A Bayesian model for quantifying errors in citizen science data: application to rainfall observations from Nepal

J.A. Eisma[1], G. Schoups[2], J.C. Davids[3], N. van de Giesen[2]

[1]Lyles School of Civil Engineering, Purdue University, West Lafayette, Indiana, USA

[2]Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands

[3]Department of Civil Engineering and College of Agriculture, California State University, Chico, California, USA

Key Points:

- A Gaussian mix of regressions explains the likelihood of citizen scientists committing errors
- Citizen scientists are sorted into communities based on characteristics and the type and frequency of errors committed
- The distribution of errors committed by citizen scientists evolves as they gain experience

Corresponding author: J.A. Eisma, `jeisma@purdue.edu`

Abstract

High quality citizen science can be instrumental in advancing science toward new discoveries and a deeper understanding of under-observed phenomena. However, the error structure of citizen scientist (CS) data must be well-defined. Within a citizen science program, the error types in submitted observations vary, and their occurrence may depend on a variety of CS-specific variables, such as motivation. This study develops a graphical Bayesian inference model of error types in CS data. The model assumes that: (1) each CS observation is subject to a specific error type, each with its own bias and noise; and (2) an observation's error type depends on the error community of the CS, which in turn relates to characteristics of the CS submitting the observation. Given a set of CS observations and corresponding ground-truth values, the model can be calibrated for a specific application, yielding (i) number of error types and communities, (ii) bias and noise of each error type, (iii) error distribution of each community, and (iv) the community to which each CS belongs. The model, applied to Nepal CS rainfall observations, identifies seven error types and sorts CSs into four model-inferred communities. In the case study, 79% of CSs committed errors in fewer than 6.3% of their observations. The remaining tended to commit unit, meniscus, and unknown errors. A CS's assigned community, coupled with the model-inferred error probability, can identify observations that require verification. With such a system, the onus of validating CS data is partially transferred from human effort to machine-learned algorithms.

1 Introduction

Communities worldwide face increasing uncertainty regarding extreme weather events engendered by climate change. Reliable weather forecasts allow a community to initiate proactive measures when anticipating an extreme event—measures that sometimes save hundreds, if not thousands of lives. Unfortunately, sparse weather data in many regions of the world inhibit coordinated response efforts of local and regional governments (Teague & Gallicchio, 2017, p. 218). Citizen science can help bridge such data gaps.

Citizen science programs, organized efforts to collect scientific data from members of the public, have become increasingly popular as advances in technology have made the data collection and submission process more accessible (Bonney et al., 2009; Newman et al., 2012). Some traditional scientists, questioning the quality of data submitted by lay members of the public, have yet to accept the legitimacy of scientific discoveries advanced by citizen scientists (Hunter, Alabri, & van Ingen, 2013; Riesch & Potter, 2014; Sheppard & Terveen, 2011). Others, however, have embraced citizen science as an effective means for increasing the spatial and temporal resolution of scientific data. Successful citizen science programs investigate the type and frequency of errors committed by program participants and develop training initiatives designed to reduce errors (Bird et al., 2014; Crall et al., 2011; Davids et al., 2019).

Most citizen scientist programs conduct quality control of the data submitted by their participants. For example, citizen scientists report when they feel an earthquake and rank its strength for the United States Geological Survey's (USGS) Did You Feel It? program. The USGS removes outliers and aggregates reported intensities at zip code or city-level after processing the data through the Community Decimal Intensity algorithm (USGS, n.d.). While the USGS's quality control measures are simple to implement and suitable for their program goals, some citizen scientist programs invest significant time and energy into assuring the quality of their data. For example, citizen scientists submit rainfall depth observations to the SmartPhones4Water-Nepal (S4W-Nepal) program. S4W-Nepal checks the value of each submitted rainfall observation against an accompanying photograph of the rain gauge and manually corrects erroneous observations (Davids et al., 2019).

Rainfall observations submitted by citizen scientists have immense potential to increase the scientific community's understanding of rain events which are, by nature, highly heterogeneous in space and time. Currently, only about 1.6% of land on Earth lies within 10 km of a rain gauge, and rain gauges are notoriously inconsistent (Kidd et al., 2017). So much so that rain gauges 4 km apart in the midwestern United States produced a correlation coefficient less than 0.5 for instantaneous rainfall (Habib, Krajewski, & Ciach, 2001). Citizen science rainfall observation programs must contend with the systematic errors inherent in measuring rainfall as well as the tendency of citizen scientists to commit measurement errors. Detailed investigations into the errors committed by citizen scientists, such as the efforts of S4W-Nepal, can help increase the utility of citizen science data and inform future program development, and is the subject of this study.

Motivated by the need to reduce the time-cost of performing quality control of citizen science data without sacrificing effectiveness, this study seeks to develop a reliable, semi-automated method for identifying citizen science observations that require additional verification. Most error analyses of citizen science data focus on identifying and removing outliers from a dataset. Trained filters flag outliers by identifying observations that do not fit within the expected range of values or classes, such as species range or allowable count (Bonter & Cooper, 2012; Wiggins, Newman, Stevenson, & Crowston, 2011). Some citizen science programs develop eligibility or trust rating procedures to identify users that are likely to submit correct observations (Delaney, Sperling, Adams, & Leung, 2008; Hunter et al., 2013). Ratings schemes that consider demographic and experience-related characteristics have potential for describing the variability in citizen science data reliability (Kosmala, Wiggins, Swanson, & Simmons, 2016). However, some individual citizen scientists do not submit enough observations to be accurately assigned a rating. To overcome such limitations, Venanzi, Guiver, Kazai, Kohli, and Shokouhi (2014) based their error analysis on four communities of citizen scientists, each with a distinctive pattern of errors. Machine learning algorithms and hierarchical, generalized linear, and mixed-effects models have also been employed by a variety of citizen science programs to identify errors (Bird et al., 2014; Venanzi et al., 2014). Despite the wide range of existing research on citizen science errors, flexible methods for analyzing errors in quantitative citizen science data remains largely unexplored.

The objective of this study is to inform quality control of quantitative citizen science data by developing a Bayesian inference model that discovers and explains the errors present in rainfall observations submitted by citizen scientists. A probabilistic graphical model was developed based on assumptions about the probabilistic relationships between citizen scientists, their characteristics, and the magnitude of errors they commit. The model identifies unique error types within the S4W-Nepal citizen scientist rainfall observations, and groups citizen scientists into communities based on their characteristics and error profile. Each community has a distinct distribution of error types and describes the likelihood that a submitted observation should be reviewed further. After testing and training, the model was applied to investigate three practical issues: multiple observations of a single rainfall event, observations submitted by citizen scientists with unknown characteristics, and the error evolution of citizen scientist data over time.

2 Study Area

SmartPhones4Water Nepal (S4W-Nepal) partners with citizen scientists across Nepal to collect rainfall observations (see Figure 1). Nepal provides an interesting background for a citizen science rainfall initiative, because of the high spatial and temporal heterogeneity in rainfall across the country. Average annual rainfall in Nepal varies from 250 mm on the leeward side of the Himalayas to over 3,000 mm in the center of the country near Pokhara, as seen in Figure 1 (Nayava, 1974). The South Asian summer monsoon brings approximately 80% of Nepal's annual precipitation during the months of June to September (Nayava, 1974). The majority of citizen scientists participating in S4W-Nepal's rainfall data collection efforts reside in the Kathmandu Valley, home to about 10% of Nepal's population (Vibhāga, 2012). While the average annual precipitation is approximately 1,500 mm in the city of Kathmandu and 1,800 mm in the surrounding hills, it is highly variable and unpredictable (Thapa, Ishidaira, Pandey, & Shakya, 2017).

