
SmokeViz: A Pseudo-Labeled Smoke Plume Dataset For Deep Learning Applications

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 The increase in the frequency of wildfires on a global scale underscores the need
2 for advancements in fire monitoring techniques for disaster management, environmental
3 protection and to mitigate negative health outcomes. This research
4 introduces an innovative, data-driven framework that leverages the semi-supervised
5 method, pseudo-labeling, to generate smoke plume annotations in geostationary
6 satellite imagery. The primary objective is to refine an existing National Oceanic
7 and Atmospheric Administration smoke dataset that provides temporal and geo-
8 graphical information on individual smoke plumes but at variable and, primarily,
9 low temporal resolution. To do this, we use deep learning and pseudo-labels to
10 pinpoint the singular, most representative, satellite image that optimally illustrates
11 the smoke annotation within the given time window. By identifying the most
12 representative imagery of smoke plumes for a given smoke annotation, the study
13 seeks to create an accurate and relevant machine learning dataset. The resulting
14 dataset is anticipated to be an instrumental tool in developing further machine
15 learning models, such as an automated system capable of real-time monitoring and
16 annotation of smoke plumes directly from streaming satellite imagery.

17

1 Introduction

18 In recent years, the escalation of wildfire incidents worldwide has become a prominent environmental
19 and public health concern. The combustion process in wildfires releases smoke containing fine
20 particulate matter (PM2.5) and harmful gases, posing severe hazards to human health and air quality.
21 These risks underscore the necessity for efficient and effective monitoring methods to mitigate the
22 adverse health impacts associated with wildfire smoke.

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

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

35 The integration of satellite imagery in wildfire smoke monitoring is not only instrumental in providing
 36 real-time data but also plays a significant role in public health planning and response. By mapping
 37 the spread and density of smoke, health authorities can issue timely warnings, implement evacuation
 38 protocols, and deploy resources effectively to mitigate health risks. Furthermore, long-term data
 39 gathered from satellite observations can aid in understanding the broader impacts of wildfire smoke
 40 on public health, influencing policy decisions and preventive measures.
 41 Currently, multi-channel thresholding is a popular method to distinguish smoke pixels from pixels
 42 containing dust, clouds or other phenomenon with similar signatures. The method uses historical,
 43 labeled data to extract optimal radiance values for each channel that corresponds with the labeled
 44 class. These methods are tuned to particular biogeographies and often have issues with generalization
 45 to new locations with varying fuel types [11].
 46 In contrast to the numerical thresholding approach, human visual inspection of satellite imagery is
 47 another commonly used method for smoke identification. Trained analyst will inspect imagery and
 48 label the smoke by hand. This method is not as scalable as an automated approach and is limited by
 49 the availability of analysts and their time.
 50 To address these challenges we can look towards innovative approaches and technological advancements
 51 in computer vision. Machine learning methods have shown potential in improving the accuracy
 52 and efficiency of satellite-based wildfire smoke detection and monitoring. For instance, SmokeNet,
 53 uses a convolutional neural network (CNN) based framework to determine if a scene of MODIS
 54 imagery contains smoke [1]. Another study also used a CNN to identify smoke on a pixel-wise basis
 55 using imagery from Himiawari-8 [7]. Additionally, Wen et al. developed a CNN architecture that
 56 takes GOES-EAST imagery as input and National Oceanic and Atmospheric Administration (NOAA)
 57 generated annotations for the target labels during training [19].
 58 The success of deep learning methods, such as CNNs, relies heavily on the availability of a large,
 59 representative dataset [17]. Existing methods use relatively small amounts of data, from 57 [18]
 60 to 6825 [19]. In contrast, benchmark datasets for image classification contain tens of thousands
 61 (CIFAR-10 and MNIST) to millions (CIFAR-100 and ImageNet) of data samples. Keeping in mind
 62 the correlation between both the quality and quantity of data with model performance, we introduce
 63 the largest known smoke dataset, SmokeViz, containing over 120,000 samples.

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

Reference	Method	# Samples	Satellite	# Channels	Level
[1]	CNN	6255	MODIS	5	image
[19]	CNN	6825	GOES-EAST	5	pixel
[7]	CNN	975	Himiwari-8	7	pixel
[18]	U-Net	47	Landsat-8	13	pixel
SmokeViz	U-Net	120,000	GOES-EAST/WEST	3	pixel

64 An approach to increase the number of labeled samples in a dataset, semi-supervised learning
 65 leverages a labeled dataset to generate new labels for an often larger, but unlabeled, dataset. Pseudo-
 66 labeling, a form of semi-supervised learning, uses labeled data to train an initial model, then runs
 67 that model on unlabeled data to predict pseudo-labels, and finally trains a new model using the
 68 pseudo-labels [8]. We introduce a variation of pseudo-labeling not to increase the size, but to increase
 69 the quality of our dataset by using the pseudo-labels to choose the best satellite image out of a given
 70 time-window to represent each smoke plume annotation.

