Journal of Hydrology 583 (2020) 124522

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

The use of support vectors from support vector machines for hydrometeorologic monitoring network analyses

W.H. Asquith[1]

U.S. Geological Survey, 8802 Urbana, Lubbock, TX, United States

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Shreedhar Maskey, Associate Editor

2000 MSC:
62-07
62H11
62J99
62P12
86A05

Keywords:
Network analyses
Support vector machine
Generalized additive model
Groundwater levels
Peak streamflow
Texas

ABSTRACT

Hydrometeorologic monitoring networks are ubiquitous in contemporary earth-system science. Network stakeholders often inquire about the importance of sites and their locations when discussing funding and monitoring design. Support vector machines (SVMs) can be useful by their assigning each monitoring site as either a support or nonsupport vector. A potentiometric surface was created from synthetic data and 800 random observation locations (sites) as an analog to a groundwater-level network. Using generalized additive models for potentiometric surface prediction, simulations show that a subsample of support vectors from the 800 sites will out perform random samples of sample size equaling the support vector count. Support vector percentages from simulation quantify the recurrence that SVMs assign each site as a support vector, and these percentages in turn measure site importance. An example application of support vector percentages identifies important monitoring sites needed to regionalize the 0.1 annual exceedance probability peak streamflow. The results indicate that 152 of 283 streamgages with support vector percentages equalling 100 percent have not operated since about 2000 and generally have much smaller drainage areas than the greater streamgage network in Texas. The drainage area disparity is an indication of historical imbalance in peak streamflow data acquisition from various stream sizes in Texas.

1. Introduction

Hydrometeorologic monitoring (observation) networks are ubiquitous in contemporary earth-system science. The networks usually collect water-resources data such as groundwater levels, surface-water streamflow, and precipitation. Networks collect data in multi-dimensional domains of covariates including space (horizontal and vertical dimensions) along with time. Somewhat less obvious covariates exist. For example, groundwater level responses in wells are affected by the given hydrogeologic framework (aquifer geometry and properties), type of well completion (construction), local and regional pumping histories, and contexts of seasonal recharge and discharge. Streamflow responses, conversely, are produced by precipitation inputs on a watershed characterized by a drainage area, channel slope, land-use patterns, soil horizons, and geologic substrata.

One of many science objectives of hydrometeorologic networks, especially surface water and meteorology, is the collection of data from which computed statistics (e.g. annual minimum water levels, 0.1 and 0.01 annual exceedance probability peak streamflows, or mean annual precipitation) can be used with predictor variables (covariates) to make estimates at unmonitored locations. Statistical techniques used range across a broad class of prediction methods from multi-linear regression (e.g. Asquith and Roussel, 2009; Williams–Sether, 2015) to sophisticated machine learning (e.g. Carlisle et al., 2016).

1.1. Reasons for network analyses

Network analyses include understanding the information content produced by a monitoring network and how the network has evolved over time. Such understanding constitutes just a part of overall network analysis and evaluation. Network analyses are important not only for scientific and statistical purposes but are crucial to the various stakeholders investing funds in and (or) planning the design of a given network.

Groundwater monitoring networks have been widely studied and remain of great interest to researchers and local stakeholders (Andricevic, 1989, 1990; Esquivel et al., 2015; Fisher, 2013; Kollat et al., 2011; Van Greer et al., 1991; Wood, 2004). Precipitation networks (e.g. Putthividhya and Tanaka, 2012) also have been widely studied and concepts such as entropy and information theory are useful (e.g. Chen et al., 2007).

E-mail address: wasquith@usgs.gov.
[1] Research Hydrologist, U.S. Geological Survey.

https://doi.org/10.1016/j.jhydrol.2019.124522
Received 29 April 2019; Received in revised form 16 October 2019; Accepted 26 December 2019
Available online 14 January 2020
0022-1694/ Published by Elsevier B.V.

Streamgage networks at regional, state, and national scales also have been widely studied (Chebbi et al., 2017; Kiang et al., 2013; Markus et al., 2003; Medina, 1987; National Research Council, 2004). In particular, Markus et al. (2003) assessed information transfer amongst the streamflows in the context of various hydrologic regimes (data regimes): low, average, and high flow conditions. The study also demonstrated an additional complexity in network analyses—often the characteristic (statistic) of interest for statistical transfer requires selection, which means that the statistic under study can influence network analysis itself.