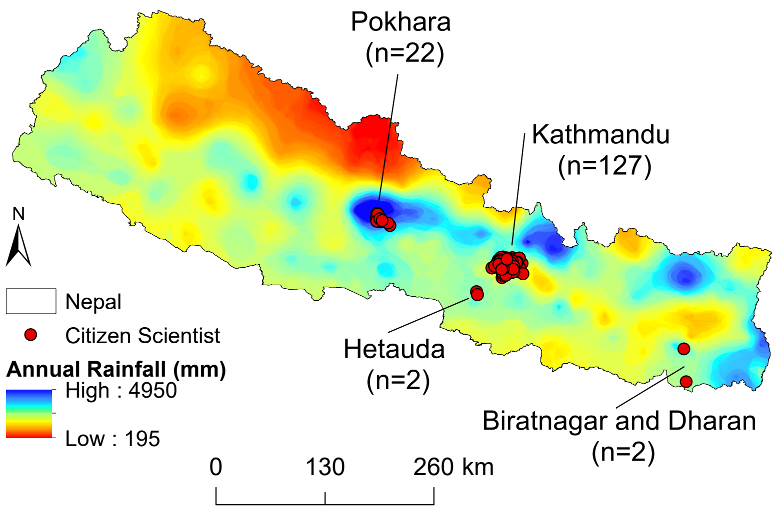

Figure 1. Locations of citizen scientists for which characteristics are known with the number of citizen scientists at specified locations shown in parentheses. Average annual rainfall shown from USAID Nepal.

3 Data

SmartPhones4Water Nepal (S4W-Nepal) recruits citizen scientists to participate in a crowdsourced rainfall observation program in Nepal. S4W-Nepal collects the submitted observations via the Open Data Kit application for smart phones. Submitted observations include geo-location data, time of measurement, citizen scientist-reported depth of rainfall in millimeters, and a photograph of the rain gauge. The program is ongoing and has collected over 24,500 observations from over 265 citizen scientists since 2016.

3.1 Rain gauges

The participants were given a rain gauge constructed by S4W-Nepal and provided instructions on the proper installation and recording of rainfall data. The rain gauges were constructed from a re-purposed clear plastic bottle with a 100 mm diameter. The bottle was filled with a few centimeters of concrete to provide stability and a level measuring surface. The lid of the bottle was cut off where the taper ends, inverted, and placed flush with the top of the bottle to reduce evaporation losses. Finally, a ruler with millimeter precision was attached to the bottle to assist the reading of the rainfall depth (Davids et al., 2019).

3.2 Citizen characteristics

During the recruitment process, S4W-Nepal recorded characteristic data for 153 citizen scientists. Characteristics recorded were: motivation (paid/volunteer), recruitment method (personal connection, random site visit, social media, outreach), age (≤18, 19-25, >25), education (<Bachelors, Bachelors, >Bachelors), place of residence (urban, semi-urban, rural), occupation (agriculture, student, other), and gender (male, female). Citizen scientist characteristics will be used here to relate individual citizen scientists with their tendency to commit errors.

4 Methods

4.1 Identification of erroneous observations

To detect erroneous rainfall observations submitted by citizen scientists, S4W-Nepal checks the value of each submitted rainfall observation against the accompanying rain gauge photograph. If they detect an error, the correct rain depth is recorded while preserving the record of the original value submitted by the citizen scientist. This allows S4W-Nepal to track the types and frequencies of errors committed by the citizen scientists. The errors that S4W-Nepal has detected are unit errors, meniscus errors, and unknown errors. Overall, approximately 9% of submitted rainfall observations are erroneous. Meniscus errors are the most common (58% of errors), followed by unknown errors (33%), and unit errors (8%) (Davids et al., 2019).

4.2 Model development

4.2.1 Assumptions and model structure

A graphical Bayesian inference model is developed based on a number of assumptions about the data being modeled. These assumptions are used to inform the relationships between the variables and ensure the model accurately represents the modeler's understanding of the physical processes that underlie the data (Winn, Bishop, Diethe, Guiver, & Zaykov, 2020). The following assumptions inform the development of the citizen science errors inference model:

1. Each citizen scientist belongs to a single community.
2. A citizen scientist's community is defined by their collective demographic and experience-related characteristics and the type and frequency of errors they have committed in prior submissions.
3. Each citizen scientist in a particular community will submit an observation with a community-specific error type distribution.
4. Each citizen scientist observation relates to an underlying true value with a systematic bias and random noise level that depends on the error type of the observation.

These assumptions are translated into the following set of equations describing the probabilistic relationship between model variables. The community C to which citizen scientist s belongs is assumed to be drawn from a discrete distribution with probability vector $ProbCommunity$ that specifies the probability of a community occurring within the citizen scientist population:

$$C_s \sim Discrete(ProbCommunity), \tag{1}$$

We use a lower case subscript to denote a random variable index (e.g. C_s indicates there is a community variable for each citizen scientist s), whereas square brackets are used to denote dependence on a random variable. The value of citizen characteristic c for citizen scientist s is assumed to be drawn from a discrete distribution with probability vector $ProbCharacteristic_c[C_s]$ that depends on the characteristic c under consideration and the community C_s the citizen scientist belongs to:

$$CitizenCharacteristic_{c,s} \sim Discrete(ProbCharacteristic_c[C_s]), \tag{2}$$

Equation 2 describes the conditional probability table between each citizen characteristic and each assigned community. Similarly, Equation 3, below, describes the conditional probability table for each error type and community. The error type $E_{s,e}$ of event e observed by citizen scientist s is assumed to be drawn from a discrete distribution with probability vector $ProbError[C_s]$ that depends on community C_s the citizen scientist belongs to:

$$E_{s,e} \sim Discrete(ProbError[C_s]), \tag{3}$$

As seen in Equations 1-3, the model assigns each citizen scientist to a single community based on their characteristics and the type and frequency of errors they commit. Next, we quantify systematic (bias) and random (noise) differences between observations and underlying true values by means of a linear regression model parameterized by an error-type specific slope a, offset b and precision (inverse variance) τ:

$$Obs_{s,e} \sim \mathcal{N}(a[E_{s,e}]True_e + b[E_{s,e}], \tau[E_{s,e}]), \tag{4}$$

where $Obs_{s,e}$ represents observed value of rainfall event e submitted by citizen scientist s, and $True_e$ is the corresponding true rainfall value for event e. Given the error type of an observation, the observed value is thus drawn from a Gaussian distribution with mean equal to an error-type specific linear function of the true value and an error-type specific variance. Square brackets indicate a, b, and τ depend on error type $E_{s,e}$. It follows that unconditionally, i.e. without knowing the error type, the relation between observed and true value is a mixture of error-type specific Gaussians, with the weight of each Gaussian in the mixture given by probability of the corresponding error type.

4.2.2 Model implementation

We implemented the probabilistic model formulated in the previous section using Microsoft Research's open source Infer.NET software framework (Minka et al., 2018). Infer.NET's framework provides adaptable tools to develop and run Bayesian inference for graphical models. The modeler must define the variables, the relationships between variables, and provide prior distributions for the variables upon which inference will be performed. Infer.NET then automatically generates a computationally efficient code for the inference algorithm. Three primary message-passing algorithms for performing inference are built into Infer.NET: expectation propagation, variational message passing, and Gibbs sampling. The model developed here employs the expectation propagation algorithm.

For implementation in Infer.NET, Equations 1-4 are translated into the factor graph shown in Figure 2. The factor graph includes observed and inferred variables, factor nodes, edges, plates, and gates. Variables are depicted by shaded or unfilled ellipses. A shaded variable is observed; an unfilled variable is inferred. Factor nodes are the small black boxes connected to variables, describing the relation between variables connected to the factor. Edges connect factor nodes to variables and identify child and parent-child relationships, as indicated by directional arrows. The value of a child variable is defined relative to the value of a parent variable. (Winn et al., 2020).

