

# Using Deep Learning to Develop an Annotated Dataset of Smoke Plumes in Satellite Imagery

Rey Koki,<sup>a, b, c</sup> Micheal McCabe,<sup>a</sup> Jebb Q. Stewart,<sup>b</sup> Christina Kumler,<sup>b, c</sup> Jed Brown,<sup>a</sup> Dhruv Kedar,<sup>d</sup> Wilfrid Schroeder<sup>e</sup>

<sup>a</sup> *Computer Science Department, University of Colorado, Boulder, Colorado*

<sup>b</sup> *NOAA's Global Systems Laboratory, Boulder, Colorado*

<sup>c</sup> *CIRES, University of Colorado, Boulder, Colorado*

<sup>d</sup> *JILA and Department of Physics, University of Colorado, Boulder, Colorado*

<sup>e</sup> *NOAA NESDIS, College Park, Maryland*

<sup>10</sup> Corresponding author: Rey Koki, rey.koki@noaa.gov

11 ABSTRACT: The increase in the frequency of wildfires on a global scale underscores the need for  
12 advancements in monitoring techniques for effective health related mitigation, disaster manage-  
13 ment and environmental protection. This research introduces an innovative, data-driven framework  
14 for creating annotated geostationary satellite imagery, specifically focused the detection of smoke  
15 plumes by leveraging the data generated by NOAA's (National Oceanic and Atmospheric Admin-  
16 istration) Hazard Mapping Systems (HMS). The primary objective is to refine the existing HMS  
17 dataset, that provides temporal and geographical information on individual smoke plumes, and  
18 then use deep learning to pinpoint the singular, most representative, GOES image that optimally  
19 illustrates the smoke annotation within the given time window. By identifying the most represen-  
20 tative imagery of smoke plumes within the given time window, the study seeks to create a highly  
21 precise and relevant training dataset. The resulting dataset is anticipated to be an instrumental tool  
22 in developing further machine learning models, such as an automated system capable of real-time  
23 monitoring and annotation of smoke plumes directly from satellite imagery.

<sup>24</sup> **1. Introduction**

<sup>25</sup> In recent years, the escalation of wildfire incidents worldwide has become a prominent environmental and public health concern. The combustion process in wildfires releases vast amounts of smoke containing fine particulate matter (PM2.5) and harmful gases, posing severe risks to human health and air quality. This scenario underscores the necessity for efficient and effective monitoring methods to mitigate the adverse health impacts associated with wildfire smoke.

<sup>30</sup> Traditionally, wildfire monitoring has relied on ground-based methods, such as forest service patrols, manned lookout towers, and aviation surveillance. While these methods provide valuable local insights, they are constrained by geographical and logistical limitations, often failing to deliver timely and comprehensive data, especially over large and remote areas. In contrast, satellite imagery offers a vantage point that overcomes these limitations, providing continuous, wide-area coverage and real-time data crucial for assessing and responding to the health risks posed by wildfire smoke.

<sup>37</sup> Satellite imagery, equipped with advanced sensors, such as the Advanced Baseline Imager on the Geostationary Operational Environmental Satellites (GOES), have revolutionized environmental monitoring. These tools enable the detailed observation of smoke plumes, their particulate density, and the extent of smoke spread. These satellite-based systems offer the capabilities to provide critical insights into the concentration and movement of smoke particulates, facilitating accurate and timely assessments of air quality.

<sup>43</sup> The integration of satellite imagery in wildfire smoke monitoring is not only instrumental in providing real-time data but also plays a significant role in public health planning and response. By mapping the spread and density of smoke, health authorities can issue timely warnings, implement evacuation protocols, and deploy resources effectively to mitigate health risks. Furthermore, long-term data gathered from satellite observations can aid in understanding the broader impacts of wildfire smoke on public health, influencing policy decisions and preventive measures.

<sup>49</sup> Currently, multi-channel thresholding is a popular method to distinguish smoke pixels from pixels containing dust, clouds or other phenomenon with similar signatures. The method uses historical, labeled data to extract optimal radiance values for each channel that corresponds with the labeled class. These methods are tuned to particular biogeographies and often have issues with generalization to new locations with varying fuel types (Randriambelo et al. 1998).

54 In contrast to the numerical thresholding approach, human visual inspection of satellite imagery  
55 is another commonly used method for smoke identification. Trained analyst will inspect imagery  
56 and label the smoke by hand. This method is not as extendable as one that is automated and is  
57 limited by the availability of analyst and their time.

58 To address these challenges we can look towards innovative approaches and technological ad-  
59 vancements in computer vision. Machine learning methods have shown promising potential in  
60 improving the accuracy and efficiency of satellite-based wildfire smoke detection and monitoring.  
61 For instance, SmokeNet, uses convolutional neural network (CNN) based framework to determine  
62 if a scene of MODIS imagery contains smoke (Ba et al. 2019). Another study also used a CNN  
63 to identify smoke on a pixel-wise basis using imagery from Himiwi-8 (Larsen et al. 2021).  
64 Additionally, Wen et al. developed a CNN architecture that takes GOES-EAST imagery as input  
65 and HMS annotations for the target labels during training(Wen and Burke 2021).