Stakeholder interests generally have scientific or statistical concerns about networks, which include questions such as "Is the (our/their) network designed to collect …" (1) data sufficiently in space? (2) data sufficiently in time (sampling intervals)? (3) data in a cost-effective, scientifically-justified manner? and (or) (4) data at a price of some information gaps? Further interests could include asking whether redundant data are being collected.

Challenges in quantitative answers to these questions further predicate what type of objective function is to be maximized (or minimized) and whether a universal quantitative objective for all stakeholders even exists. Some issues to consider are financial constraints or prediction uncertainties of statistical estimates projected for unmonitored locations of interest.

1.2. Study purpose, design, and organization

This study investigates a novel use of support vector machines (SVMs) (Bishop, 2006; Steinwart and Christmann, 2008) to support hydrometeorologic monitoring network analyses. A hypothetical observational groundwater-well network and statistical methods are used to predict a known 2-dimensional potentiometric surface. The study is oriented around simulations to pinpoint the most and least informative sites in a network through their identification as support vectors by the SVM. Two specific objectives are (1) to show that sites in a network identified as SVM support vectors contain more information than random sites in the network and (2) to show that large prediction differences from SVMs and regression-like methods could be used to identify information gaps in the network.

Mathematics and implementation overview of statistical methods are presented in Section 2 and include GAMs (Section 2.1) and SVMs (Section 2.2). The primary data for this study are synthetic and are described in Section 3. A 1-dimensional prediction example in Section 4 communicates several geometrical nuances of these methods sets up a more thorough study of surface estimation (Section 5) using the synthetic data. Section 5.1 formally frames a question as to how support vectors could be useful for network analyses. Experiments with GAMs and SVMs using all the data are conducted (Section 5.2), which are followed by study of support vector subsamples (Section 5.3). Further experiments based on simulation of SVMs are presented in Section 5.4 followed by closing discussion (Section 5.5). An example application (Section 6) for a surface-water network analysis in Texas identifies sites particularly important for the estimation of the 0.1 annual exceedance probability peak streamflow, and lastly, conclusions are made in Section 7.

Computations were made in the R language (v.3.6.1; x86_64-apple-darwin15.6.0 [64-bit]) (R Development Core Team, 2019), principle external packages are cited, and Supplemental Information accompanying this paper provides the source code. Specific functions or arguments that are part of the source code are in a `monospaced` typeface. Effort to promote reproducibility of results is made through seed setting on pseudo-random number generators, but the numerical incongruities could remain across computer platforms and R as well as R package versions because of a dependence herein on simulation.

2. Methods for statistical prediction

Surface estimation methods often include multi-linear regression (Faraway, 2005; Faraway, 2006), inverse-distance weighting (Brunsdon Comber, 2015; Davis, 2002), geostatistical methods of kriging and its variants (Chebbi et al., 2017; Davis, 2002; Olea, 1999; Tonkin and Larson, 2002), and machine learning (Kuhn and Johnson, 2016).

In this study, GAMs (Stasinopoulos et al., 2017; Wood, 2017) were chosen for spatial prediction because they offer sufficient flexibility, are fast, and are readily extendable into higher covariate dimensions. Because they are analogous to regression, GAMs, in contrast to SVMs, offer familiar diagnostics including adjusted R-squared, residual standard errors, standard errors of prediction, and p-values. SVMs were also used in a regression mode. A general feature of SVMs is that certain observations are retained-by-the-model whereas other observations are not. The in-and-out binary classification of individual data points by SVMs is of interest to this study.

2.1. Generalized additive models (GAMs)

GAMs are flexible and capable of mimicking complex and curvilinear patterns in data (Harwell and Tibshirani, 1990; Hastie et al., 2008; Wood, 2017). GAMs model a response variable using an additive combination of various parametric terms and smooth terms (smooth functions) of predictor variables. The incorporation of smooth functions is an advantage of a GAM over simpler multi-linear, least-squares regression and similar methods because appropriately configured smooth functions can adapt to nonlinear relations.

A general form of a GAM is

$$z_i = \beta_o + \mathbf{X}_i\Theta + f(q_i) + \cdots + \epsilon_i, \tag{1}$$

where z_i is the response variable that is the ith observation, β_o is an intercept, $\mathbf{X}_i$ is a vector for strictly parametric and suitably transformed predictor variables, Θ is a one-column parameter matrix of length equaling the number of parametric predictors, f is a smooth function that has arguments estimated automatically for the predictor variable q_i, the $\cdots$ (three dots) represent additional smooth terms as needed, and ϵ_i are errors (residuals) taken as independent and identically distributed (i.i.d) with zero mean. A distribution family, such as Gaussian, binomial, or Poisson, can be chosen as needed for the model of these errors. Specifically for this study, z_i is an observation of the synthetic data for a given location.