Plates. Plates are the large gray boxes surrounding portions of the factor graph. Plates are a simplified way to express repeated structures. The number of times said structure will be repeated is based on the index variable shown in the bottom right corner of the plate (Winn et al., 2020). For example, in Figure 2, the structure within the characteristics plate is repeated nine times, because the model considers nine different CS characteristics: motivation, recruitment, age, education, place of residence, occupation, gender, performance, and experience.

Gates. Gates are indicated by a dashed box, as seen around the Regression factor node in Figure 3. Gates essentially act as a switch, turning on and off depending on the value of the selector variable, which is the error type here (Minka & Winn, 2008). When gates are used to define a distribution, that distribution is mixed, as in Equation 4. Infer.NET approximates mixture distributions as a single mode distribution, which will be discussed further in Section 5.4.

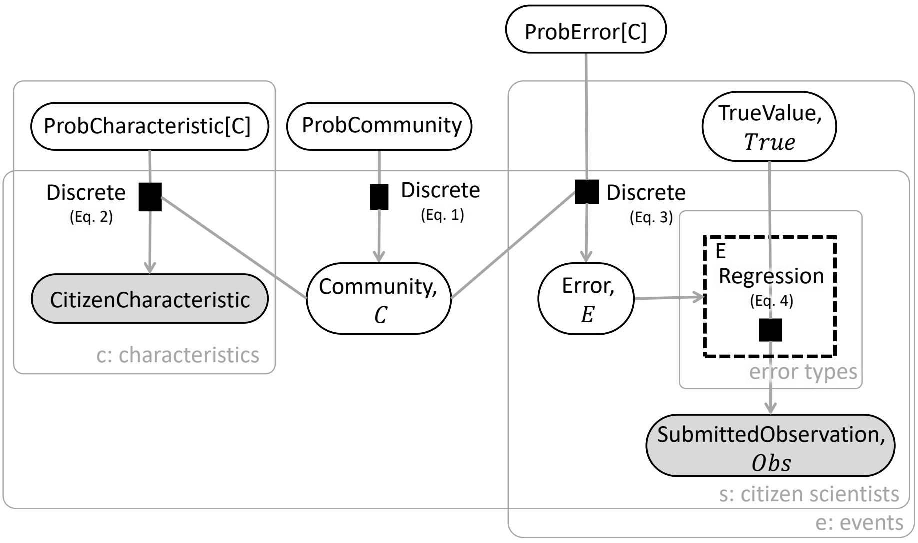

Figure 2. The citizen science error model depicted as a factor graph. A factor node is represented by small filled box. A variable is named in an oval, with shading identifying observed variables. Edges depict parent-child relationships. A gate is represented by a dashed box. Plates are represented by gray rectangles with rounded corners. Symbols adopted from Winn et al. (2020).

4.2.3 Training and testing the model

The inference model was trained and tested to ensure model performance was consistent across different groups of data. During training and testing, the following characteristics were known for each citizen scientist: motivation, recruitment, age, education, place of residence, occupation, gender, performance, and experience. The first seven characteristics were recorded by S4W-Nepal and are explained in Section 3. The last two characteristics, performance and experience, were defined based on the observations submitted by each citizen scientist. Performance is simply the percentage of observations submitted by a citizen scientist that did not require correction. A performance of 90% indicates that 90% of that citizen scientist's submitted observations matched the true value shown in the associated photograph. Experience is a count of how many observations a citizen scientist submitted through the 2018 monsoon season. Performance and experience rates were split into three levels based on the distribution of values.

Splitting the data. Rainfall observations submitted by citizen scientists with known characteristics from 2016 to 2018 were randomly split into a training data set and a testing data set. The training set consisted of 92% of available observations, representing 6,091 observations submitted by 152 citizen scientists. The citizen scientists in the training set submitted anywhere from 1 to 159 observations, with the average number of submissions being 43.5. The testing set consisted of the remaining 8% of available observations, representing 528 observations from 110 citizen scientists. The citizen scientists in the testing set submitted anywhere from 1 to 159 observations, with the average number of submissions being 57.4. Of the 110 citizen scientists in the testing set, 109 were also in the training set. Note that the individual observations were unique between the groups.

Training the model. Before training the model, prior distributions were set for the variables that will be inferred. Uniform prior distributions were set for the citizen characteristics (see Equation A.1), community (see Equation A.2), and error (see Equation A.3). The prior distribution for the true value parameter was a Gaussian distribution with a mean equal to the value of the submitted observation and a large variance (see Equation A.4). The prior distributions for the Gaussian mixture parameters (a, b, and τ) were assigned based on the magnitude of errors reasonably expected for rain gauge observations.

While running the model in the training phase, the characteristics for each citizen scientist, the submitted observations, and the true values were set as observed variables in the model. The community for each citizen scientist, the error type for each submitted observation, the conditional probability tables for each characteristic and error type, and parameters for the Gaussian mixture were inferred (see Equations 2-4 and Figure 2). The training phase provided posterior distributions that were then used while testing the model.

Testing the model. To test the model, prior distributions for unobserved variables were set to the associated posterior distribution calculated by Infer.NET during training. The characteristics for each citizen scientist and the values of the submitted observations were observed. The model inferred the community for each citizen scientist, the probable error type for each observation, and provided a posterior distribution for the true value of the submitted observation. The performance of the model was assessed based on the whether the inferred posterior distribution of true value covered the true value identified in the accompanying photograph submitted by the citizen scientist and whether the mode of the true value posterior matched the actual true value.

5 Results and Discussion

5.1 Number of communities and error types

To select the appropriate number of communities to capture the differences among the citizen scientists, model evidence was used. Model evidence indicates which model best explains the data relative to the model's complexity (MacKay, 2003, p. 343-386). Too many communities may lead to overfitting, whereas too few communities may lead to underfitting. The model evidence automatically makes this trade-off and identifies the optimal number of communities. Model evidence was computed for models with one to ten communities. The number of communities that produced the largest model evidence was selected as the correct number of communities for the model and data. Similarly, model evidence was used to determine how many error types were present in the data. Model evidence was computed for one to twelve error types while using the optimal number of communities. The number of error types that resulted in the largest model evidence was selected as the number of error types for the model and data. After selecting the number of error types, model evidence was again checked to verify that the optimal number of communities remained constant.

Model evidence indicated that there are four communities and seven error types present in the data, given the model structure (see Figure 3). In comparison, S4W-Nepal identified four error types in the data based on visual inspection of the submitted observations. The inference model, however, is a much more powerful tool for uncovering nuances in the data than graphical techniques. Therefore, the number of communities and error types inferred from the model were used for the remaining analysis. The model developed here and model evidence are, together, a powerful tool for identifying distinct error types in quantitative citizen science observations.