66 The success of deep learning methods, such as CNNs, relies heavily on the availability of a large,  
67 representative dataset (Sun et al. 2017). Existing methods use relatively small amounts of data,  
68 from 57 (Wang et al. 2022) to 6825 (Wen and Burke 2021). In contrast, benchmark datasets for  
69 image classification contain tens of thousands (CIFAR-10 and MNIST) to millions (CIFAR-100 and  
70 ImageNet) of data samples. Keeping in mind the correlation between both the quality and quantity  
71 of data with model performance, we introduce the largest known smoke dataset containing over  
72 100,000 independent samples.

73 TABLE 1. Comparison of different studies including method used, dataset size, satellite source, number of  
74 channels used and if the detection is done at a pixel or image level.

Reference	Method	Number of Images	Satellite	# Channels	Detection Level
(Wen and Burke 2021)	CNN		GOES-EAST	6825	pixel
(Ba et al. 2019)	CNN	6255	MODIS	5	image
Larsen et al. (2021)	CNN	975	Himiwi-8	7	pixel
(Wang et al. 2022)	U-Net	47	Landsat-8	13	pixel
SmokeViz (2023)	U-Net	60,000	GOES-EAST, GOES-WEST	3	pixel

75 **2. Methods**

76 *Dataset*

77 Dataset development came in three stages. First, we developed a rough dataset of satellite  
78 imagery by accounting for light scattering to make an educated guess on which image would best  
79 represent the analyst annotation. Second, we used that dataset to train a model that would identify  
80 smoke in satellite imagery. Third, we use that model to determine the optimal satellite image to  
81 represent each annotation.

82 **SMOKE LABELS**

83 The National Environmental Satellite, Data and Information Service (NESDIS) and National  
84 Oceanic and Atmospheric Administration (NOAA) manage environmental satellite programs such  
85 as the Hazard Mapping System (HMS) (McNamara et al. 2004; Schroeder et al. 2008). The HMS  
86 program is an operational system that uses an aggregation of satellite data to generate active fire  
87 and smoke data that is used in applications such as air quality assessments and serves as verification  
88 and validation for NOAA's smoke forecasting model, HYSPLiT, (Rolph et al. 2009). To train our  
89 model, we implement a supervised learning framework that uses the HMS analyst smoke product  
90 as truth labels during the model training process.

91 HMS smoke analysis data gives the coordinates of the smoke perimeter and classifies the smoke  
92 by density within a given time window. The time windows can range from instantaneous (same  
93 start/end time) to lengths of 5 hours. While the bounds of the smoke annotations can change  
94 within the larger time spans, the analyst is making an approximation that should reflect the smoke  
95 coverage over the duration of the window. The density information is qualitatively determined by  
96 the analyst based on smoke opacity and categorized as either light, medium or heavy as seen in  
97 figure 1a.

103 **THERMOMETER ENCODING SMOKE DENSITIES**

104 One of the challenges introduced with using human generated qualitative smoke densities was  
105 that, as seen in figure 1b and 1c, there are variations in what is labeled as heavy or light density  
106 smoke. More generally, reproducing qualitative metrics with quantitative algorithms is a chal-  
107 lenging problem, but we apply mathematical approaches that mitigate some of the underlying

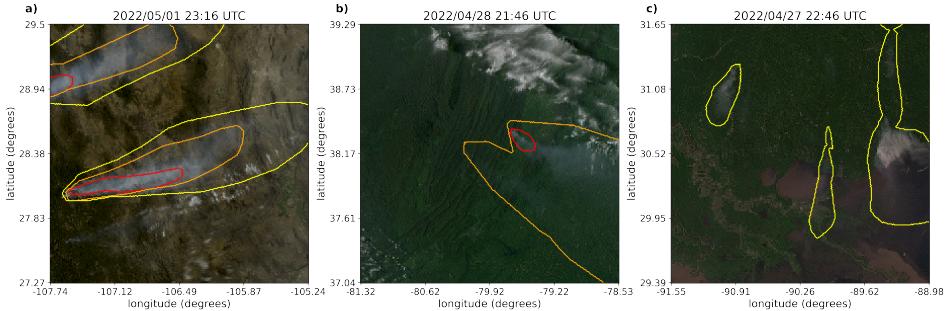


FIG. 1. Satellite imagery captured by GOES-EAST within a few days of each other. The yellow, orange and red contours indicate the extent of Light, Medium and Heavy smoke. a) shows a canonical example of a smoke plume. b) and c) show variations in the qualitatively determined density labels. b) we show Medium and Heavy densities of smoke that, upon visual inspection, could be interpreted as less dense than portions of the Light density smoke labeled in c).

complications of our specific problem. Despite the fact that the smoke densities introduce qualitative complexities, we decided that the density approximations were important to use in our dataset because of the differences in signatures the densities produce. Within the satellite imagery, the appearance of a light density smoke plume will look significantly different than a heavy density smoke plume as seen in figure 1. Additionally, a light density smoke plume is expected to be more challenging to detect since it is easier for it to be misclassified as not smoke. During the training process, the separate density categories allows us to deferentially weight the penalization given to the model for incorrect classifications based on category. For example, the model can be given a small penalization for misclassifying light smoke as not smoke while given a higher penalization for misclassifying heavy smoke as not smoke.

