# Who smells? Forecasting taste and odor in a drinking water reservoir with phytoplankton genus data

*Michael J. Kehoe*[*1], Kwok P. Chun[2], Dan Conrad[3], Helen M. Baulch[1]*

1. School of Environment and Sustainability, Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

2. Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

3. Buffalo Pound Water Treatment Plant, Buffalo Pound, Moose Jaw, Saskatchewan, Canada

KEYWORDS taste and odor, forecasting, random forest, water treatment

## ABSTRACT

Taste and odor problems can impede public trust in drinking water, and impose major costs on water utilities. The ability to forecast taste and odor events in source waters, in advance, is shown for the first time in this paper. This could allow water utilities to adapt treatment, and where effective treatment is not available, consumers could be warned. A unique 24-year time series, from an important drinking water reservoir in Saskatchewan, Canada, is used to develop forecasting models of odor using physical, chemical, and biotic predictors. We demonstrate, using linear regression and random forest models, that odor events can be forecast at 0-26 week time lags, and that the models are able to capture a significant increase in odor threshold number in the mid-1990s. Models with a fortnight time-lag show high predictive capacity ($R^2 = 0.71$ for random forest; 0.52 for linear regression). Predictive skill declines for time lags from 0 to 15 weeks, then increases again, to $R^2$ values of XX (random forest) and XX (linear regression) at a 26-week lag. Results of the random forest model demonstrate that phytoplankton taxonomic data outperform chlorophyll a in terms of predictive importance. .

## INTRODUCTION

Few people want to drink smelly water that tastes funny. It is therefore unsurprising that taste and odor compounds in drinking water have significant social and economic effects (Vaughn, 1967). Smell and taste are the primary ways people assess the quality and safety of their drinking water, and as such, the occurrence of taste and odor compounds (TOCs) in treated water can erode public confidence in drinking water safety (McGuire, 1995). Furthermore, TOC's may reflect anthropogenic degradation of water quality (Watson, 2004). While no national estimate of the cumulative costs of taste and odor issues exists, it is estimated that consumers shifting to bottled water, associated with TOCs, could cost the US economy more than $813 million annually (Dodds et al, 2008), and the cost associated with treatment of these problems could be much higher (Srinivasan and Sorial 2011). In Buffalo Pound Lake, a drinking water reservoir in Saskatchewan, Canada, removal of TOCs is estimated to cost $$/annum, or X% of the annual treatment costs (REF). Besides affecting drinking water, TOCs also spoil the taste of fish – creating major issues in the aquaculture industry (Tucker, 2000).

Taste and odor compounds are produced by a number of different phytoplankton and bacteria genera. Cyanobacteria and actinobacteria produce two of the most problematic TOCs: geosmin (trans-1, 10-dimethyl-trans-9-decalol) and MIB (2-methylisoborneoyl). Both compounds are recalcitrant to common treatment options (Srinivasan and Sorial, 2011) and can be detected at extremely low concentrations (Suffet et al, 1999). Amongst the cyanobacteria, only filamentous genera have been found to produce geosmin and MIB (Juttner & Watson, 2007). Like

cyanobacteria, only a subset of actinobacteria produce geosmin and MIB (Zaitlan & Watson, 2006). Actinobacteria are commonly assumed to contribute to aquatic taste and odor via runoff from soils into surface waters. However, actinobacteria can also function within aquatic environments and, for example, can produce MIB by using phytoplankton as a carbon source (Sugiura et al, 1994). Diatoms and chrysophytes can also produce taste and odor compounds. In their case, the TOCs are produced as a result of enzymatic degradation of polyunsaturated fatty acids via bacteria, when the algae die (Watson, 2002). These TOCs are more easily degraded than geosmin and MIB and as a result, tend to be a lesser issue for drinking water treatment. As with cyanobacteria and actinomycetes, only particular species of diatoms and chrysophytes are associated with TOCs. One notable species of diatom producing TOCs is *Cyclotella* -- a common constituent of spring blooms in temperate lakes which can produce sulfur-based TOC's. Beyond these microbial sources there are a number of other compounds which can cause taste and odor issues. These include pesticides and other pollutants, and chemicals used in treatment (Young et al, 1996).

The development of models for predicting or forecasting taste and odor events has been hampered by a lack of long-term time series. To date most studies have used linear regression models that contain common parameters associated with phytoplankton productivity (e.g., chlorophyll a, turbidity/water transparency, and total phosphorus) to predict concentrations of taste and odor compounds (Smith et al 2002, Mau et al 2004, Sugiura et al 2004, Christensen et al 2006, Dzialowski et al 2009). This linear modelling approach has been extended to non-linear models, which include a broader range of parameters, including microbial abundance data (Parinet et al 2010, Parinet et al 2013). Most recently, non-linear models have been developed

that include detailed measurements of hydrodynamics and phytoplankton data with a view to incorporation in hydrodynamic models (Bruder et al 2014). However, all of these models are based on short-term datasets which can make assessment of model performance and relationships among variables difficult.

In this paper, we develop, for the first time, forecasting models of odor in a drinking water source. The largest yet published time series of odor dynamics is used to calibrate and validate random forest and linear regression models. The modelling objective was to predict odor threshold number a fortnight in advance as this is a timescale which allows preparation of treatment options and public warning if needed. The time scales upon which odor can be predicted is also assessed. Finally, the uncertainty of the models predictions is also quantified.