5.2 Error analysis

Parameters for the error-specific linear regressions were inferred for the seven error types in the submitted rainfall observations (see Table 1). The inferred parameters included the mean and precision, τ, of the Gaussian distribution, where the mean is based

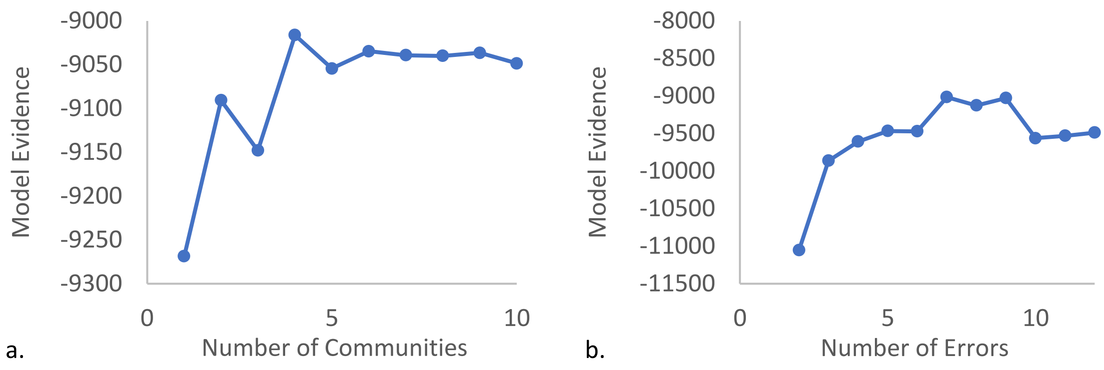

Figure 3. Model Evidence for selecting a. number of communities and b. number of error types present in the data given the model structure.

on a linear regression a, b, and $True$ as shown in Equation 4. Five of the seven error types align well with the error types identified by Davids et al. (2019): none, unit, meniscus, big meniscus, and unknown. Davids et al. (2019) only identified one meniscus error type, but the model separated this type of error into meniscus (b=2.00 mm) and big meniscus (b=3.78 mm) error types. Meniscus errors occur when a citizen scientist reports the top of a concave meniscus rather than the bottom of the meniscus. Unit errors indicate instances where a citizen scientist submitted an observation in units of centimeters rather than millimeters, resulting in a unit error slope, a, of 0.10. Unknown errors do not present a discernible pattern that would explain their origin, as indicated by the low inferred precision (0.01) for this error type.

The inference model identified two error types that were overlooked during the Davids et al. (2019) analysis of errors in the Nepal citizen science data: slope outliers and intercept outliers. Slope outliers signify a case where the citizen scientist's reported observation was approximately ten times greater than the true value evident in the accompanying photograph of the rainfall gauge. Intercept outliers occur when a citizen scientist submits an observation that is about ten millimeters less than the true value identified by S4W-Nepal during quality control. The underlying cause of outlier errors is unclear, but these outliers can likely be attributed to typos (e.g. adding an additional zero) or general carelessness on the part of the citizen scientist. Of the 6,091 observations included in the training data, only 9 were labelled as slope (n=2) or intercept (n=7) outliers.

Table 1. Inferred regression parameters for the different error types

Error Type	Slope, a	Intercept, b	Precision, τ
None	1.0	0.0	55750.5
Unit	0.1	0.1	37.1
Meniscus	1.0	2.0	1906.9
Big Meniscus	1.0	3.8	0.8
Unknown	0.9	2.4	0.00
Slope Outlier	10.3	-0.4	5.0
Intercept Outlier	1.6	-10.4	3.2

5.2.1 Error distribution within communities

The distribution of errors committed by citizen scientists varied depending on the assigned community, as seen in Table 2. Each community was named based on its respective error distribution: Few, Few-MUn, Mensicus, and Unit-MUn. The Few community commits very few errors—only 2.8% of submitted observations are erroneous. Of the erroneous submissions, members in the Few community are most likely to commit small and big meniscus errors (2.0%). The Few-MUn community also commits relatively few errors but does so at a rate of 6.3%. Members of the Few-MUn community are almost equally likely to commit small and big meniscus errors (3.1%) and unknown errors (2.8%). The two remaining communities, Meniscus and Unit-MUn, are much more prone to submitting erroneous rainfall observations. The Meniscus community submits erroneous observations at a rate of 21.4%. These observations are largely erroneous due to citizen scientists reading the meniscus of the water incorrectly (19.3%). Lastly, the Unit-MUn community commits the most errors, with 27.4% of its observations requiring correction. While the Unit-MUn community commits primarily unit errors (10.8%), meniscus (7.2%) and unknown (7.7%) errors still claim a large portion of the erroneous submissions. Members of the Unit-MUn community are prone to committing a wide variety of errors.

The Few community members have a high degree of scientific literacy and generally take great care in submitting their observations. The Few-MUn community members likely also have high scientific literacy but are occasionally careless. Citizen scien-

Table 2. Distribution of errors committed by citizen scientists in each community

Community	None	Unit	Meniscus	Big Meniscus	Unknown	Slope Outlier	Intercept Outlier
Few (0.54)	0.972	0.001	**0.015**	0.005	0.005	0.000	0.001
Few-MUn (0.25)	0.937	0.003	**0.021**	0.010	**0.028**	0.001	0.002
Meniscus (0.16)	0.786	0.006	**0.083**	**0.110**	0.011	0.002	0.001
Unit-MUn (0.05)	0.726	**0.108**	0.038	0.035	0.077	0.003	0.012

Note : The probability of each community is shown in parentheses after the community name. Bold values indicate the most common error type(s) for each community.

tists that were initially error prone but were able to correct their misunderstandings based on the feedback provided by S4W-Nepal could also be assigned to the Few-MUn community. The Meniscus community largely misunderstands how to correctly read the depth of water in the rain gauge. The Unit-MUn community has several misunderstandings that cross multiple error types, therefore leading citizen scientists in this community to commit a random mix of errors.

The distribution of errors within each community is a useful tool not only for selecting which submitted observations might require verification, but also for identifying opportunities to improve the overall accuracy of submitted observations. Citizen science project organizers can use targeted training to help specific communities improve their performance or to maintain their motivation for submitting frequent observations (Budde et al., 2017; Sheppard & Terveen, 2011). For example, S4W-Nepal could occasionally send feedback messages to the meniscus community members reminding them to read the rainfall depth from the bottom of the meniscus. As another example, members in the Few community might positively respond to feedback messages acknowledging their strong record of accurate observations. After receiving such feedback, the Few community might be motivated to continue active participation in the citizen science initiative. Knowing the error structure of observations submitted by different communities can help improve the overall effectiveness of citizen science programs.

5.3 Community composition

The model grouped citizen scientists into four distinct communities with a unique combination of characteristics and probability of committing errors. The Few community is the largest with 54% of citizen scientists in the training group assigned to this community (see Table 2). The Unit-MUn community is the smallest with only 5% of citizen scientists classified into this group. The remaining citizen scientists are grouped into the Few-MUn (25%) and Meniscus (16%) communities. Overall, only 21% of participating citizen scientists are likely to commit errors in more than 6.3% of their submitted observations.

The probability that a citizen scientist will belong to a specific community depends, in part, on the unique characteristics of that citizen scientist. Figure 4 provides the inferred posterior probability that a citizen scientist with a particular characteristic would belong to each community, offering insight into the characteristic composition of each community. Singular characteristics may have a large impact on the tendency of a citizen scientist to commit errors, and therefore to be assigned to a specific community. However, it is also true that any combination of characteristics could contribute to the probability of a citizen scientist being assigned to a community. In some cases, citizen scientists are likely to possess a similar combination of characteristics, which surfaces in the community distributions. For example, Figure 4 indicates that citizen scientists recruited during a random visit, older than 25 years of age, holding less than a bachelor's degree, and with an "other" occupation have a similar community distribution. Twenty percent of the citizen scientists older than 25 years of age were also recruited during a random visit, have less than a bachelor's degree, and have an "other" occupation. While community assignment trends for singular characteristics can be enlightening, the impact of multiple citizen scientists with a similar combination of characteristics must be acknowledged.