In addition to the densities being ordered and categorical, the differences in opacity is undefined by any metric such as particulates per square meter, so we can treat the density labels as ordinal instead of just categorical data. This allows us to use thermometer encoding, which leverages the idea that heavy density smoke includes both medium and light density smoke, that heavy density smoke is closer to medium than it is to light and automatically weights the loss functions and incorporates the ranked ordering of the densities. As seen in Table 2, one-hot encoding, commonly used for categorical data, doesn't take ordinal properties of the data into consideration.

TABLE 2. A comparison of one-hot encoding used for categorical data to thermometer encoding for ordinal data.

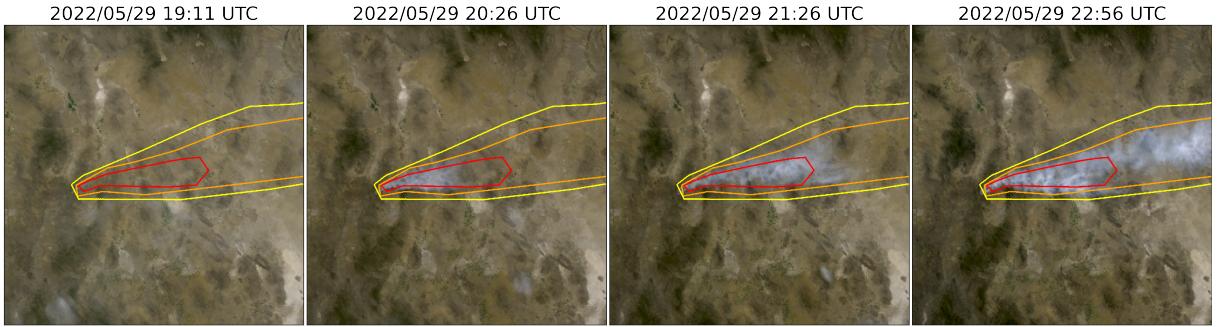
category	one-hot	thermometer
No Smoke	[0 0 0]	[0 0 0]
Light	[0 0 1]	[0 0 1]
Medium	[0 1 0]	[0 1 1]
Heavy	[1 0 0]	[1 1 1]

## 125 TIME WINDOWS FOR SMOKE ANNOTATIONS

126 In order to take into account movement characteristics to help identify smoke, analysts use  
 127 multi-frame animations of the satellite imagery. The resulting annotations often have large time  
 128 windows over multiple hours to represent one smoke plume. Since their goal is to show the general  
 129 coverage over that time span, often the smoke boundaries don't match up with the satellite imagery  
 130 over the entire time window 2. One way to approach this problem would be to use all the satellite  
 131 images the analysts used as input. Since the timespans are non-uniform, this would vary the length  
 132 in imagery inputs into the model, which would be difficult with a CNN architecture. Moreover,  
 133 this would require a large amount of additional memory and computational resources. Instead  
 134 of using the original analysts' many satellite image inputs to one annotated output, we decide to  
 135 develop a one-to-one input-to-output by finding the best singular satellite image input to represent  
 136 the annotation. As discussed in the next section, we do this by making physics based choices on  
 137 which satellite and timestamp would give the optimal angle between the sun and satellite that would  
 138 produce the strongest smoke signature for the geolocation and timestamp of the smoke plume.

## 142 SATELLITE IMAGERY

143 The Geostationary Operational Environmental Satellites (GOES) are operated by the NOAA and  
 144 NESDIS support meteorology research and forecasting for the United States. We use the latest  
 145 operational satellites, GOES-16 (EAST), 17 and 18 (WEST) that carry the Advanced Baseline  
 146 Imager (ABI), that measure 16 bands between the visible and infrared wavelengths. In improvement  
 147 to the GOES predecessors, imagery is collected every 5 minutes for the contiguous United States  
 148 and every 10 minutes for the full disk. We use bands 1-3 (Table 3) as input to Satpy's composite  
 149 algorithm to develop a true color image representation, similar to what is used as input by HMS  
 150 analysts (Raspaud et al. 2018) and (Bah et al. 2018).

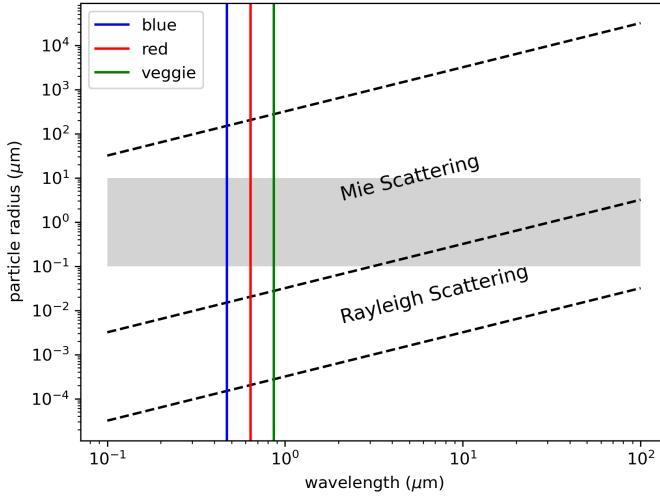


139 FIG. 2. True Color GOES imagery from May 2022, Southeast New Mexico ( $31^{\circ}$  Latitude,  $-100^{\circ}$  Longitude)  
 140 during the start of the Foster Fire. The HMS annotations for the smoke outlines shown here spanned from  
 141 19:10-23:00 UTC but visually match the smoke location for the last part of the