## MATERIALS AND METHODS

### Study site and data description

Buffalo Pound Lake (Saskatchewan, Canada) (Figure 1) is a eutrophic reservoir that supplies drinking water to approximately 1/4 of the population of Saskatchewan. The lake is shallow (4 m), narrow (5 km), and long (35 km). Originally a natural lake, the installation of a dam in 1939 means it is best characterized as a reservoir. Motivated in part by persistent issues with taste and

odor, the Buffalo Pound water treatment plant has been monitoring a range of water quality parameters since 1977. Collected weekly, the data includes odor threshold number (OTN) along with standard water quality parameters such as chlorophyll a, total phosphorus, temperature and turbidity (methods summarized in table XX). The data also includes weekly phytoplankton count data identified to the genus level. This afforded an opportunity to identify whether phytoplankton genus abundance provides more skillful predictors than the more commonly used biogeochemical parameters. In this study, we restrict our analysis to periods where complete weekly data for the parameters of interest were available. This meant that possible predictors, nitrate and ammonia, were excluded from the models (due to variation in measurement frequency). Furthermore some phytoplankton data is missing and so some years were excluded due. Nonetheless, despite these shortcomings, 1251 weeks (24 years) of data were left with which to calibrate and validate forecasting models.

**Odor threshold number**

Odor threshold number (OTN) is an indicator of water odor, determined using serial dilutions with odor-free water (ASTM D1292-10). Multiple trained individuals (an odor panel) are asked to report the first dilution at which the odor can be detected. Despite analytical advances that now allow the detection of individual TOC compounds at very low concentrations, OTN remains a common method for determining the magnitude of taste and odor compounds (Rigal et al, 1995). Advantages of this method include low cost, simplicity (with no complex instrumentation requirements), and generality – that is -- all compounds perceptible to the odor panel are

reported, rather than having to test for the tens of compounds that can cause taste and odor (Young et al, 1996). Furthermore, if water is from a common source and proper standardized procedures are followed, variation in human perception is similar to the variation of direct chemical analysis (Bousquet 1983). As noted, the development of odor models has been hampered by short time series, which is related to high cost, relatively recent development of analytical techniques, and methodological changes. As a result, OTN data constitute a valuable long-term source of information on odor problems where records and consistent methodology have been maintained.

Temperature, total phosphorus, turbidity, Chlorophyll a, algal genus.!!!! Hi Helen, just undeleted these so you have a list for the table.

**Model Development**

The dataset was filtered to exclude parameters, and time periods for which weekly data were not available. Due to missing phytoplankton data the periods 1985-1987 and 1993-1995 were omitted from the model. This left 1251 weeks with which to construct the models. The following 9 predictors were then chosen for our model-based analyses: Chl a, turbidity, total phosphorous, and algal count data for the following taxa: *Anabaena* sp., *Aphanizemenon* sp./*Oscillatoria* sp., *Chlorella* sp., *Cyclotella* sp. and *Astrionella* sp. *Aphanizemenon* sp. and *Oscillatoria* sp. data were combined because the data record had them sometimes recorded separately and sometimes together. Linear (regression) and non-linear models (random forests)

were calibrated and validated (90-10% split) on randomized subsets of the total dataset. This calibration/validation process was repeated 1000 times in order to quantify the uncertainty resulting from the choice of calibration and validation dataset. This randomized calibration and validation was conducted at time-lags of 0 to 26 weeks to quantify how the predictive performance of the models varied with the forecast time-lag and determine whether the predictive importance of different variables changed over time. Finally, a student t-test was carried out to test reports of a significant increase in OTN since approximately 1997.

The model construction methodology contained elements which were similar for both the linear regression and random forest models, as well as some which were different. In what follows the linear regression and random forest models are described, then the general procedure used to calibrate, validate and measure model performance is explained in detail.

The linear regression model constructed according to equation 1:

$$y(t) = \sum_{i=1}^{n=9} x_i(t)\,\alpha_i + \beta \quad \text{(equation 1)}$$

Where $y(t)$ is the predicted OTN values, $\beta$ is bias, and $\alpha_i$ are the respective regressors of each of the $n$=9 predictor variables $x_i(t)$. Uncertainty in model predictions was calculated at the 95% confidence level. The primary purpose of using a linear model was to provide a baseline against which to compare the non-linear random forest model. The R package 'lm' (R core team, 2014) was the implementation used for all linear regression modelling.

Random forests are a machine learning method which constructs a non-linear function based on the mean response of an ensemble of simpler decision tree models (Breiman, 2001). Specifically, a large number of decision tree models are independently constructed on randomly selected subsets of a training data set. Each of these simple trees is biased towards predicting their own particular training data and make poor predictors of the total dataset. However, when the mean prediction of a large number of these random trees (forest) is calculated they produce good predictions, and are increasingly being used to model and understand environmental systems (Cutler, 1998, Kehoe et al 2010). The great benefit of this approach over other machine learning methods is generality. Increasing model complexity by adding greater numbers of trees does not lead to a model which perfectly fits the training data, rather the models predictive performance tends to asymptote with a general diminishing return in predictive performance as more trees are added (Breiman 2001). This feature means that random forests require no assumptions as to the complexity of the data on the part of the modeler and so are very useful for discovering hidden relationships with data, free from a priori assumptions. Furthermore, thanks to the ensemble approach there is a natural way of estimating predictive uncertainty. Because each tree makes a prediction, estimation of uncertainty is straightforward. Here uncertainty is reported as +/-2σ for each prediction. A further advantage of random forests is that they are able to also provide information on the relative importance of different predictors. This is done by considering how prediction accuracy changes when a given parameter is excluded from the model (here, calculated as the average reduction in mean square error). All random forest models were developed within R statistical software with the 'randomForest' package (Liaw and Weiner, 2002).