Motivation. Citizen scientists motivated by payment are more likely to commit errors than volunteer citizen scientists. This, however, does not necessarily indicate that paying a citizen scientist reduces their accuracy. Most paid citizen scientists (92%) live in rural or semi-urban areas and were recruited through random visits. Conversely, many volunteers were recruited through social media, personal connections, and outreach programs organized at secondary schools and universities. The scientific literacy of paid cit-

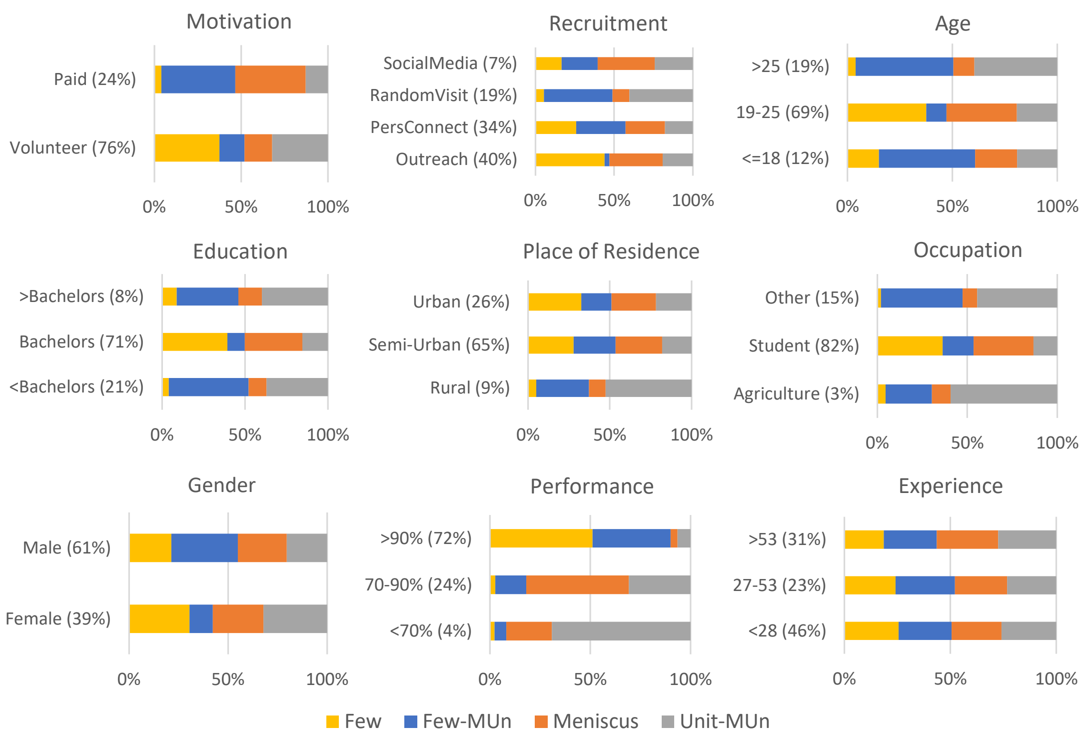

Figure 4. Community composition for each characteristic. The percentage of participating citizen scientists with the associated characteristic is shown in parenthesis.

izen scientists is likely lower than their unpaid counterparts (Davids et al., 2019). Despite this, paid citizen scientists committed fewer random errors (Unit-MUn community) than unpaid citizen scientists.

Recruitment. Citizen scientists recruited via outreach are the most likely to be in the Few community, while those recruited through random visits are almost equally likely to be assigned to the Few-MUn or Unit-MUn communities. Little differentiates the community assignments of citizen scientists recruited via social media or through personal connections. Interestingly, however, those recruited through social media were the least likely to be in the Few or Few-MUn communities, despite being the most active unpaid citizen scientists per Davids et al. (2019). This indicates that a high rate of submitting observations is not necessarily correlated with the accuracy of those observations.

Age. Citizen scientists between the ages of 19 and 25 were the most likely to be in the Few community. The community distributions for those outside of this age range were similar, but those 18 and younger were slightly more likely to be in the Few community and less susceptible to random errors than their counterparts older than 25 years of age. Those older than 25 years are the farthest removed from formal education and are more likely to have many responsibilities, reducing the time and care they can dedicate to collecting and submitting rainfall observations. This trend of less reliable older citizen scientists may be unique to citizen science projects in developing countries. Adult workers in developing countries generally have less leisure time to pursue non-work-related activities than those in developed countries (Jones & Klenow, 2016).

Education. The community distribution for different educational levels largely mirrors the trend seen in the age community distributions, with one exception. Citizen scientists with less than a bachelor's degree are more prone to committing random errors than citizen scientists that are younger than 18 years of age. The community distribution for those with the highest level of education and those with the lowest level of education are almost identical, indicating that education alone does not result in more accurate citizen science observations.

Place of residence. Citizen scientists in urban and semi-urban areas are almost equally likely to be assigned to any of the four communities. However, citizen scientists living in rural areas are much more error prone. Those living in rural areas may have the lowest scientific literacy. Also, only 9% of the participating citizen scientists lived in rural areas, so this small sample may not be representative.

Occupation. Students are equally likely to be in the Few and Mensicus communities. Conversely, citizen scientists with an "other" occupation are equally likely to be in the Few-MUn and Unit-MUn communities. Of the three occupation categories recorded, those in the agriculture sector are the most likely to commit many random errors (Unit-MUn community). However, like those in rural areas, agriculture workers only make up 3% of the citizen scientists involved in the project. This, again, may not be a representative sample.

Gender. Overall, men are less likely to submit erroneous observations than women, with over half of the men being assigned to the Few or Few-MUn communities. However, women are more likely to be in the Few community than men. This trend is an indication that scientifically literate women may take more care than men in submitting observations.

Performance. Unsurprisingly, citizen scientists that submit correct observations more than 90% of the time are most likely to be in the Few or Few-MUn error communities. Citizen scientists with a performance level between 70 and 90% are likely to be in the Meniscus community. The poorest performers (<70%) are generally assigned to the Unit-MUn community.

Experience. No trend is particularly evident in the community distributions for citizen scientists at different experience levels. Citizen scientists with a high participation rate generally have the same likelihood of being in any community as those that submit fewer observations. However, those with a high level of participation are slightly less likely to be in the Few or Few-MUn communities, simply because they have more opportunities to commit errors.

5.4 Testing the model's ability to infer the true value of a submitted observation

In addition to providing insight into the error structure of the submitted observations and the relationship between citizen scientist characteristics and error tendencies, the model provides information about the true value of submitted observations. Testing the model reveals the ability of the model to infer a previously unknown true value based solely on the value of the submitted observation and the characteristics of the citizen scientist. In most cases, the actual true value of the submitted observation falls within the range of the posterior distribution inferred for the true value variable as seen in Figure 5. However, as Figure 5b,c show, the mode of the Infer.NET posterior distribution is not always a good estimate of the actual true value.

To increase the computational efficiency of an inference algorithm that sometimes needs to consider thousands of variables, Infer.NET approximates a multi-mode posterior distribution with a single-mode distribution (Minka et al., 2018) by minimizing the Kullback-Leibler divergence between the two (Minka, 2005). In many applications, this method works very well. However, here, the mixture distribution covers values ranging from 10% (unit error) of the true value up through 1,000% (outlier error) of the true value. Such a wide range of possible true values results in an Infer.NET predicted true value posterior with high variance and a mode that is often shifted left or right of the true value (see Figure 5). Informing the true value prior distribution with the value of the submit-

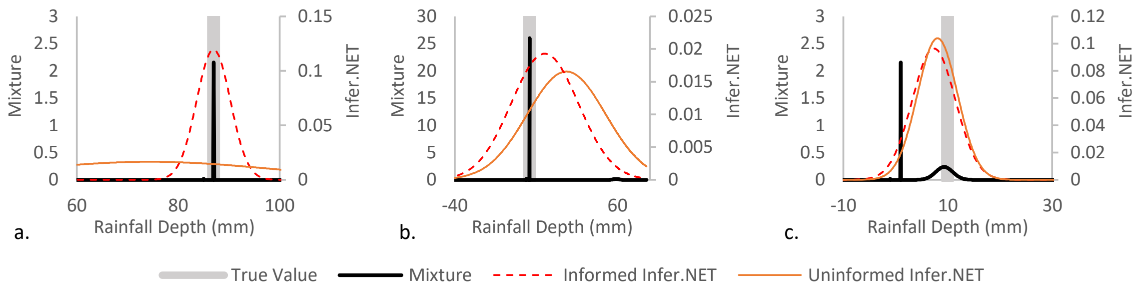

Figure 5. Exact Gaussian mixture distributions for the true value posterior and the Infer.NET-estimated Gaussian distribution of the true value posterior based on informed and uninformed priors; a. Infer.NET correctly infers true value; b. Infer.NET incorrectly infers true value, and Gaussian mixture correctly identifies true value; c. Infer.NET and Gaussian mixture incorrectly identify true value

ted observation rather than using an uninformed prior can shift the mode of Infer.NET's predicted true value posterior toward the actual true value, but this is not always the case (see Figure 5c).