71 2 Methods

72 Dataset

73 The initial data source, discussed in further detail in the next section, is uniquely characterized by
 74 each annotation having corresponding imagery ranging between 1-60 frames, where each frame
 75 captures 5 minutes of exposure. Additionally, we have two satellite sensors, GOES-EAST and
 76 GOES-WEST, doubling the number of frames for a single annotation. We apply pseudo-labeling to

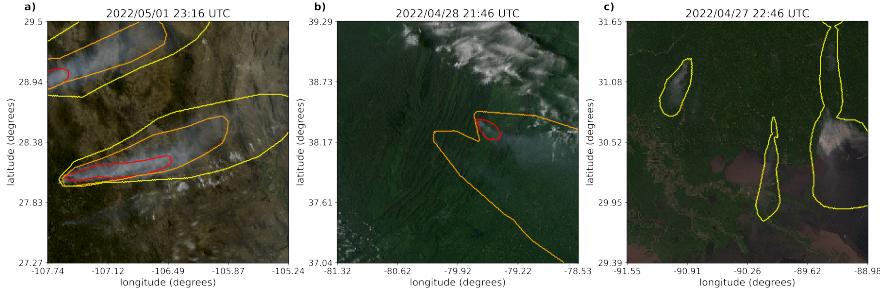


Figure 1: Satellite imagery captured by GOES-EAST within a few days of each other. The yellow, orange and red contours indicate the extent of Light, Medium and Heavy smoke. a) shows a canonical example of a smoke plume. b) and c) show variations in the density labels. b) we show Medium and Heavy densities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

77 develop a dataset that has a one-to-one annotation-to-image ratio, where we choose the best satellite
 78 image that represents where the smoke is located in the analyst annotation.

79 Dataset development came in three stages. First, we use the physics of light scattering to determine
 80 which singular satellite image would be in the optimal configuration for smoke detection. Second, we
 81 used that dataset to train an initial model that will identify smoke in satellite imagery. Third, we use
 82 that initial model to label each satellite image in a given annotation’s time-window and the optimal
 83 satellite image is chosen based on which image’s pseudo-labels has the greatest overlap with the
 84 analyst annotation for the given location and densities of smoke.

85 Smoke Labels

86 The National Environmental Satellite, Data and Information Service (NESDIS) and NOAA manage
 87 environmental satellite programs such as the Hazard Mapping System (HMS) [9, 16]. The HMS
 88 program is an operational system that uses an aggregation of satellite data to generate active fire and
 89 smoke data that is used in applications such as air quality assessments and serves as verification and
 90 validation for NOAA’s smoke forecasting model, HYSPLIT, [14]. To train our model, we implement
 91 a supervised learning framework that uses the HMS analyst smoke product as truth labels during the
 92 model training process.

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

99 Thermometer Encoding Smoke Densities

100 One of the challenges introduced with using human generated qualitative smoke densities was that, as
 101 seen in figure 1b and 1c, there are variations in what is labeled as heavy or light density smoke. More
 102 generally, reproducing qualitative metrics with quantitative algorithms is a challenging problem, but
 103 we apply mathematical approaches that mitigate some of the underlying complications of our specific
 104 problem. Despite the fact that the smoke densities introduce qualitative complexities, we decided
 105 that the density approximations were important to use in our dataset because of the differences in
 106 signatures the densities produce. Within the satellite imagery, the appearance of a light density
 107 smoke plume will look significantly different than a heavy density smoke plume as seen in figure 1.
 108 Additionally, a light density smoke plume is expected to be more challenging to detect since it is easier
 109 for it to be misclassified as not smoke. During the training process, the separate density categories
 110 allows us to deferentially weight the penalization given to the model for incorrect classifications