The $\mathbf{X}_i\Theta$ term is the multi-linear parametric regression (Faraway, 2005) component of a GAM and is not considered further in this study. A GAM is fit, by the defaults provided in Wood (2019), using "penalized iterative [reweighted] least squares (PIRLS)" (p. 180–182 Wood, 2017) and a generalized cross validation (GCV) (a type of leave-one-out) score for the smoothness controls on the smooth functions (p. 171 Wood, 2017).

For this study, the GAM formulation is

$$z_i = \beta_o + f(x_i, y_i) + \epsilon_i, \tag{2}$$

where z_i again is the response variable for the ith observation, β_o is an intercept, x_i is an easting horizontal coordinate, y_i, is a northing horizontal coordinate, and $f(x_i, y_i)$ is a 2-dimensional smooth. The `gam()` function from Wood (2019) was used for this study, the error distribution family was Gaussian (normal) (default), and the $f(x_i, y_i)$ used a thin-plate regression spline. GAMs by their nature have some capacity to reasonably extrapolate somewhat away from the data in contrast to SVMs.

2.2. Support vector machines (SVMs)

SVMs are a type of machine-learning approach to prediction in which complex linear combinations of specific data points (the support vectors) and attendant weights are used to define a hyperplane through the data (entire data set). SVMs (Kuhn and Johnson, 2016; Hastie et al., 2008; Steinwart and Christmann, 2008), like GAMs, also are flexible and capable of mimicking curvilinear patterns in the data, but mathematically SVMs are very different.

SVM description is best given as an analogy to linear regression, although SVMs are commonly thought of as binary classifiers. SVMs

also have a history of use in novelty detection (outlier identification) (Clifton et al., 2014; Pimentel et al., 2014). SVMs are also a type of robust regression in which squared residuals are not used for fitting as they are in least squares regression (Faraway, 2005). As a result, data points with large residuals have limited effects on the fitted SVM (Kuhn and Johnson, 2016 p. 153). A curious and distinct (identifying) feature of SVMs is that the data points residing near predictions—residing in the "prediction tube"—by the SVM will actually have no effect. No effect means that the weights for some data points are zero, and such observations are not required to "support" the model. Some control on how large residuals can be with no effect on the SVM is influenced by a so-called "epsilon" setting. It is this either in-and-out binary classification that is critical to this study because the binary distinction is a simple and approachable concept to describe to stakeholders with various disciplinary backgrounds.

The fit phase of an SVM is based on minimization of an error function subject to a penalty as the model progresses towards overfitting. Following the description from Kuhn and Johnson (2016), the coefficients (β_j) of a fitted SVM minimize the f_o objective function for a model having p number of predictors:

$$f_o(y_i;\mathscr{C},\mathscr{L}_\epsilon) = \mathscr{C}\sum_{i=1}^{n}\mathscr{L}_\epsilon(z_i-\hat{z}_i;\epsilon^{\text{svm}}) + \sum_{j=1}^{p}\beta_j^2, \tag{3}$$

where $\mathscr{C}$ is a penalty (cost) parameter that is responsible for penalizing large residuals ($\epsilon_i = z_i - \hat{z}_i$) for the n sample of z_i and predictions . The $\mathscr{C}$ is a multiplicative factor on an $\mathscr{L}_\epsilon$ error function (a basis function) to increase or decrease the importance of errors. The $\mathscr{L}_\epsilon$ can be a number of so-called kernels including linear kernel, polynomial kernel, Gaussian radial basis functions (RBFs), and others. The RBF represents a normalized Euclidean distance metric (Bishop, 2006 chap. 6) and RBF use is prevalent.

The $\mathscr{L}_\epsilon$ has an epsilon setting (ϵ^{svm}) in SVM implementation (Karatzoglou et al., 2018) that controls the width of the prediction tube, which influences the inclusion of data near the SVM hyperplane (the regression). Like the method of least squares (Faraway, 2005), large ϵ_i increase the objective function being used, but in contrast to least squares, large β_j also act to increase the objective function.