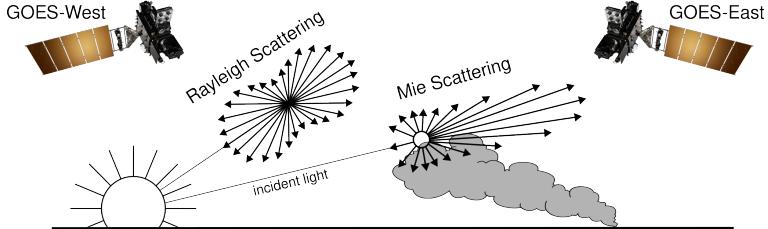
151 TABLE 3. To create a true color image, we use the following bands from the Advanced Baseline Imager Level  
 152 1b CONUS (ABI-L1b-RadC) product.

band	description	center wavelength	spatial resolution (km)
C01	blue visible	0.47	1
C02	red visible	0.64	0.5
C03	veggie near infrared	0.865	1

153 We used a physics-informed approach in selecting the initial dataset for training our model.  
 154 Rather than use the cumulative data from GOES-WEST and GOES-EAST images, we select one or  
 155 the other based on the solar zenith angle. For smoke identification, such an approach can achieve a  
 156 much higher signal-to-noise than imaging the earth's surface from an arbitrary angle. The elastic  
 157 scattering of light is key to this approach - while the atmosphere is composed of molecules with  
 158 size  $< 1$  nm, smoke particles can vary from  $100$  nm -  $10 \mu\text{m}$  in diameter,  $d$ . The GOES ABI  
 159 covers spectral bands from  $0.47 \mu\text{m}$  -  $13.3 \mu\text{m}$ , so atmospheric and smoke particle sizes occupy  
 160 two very different regimes with respect to the imaging wavelength  $\lambda$ , as shown in figure 3. In  
 161 the extreme limit of  $\lambda \gg d$ , the physics of scattering of light off a small sphere is captured by  
 162 Rayleigh scattering. This process has two critical consequences: (1) the scattering cross section of  
 163 light is strongly wavelength dependent (scaling with  $\lambda^{-4}$ ), meaning that photons with wavelength  
 164 closer to the ultraviolet are scattered more strongly than infrared photons. (2) the scattering cross



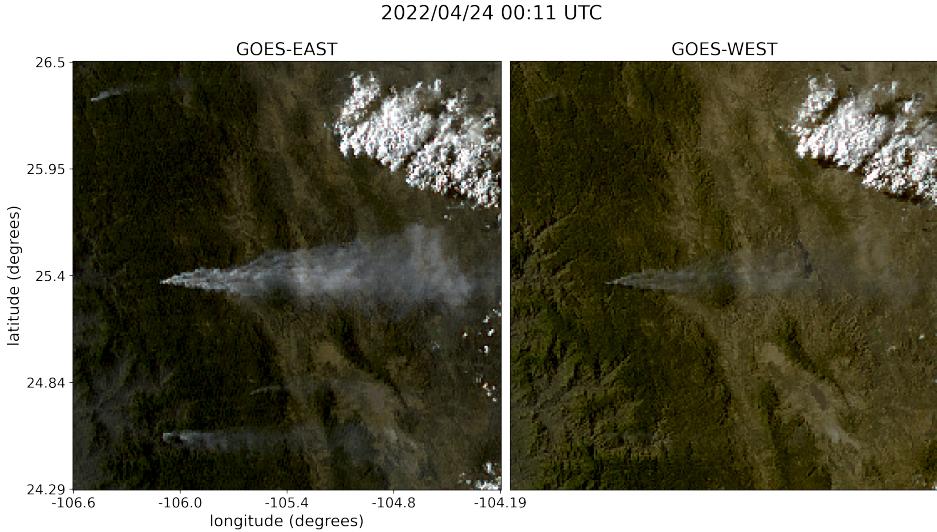
169 FIG. 3. Relationship between the size of a particle, the wavelength of light interacting with the particle and  
 170 the type of scattering behavior induced by that interaction. The dotted lines represent rough estimates of the  
 171 boundaries between the scattering regimes (Petty 2006). The gray area represents the range of particle radius  
 172 relevant to smoke particulate matter.



173 FIG. 4. If the particle size is  $< \frac{1}{10}$  the wavelength of the interacting light, then the primary scattering will  
 174 be Rayleigh. Mie scattering is the predominant scattering mechanism when the particle size is larger than  
 175 wavelength of light.

176 section scales with an angular dependent cross section of  $(1 + \cos^2 \theta)$ . Scattered photons follow the  
 177 emission distribution of a radiating dipole, scattering more strongly in the forward and backwards  
 178 directions ( $\theta = 0, \pi$ ) than orthogonal to the direction of propagation ( $\theta = \pi/2, 3\pi/2$ ), see figure 4  
 for Rayleigh scattering schematic.