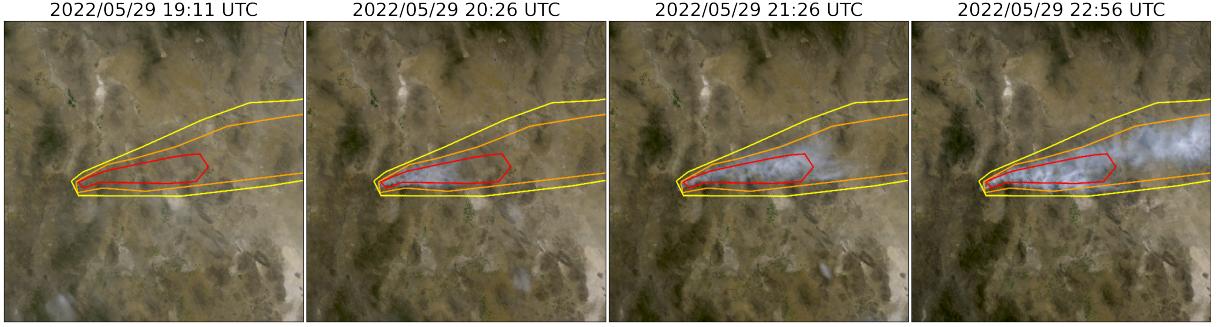


Figure 2: True Color GOES imagery from May 2022, Southeast New Mexico (31°N , 100°W) during the start of the Foster Fire. The HMS annotations for the smoke outlines shown here spanned from 19:10–23:00 UTC but visually match the smoke location for the last part of the timewindow.

111 based on category. For example, the model can be given a small penalization for misclassifying light
 112 smoke as not smoke while given a higher penalization for misclassifying heavy smoke as not smoke.
 113 In addition to the densities being ordered and categorical, the differences between the density
 114 categories are not evenly distributed by a metric, such as particulate matter per square meter. The
 115 intervals between densities being unknown along with the hierarchical nature of the density labels
 116 makes the labels ordinal instead of just categorical. This data property allows us to use thermometer
 117 encoding, which leverages the idea that heavy density smoke includes both medium and light density
 118 smoke, that heavy density smoke is closer to medium than it is to light and automatically weights
 119 the loss functions and incorporates the ranked ordering of the densities. As seen in Table 2, one-
 120 hot encoding, commonly used for categorical data, doesn't take ordinal properties of the data into
 121 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]

122 Time Windows For Smoke Annotations

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

Table 3: To create a true color image, we use the following bands from the ABI Level 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

136 Satellite Imagery

137 The Geostationary Operational Environmental Satellites (GOES) are operated by the NOAA and
 138 NESDIS support meteorology research and forecasting for the United States. We use the latest
 139 operational satellites, GOES-16 (EAST), 17 and 18 (WEST) that carry the ABI, that measure 16
 140 bands between the visible and infrared wavelengths. In improvement to the GOES predecessors,
 141 imagery is collected every 5 minutes for the contiguous United States and every 10 minutes for the
 142 full disk. We use bands 1-3 (Table 3) as input to Satpy's composite algorithm to develop a true color
 143 image representation, similar to what is used as input by HMS analysts [12] and [2].

144 We used a physics-informed approach in selecting the initial dataset for training our model. Rather
 145 than use the cumulative data from GOES-WEST and GOES-EAST images, we select one or the other
 146 based on the solar zenith angle. For smoke identification, this approach can achieve a much higher
 147 signal-to-noise than imaging the earth's surface from an arbitrary angle. The elastic scattering of
 148 light is the primary mechanism to account for - while the atmosphere is composed of molecules with
 149 size $< 1\text{nm}$, smoke particles can vary from $100\text{ nm} - 10\mu\text{m}$ in diameter, d . The GOES ABI covers
 150 spectral bands from $0.47\mu\text{m} - 13.3\mu\text{m}$, so atmospheric and smoke particle sizes occupy two very
 151 different regimes with respect to the imaging wavelength λ , as shown in figure 3. In the extreme limit
 152 of $\lambda \gg d$, the physics of scattering of light off a small sphere is captured by Rayleigh scattering. This
 153 process has two critical consequences: (1) the scattering cross section of light is strongly wavelength
 154 dependent (scaling with λ^{-4}), meaning that photons with wavelength closer to the ultraviolet are
 155 scattered more strongly than infrared photons. (2) the scattering cross section scales with an angular
 156 dependent cross section of $(1 + \cos^2\theta)$. Scattered photons follow the emission distribution of a
 157 radiating dipole, scattering more strongly in the forward and backwards directions ($\theta = 0, \pi$) than
 158 orthogonal to the direction of propagation ($\theta = \pi/2, 3\pi/2$), see figure 4 for Rayleigh scattering
 159 schematic.