In the regression analogy and given that vector of predictors $X_i = (x_{1i}, x_{2i}, \cdots, x_{pi})$ for the i having size n, SVM mathematics can be shown. Using real-world examples, the vector of predictors could represent (1) the coordinates of a groundwater well and current monthly rainfall for the month of the water-level measurement or (2) the drainage area, channel slope, mean annual precipitation, and percent grassland area of a watershed for the ith mean annual streamflow observation. An estimate (prediction) for a sample z_i uses a linear combination of model parameters (coefficients, β_j) and is applied to the new sample of the p predictors (u_j) for $j \in (1, 2, \cdots, p)$. A prediction ($\hat{z}$) can be written with the intercept term (β_0) as

$$\begin{aligned}\hat{z} &= \beta_0 + \beta_1 u_1 + \cdots + \beta_p u_p \\ &= \beta_0 + \sum_{j=1}^{p}\beta_j u_j.\end{aligned} \tag{4}$$

The β_j parameters are chosen such that the sum of squared errors (SSE) is minimized:

$$\text{SSE} = \sum_{i=1}^{n}(\hat{z}_i - z_i)^2. \tag{5}$$

The linear SVM uses these same mathematics when used for estimation, but the β_j estimates importantly can be written as functions of a set of additional, but unknown, hyperparameters (α_i) and the sample data of size n, denoted as x_{ij} so that

$$\begin{aligned}\hat{z} &= \beta_0 + \beta_1 u_1 + \cdots + \beta_p u_p = \beta_0 + \sum_{j=1}^{p}\beta_j u_j = \beta_0 + \sum_{j=1}^{p}\sum_{i=1}^{n}\alpha_i x_{ij} u_j \\ &= \beta_0 + \sum_{i=1}^{n}\alpha_i\left(\sum_{j=1}^{p} x_{ij} u_j\right).\end{aligned} \tag{6}$$

The final form shows weighted linear combinations of training data and the new explanatory data used to make predictions. The matrix algebra can be rewrite the training data and the new explanatory data as a linear "kernel function" (p. 155, Kuhn and Johnson, 2016). With notational adjustment, the function itself can be replaced with a nonlinear kernel that extends the SVM into nonlinear regression.

There are as many α_i parameters as there are data points. Compared to conventional regression, the SVM is highly over parameterized as judged by the tenet of conventional regression that the model should have vastly fewer parameters than data points (p. 154 Kuhn and Johnson, 2016). But the penalty or cost used for SVM fit compensates for this undesirable situation, and in practice many of the α_i have a value of zero. Those observations (x_i and y_i) with $\alpha_i \neq 0$ support the model (the "support vectors"), which means that these and only these observations are required during the SVM's prediction phase. Geometrically, the values closest to the predictions have weights $\alpha_i = 0$, and these observations are plausibly less informative. To clarify a point of common confusion, the support vectors are classified during the fit or training phase of the SVM and are not a decision made by the human user.

Thus, the SVM is known as a sparse model. The origin of the sparsity of SVMs is succinctly stated by Bishop (2006): "The hyperplane is defined by the locations of the support vectors. Other data points (the nonsupport vectors) can be moved freely without changing the decision boundary." In regression, this means that the solution of the SVM is independent of these other data points (nonsupport vectors) and is thus formed by those observations that are in a sense furthest away from central tendencies of the covariates conditioned on the value of the response variable.

In contrast with GAMs, SVMs might be less utilitarian as extrapolation increases away from the training data when given new data on which to make predictions. As extrapolation is encountered, the SVM regresses to the global mean of the training data. The `ksvm()` function from Karatzoglou et al. (2018) and further described by Karatzoglou et al. (2006) was used for this study with a default radial basis kernel as Gaussian (normal). Eight kernels are provided by Karatzoglou et al. (2018), but users could provide their own. The sensitivity results to the choice of kernel were not assessed in this study, and the default settings of $\mathscr{C} = 1$ and $\epsilon^{\text{svm}} = 0.1$ were used unless otherwise stated.

3. Synthetic data as an analog to a groundwater-level network

Synthetic data were created for this study to investigate the ability of SVMs to estimate a 2-dimensional potentiometric surface of a water table in an unconfined aquifer. The synthetic data are conceptually consistent with that data collected from a common groundwater-level network. Statistical estimation of groundwater levels at unmonitored locations in time and space could accelerate development of numerical simulation models and assist the change-in-paradigm thinking proposed by White (2017).