178 The same general procedure was followed for the development of both the linear regression and
179 random forest models. For each model type a random 90% subset of the available dataset was
180 chosen to be for calibration with the remaining 10% reserved for validation. Models were then
181 calibrated and validated using procedures for each which are outlines below. The $R^{2:}$

182 $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$ (equation 2)

183

184 then being calculated and recorded. This process was then repeated 1000 times in order to
185 quantify the uncertainty related with the choice of calibration and validation data sets but also to
186 gain insight into the uncertainty in the importance measure of the random forest algorithm.

187

188 **RESULTS**

189 Odor threshold number varied both annually and inter-annually over the data record (Figure 1).
190 In particular, OTN was significantly higher in 1998-2011 compared to 1980-1997 ($p<0.001$), and
191 although the most significant annual peak is in August, April is when OTN first increases after
192 its winter minimum. This coincides with the ice-melt period and the spring phytoplankton
193 bloom.

194 Both random forest and linear regression modeling techniques led to good model fits for the
195 validation data (Figure 2). The random forest was the better of the two at each time lag (based
196 on $R^2$); however, the linear regression model predicts the calibration and validation datasets to
197 similar accuracy (within the 95% confidence levels) at every time lag. The random forest model

198 tended to over-predict the calibration dataset compared to the validation dataset at each time lag
199 (Figure 2). Both models showed greater uncertainty in prediction of validation dataset compared
200 to the prediction of calibration dataset (Figure 2).

201 The random forest model ($R^2$=0.71) was a better predictor on average of the validation dataset
202 compared to the linear regression model ($R^2$ =0.52) at the fortnightly time lag (Figure 2).
203 However, the non-linear random forest model predicts the calibration dataset more accurately
204 than the validation data set. The linear model in comparison predicts both the calibration and
205 validation dataset to a similar level of accuracy (Figure 2). This shows that fair comparison of
206 the predictive ability of different models should only be done on datasets not used in model
207 training or parameter estimation. Also, as noted above, the uncertainty bounds on sensitivity due
208 to choice of calibration data set are small compared to the linear regression model. In addition to
209 fitting the data well, the random forest and linear regression model (not shown) are both able to
210 capture interannual and decadal variation in odor threshold number including a significant
211 increase in the mid 1990's.

212 The least important predictor in the random forest model was almost always chlorophyll a,
213 with the exception of the 0-3 week time lags when total phosphorus and *Astrionella* sp. were the
214 least important. *Chlorella* sp., *Cyclotella* sp., and *Aphanizemenon* sp. /*Oscillatoria* sp. were
215 consistently better predictors than chlorophyll *a*. *Anabaena* sp. was more important at most time
216 scales except between 11-13 week time-lag where is was similar or equal in predictive
217 importance to chlorophyll a (both approximately 0.1% MSE). Finally *Astrionella* sp. was better
218 at every time lag except from 0-3 weeks where it became progressively worse and is only able to
219 reduce mean square error on average by 0.07% compared to 0.13% for Chlorophyll a at 0 week
220 time lag.

The variation in relative importance of different predictors over time revealed a number of patterns. Temperature had two peaks in predictive importance at 8 weeks (0.3% MSE) and 26 weeks (0.45% MSE). Total phosphorus decreased in importance as time lag declined, from a high of 0.21% MSE at 26 weeks' time lag to a low of 0.09% at 0 weeks' time lag. Except for a small decline from 26 week to 24 week time lag, turbidity, increased in importance as time lag reduced and was the second most important predictor behind *Chlorella* sp. at time lags shorter than 3 weeks. *Anabaena* sp. had peaks in predictive importance at 20 week and 0-5 week time lag. Finally *Aphanizemenon* sp. /*Oscillatoria* sp. (6 weeks) and *Astrionella* sp. (15 weeks) had unimodal peaks in relative importance.

## DISCUSSION

This study reports the first fully validated and operational forecasting model of odor for a drinking water reservoir. In particular, it is first model to investigate predictive performance at different time lags; previous models only predict odor events at the present time. Critically, the model is able to capture a large ten year long increase in odor threshold which begins in 1998; this appears to be related to an increase in abundance of filamentous cyanobacteria and *Chlorella*. In addition, the model captures a recent (2008-2012) decline and plateau in OTN. The ability to predict odor at long time lags, as demonstrated here, suggests that long-term ecological processes are important in driving odor production.

Due to the long time series of weekly OTN data, rigorous calibration and validation of models was possible. The large data set also meant the uncertainty in model predictions and associated statistics, such as $R^2$ and importance, could be evaluated. Previous modelling studies on odor have been based on relatively short time series of common odor causing compounds: geosmin and MIB. As a result, more limited data has been available for model calibration, and model validation has only been rarely performed (see table S1).