160 The significance of these scalings is that the observer, or detector, will receive blue photons in most
 161 directions orthogonal to the source. Equivalently, photons traveling colinearly with line of sight
 162 to the emission source will mostly have wavelengths in the infrared band. In the converse regime
 163 of $d > \lambda$, the elastic scattering of light against matter is modeled through Mie scattering. Unlike
 164 Rayleigh scattering, Mie scattering is largely wavelength independent and has a more complicated
 165 radiation pattern where the cross section has a maximal amplitude in the forward direction. An
 166 observer downstream of this scatterer will collect more photons than one positioned directly behind
 167 it. In the context of smoke identification, a sunrise or sunset will lead to a higher Mie scattered signal
 168 in GOES-WEST and GOES-EAST respectively, as shown with a smoke plume producing a stronger
 169 signal in GOES-EAST imagery near sunset in figure 2.

170 Smoke identification therefore amounts to extracting a signal of $d > \lambda$ photons from the $\lambda \gg d$
 171 background. Positioning a detector along line of sight to the scatterer will result in a higher signal
 172 from smoke particles (figure 4). Filtering the imaged wavelength can enhance this signal; photons
 173 collected in the blue spectrum will have a naturally lower background along the line of sight to the
 174 illumination source do their high level of Rayleigh scattering as. Therefore, as demonstrated in figure
 175 6, this configuration results in the highest signal to noise imaging for smoke particles.

176 Based on these criteria, the optimal strategy is to pull data from GOES-WEST right after sunrise
 177 and from GOES-EAST right before sunset. Another consideration to account for was that when the
 178 sun is in optimal alignment with the satellite for detecting smoke also coincides with the maximal
 179 amount of atmosphere the light travels through. This is shown in figure 7, where the noise introduced
 180 by higher amounts of atmospheric interactions can obfuscate the signal from the smoke, despite the
 181 smoke signal being at its highest. This phenomenon is even more prominent the further the smoke
 182 is longitudinally from the satellite since the light must travel through more atmosphere between

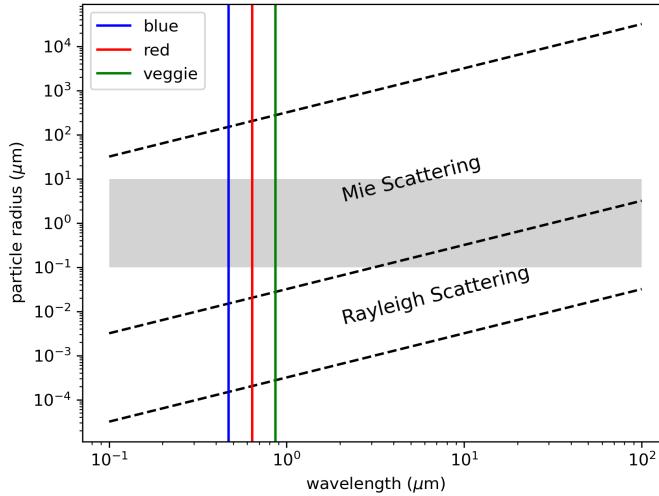


Figure 3: Relationship between the size of a particle, the wavelength of light interacting with the particle and the type of scattering behavior induced by that interaction. The dotted lines represent rough estimates of the boundaries between the scattering regimes [10]. The gray area represents the range of particle radius relevant to smoke particulate matter.

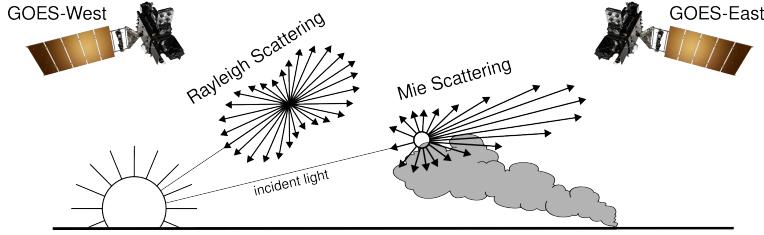


Figure 4: If the particle size is $< \frac{1}{10}$ the wavelength of the interacting light, then the primary scattering will be Rayleigh. Mie scattering is the predominant scattering mechanism when the particle size is larger than wavelength of light.

183 scattering off the smoke and reaching the detector. Additionally shown in figure 7, and is especially
 184 evident for data close to sunrise, when the time window is large, the smoke has often not dispersed
 185 to the extent of the analysts' annotation boundaries. We consider the atmospheric interaction noise
 186 in our algorithms to develop the dataset by choosing a lag time between sunrise and optimal image
 187 timestamp as a function of longitude.

188 The resulting algorithm used atmospheric properties and light scattering physics to make an estimate
 189 of which singular satellite image within the analyst time-window would give the best representation of
 190 the smoke plume label. That dataset was then used to train a model that would generate pseudo-labels
 191 for every image within the time-window and choose the image with the highest alignment between
 192 smoke in the image and annotation.