The experimental surface $Z(X, Y)$ for spatial locations X and Y is defined as

$$\begin{aligned}z_i(x_i, y_i) = &+20\sin(+1.75\pi(x_i - y_i)/100) + 20\cos(-1.50\pi(y_i)/100) \\ &+ 0.001x_i^2 - 0.1y_i + \mathscr{N}(i;0, 2),\end{aligned} \tag{7}$$

where z_i is a water level for x_i as an easting (horizontal) coordinate and y_i as a northing (vertical) coordinate for the ith observation. The $\mathscr{N}(i;\mu = 0, \sigma = 2)$ is Gaussian (normally) distributed noise with a mean (μ) of zero and a standard deviation (σ) of 2, which was arbitrarily chosen. It is desirable for the surface to have sufficient complexity to

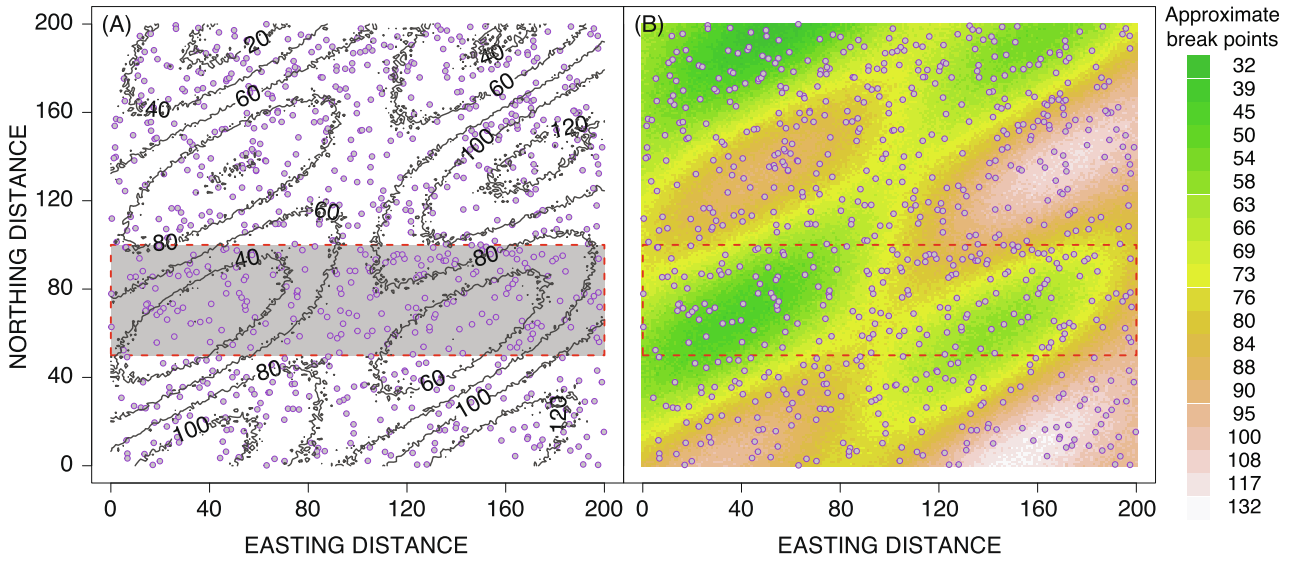

Fig. 1. A 10-unit contour interval (A) and terrain-colored raster (B) representation of the synthetic 200 × 200 gridded potentiometric surface along with the 800 locations (purple circles) to be used for later statistical estimation. The outlined subregion is used for analysis described in Section 4 (Figs. 2–4).

stress statistical prediction. The noise component is added to provide some mimicry of real-world conditions.

In practical applications, the continuous surface is unknown and is instead observed at discrete and irregular points. For this study, 800 random observation locations were generated using a seed on the random number generator set to 62 (`set.seed(62)`, an arbitrary value) prior to prediction of the X coordinates, the Y coordinates, and the Z from the (X, Y). Declaration of the seed is to foster some reproducibility of results with the code base provided in the Supplementary materials that accompany this paper.

Depictions of the potentiometric surface are made by the `contour()` (Fig. 1A) and `image()` (Fig. 1B) functions by a unit incremented grid of $(i, j) \in 1, 2, \cdots, 200$ to become $x_i = i$ and $y_j = j$ (easting and northing, respectively). No statistical prediction is intended to be represented in the figure—just simple visualization. The raggedness of the contour lines represents the added noise component. The highest contours are near the lower right, whereas the lowest contours are near the upper left in both figures. The 800 locations from random values for the X and Y coordinates and then the associated $Z(X, Y)$ become a set of "observational data" for this study where the locations are analogs to groundwater wells. The X, Y, and Z are used to construct and use predictive statistical models described in Sections 4 and 5.