Given the complexity of odor production processes (Bruder et al. 2014), it is likely that different predictive models will have to be developed for different systems; however, key insights can be gleaned by comparing model predictors across ecosystems. To date, we have identified 13 models associated with 7 studies that had the explicit goal of prediction of odor, although all, with the exception of this study, are restricted to real-time prediction (i.e. present day or week). Of these studies, 6 of the 13 models included temperature as a predictor variable, 9 of the 13 included a nutrient, and only 3 included chlorophyll (Table S1). The two studies to date using algal taxonomy data (Bruder et al. 2014; this study) had the highest predictive capacity of models reported to date. Results of importance analyses (Figure 4) demonstrated that phytoplankton taxonomic data had far greater utility than measurements of chlorophyll a, and, despite the time-consuming nature of algal counts, where odor prediction, and understanding long-term ecological change are important (as they are in many bloom-affected reservoirs), clear value of taxonomic data is demonstrated.

*Anabaena* sp., *Oscillatoria* sp., *Aphanizemenon* sp. were all important predictors of odor. They are all cyanobacteria, and all have species which have been proven to produce geosmin.

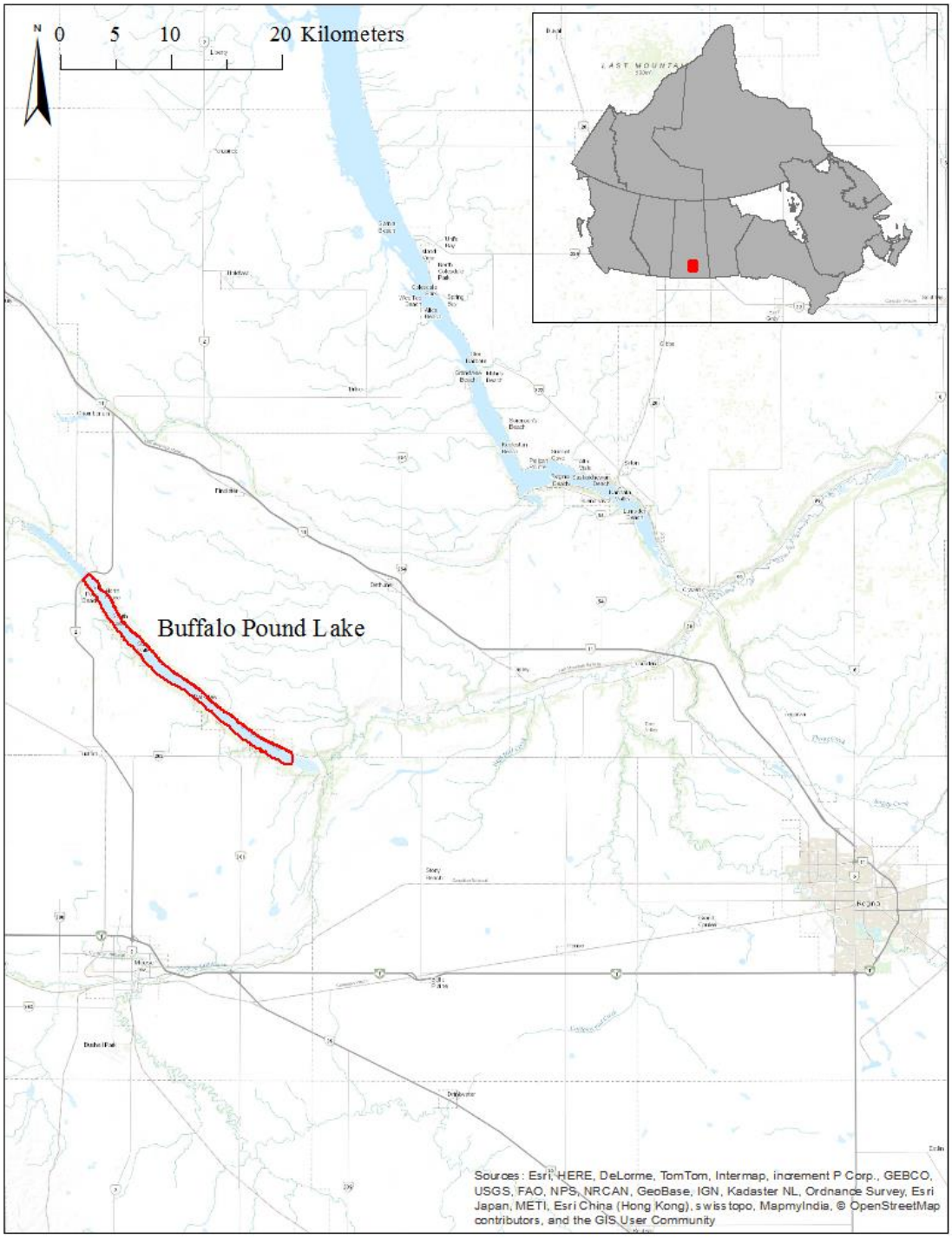
*Astrionella* sp and *Cyclotella* sp. are associated with odorous alkenes which are produced when they decay after death (Watson 2004). However, the dominance of *Chlorella* sp. as a predictor in our modelling results is surprising given it has not been demonstrated to produce odor. We have two hypotheses. The first is simply that Buffalo Pound Lake contains a strain of *Chlorella* sp. which may actively produce TOCs, or lead to their production upon death and decay. The second is that abundance of *Chlorella* sp. is a proxy for actinobacterial activity. It has been shown the actinobacteria are able to use carbon fixed by *Chlorella* sp. to produce taste and odor compounds (Sugiura et al 1994).