193 Machine Learning Model

194 We implement a deep learning architecture that uses the encoder from the ResNet model [5] and a
 195 semantic segmentation classifier from the U-Net model [15]. Transfer learning has shown to reduce
 196 the time and resources needed to train a model by leveraging information from pre-trained models
 197 [20], [13]. We initialize the values of our model weights using the pre-trained values originally
 198 trained on the ImageNet dataset [3], containing 1.2 million images and 1000 categories. Our model
 199 was developed using the Segmentation Models PyTorch package [6] that was written as a high level
 200 API for implementing models for semantic segmentation problems. We input 256x256x3 snapshots

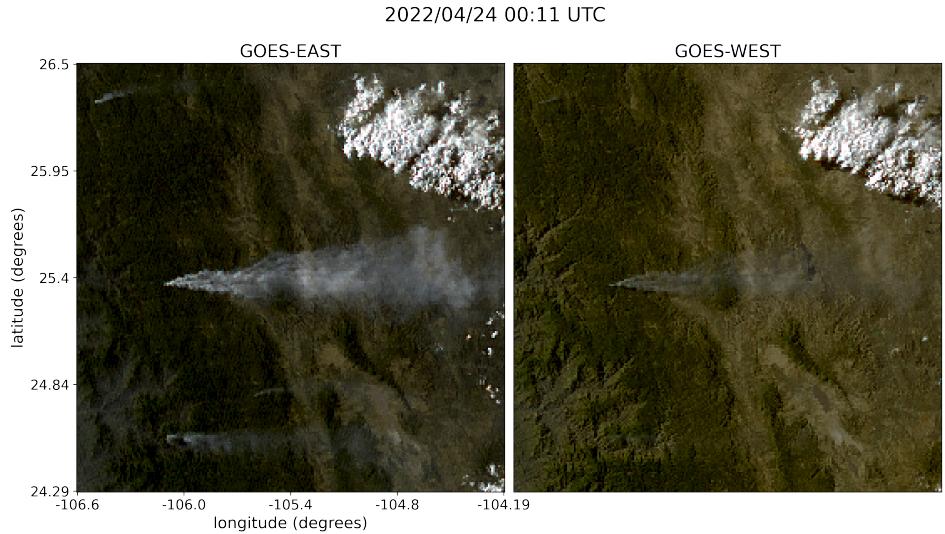


Figure 5: True Color GOES-EAST (left) and GOES-WEST (right) imagery from April 24th, 2022. The images were taken about 1.5 hours before sunset for this geolocation and time of year (01:43 UTC).

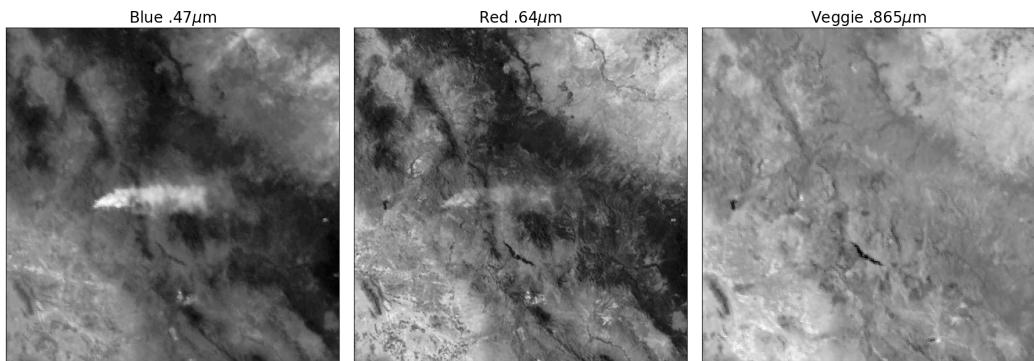


Figure 6: The three bands of GOES-EAST data are the raw input to generate the True Color image shown in figure 5. There appears to be a higher signal-to-noise ratio for smoke detection as the wavelength, λ , of light being measured decreases.

