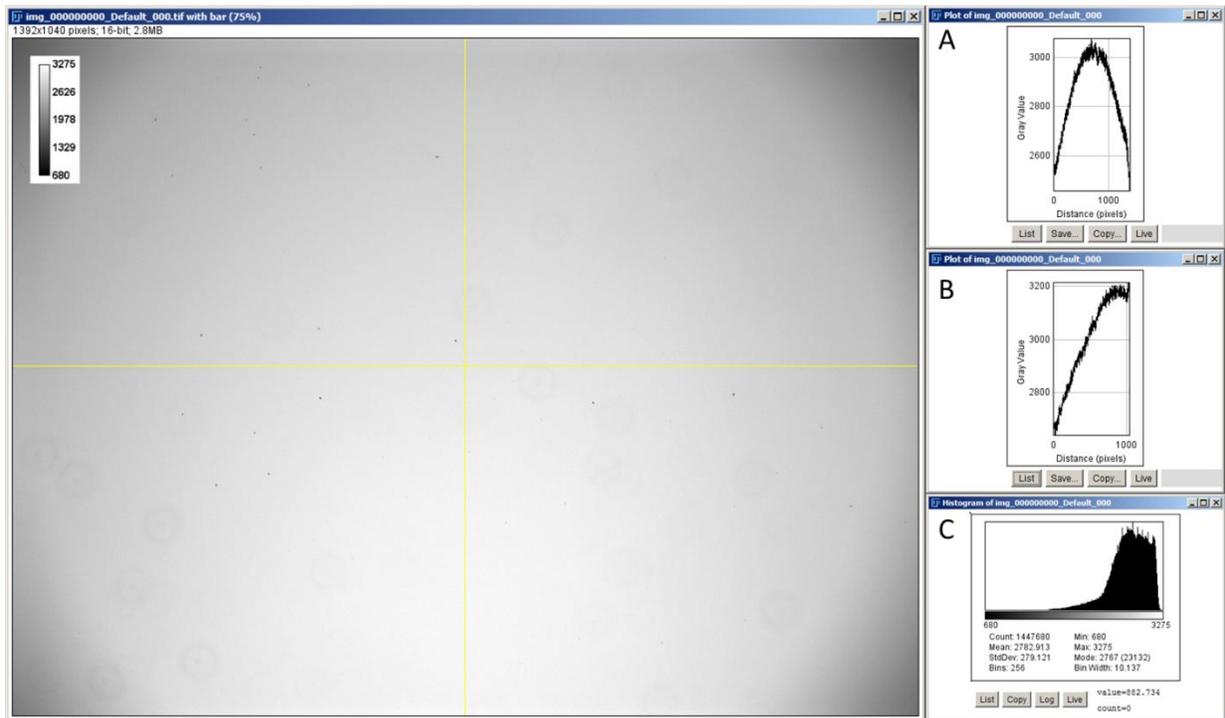
**Figure S3.** The detection threshold ( $t_{DT}$ ), saturation exposure time ( $t_{SAT}$ ), the rate at which counts were detected from each  $10 \mu\text{m}^2$  of the glass ( $r_{glass}$ ), and the mean intensity across the field using a 1280 ms exposure time are plotted for nine consecutive runs of the performance procedure for the Schott 475 GG filter glass (A-D) and the uranyl-ion-doped glass (E-H).



**Figure S4.** The average values of the benchmarked parameters were calculated from control charting performed over three days with six runs on each day in both 2013 and 2014 ( $n = 18$  for each bar). Error bars represent the 95% confidence intervals computed over the 18 runs of the performance benchmarking protocol. Intensity response (far right) is the rate at which counts were detected from each  $10 \mu\text{m}^2$  of the glass ( $r_{\text{glass}}$ ).

#### 4. Considerations on the Effect of a Non-uniform Field on the Benchmarked Parameters

The intensity across the field is not uniform on most widefield fluorescence microscopes. An example of field non-uniformity is shown in Figure S5 and is typical of the field non-uniformity of the images used in this study. When the field is non-uniform, the performance parameters measured by the procedure described in this paper will be spatially dependent. As an illustration of this effect, consider an image where the field is brighter in the center than at the edges. Less time will be required to reach the detection threshold for the same object in the center compared with the edges of the field. Likewise, the saturation exposure time will be shorter in the center compared to the edges of the field. To facilitate the benchmarking of a microscope with spatially dependent intensity across the field, the MicroManager script has been written to use “regions-of-interest” so that specific parts of the microscope field can be selected and benchmarked independently.



**Figure S5.** Image of the Schott 475 GG Schott glass using an exposure time of 2970 ms ( $=t_{Sat}$ ). Image dimensions are 0.873 mm x 0.652 mm (1392 pixels x 1040 pixels). A greyscale intensity calibration bar is shown in the upper left of the image. The panels on the right show (A) a horizontal line scan of the greyscale values corresponding to the horizontal line in the image, (B) vertical line scan of the greyscale values corresponding to the vertical line in the image, and (C) histogram of greyscale values in the image.