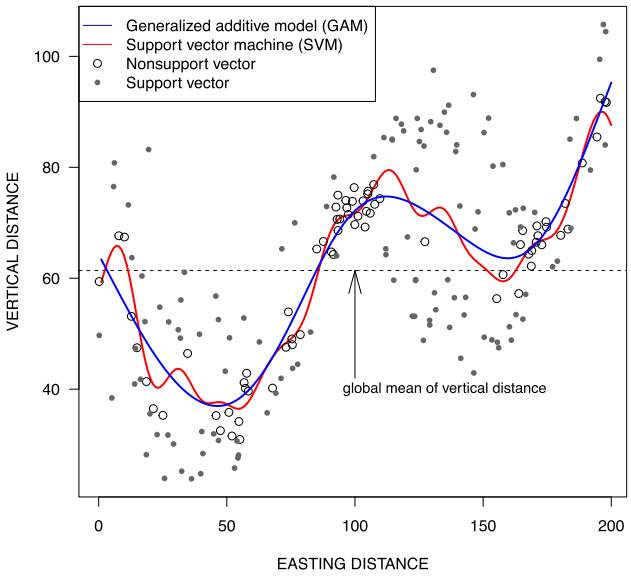

Fig. 2. GAM and SVM model predictions in the a northing coordinate partition between 50 and 100 (Fig. 1) along with support vectors, nonsupport vectors, and global mean.

4. A one-dimensional example of GAM and SVM geometry

To assist in understanding what support vectors are, it is useful to compare and contrast SVM and GAM performance in a 1-dimensional problem (one predictor variable). To this end, a left to right partition in Fig. 1 was extracted between 50 and 100 in the northing (vertical) coordinates. The extracted data are shown in Fig. 2.

The GAM uses all of the 800 locations shown, whereas the SVM uses just the support vectors as shown although all 800 locations are fed to the SVM algorithms. Distinction is made (Fig. 2) between the support vectors (filled circles) and nonsupport vectors (open circles). The nonsupport vectors are the data within the prediction tube defined by the ϵ^{svm} parameter (Karatzoglou et al., 2018), which was set at `epsilon = 0.3` (nondefault).

The overall pattern and local curvature of the GAM and SVM predictions are similar (Fig. 2). The GAM generally shows smoother predictions than the SVM. However, the little hooks (bends) in the SVM solutions at the extreme far left and extreme far right are related to the SVM becoming adherent to local information and then swinging back towards the global mean of the data shown.

To further demonstrate potential extrapolation weakness in SVMs, the data (Fig. 2) were then subsetted (Fig. 3) to create a large information gap between the easting range of 100 to 150. The GAM reasonably extrapolates through the gap, but the SVM swings toward the global mean of the data. (Close scrutiny of the support and nonsupport vectors between Figs. 2 and 3 shows reclassification by the SVM of some data points.)

In the context of monitoring network analysis, one topic of interest is information gap detection. Because the SVM must regress (swing) towards the mean as extrapolation is progressively encountered and other methods, such as the GAM, behave differently, a range of the largest of absolute differences in predictions between a GAM and an SVM could be used. For example, consider a study domain of gaged and ungaged (unmonitored) streamflow sites, large GAM and SVM differences predicted for either site type could forecast information gaps in the network. Stakeholders might then choose to add new monitoring or reactivate discontinued sites within those areas with gaps.

SVMs, at least as implemented by Karatzoglou et al. (2018), have an inherent probabilistic nature. This is demonstrated by one more

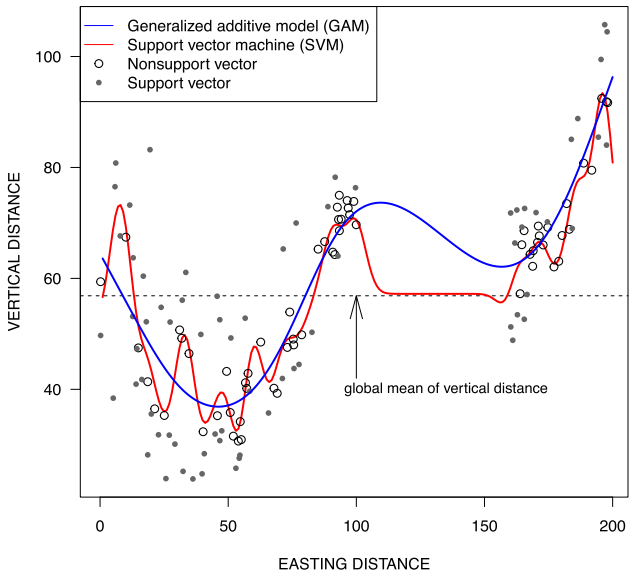

Fig. 3. GAM and SVM fits to data from Fig. 2 subsetted to demonstrate potential extrapolation weakness in an SVM compared to a GAM across a data gap, support vectors, nonsupport vectors, and global mean.