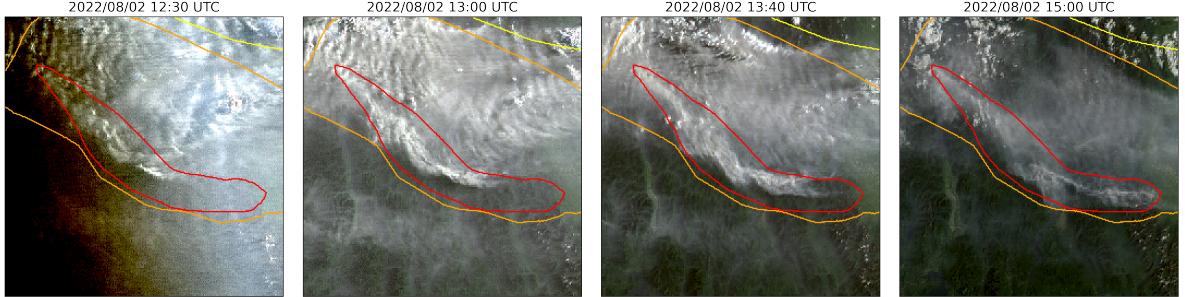


Figure 7: A smoke annotation projected onto GOES-WEST imagery from August 2022 that spans from 11:00 UTC to 15:00 UTC, sunrise on August 2nd, 2022 at coordinates ($49^{\circ}24'N$, $115^{\circ}29'W$) was 12:15 UTC. 15 minutes after the calculated sunrise, the imagery contains a noticeably higher levels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

201 of True Color GOES imagery that contains smoke and output a $256 \times 256 \times 3$ classification map that
 202 predicts if a pixel contains smoke and if so, what the density of that smoke is. As mentioned earlier,
 203 we apply the thermometer encoding shown in table 2 to encode the smoke densities and apply binary
 204 cross entropy as the loss function per density of smoke.

205 The original dataset developed using the Mie algorithm contained over 120,000 samples. To train our
 206 model, we split the dataset into training (95,000 samples), validation (12,000 samples) and testing
 207 (12,000) datasets. Training data contains data from the years 2018, 2019, 2020, 2021 and 2023 while
 208 the data from 2022 is split into validation and testing data by taking data from alternating 10 days of
 209 the year. Splitting 2022 data by 10 days allowed us to leave more full years of data for the training
 210 set and allowed each dataset to show yearlong trends while trying to keep the datasets independent
 211 from one another.

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

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

216 To determine which image best represents the analyst annotation, we gather all the satellite imagery
 217 for the given time window and run them through the machine learning model. The output of the
 218 model gives a prediction on if there is smoke in the image, and if there is smoke, where the smoke is
 219 in that image and what the density of that smoke is. The model output generates pseudo-labels for
 220 each density of smoke that are compared to the analyst annotations. To compare the pseudo labels and
 221 analyst labels, we calculate the IoU using the total set of pixels for the pseudo-labels at that density of
 222 smoke and the entire set of pixels for the analyst labels for a particular smoke density in each image.
 223 The image with the highest IoU score is chosen as the image that best represents the analyst smoke
 224 annotation. Generally, a confidence threshold value is defined to decide if a pseudo-label should to
 225 be included in a dataset [4]. We chose a confidence threshold that would include the sample in the
 226 dataset if the maximum IoU value was over 0.1.

227 Results

228 To interpret the performance of our trained model, we report the IoU metrics in table 4 that were
 229 computed by running the model on the Mie algorithm derived dataset and the pseudo-labeled dataset.

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

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

230 For each density, we calculate the IoU using the total set of pixels that the model predicts as that
 231 density of smoke and the entire set of pixels labeled by the analyst as a particular smoke density over
 232 all imagery contained in the testing dataset. Additionally, we compute the overall IoU for all densities
 233 by first computing the number of pixels that intersect their correct density and divide that by the total
 234 number of pixels that make up the union of model predicted and analyst labeled smoke as shown in
 235 equation 2.

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

236 3 Limitations

237 **TO DO:** Address how the pseudo-labeled dataset is obviously going to have a better IoU score.

238 4 Conclusion

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

248 To make improvements on the dataset's reliability, we apply a machine learning model trained on the
 249 Mie-derived dataset to select the satellite image within the time-frame that best overlaps with the
 250 analyst's annotation. A notable illustration of the improvements introduced by the machine learning
 251 method is evident in Figure 8. The annotation associated with this example encompasses five hours of
 252 imagery and we show the images that are chosen by each of our two methods. While the Mie algorithm
 253 tries to optimize for the highest possible signal-to-noise, which is the image closest to sunrise, our
 254 machine learning algorithm chooses the image that maximizes the overlap of smoke predicted by the
 255 model with the analyst's annotation.

256 The result of this study is a curated dataset that can be used to train machine learning models for
 257 various wildfire smoke applications. The end goal is to produce a robust and reliable machine learning
 258 based approach for detecting wildfires using satellite imagery. That information can be used for
 259 wildfire monitoring and as data provided to public health officials for air quality assessments.