179 The significance of these scalings is that the observer, or detector, will receive blue photons in  
 180 most directions orthogonal to the source. Equivalently, photons traveling colinear with line of sight  
 181 to the emission source will mostly have wavelengths in the infrared band. In the converse regime

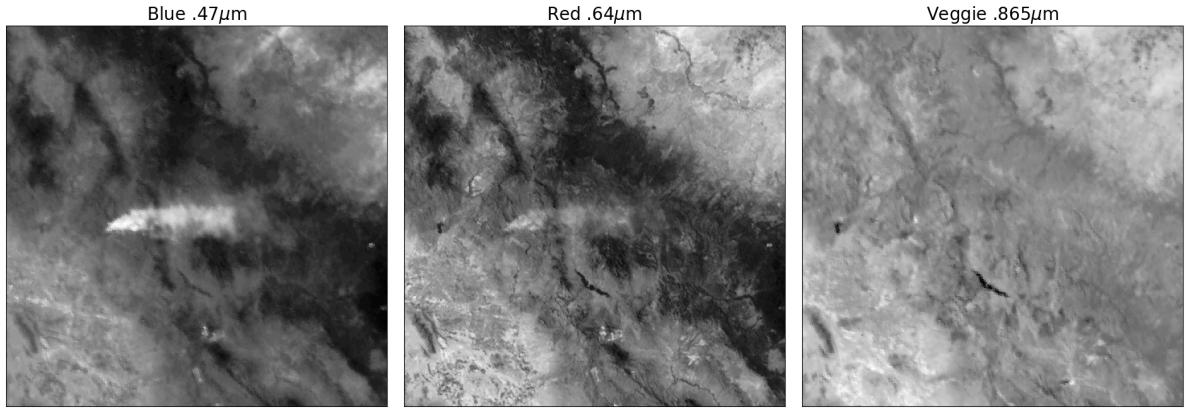


186 FIG. 5. True Color GOES-EAST (left) and GOES-WEST (right) imagery from April 24<sup>th</sup>, 2022. The images  
 187 were taken about 1.5 hours before sunset for this geolocation and time of year (01:43 UTC).

188 of  $d \gtrsim \lambda$ , the elastic scattering of light against matter is modeled through Mie scattering. Unlike  
 189 Rayleigh scattering, Mie scattering is largely wavelength independent and has a more complicated  
 190 radiation pattern where the cross section has a maximal amplitude in the forward direction. An  
 191 observer downstream of this scatterer will collect more photons than one positioned directly behind  
 192 it. In the context of smoke identification, a sunrise or sunset will lead to a higher Mie scattered  
 193 signal in GOES-WEST and GOES-EAST respectively, as shown with a smoke plume producing a  
 194 stronger signal in GOES-EAST imagery near sunset in figure 2.

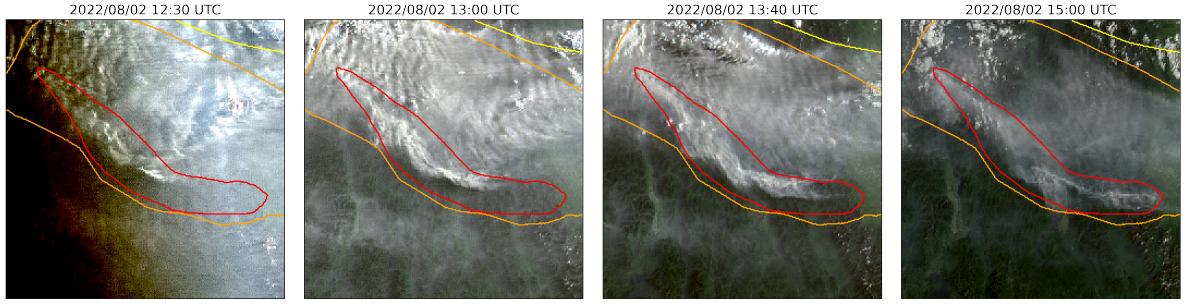
195 Smoke identification therefore amounts to extracting a signal of  $d \gtrsim \lambda$  photons from the  $\lambda \gg d$   
 196 background. Positioning a detector along line of sight to the scatterer will result in a higher signal  
 197 from smoke particles (figure 4). Filtering the imaged wavelength can enhance this signal; photons  
 198 collected in the blue spectrum will have a naturally lower background along the line of sight to the  
 199 illumination source due to their high level of Rayleigh scattering as. Therefore, as demonstrated in  
 200 figure 6, this configuration results in the highest signal to noise imaging for smoke particles.

201 After these realizations, we concluded that the optimal times were to grab data from GOES-  
 202 WEST right after sunrise and from GOES-EAST right before sunset. Another consideration we  
 203 needed to account for was that when the sun is in optimal alignment with the satellite for detecting  
 204 smoke also coincides with the maximal amount of atmosphere the light travels through. This is



194 FIG. 6. The three bands of GOES-EAST data are the raw input to generate the True Color image shown in  
 195 figure 5. There appears to be a higher signal-to-noise ratio for smoke detection as the wavelength,  $\lambda$ , of light  
 196 being measured decreases.