Forecasting models, like the one presented here, can help inform planning for water treatment utilities, allowing optimization of treatment, and helping to minimize costs. Addition of activated carbon is the primary way taste and odor problems are managed (REF); however, this process is expensive, and there is the potential for cost savings with advance warning. Prediction also means the public can be warned of possible taste and odor issues ahead of time - increasing confidence in water treatment organizations. Models like the ones presented here allow managers to assess the likelihood and expected magnitude of odor problems well in advance. In particular the approach taken here would allow a manager to start dealing with, and tracking, odor problems 6 months in advance.

283

284 FIGURES

285 **Figure 1.** Map of Buffalo Pound Lake and its location within Canada.




286

287

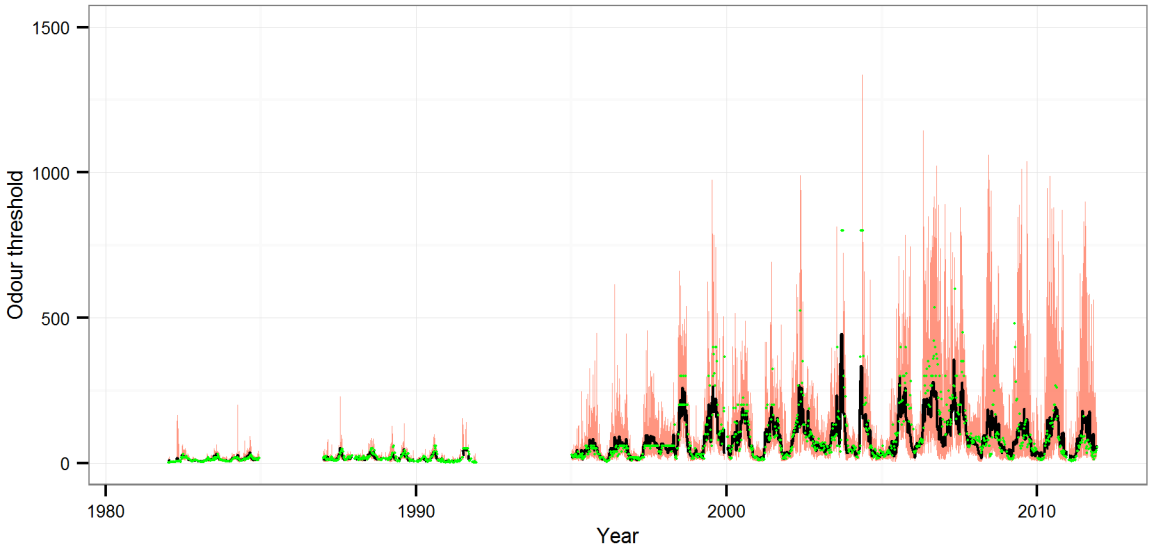
288

289

290 **Figure 2.** Time series plot of odor threshold number data (bright green dots), model predictions

291 (black line) and 95% confidence interval of predictions (coral ribbon).

292




293

294

**Figure 3.** Predictive accuracy as measured by $R^2$ for different time-lags. The light purple ribbon is the random forest model $R^2$ performance on validation data (dashed-line is the mean value, ribbon is ±2σ from mean) while the dark purple ribbon is the random forest model $R^2$ performance on calibration data (solid-line is the mean value, ribbon is ±2σ from mean). The light coral ribbon is the linear model $R^2$ performance on validation data (dashed-line is the mean value, ribbon is ±2σ from mean) while the dark coral ribbon is the linear model $R^2$ performance on calibration data (solid-line is the mean value, ribbon is ±2σ from mean).