260 5 Acknowledgments and Disclosure of Funding

261 This work was partially supported by the NOAA Global Systems Laboratory and Cooperative Institute
 262 for Research in Environmental Sciences at the University of Colorado Boulder. We thank Wilfrid
 263 Schroeder and the Hazard Mapping Systems team for giving guidance on how they created their

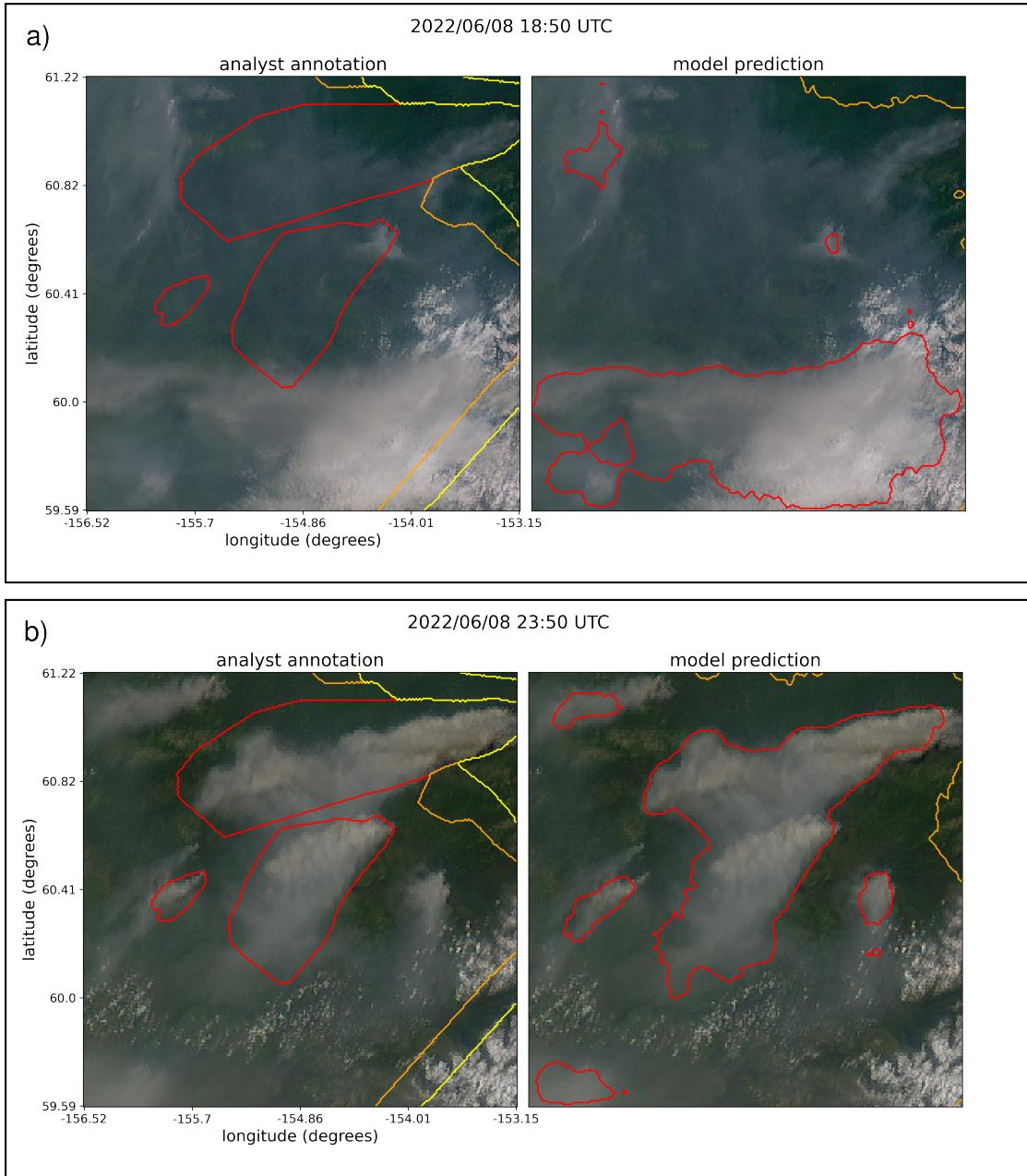


Figure 8: GOES-WEST imagery showing smoke on June 8th, 2022 in Alaska where at the coordinates (61°03'N, 156°07'W), daylight was between 12:43-7:53 UTC. The smoke annotations displayed span from 18:50 to 23:50 UTC. a) shows the imagery that was selected using the Mie algorithm, which optimizes for the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

264 smoke plume dataset. This work utilized the Alpine high performance computing resource at the
265 University of Colorado Boulder. Alpine is jointly funded by the University of Colorado Boulder, the
266 University of Colorado Anschutz, Colorado State University, and the National Science Foundation
267 (award 2201538).