While Infer.NET's predicted true value posterior distribution often does not estimate the actual true value very well, the mode of the exact Gaussian mixture posterior often estimates the actual true value quite well (see Figure 5a,b). In other cases, as shown in Figure 5c, the actual true value is equal to the value at a local peak. Of the error types, the unit error presents the most difficulty when estimating the true value based on the mode of the mixture distribution. This is attributed to the small values of observations submitted with a unit error coupled with the high precision of the no error contribution to the Gaussian mix. For example, Figure 5c depicts an instance where a citizen scientist committed a unit error by submitting an observation of 1 mm while the true value was 10 mm. The inferred posterior error distribution for this observation indicated that there was a 97.3% probability that the submitted observation had a unit error, and a 2.3% probability that it had no error. Despite this discrepancy in error probabilities, the mode of the mixture distribution still presents at the no error value (1 mm), because of the high precision associated with the none error type (see Table 1). The Gaussian mixture posterior distribution calculated from Infer.NET's posterior distributions of the error, regression, and precision parameters provides a more accurate estimate of the true value of a submitted observation than the approximate Gaussian posterior distribution obtained by Infer.NET.

5.5 Further model applications

The trained model was tested for three different unique applications that provide insight into the utility of the model in practical applications and the error structure of citizen science data over time.

5.5.1 Multiple observations of a single event

Analyzing multiple observations of a single rainfall event can improve the accuracy of the predicted true value of rainfall. To determine which submitted observations constituted multiple observations of a single event, k-means clustering was employed (Hadi, Yudistira, Anggraeni, & Hasan, 2018). K-means clustering of the observations submitted on a single day was performed on the dimensions of latitude, longitude, elevation, time of day, and the value of the submitted observation (Hadi et al., 2018). The number of clusters, k, or single events was determined by calculating the Pseudo-F statistic for k values ranging from 1 to 15. Once the single events (clusters) were identified, the true value prior distribution for an event was set to a Gaussian distribution with a mean and variance equal to those of the corresponding cluster. If each observation in a cluster actually refers to the same underlying event, a Gaussian distribution estimated from individual posteriors would provide a reliable true value posterior for that event. Thus, to quantify the uncertainty in predicting the true value of the event, a true value Gaussian distribution was estimated from the true value posterior distributions for individual observations in the event.

K-means clustering determined that the 22 rainfall observations submitted on May 30, 2019 were observations of 9 distinct events (see Table 3). The inferred true value posteriors for each event often failed to cover the range of submitted observations and tended to skew towards the value of one of the submitted observations. Table 3 shows that the number of observations submitted for each event ranged from 1 to 4—likely too few observations to accurately predict the actual true value of the event. The prediction of the true value for most events is highly uncertain, as evidenced by variances up to 132 mm^2 from the Gaussian distributions estimated from the true value posteriors inferred for in-

Table 3. True Value Estimated from Multiple Observations

		Gaussian Estimator	
Event	Size	Mean	Variance
1	3	43.1	132.0
2	4	35.0	71.9
3	1	37.5	2.9
4	2	15.9	1.9
5	4	26.4	92.7
6	2	113.4	43.8
7	1	61.9	1.1
8	2	28.8	20.2
9	3	20.6	67.7

dividual observations. More than 4 observations of a single event are likely needed to narrow down the prediction of the true value of a rainfall event, especially in a region, such as Nepal, where rainfall is highly heterogeneous in space and time.

5.5.2 Citizen scientists with unknown characteristics

As citizen scientist programs expand, recording complete characteristics data for each participating citizen scientist can become challenging. The model's ability to infer the correct community for citizen scientists with unknown characteristics and the correct true value for the observations they submit was investigated. The characteristics for each unknown citizen scientist were drawn from a discrete distribution estimated from the characteristics data of citizen scientists observed during training. The community for each citizen scientist and the true values of their submitted observations were inferred and compared to the communities and true values inferred when the characteristics were known precisely.

The model performed quite well while inferring the community of unknown citizen scientists and the true values of observations submitted by unknown citizen scientists. Known citizen scientist communities were correctly predicted 11.8% more than unknown citizen scientist communities. The coefficient of determination between the actual true values and predicted true values was 0.015 higher for known citizen scientists than for unknown citizen scientists. While the predicted true values for known and unknown citizen scientists were similar, the uncertainty of the true values predicted from observations submitted by unknown citizen scientists was higher. The average variance of the inferred true value posteriors was 68.5 mm^2 for unknown citizen scientists and 53.8 mm^2 for known citizen scientists. Overall, the value of submitted observations has greater influence on the inferred true values of rainfall than the characteristics of the associated citizen scientist. While knowing the characteristics of all citizen scientists increases the accuracy of predicting the true value of submitted observations, it is not essential.

5.5.3 Evolution of error structure within communities

The change in error distribution over time within each community was studied. The observations submitted by citizen scientists with known characteristics were divided into years 2017, 2018, and 2019. The same communities assigned to each citizen scientist during training were assigned, and the a, b, and τ for each error type inferred during training were made static. In addition, a uniform prior was set for the community error distributions to reduce skew in the posterior distribution. Then, the inference model was run to infer the error distribution for each community during each year.

The probability that a citizen scientist in each community would commit a type of error changed from the 2017 to 2018 to 2019 S4W-Nepal program years (see Figure 6). In 2017, only 16 citizen scientists for whom characteristics are known submitted observations (see Table 4). The 2017 community error distributions, particularly the Meniscus and Unit-MUn communities, are highly uncertain due to the small sample size. Overall, citizen scientists became increasingly active as S4W-Nepal's program progressed through the years. Citizen scientists submitted an average of just over 8 observations in 2017, growing to nearly 80 by 2019. In the first full year of rainfall submissions (2017), most citizen scientists were assigned to the Few-MUn community. In the following two years, active citizen scientists were most often in the Few community, followed by the Few-MUn community. In all three years of S4W-Nepal's program, the Unit-MUn community represented the smallest fraction of active citizen scientists.

As S4W-Nepal gained experience in operating a citizen science program, the participating citizen scientists also gained skills in collecting and submitting accurate rain-

Table 4. Yearly Observations and Community Sizes

	2017	2018	2019
	Number of Observations		
Min.	1	1	1
Max.	30	216	409
Average	8.1	46.7	80.0
Std. Dev.	9.6	47.6	93.0
Total	130	6,916	4878
Community	Probability (Count)		
Few	0.31 (5)	0.55 (82)	0.41 (25)
Few-MUn	0.50 (8)	0.23 (24)	0.30 (18)
Meniscus	0.13 (2)	0.17 (25)	0.25 (15)
Unit-MUn	0.06 (1)	0.05 (7)	0.05 (3)

Note : The number of citizen scientists in each community is shown in parentheses.

fall observations. The Few-MUn and Meniscus communities had an increasing probability of submitting correct observations in each year after 2017 (see Figure 6). This trend also holds for the Few and Unit-MUn communities for 2018, but both communities saw a decrease in the probability of submitting correct observations in 2019. As the years progressed, all communities submitted successively fewer meniscus and big meniscus errors. Similarly, unit errors tended to decrease or remain the same as citizen scientists gained experience. Interestingly, while meniscus type errors and unit errors decreased over time, 2019 saw relatively high rates of unknown errors. The reason for an increase in unknown errors is difficult to diagnose but may be due to an evolution in the magnitude of errors committed. For example, if the regression parameters for this analysis are inferred rather than held constant, the unknown error b decreases from 2.6 in 2017 to 1.5 in 2019. The error structure of observations submitted by citizen scientists is evolving as both S4W-Nepal and the participating citizen scientists gain experience, a common trend in citizen science programs (Kosmala et al., 2016).