201 shown in figure 7, where the noise introduced by higher amounts of atmospheric noise can make  
 202 the signal from the smoke difficult to discern despite the smoke signal being at its highest. This  
 203 phenomenon is even more prominent the further the smoke is longitudinally from the satellite since  
 204 there will be more atmosphere to travel through between the light interacting with the smoke and  
 205 reaching the detector. Additionally shown in figure 7, and is especially evident for data close to  
 206 sunrise, when the time window is large, the smoke has often not dispersed to the extent of the  
 207 analysts' annotation boundaries. We consider the atmospheric interaction noise in our algorithms  
 208 to develop the dataset by choosing a lag time between sunrise and optimal image timestamp as  
 209 a function of longitude, but that only served as an approximation that improved our dataset but  
 210 still produced poor quality data in certain circumstances. There were additional variables, such  
 211 as latitude and time of year that were becoming unwieldy to the point that we decided to consider  
 212 alternative options since even if we were able to pinpoint an algorithm that accounted for the  
 213 atmospheric noise we would not be able to account for the smoke's alignment to the annotation.  
 214 Instead, as we will discuss in further detail, we use the current dataset to train a machine learning



217 FIG. 7. A smoke annotation projected onto GOES-WEST imagery from August 2022 that spans from 11:00  
 218 UTC to 15:00 UTC, sunrise on August 2nd, 2022 at coordinates ( $49^{\circ}24'N$ ,  $115^{\circ}29'W$ ) was 12:15 UTC. 15  
 219 minutes after the calculated sunrise, the imagery contains a noticeably higher levels of noise than the subsequent  
 220 imagery that's light travels through (thus interacts with) less atmosphere as the sun rises and the angle between  
 221 GOES-WEST and the sun decreases.

215 model that would then be able to go back to the original satellite imagery options and choose which  
 216 image best overlaps with the analyst annotation.

## 222 MACHINE LEARNING MODEL

223 We implement a deep learning architecture that uses the encoder from the ResNet model He et al.  
 224 (2015) and a semantic segmentation classifier from the U-Net model Ronneberger et al. (2015).  
 225 Transfer learning has shown to reduce the time and resources needed to train a model by leveraging  
 226 information from pre-trained models Yosinski et al. (2014), Razavian et al. (2014). We initialize  
 227 the values of our model weights using the pre-trained values originally trained on the ImageNet  
 228 dataset Deng et al. (2009), containing 1.2 million images and 1000 categories. Our model was  
 229 developed using the Segmentation Models PyTorch package Iakubovskii (2019) that was written  
 230 as a high level API for implementing models for semantic segmentation problems. We input  
 231 256x256x3 snapshots of True Color GOES imagery that contains smoke and output a 256x256x3  
 232 classification map that predicts if a pixel contains smoke and if so, what the density of that smoke  
 233 is. As mentioned earlier, we apply the thermometer encoding shown in table 2 to encode the smoke  
 234 densities and apply binary cross entropy as the loss function per density of smoke.

235 The original dataset developed using the Mie algorithm contained over 120,000 samples. To  
236 train our model, we split the dataset into training (95,000 samples), validation (12,000 samples)  
237 and testing (12,000) datasets. Training data contains data from the years 2018, 2019, 2020, 2021  
238 and 2023 while the data from 2022 is split into validation and testing data by taking data from  
239 alternating weeks of the year. Splitting 2022 data by week allowed us to leave more full years of  
240 data for the training set and allowed each dataset to show yearlong trends while trying to keep the  
241 datasets independent from one another.

242 We trained a model over 20 epochs and then used that model to develop the dataset by determining  
243 which satellite image provided the best Intersection over Union (IoU) value. The IoU metric is  
244 given by the ratio of area of overlap to the area of union as defined in equation 1, where A and B  
245 are the truth labels and the model's predictions.

$$IoU = \frac{|A \cap B|}{|A| \cup |B|} \quad (1)$$

246 To determine which image best represents the analyst annotation, we gather all the satellite  
247 imagery for the given time window and run them through the machine learning model. The output  
248 of the model gives a prediction on if there is smoke in the image, if so, where the smoke is in that  
249 image and what the density of that smoke is. For each density predicted, we calculate the IoU  
250 using the total set of pixels that the model predicts as that density of smoke and the entire set of  
251 pixels labeled by the analyst as a particular smoke density for each image. The image with the  
252 highest IoU score is chosen as the image that best represents the analyst smoke annotation. If the  
253 maximum IoU value was under .1, we did not include that annotation in the dataset..

## 254 Results

255 To interpret the performance of our trained model, we report the IoU metrics in table 4 that were  
256 computed by running the model on the Mie algorithm derived dataset and the ML derived Dataset.  
257 For each density, we calculate the IoU using the total set of pixels that the model predicts as that  
258 density of smoke and the entire set of pixels labeled by the analyst as a particular smoke density  
259 over all imagery contained in the testing dataset. Additionally, we compute the overall IoU for all  
260 densities by first computing the number of pixels that insect their correct density and divide that

261 by the total number of pixels that make up the union of model predicted and analyst labeled smoke  
262 as shown in equation 2.

$$IoU_{overall} = \frac{\sum_{i=light}^{heavy} |A_i \cap B_i|}{\sum_{i=light}^{heavy} |A_i| \cup |B_i|} \quad (2)$$

TABLE 4. IoU results per density of smoke and over all densities.

category	IoU Mie Dataset	IoU ML Dataset
Light	0.394	0.551
Medium	0.283	0.392
Heavy	0.233	0.290
Overall	0.365	0.510