268 References

- 269 [1] R. Ba, C. Chen, J. Yuan, W. Song, and S. Lo. Smokenet: Satellite smoke scene detection using
270 convolutional neural network with spatial and channel-wise attention. *Remote Sensing*, 11(14):
271 1702, 2019.
- 272 [2] M. Bah, M. Gunshor, and T. Schmit. Generation of goes-16 true color imagery without a green
273 band. *Earth and Space Science*, 5(9):549–558, 2018.
- 274 [3] J. Deng, K. Li, M. Do, H. Su, and L. Fei-Fei. Construction and Analysis of a Large Scale Image
275 Ontology. Vision Sciences Society, 2009.
- 276 [4] R. E. Ferreira, Y. J. Lee, and J. R. Dórea. Using pseudo-labeling to improve performance of
277 deep neural networks for animal identification. *Scientific Reports*, 13(1):13875, 2023.
- 278 [5] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.
- 279 [6] P. Iakubovskii. Segmentation models pytorch. https://github.com/qubvel/segmentation_models.pytorch, 2019.
- 280 [7] A. Larsen, I. Hanigan, B. J. Reich, Y. Qin, M. Cope, G. Morgan, and A. G. Rappold. A deep
281 learning approach to identify smoke plumes in satellite imagery in near-real time for health risk
282 communication. *Journal of exposure science & environmental epidemiology*, 31(1):170–176,
283 2021.
- 284 [8] D.-H. Lee. Pseudo-label : The simple and efficient semi-supervised learning method for deep
285 neural networks. *ICML 2013 Workshop : Challenges in Representation Learning (WREPL)*, 07
286 2013.
- 287 [9] D. McNamara, G. Stephens, M. Ruminski, and T. Kasheta. The hazard mapping system (hms) -
288 noaa’s multi-sensor fire and smoke detection program using environmental satellites. *Conference
289 on Satellite Meteorology and Oceanography*, 01 2004.
- 290 [10] G. Petty. *A First Course in Atmospheric Radiation*. Sundog Pub., 2006. ISBN 9780972903318.
- 291 [11] T. Randriambelo, S. Baldy, M. Bessafi, M. Petit, and M. Despinoy. An improved detection
292 and characterization of active fires and smoke plumes in south-eastern africa and madagascar.
293 *International Journal of Remote Sensing*, 19(14):2623–2638, 1998.
- 294 [12] M. Raspaud, D. Hoese, A. Dybbroe, P. Lahtinen, A. Devasthale, M. Itkin, U. Hamann, L. Ø.
295 Rasmussen, E. S. Nielsen, T. Leppelt, et al. Pytroll: An open-source, community-driven python
296 framework to process earth observation satellite data. *Bulletin of the American Meteorological
297 Society*, 99(7):1329–1336, 2018.
- 298 [13] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: an
299 astounding baseline for recognition, 2014.
- 300 [14] G. D. Rolph, R. R. Draxler, A. F. Stein, A. Taylor, M. G. Ruminski, S. Kondragunta, J. Zeng,
301 H.-C. Huang, G. Manikin, J. T. McQueen, et al. Description and verification of the noaa smoke
302 forecasting system: the 2007 fire season. *Weather and Forecasting*, 24(2):361–378, 2009.
- 303 [15] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image
304 segmentation, 2015.
- 305 [16] W. Schroeder, M. Ruminski, I. Csizar, L. Giglio, E. Prins, C. Schmidt, and J. Morisette.
306 Validation analyses of an operational fire monitoring product: The hazard mapping system.
307 *International Journal of Remote Sensing*, 29(20):6059–6066, 2008.

- 309 [17] C. Sun, A. Shrivastava, S. Singh, and A. Gupta. Revisiting unreasonable effectiveness of data in
310 deep learning era, 2017.
- 311 [18] Z. Wang, P. Yang, H. Liang, C. Zheng, J. Yin, Y. Tian, and W. Cui. Semantic segmentation and
312 analysis on sensitive parameters of forest fire smoke using smoke-unet and landsat-8 imagery.
313 *Remote Sensing*, 14(1):45, 2022.
- 314 [19] J. Wen and M. Burke. Wildfire smoke plume segmentation using geostationary satellite imagery.
315 *ArXiv*, abs/2109.01637, 2021. URL [https://api.semanticscholar.org/CorpusID:
316 237416777](https://api.semanticscholar.org/CorpusID:237416777).
- 317 [20] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural
318 networks?, 2014.