The benchmarked parameters characterize the average response over all of the pixels in the field or the selected region-of-interest. The saturation exposure time,  $t_{Sat}$ , is expected to depend on the degree of non-uniformity of the field because brighter parts of the field will saturate faster than dimmer parts of the field. As soon as any part of the field begins to saturate, the linearity of the  $\sigma_t^2$  versus exposure time is expected to diminish. However, the detection threshold,  $t_{DT}$ , and the factor used for intensity normalization,  $r_{glass}$ , are not expected to depend on how intensities are distributed across the field or region-of-interest. The reason for this follows.

$t_{DT}$  and  $r_{glass}$  only depend on  $I_t$ ,  $\sigma_t^2$ , and  $A$  (see Eqns. 4-8, and Figure 1) which we show below are all independent of the distribution of intensities across the field, and therefore  $t_{DT}$  and  $r_{glass}$  are, in principle, independent of the distribution of intensities across the field.

$I_t$  is the average intensity (Eqn. 1) and within the linear dynamic range of the camera, the average intensity will be independent of the distribution of intensities across the field.

In the derivation below,  $\sigma_t^2$  is also shown to be independent of the distribution intensities.

To begin, the expected value for  $\sigma_t^2$  is defined as (same as Eqn. 2):

$$E[\sigma_t^2] = \frac{E[\sum_{i=1}^{Np} (X1_i - X2_i)^2]}{2 \times Np} \quad (S1)$$

where  $E$  denotes the expected value. The linearity of expectation rule states that the expectation value of a sum of random variables is the sum of the expected value of the variables. This can be used to rewrite Eqn. S1 as,

$$E[\sigma_t^2] = \frac{\sum_{i=1}^{Np} E[(X1_i - X2_i)^2]}{2 \times Np} \quad (S2).$$

For reference, we write below the computation for the statistical variance of a random variable as the sum of the means of the squares minus the sum of the squares of the means,

$$\sum_{i=1}^{Np} Var((X1_i - X2_i)) = \sum_{i=1}^{Np} (E[(X1_i - X2_i)^2] - E[X1_i - X2_i]^2) \quad (S3).$$

Using Eqn. S3, Eqn. S2 can be written as

$$E[\sigma_t^2] = \frac{\sum_{i=1}^{Np} (Var((X1_i - X2_i)) + E[X1_i - X2_i]^2)}{2 \times Np} \quad (S4)$$

which can be rewritten as

$$E[\sigma_t^2] = \frac{\sum_{i=1}^{Np} (Var(X1_i) + Var(X2_i) - 2Cov(X1_i, X2_i) + [E(X1_i) - E(X2_i)]^2)}{2 \times Np} \quad (S5)$$

by using the formula for the variance of the difference between two random variables and applying the linearity of expectation rule. With the assumption that the pixels on the CCD are independent, the covariance term in the numerator goes to 0 and because photobleaching is negligible during the acquisition of the image pairs,  $[E(X1_i) - E(X2_i)]^2$  in the numerator goes to zero and Eqn. S5 simplifies to

$$E[\sigma_t^2] = \frac{\sum_{i=1}^{Np} (Var(X1_i) + Var(X2_i))}{2 \times Np} \quad (S6).$$

Now because  $\sum_{i=1}^{Np} Var(X1_i) = \sum_{i=1}^{Np} Var(X2_i)$ , Eqn. S6 simplifies to

$$E[\sigma_t^2] = \frac{E[\sum_{i=1}^{Np} (X1_i - X2_i)^2]}{2 \times Np} = \frac{2 \sum_{i=1}^{Np} Var(X1_i)}{2 \times Np} = \frac{\sum_{i=1}^{Np} Var(X1_i)}{Np} = \bar{\sigma}^2 \quad (S7)$$

where  $\bar{\sigma}^2$  is the average of the pixel variances. Since  $\sigma_t^2$  is only a function of the average variance, the distribution of intensities across the field will not influence the image variance that is computed during the benchmarking procedure.

Finally,  $A$ , the number of pixels per  $10 \mu\text{m}^2$ , is only a calibration factor. Thus,  $t_{DT}$ , and the factor used for intensity normalization,  $r_{glass}$ , are not expected to depend on how intensities are distributed across the field or region-of-interest.

Intuitively, the above can be summarized by imagining a simple  $3 \times 3$  CCD detector and comparing the results for the average intensity and the average variance that one would obtain in two cases:

- A) The intensity is uniformly distributed across the field with an average intensity of 1.
  - a. Average intensity = 1, average variance = a constant,  $c \times 1$ .
- B) All of the intensity is confined to the center pixel with intensity 9.
  - a. Average intensity = 1, average variance = a constant,  $(c \times 9) / 9$

In both cases, the average per pixel intensity and the average variance are the same.