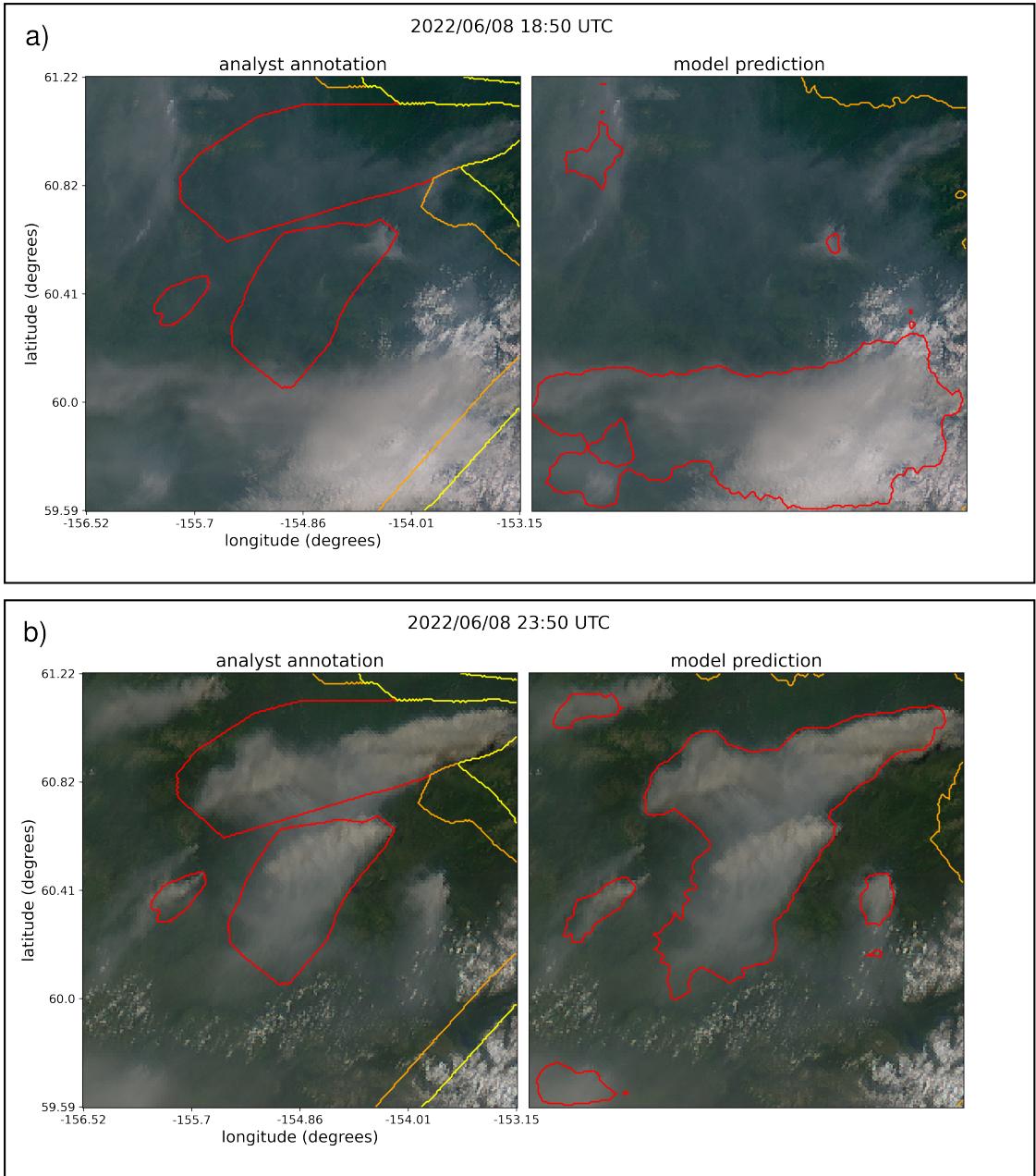
### 263 3. Conclusion

264 In this study, we have refined an existing dataset originally curated by the HMS team, transforming  
265 it from a many-to-one imagery-to-annotation format to a more concise one-to-one satellite image-  
266 to-annotation dataset. The initial HMS dataset primarily gave a approximation of where smoke  
267 had been present for a given time window, though it did not confirm the actual existence of smoke  
268 in the pixels of the selected images. Due to that nature of the HMS dataset, our Mie derived dataset  
269 gave an approximation of when we'd best be able to measure the smoke signal but did not factor  
270 in if the smoke was actually present in the selected image. This discrepancy can be detrimental  
271 when training a machine learning model, as it may penalize accurate predictions and inadvertently  
272 introduce biases towards misclassifying noise as meaningful signal.

273 To make improvements on the dataset's reliability, we apply a machine learning model trained  
274 on the Mie-derived dataset to select the most appropriate satellite images within the timeframe  
275 specified by the analyst's annotation. A notable illustration of the improvements introduced by  
276 the machine learning method is evident in Figure 8. The annotation associated with this example  
277 encompasses five hours of imagery and we show the images that is chosen by each of our two  
278 methods. While the Mie algorithm tries to optimize for the highest possible signal-to-noise, which

<sup>279</sup> is the image closest to sunrise, our machine learning algorithm chooses the image that maximizes  
<sup>280</sup> the overlap of smoke predicted by the model with the analyst's annotation.