1-dimensional experiment. The SVM fit (Fig. 3 [red line] and Fig. 4 [green line]) and 300 repeated SVM fits (Fig. 4 [transparent red lines]) to the data show a range of outcomes. Modestly different solutions do occur and are attributable in part to different combinations of support and nonsupport vectors, but some trajectories seem more favorable than others. An SVM, as it extrapolates beyond the data, swings toward the global mean (especially seen in Fig. 3). A given trajectory is influenced by the pseudo-random number generator seed in force before the SVM is fit by the `ksvm()` function from Karatzoglou et al. (2018).

5. Measuring site importance using support vectors

5.1. Site importance for network analyses

Using functions from Karatzoglou et al. (2018), support and nonsupport vectors can change between successive SVM fits to the same

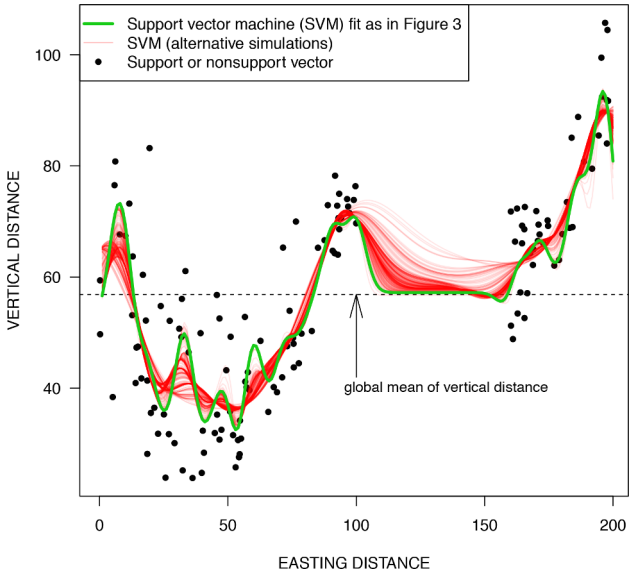

Fig. 4. SVM fit from Fig. 3 shown with 300 repeated SVM simulations to the same data but showing various outcomes with no symbol distinction made between support vectors and nonsupport vectors.

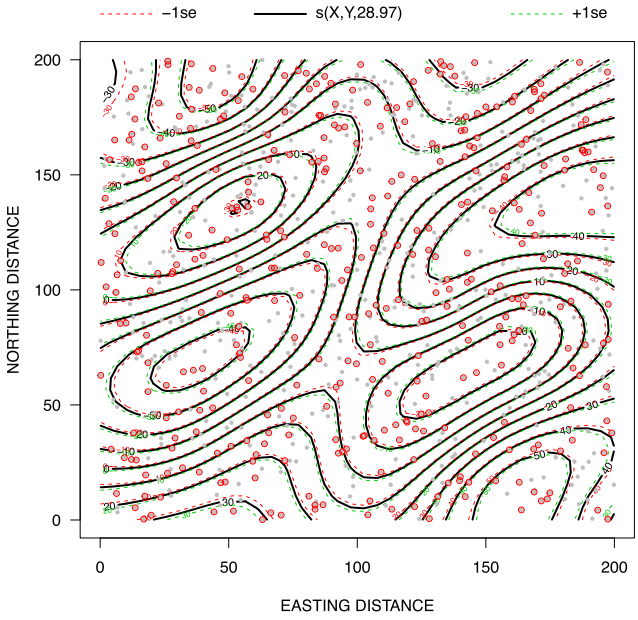

Fig. 5. Output of Wood (2019) model plotting (`mgcv::plot.gam()`) for a GAM using a smooth on easting and northing for the 800 locations (grey circles) with support vectors of the SVM (red circles) for estimation of the surface including the smooth estimate ("s(X,Y,28.97)") (10-unit contour interval) and ± standard error ("se") estimates.

dataset. A hypothesis can now be stated based on support vectors containing relatively more information than nonsupport vectors in the form of a question: Does a sample of size n identified by the support vectors or their frequency of identification, contain more information than a random sample of size n drawn from the observational network containing m sites? Or alternatively, are support vectors, though nonrandom, a type of super subsample when subsetted from the original data?