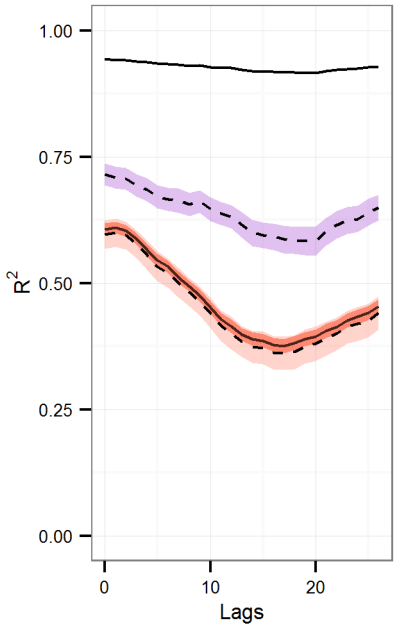
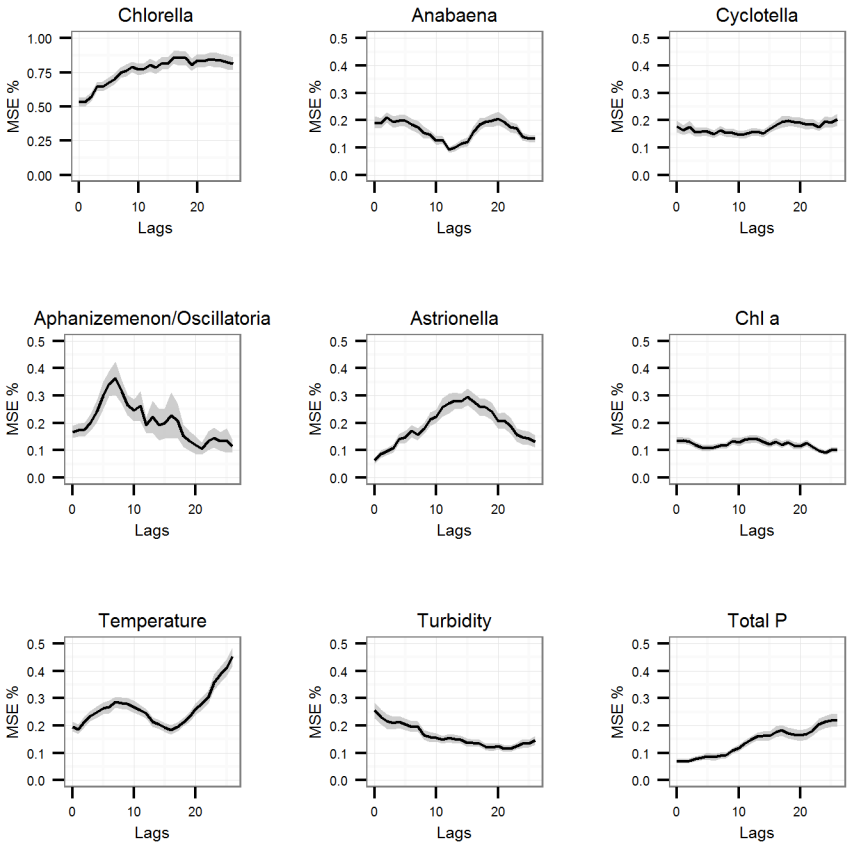


**Figure 4.** Ribbon plots of the relative importance of different predictors at different time-lags. Importance is measured as the mean reduction in mean square error as a result of using the predictor. Model was run 5000 times on randomized selections of calibration validation data set to generate estimates of 95% confidence interval as ± 2σ from mean.



| Study | Response Variable | Prediction Variables | Model type | Calibration $R^2$ (data number) | Validation $R^2$ (data number) |
|---|---|---|---|---|---|
| Christensen 2006 | Geosmin | Turbidity (FNU), specific conductance | Linear | 0.71(18) | NA(NA) |
| Smith 2002 | Geosmin | Chlorophyll *a* | Linear | 0.72(6) | NA(NA) |
| Mau 2004 | Geosmin | Secchi depth, specific conductance, turbidity (NTU) | Linear | 0.70(16) | NA(NA) |
| Sugiura, 2004 | MIB | Water temperature,silicic acid, chemical oxygen demand | Linear | 0.59(64) | 0.26(32) |
| Sugiura 2004 | MIB | Water temperature,silicic acid, chemical oxygen demand | Artificial neural network | 0.65(64) | 0.27(32) |
| Sugiura, 2004 | Geosmin | Total phosphorus, chemical oxygen demand, dissolved oxygen | Linear | 0.45(64) | 0.28(32) |
| Sugiura 2004 | Geosmin | Total phosphorus, chemical oxygen demand, dissolved oxygen | Artificial neural network | 0.49(64) | 0.42(32) |
| Dzialowski 2009 | Geosmin | secchi disk | Linear | 0.24(57) | NA(NA) |
| Dzialowski 2009 | Geosmin | orthophosphate | Linear | 0.25(57) | NA(NA) |
| Bruder 2014 | Geosmin | *Pseudanabaena* spp., diatoms, *Planktrothrix agardhii*, water temperature, salinity, and TKN | Neuro-fuzzy | 0.83(102) | 0.49 (10) |
| Bruder 2014 | MIB | *Pseudanabaena* spp., diatoms, *Planktrothrix agardhii*, water temperature, salinity, and TKN | Neuro-fuzzy | 0.82(102) | 0.70 (7) –three data points removed from validation set |
| **This study** | **Odor threshold number** | **Water temperature, total phosphorus, Chlorophyll *a*, *Anabaena* sp., Aphanizemenon sp. /Oscillatoria sp., *Chlorella* sp., *Astrionella* sp., *Cyclotella* sp.,** | **Linear** | **0.54 (875)** | **0.52 (376)** |
| **This study** | **Odor threshold number** | **Water temperature, total phosphorus, Chlorophyll *a*, *Anabaena* sp.,** | **Random forest** | **0.94 (875)** | **0.71 (376)** |

| | | | | | |
|---|---|---|---|---|---|
| | | **Aphanizemenon sp. /Oscillatoria sp., *Chlorella* sp., *Astrionella* sp., *Cyclotella* sp.,** | | | |

311

312

313

314

315

316 TABLES. Each table must have a brief (one phrase or sentence) title that describes its contents.
317 The title should follow the format "**Table 1.** Table Title" (Word Style "VD_Table_Title"). The
318 title should be understandable without reference to the text. Put details in footnotes, not in the
319 title (use Word Style "FE_Table_Footnote"). Define nonstandard abbreviations in footnotes.