S4W-Nepal uses various training techniques and feedback methods to increase the scientific literacy of citizen scientists (Davids et al., 2019). Their methods have been effective in reducing the magnitude and frequency of errors committed by the citizen scientists. Perhaps the best evidence for this change is the reduction in meniscus and big meniscus errors committed by citizen scientists in the Meniscus community. From 2018 to 2019, the probability of meniscus or big meniscus errors in the Meniscus community decreased from 17.9 to 4.2%. Similarly, unit errors committed by those in the Unit-MUn community decreased from 10.2% in 2018 to 5.8% in 2019. While a trend in reduced meniscus and unit errors over two years is promising, additional analysis after multiple years of collecting citizen scientist observations would provide more conclusive evidence for increased scientific literacy of the participants.

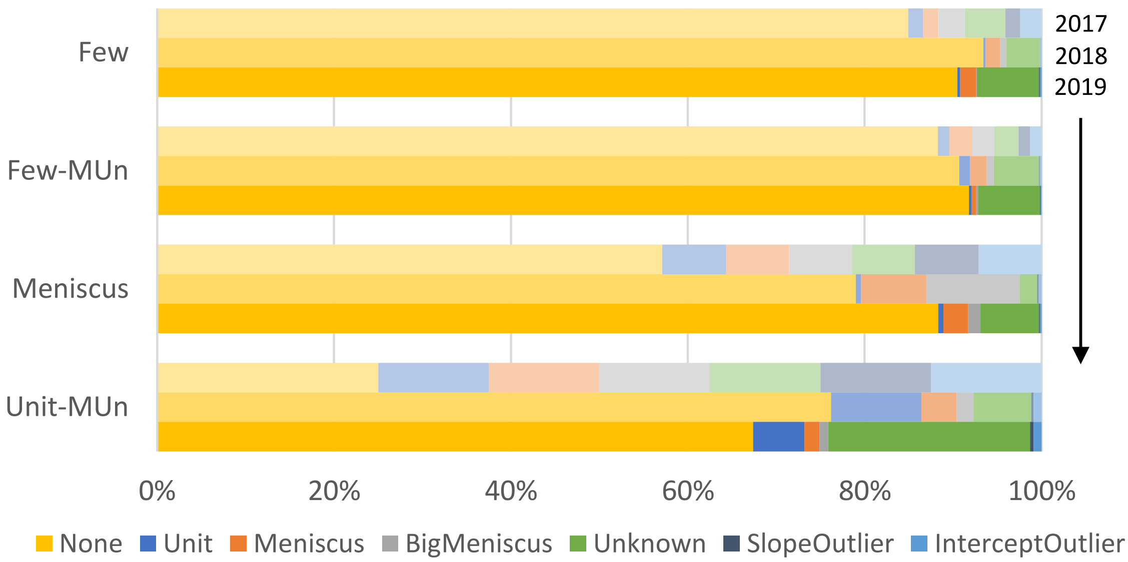

Figure 6. Change in the distribution of errors for each community over time. Note that the 2017 error distributions for the Meniscus and Unit-MUn communities are poorly informed due to the low number of active citizen scientists assigned to those communities.

6 Summary and conclusions

This study developed a Bayesian inference model to investigate the type and frequency of errors present in citizen science data. The model assigns citizen scientists to a community based on the characteristics of the citizen scientist and their tendency to submit erroneous observations. This helps to target manual corrections of CS data. The model then infers a posterior distribution of the true value of a submitted observation from the value of the observation and the community of the participating citizen scientist. Designed thus, the model can be adapted to a wide array of citizen science datasets.

Analysis of the error structure in citizen scientist rainfall observations revealed that individuals fall into one of four error patterns: not error prone, mostly not error prone, meniscus error prone, and random or various error prone. While the Bayesian inference model developed here used communities to relate citizen scientist characteristics to error tendencies, the magnitude and type of errors committed is the crux of every community assignment. The distribution of characteristics within each community is useful for investigating potential reasons for committing errors rather than for identifying individuals who might be particularly error prone.

The Bayesian inference model developed using Infer.NET's software framework uncovered seven error types and their probability distribution within each of the four error-based communities. The community assignments are a useful tool for discerning which citizen scientists are more likely to submit erroneous observations that require further review. In addition, community-specific training and feedback messages could be a powerful tool for increasing the quality and frequency of submissions.

While the Bayesian inference model was unable to regularly predict the true value of a submitted observation, the model did extrapolate useful error probabilities for each observation. These error probabilities, in conjunction with the model's inferred error-specific regression and precision parameters, can be used to calculate a true Gaussian mixture distribution that predicts the true value of submitted observations with more accuracy than Infer.NET's single-mode true value prediction. As citizen science programs expand to include multiple participants submitting observations of a single event, the model's ability to predict the true value for that event will likely increase. However, the model's potential may be limited in regions where the target parameter is highly heterogeneous in space and time.

As a graphical, assumption-based Bayesian inference model, the citizen science error model presented here has immense potential for adaptation to other citizen science programs with diverse data types. The implementation of error-based communities provides a simple, yet effective method for tracking changes in the types and frequency of errors committed by citizen scientists. The communities also provide targeted training and feedback opportunities to improve citizen science data at the point of collection, rather than at the point of correction. Improving the quality of citizen science data at every step enables increasingly more citizen scientist-supported decision-making and discoveries.

A Prior Distributions

The prior distribution for each inferred model variable was a uniform Dirichlet distribution, with the exception of the true value prior. The prior distribution for true value was a Gaussian distribution with a mean equal to the value of the submitted observation and the variance set to 600. The variance for the true value prior was selected based on the variance of the entire true value dataset.

$$ProbCharachteristic_c[C] \sim Dirichlet(Uniform), \tag{A.1}$$

$$ProbCommunity \sim Dirichlet(Uniform), \tag{A.2}$$

$$ProbError[C] \sim Dirichlet(Uniform), \tag{A.3}$$

$$True_e \sim \mathcal{N}(Obs_{s,e}, 600), \tag{A.4}$$

Notation

$\boldsymbol{Dirichlet}$ Dirichlet distribution

$\boldsymbol{Discrete}$ Discrete distribution

$\boldsymbol{\mathcal{N}}$ Gaussian distribution

$\boldsymbol{c}$ characteristic

$\boldsymbol{s}$ citizen scientist

$\boldsymbol{E}$ error type

$\boldsymbol{e}$ event

$\boldsymbol{C}$ Community

$\boldsymbol{E}$ Error type

Obs SubmittedObservation

True TrueValue

Acknowledgments

The dataset analyzed for this study can be accessed in the Supplementary Material published by Davids et al. (2019). This research has been supported by the National Science Foundation, Division of Graduate Education (grant no. DGE-1333468) and the Dutch Research Council. Data collection and quality control was supported by the Swedish International Development Agency (grant no. 2016-05801) and by SmartPhones4Water (S4W). The authors declare that they have no conflict of interest. The authors would like to thank S4W's Saujan Maka for instrumental guidance.