<sup>287</sup> The result of this study is a curated dataset that can be used to train machine learning models  
<sup>288</sup> for various wildfire smoke applications. The end goal is to produce a robust and reliable machine  
<sup>289</sup> learning based approach for detecting wildfires using satellite imagery. That information can  
<sup>290</sup> be used for wildfire monitoring and as data provided to public health officials for air quality  
<sup>291</sup> assessments.



281 FIG. 8. GOES-WEST imagery showing smoke on June 8th, 2022 in Alaska where at the coordinates (61°03'N,  
 282 156°07'W), daylight was between 12:43-7:53 UTC. The smoke annotations displayed span from 18:50 to 23:50  
 283 UTC. a) shows the imagery that was selected using the Mie algorithm, which optimizes for the image closest  
 284 to sunrise. b) shows the imagery chosen by the ML algorithm that had the highest IoU score. The IoU scores  
 285 are similar for the low and medium density smoke, but the high density smoke IoU for a) is .01 while b) is  
 286 significantly higher at .59.

292 *Acknowledgments.*

293 *Data availability statement.*

294 **References**

- 295 Ba, R., C. Chen, J. Yuan, W. Song, and S. Lo, 2019: Smokenet: Satellite smoke scene detection  
296 using convolutional neural network with spatial and channel-wise attention. *Remote Sensing*,  
297 **11** (14), 1702.
- 298 Bah, M., M. Gunshor, and T. Schmit, 2018: Generation of goes-16 true color imagery without a  
299 green band. *Earth and Space Science*, **5** (9), 549–558.
- 300 Deng, J., K. Li, M. Do, H. Su, and L. Fei-Fei, 2009: Construction and Analysis of a Large Scale  
301 Image Ontology. Vision Sciences Society.
- 302 He, K., X. Zhang, S. Ren, and J. Sun, 2015: Deep residual learning for image recognition.  
303 1512.03385.
- 304 Iakubovskii, P., 2019: Segmentation models pytorch. GitHub, [https://github.com/qubvel/segmentation\\_models.pytorch](https://github.com/qubvel/segmentation_models.pytorch).
- 305 Larsen, A., I. Hanigan, B. J. Reich, Y. Qin, M. Cope, G. Morgan, and A. G. Rappold, 2021: A deep  
306 learning approach to identify smoke plumes in satellite imagery in near-real time for health risk  
307 communication. *Journal of exposure science & environmental epidemiology*, **31** (1), 170–176.
- 308 McNamara, D., G. Stephens, M. Ruminski, and T. Kasheta, 2004: The hazard mapping system  
309 (hms) - noaa's multi-sensor fire and smoke detection program using environmental satellites.  
310 *Conference on Satellite Meteorology and Oceanography*.
- 311 Petty, G., 2006: *A First Course in Atmospheric Radiation*. Sundog Pub.
- 312 Randriambelo, T., S. Baldy, M. Bessafi, M. Petit, and M. Despinoy, 1998: An improved detection  
313 and characterization of active fires and smoke plumes in south-eastern africa and madagascar.  
314 *International Journal of Remote Sensing*, **19** (14), 2623–2638.

- 316 Raspaud, M., and Coauthors, 2018: Pytroll: An open-source, community-driven python framework  
317 to process earth observation satellite data. *Bulletin of the American Meteorological Society*,  
318 **99** (7), 1329–1336.
- 319 Razavian, A. S., H. Azizpour, J. Sullivan, and S. Carlsson, 2014: Cnn features off-the-shelf: an  
320 astounding baseline for recognition. 1403.6382.
- 321 Rolph, G. D., and Coauthors, 2009: Description and verification of the noaa smoke forecasting  
322 system: the 2007 fire season. *Weather and Forecasting*, **24** (2), 361–378.
- 323 Ronneberger, O., P. Fischer, and T. Brox, 2015: U-net: Convolutional networks for biomedical  
324 image segmentation. 1505.04597.
- 325 Schroeder, W., M. Ruminski, I. Csiszar, L. Giglio, E. Prins, C. Schmidt, and J. Morisette, 2008:  
326 Validation analyses of an operational fire monitoring product: The hazard mapping system.  
327 *International Journal of Remote Sensing*, **29** (20), 6059–6066.
- 328 Sun, C., A. Shrivastava, S. Singh, and A. Gupta, 2017: Revisiting unreasonable effectiveness of  
329 data in deep learning era. 1707.02968.
- 330 Wang, Z., P. Yang, H. Liang, C. Zheng, J. Yin, Y. Tian, and W. Cui, 2022: Semantic segmentation  
331 and analysis on sensitive parameters of forest fire smoke using smoke-unet and landsat-8 imagery.  
332 *Remote Sensing*, **14** (1), 45.
- 333 Wen, J., and M. Burke, 2021: Wildfire smoke plume segmentation using geostationary satellite  
334 imagery. *ArXiv*, **abs/2109.01637**, URL <https://api.semanticscholar.org/CorpusID:237416777>.
- 335 Yosinski, J., J. Clune, Y. Bengio, and H. Lipson, 2014: How transferable are features in deep  
336 neural networks? 1411.1792.