320 Use tables (Word Style "TC_Table_Body") when the data cannot be presented clearly as
321 narrative, when many precise numbers must be presented, or when more meaningful
322 interrelationships can be conveyed by the tabular format. Do not use Word Style
323 "TC_Table_Body" for tables containing artwork. Tables should supplement, not duplicate, text
324 and figures. Tables should be simple and concise. It is preferable to use the Table Tool in your
325 word-processing package, placing one entry per cell, to generate tables.

326 **Displayed equations** can be inserted where desired making sure they are assigned Word Style
327 "Normal". Displayed equations can only be one column wide. If the artwork needs to be two
328 columns wide, it must be relabeled as a figure, chart, or scheme and mentioned as such in the
329 text.

## ASSOCIATED CONTENT

(Word Style "TE_Supporting_Information"). **Supporting Information**. A brief statement in non-sentence format listing the contents of material supplied as Supporting Information should be included, ending with "This material is available free of charge via the Internet at http://pubs.acs.org." For instructions on what should be included in the Supporting Information, as well as how to prepare this material for publication, refer to the journal's Instructions for Authors.

## AUTHOR INFORMATION

### Corresponding Author

*(Word Style "FA_Corresponding_Author_Footnote"). * (Word Style "FA_Corresponding_Author_Footnote"). Give contact information for the author(s) to whom correspondence should be addressed.

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally. (match statement to author names with a symbol)

### Funding Sources

Any funds used to support the research of the manuscript should be placed here (per journal style).

### Notes

Any additional relevant notes should be placed here.

ACKNOWLEDGMENT

(Word Style “TD_Acknowledgments”). Generally the last paragraph of the paper is the place to acknowledge people, organizations, and financing (you may state grant numbers and sponsors here). Follow the journal’s guidelines on what to include in the Acknowledgments section.

ABBREVIATIONS

CCR2, CC chemokine receptor 2; CCL2, CC chemokine ligand 2; CCR5, CC chemokine receptor 5; TLC, thin layer chromatography.

REFERENCES

References

1. ASTM D1292-10 Standard Test Method for Odor in Water, In Anonymous ; ASTM International: West Conshohocken, PA, .

2. Blevins, W.; Schrader, K.; Saadoun, I. Comparative physiology of geosmin production by< i> Streptomyces halstedii</i> and< i> anabaena</i> sp. Water Science and Technology 1995, 31 (11), 127-133.

3. Bousquet, G.; Ouvrard, J.; Rigal, S.; Vilagines, R. Statistical evaluation of the blind test method for water quality control. Water Science & Technology 1983, 15 (6-7), 35-46.

4. Bousquet, G.; Ouvrard, J.; Rigal, S.; Vilagines, R. Statistical evaluation of the blind test method for water quality control. Water Science & Technology 1983, 15 (6-7), 35-46.

5. Bruder, S.; Babbar-Sebens, M.; Tedesco, L.; Soyeux, E. Use of fuzzy logic models for prediction of taste and odor compounds in algal bloom-affected inland water bodies. Environ. Monit. Assess. 2014, 186 (3), 1525-1545.

6. Christensen, V.G.; Graham, J.L.; Milligan, C.R.; Pope, L.M.; Ziegler, A.C. Water Quality and Relation to Taste-and-odor Compounds in the North Fork Ninnescah River and Cheney Reservoir, South-central Kansas, 1997–2003. United States Geological Survey Scientific Investigations Report 2006–5095 2006, 5095, 1-43.

7. Dodds, W.K.; Bouska, W.W.; Eitzmann, J.L.; Pilger, T.J.; Pitts, K.L.; Riley, A.J.; Schloesser, J.T.; Thornbrugh, D.J. Eutrophication of US freshwaters: analysis of potential economic damages. Environ. Sci. Technol. 2008, 43 (1), 12-19.

8. Dzialowski, A.R.; Smith, V.H.; Huggins, D.G.; deNoyelles, F.; Lim, N.; Baker, D.S.; Beury, J.H. Development of predictive models for geosmin-related taste and odor in Kansas, USA, drinking water reservoirs. Water Res. 2009, 43 (11), 2829-2840.

9. Giglio, S.; Chou, W.; Ikeda, H.; Cane, D.; Monis, P. Biosynthesis of 2-methylisoborneol in cyanobacteria. Environ. Sci. Technol. 2010, 45 (3), 992-998.

10. Huisman, J.; Matthijs, H.C.; Visser, P.M. Harmful cyanobacteria. Springer: 2005; Vol. 3, .

11. Jardine, C.G.; Gibson, N.; Hrudey, S.E. Detection of odour and health risk perception of drinking water. Water Science and Technology 1999, 40 (6), 91-98.

12. Juttner, F. and Watson, S.B. Biochemical and ecological control of geosmin and 2-methylisoborneol in source waters. Appl. Environ. Microbiol. 2007, 73 (14), 4395-4406; 10.1128/AEM.02250-06.

13. Kadota, H. and Ishida, Y. Production of volatile sulfur compounds by microorganisms. Annual Reviews in Microbiology 1972, 26 (1), 127-138.

14. Liaw, A. and Wiener, M. Classification and regression by randomForest. R news 2002, 2 (3), 18-22.

15. Mau, D.; Ziegler, A.; Porter, S.; Pope LM. Surface-water-quality Conditions and Relation to Taste-and-odor Occurrences in the Lake Olathe Watershed, Northeast Kansas, 2000–02 United States Geological Survey Scientific Investigations Report 2004–5047 2004, 5047, 1-95.