References

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., … Frusher, S. (2014, May). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, *173*, 144–154. Retrieved 2020-05-02, from `https://linkinghub.elsevier.com/retrieve/pii/S0006320713002693` doi: 10.1016/j.biocon.2013.07.037

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009, December). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, *59*(11), 977–984. Retrieved 2020-05-02, from `https://academic.oup.com/bioscience/article-lookup/doi/10.1525/bio.2009.59.11.9` doi: 10.1525/bio.2009.59.11.9

Bonter, D. N., & Cooper, C. B. (2012, August). Data validation in citizen science: a case study from Project FeederWatch. *Frontiers in Ecology and the Environment*, *10*(6), 305–307. Retrieved 2020-05-03, from `http://doi.wiley.com/10.1890/110273` doi: 10.1890/110273

Budde, M., Schankin, A., Hoffmann, J., Danz, M., Riedel, T., & Beigl, M. (2017, September). Participatory Sensing or Participatory Nonsense?: Mitigating the Effect of Human Error on Data Quality in Citizen Science. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1–23. Retrieved 2020-05-03, from https://dl.acm.org/doi/10.1145/3131900 doi: 10.1145/3131900

Crall, A. W., Newman, G. J., Stohlgren, T. J., Holfelder, K. A., Graham, J., & Waller, D. M. (2011, December). Assessing citizen science data quality: an invasive species case study: Assessing citizen science data quality. *Conservation Letters*, *4*(6), 433–442. Retrieved 2020-05-02, from http://doi.wiley.com/10.1111/j.1755-263X.2011.00196.x doi: 10.1111/j.1755-263X.2011.00196.x

Davids, J. C., Devkota, N., Pandey, A., Prajapati, R., Ertis, B. A., Rutten, M. M., ... van de Giesen, N. (2019, March). Soda Bottle Science—Citizen Science Monsoon Precipitation Monitoring in Nepal. *Frontiers in Earth Science*, *7*, 46. Retrieved 2020-04-23, from https://www.frontiersin.org/article/10.3389/feart.2019.00046/full doi: 10.3389/feart.2019.00046

Delaney, D. G., Sperling, C. D., Adams, C. S., & Leung, B. (2008, January). Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions*, *10*(1), 117–128. Retrieved 2020-05-03, from http://link.springer.com/10.1007/s10530-007-9114-0 doi: 10.1007/s10530-007-9114-0

Habib, E., Krajewski, W. F., & Ciach, G. J. (2001). Estimation of rainfall interstation correlation. *Journal of Hydrometeorology*, *2*, 621–629. doi: 10.1175/1525-7541(2001)002⟨0621:EORIC⟩2.0.CO;2

Hadi, A. F., Yudistira, I., Anggraeni, D., & Hasan, M. (2018, June). The Geographical Clustering of The Rainfall Stations on Seasonal GSTAR Modeling for Rainfall Forecasting. *Journal of Physics: Conference Series*, *1028*, 012238. Retrieved 2020-05-03, from https://iopscience.iop.org/article/10.1088/1742-6596/1028/1/012238 doi: 10.1088/1742-6596/1028/1/012238

Hunter, J., Alabri, A., & van Ingen, C. (2013, February). Assessing the quality and trustworthiness of citizen science data. *Concurrency and Computation: Practice and Experience*, *25*(4), 454–466. Retrieved 2020-05-03, from http://doi.wiley.com/10.1002/cpe.2923 doi: 10.1002/cpe.2923

Jones, C. I., & Klenow, P. J. (2016, September). Beyond GDP? Welfare across Countries and Time. *American Economic Review*, *106*(9), 2426–2457. Retrieved 2020-04-26, from http://pubs.aeaweb.org/doi/10.1257/

`aer.20110236` doi: 10.1257/aer.20110236

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017, January). So, How Much of the Earth's Surface *Is* Covered by Rain Gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78. Retrieved 2020-05-02, from `http://journals.ametsoc.org/doi/10.1175/BAMS-D-14-00283.1` doi: 10.1175/BAMS-D-14-00283.1

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016, December). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, *14*(10), 551–560. Retrieved 2020-05-03, from `http://doi.wiley.com/10.1002/fee.1436` doi: 10.1002/fee.1436

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms.* Cambridge: Cambridge University Press.

Minka, T. (2005). *Divergence measures and message passing* (Technical Report No. TR-2005-173). Microsoft Research.

Minka, T., & Winn, J. (2008). Gates. *Advances in Neural Information Processing Systems 21*, 1073–1080.

Minka, T., Winn, J., Guiver, J., Zaykov, Y., Fabian, D., & Bronskill, J. (2018). *Infer.NET 0.3.* Microsoft Research Cambridge. Retrieved from `http://dotnet.github.io/infer`

Nayava, J. L. (1974, December). Heavy monsoon rainfall in Nepal. *Weather*, *29*(12), 443–450. Retrieved 2020-04-23, from `http://doi.wiley.com/10.1002/j.1477-8696.1974.tb03299.x` doi: 10.1002/j.1477-8696.1974.tb03299.x

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., & Crowston, K. (2012, August). The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, *10*(6), 298–304. Retrieved 2020-05-02, from `http://doi.wiley.com/10.1890/110294` doi: 10.1890/110294

Riesch, H., & Potter, C. (2014, January). Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science*, *23*(1), 107–120. Retrieved 2020-05-02, from `http://journals.sagepub.com/doi/10.1177/0963662513497324` doi: 10.1177/0963662513497324

Sheppard, S. A., & Terveen, L. (2011). Quality is a verb: the operationalization of data quality in a citizen science community. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration - WikiSym '11* (p. 29). Mountain View, California: ACM Press. Retrieved 2020-05-03, from http://dl.acm.org/citation.cfm?doid=2038558.2038565 doi: 10.1145/2038558.2038565

Teague, K. A., & Gallicchio, N. (2017). *The evolution of meteorology: a look into the past, present, and future of weather forecasting.* Hoboken, NJ: John Wiley & Sons, Inc.

Thapa, B. R., Ishidaira, H., Pandey, V. P., & Shakya, N. M. (2017, February). A multi-model approach for analyzing water balance dynamics in Kathmandu Valley, Nepal. *Journal of Hydrology: Regional Studies*, *9*, 149–162. Retrieved 2020-04-23, from https://linkinghub.elsevier.com/retrieve/pii/S2214581816303342 doi: 10.1016/j.ejrh.2016.12.080

USGS. (n.d.). *DYFI Scientific Background.* Retrieved 2020-05-05, from https://earthquake.usgs.gov/data/dyfi/background.php

Venanzi, M., Guiver, J., Kazai, G., Kohli, P., & Shokouhi, M. (2014). Community-based bayesian aggregation models for crowdsourcing. In *Proceedings of the 23rd international conference on World wide web - WWW '14* (pp. 155–164). Seoul, Korea: ACM Press. Retrieved 2020-05-03, from http://dl.acm.org/citation.cfm?doid=2566486.2567989 doi: 10.1145/2566486.2567989

Vibhāga, N. K. T. (2012). *National Population and Housing Census 2011: National report* (Vol. 1). Government of Nepal, National Planning Commission Secretariat, Central ....

Wiggins, A., Newman, G., Stevenson, R. D., & Crowston, K. (2011, December). Mechanisms for Data Quality and Validation in Citizen Science. In *2011 IEEE Seventh International Conference on e-Science Workshops* (pp. 14–19). Stockholm, Sweden: IEEE. Retrieved 2020-05-03, from http://ieeexplore.ieee.org/document/6130725/ doi: 10.1109/eScienceW.2011.27

Winn, J., Bishop, C., Diethe, T., Guiver, J., & Zaykov, Y. (2020). *Model-based machine learning* (early access ed.). online: Microsoft Research. Retrieved from www.mbmlbook.com