16. Naes, H.; Utkilen, H.; Post, A. Factors Influencing Geosmin Production by the Cyanobacterium< i> Oscillatoria brevis</i>. Water Science & Technology 1988, 20 (8-9), 125-131.

17. Naes, H. and Post, A.F. Transient states of geosmin, pigments, carbohydrates and proteins in continuous cultures of Oscillatoria brevis induced by changes in nitrogen supply. Arch. Microbiol. 1988, 150 (4), 333-337.

18. Parinet, J.; Rodriguez, M.J.; Sérodes, J. Influence of water quality on the presence of off-flavour compounds (geosmin and 2-methylisoborneol). Water Res. 2010, 44 (20), 5847-5856.

19. Parinet, J.; Rodriguez, M.J.; Sérodes, J. Modelling geosmin concentrations in three sources of raw water in Quebec, Canada. Environ. Monit. Assess. 2013, 185 (1), 95-111.

20. Rashash, D.; Dietrich, A.; Hoehn, R.; Parker, B. The influence of growth conditions on odor-compound production by two chrysophytes and two cyanobacteria. Water Science and Technology 1995, 31 (11), 165-172.

21. Rigal, S. Odour and flavour in waters: quantitative method for a new European standard. Water Science and Technology 1995, 31 (11), 237-242.

412 22. Saadoun, I.M.; Schrader, K.K.; Blevins, W.T. Environmental and nutritional factors affecting
413 geosmin synthesis by< i> Anabaena</i> SP. Water Res. 2001, 35 (5), 1209-1218.

414 23. Schrader, K.K. and Blevins, W.T. Geosmin-producing species of Streptomyces and Lyngbya
415 from aquaculture ponds. Can. J. Microbiol. 1993, 39 (9), 834-840.

416 24. Smith, V.H.; Sieber-Denlinger, J.; deNoyelles Jr, F.; Campbell, S.; Pan, S.; Randtke, S.J.;
417 Blain, G.T.; Strasser, V.A. Managing taste and odor problems in a eutrophic drinking water
418 reservoir. Lake Reserv. Manage. 2002, 18 (4), 319-323.

419 25. Spaulding, C.H. Quantitative Determination of Odor in Water. Am.J.Pub.Health 1931, 21,
420 1038.

421 26. Srinivasan, R. and Sorial, G.A. Treatment of taste and odor causing compounds 2-methyl
422 isoborneol and geosmin in drinking water: A critical review. Journal of Environmental Sciences
423 2011, 23 (1), 1-13.

424 27. Suffet, I.H.(.; Khiari, D.; Bruchet, A. The drinking water taste and odor wheel for the
425 millennium: beyond geosmin and 2-methylisoborneol. Water Science and Technology 1999, 40
426 (6), 1-13.

427 28. Sugiura, N.; Inamori, Y.; Hosaka, Y.; Sudo, R.; Takahashi, G. Algae enhancing musty odor
428 production by actinomycetes in Lake Kasumigaura. Hydrobiologia 1994, 288 (1), 57-64.

429 29. Sugiura, N.; Utsumi, M.; Wei, B.; Iwami, N.; Okano, K.; Kawauchi, Y.; Maekawa, T.
430 Assessment for the complicated occurrence of nuisance odours from phytoplankton and
431 environmental factors in a eutrophic lake. Lakes & Reservoirs: Research & Management 2004, 9
432 (3-4), 195-201.

30. Team, R.C. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012. Open access available at: http://cran.r-project.org 2014, .

31. Tucker, C.S. Off-flavor problems in aquaculture. Rev. Fish. Sci. 2000, 8 (1), 45-88.

32. Uwins, H.; Teasdale, P.; Stratton, H. A case study investigating the occurrence of geosmin and 2-methylisoborneol (MIB) in the surface waters of the Hinze Dam, Gold Coast, Australia. Water Science & Technology 2007, 55 (5), 231-238.

33. Vaughn, J.C. Tastes and Odors in Water Supplies. Environ. Sci. Technol. 1967, 1 (9), 703-709.

34. Watson, S.B. Aquatic taste and odor: a primary signal of drinking-water integrity. Journal of Toxicology and Environmental Health, Part A 2004, 67 (20-22), 1779-1795.

35. Watson, S.B. Cyanobacterial and eukaryotic algal odour compounds: signals or by-products? A review of their biological activity. Phycologia 2003, 42 (4), 332-350.

36. Watson, S.B.; Ridal, J.; Boyer, G.L. Taste and odour and cyanobacterial toxins: impairment, prediction, and management in the Great Lakes. Can. J. Fish. Aquat. Sci. 2008, 65 (8), 1779-1796.

37. Xie, Y.; He, J.; Huang, J.; Zhang, J.; Yu, Z. Determination of 2-methylisoborneol and geosmin produced by streptomyces sp. And Anabaena PCC7120. J. Agric. Food Chem. 2007, 55 (17), 6823-6828.

38. Young, W.; Horth, H.; Crane, R.; Ogden, T.; Arnott, M. Taste and odour threshold concentrations of potential potable water contaminants. Water Res. 1996, 30 (2), 331-340.

453 39. Zaitlin, B. and Watson, S.B. Actinomycetes in relation to taste and odour in drinking water:
454 Myths, tenets and truths. Water Res. 2006, 40 (9), 1741-1753

455