5.2. Whole model perspective

Using the synthetic data (Section 3), a GAM and SVM (seed set to 1) were fit to the 800 locations. Not accounting for the computed intercept for the GAM as 800, the $f(x_i, y_i)$ smooth for the GAM by the `plot.gam()` function from Wood (2019) is shown in Fig. 5. The standard errors of the contours also are shown. The figure clearly mimics some of the hills and valleys in the data (Fig. 1). Out of the 800 locations, there are the 369 SVM support vectors shown as red circles. A comparison between the GAM and SVM predictions at the 800 locations (Fig. 6) shows general overall agreement.

The Nash–Sutcliffe model efficiency (NSE) and root-mean-square error (RMSE) statistics of the entire sample (whole model, no cross validation) are listed in Table 1. The GAM apparently is not affected prior to fitting by the seed of the random number generator. The SVM has systematically lower RMSE than the GAM, which is indicative of its settings (defaults for `ksvm()`) fitting tighter than the GAM to the data. Of interest is the changing number of support vectors for the SVM, which ranged from 245 to 369 for the four seeds.

5.3. Using support vectors and nonsupport vectors as subsamples

To understand the importance of support vectors as subsamples from the original data sample, it is useful to fit GAMs to a sample comprised of the locations identified as support vectors and then again using the complement sample of the nonsupport vectors. For evaluation, predictions throughout the 200 × 200 grid (Fig. 1B) were used with the known $Z(X, Y)$ to compute 40, 000 residuals. Because the 800 locations realistically do not exist at grid intersections, the NSEs and

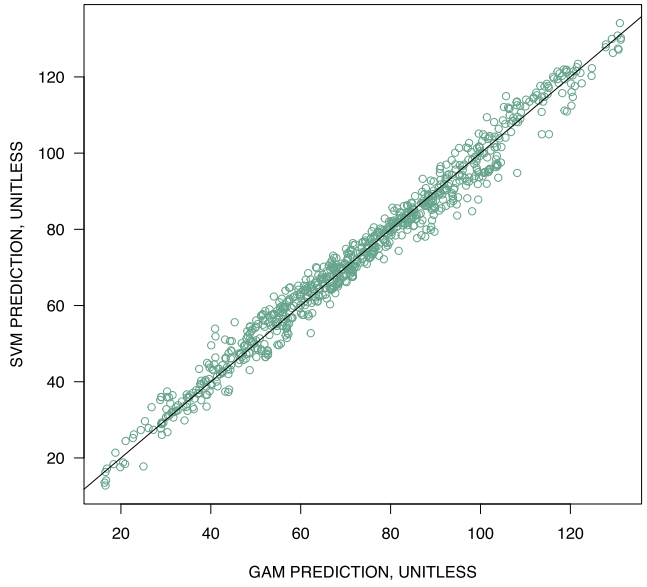

Fig. 6. Comparison of GAM and SVM predictions at the 800 locations of the synthetic dataset with an equal value line.

Table 1
Nash-Sutcliffe model efficiency and root-mean-square error statistics for the 800 used in the GAM and SVM modeling approaches.

Seed[a]	GAM		SVM		
	NSE	RMSE	NSE	RMSE	NumSV
1	0.968	4.557	0.986	3.033	369
2	–	–	0.990	2.518	290
3	–	–	0.992	2.264	245
4	–	–	0.990	2.595	302

[Seed, the integer used to set the random number generated before passing through GAM and SVM construction; GAM, generalized additive model; SVM, support vector machine for a seed of 1 (unity); NSE, Nash-Sutcliffe model efficiency; RMSE, root-mean-square error; NumSV, number of support vectors; –, results do not change from seed equaling 1.]

[a] The seed on the pseudo-random number generator the generated the underlying 800 locations (Fig. 1) was set to 62. The SVM algorithms (Karatzoglou et al., 2018) have an inherent probabilistic feature; therefore, those SVMs change somewhat even though the dataset fed to each SVM iteration does not because the same data were fed to the GAM algorithms.

RMSEs listed in Table 2 are a form of cross validation. A GAM created using the 369 support vector subsample predicting on the grid out performs the GAM created using 431 nonsupport vectors.

Comparison of results for NSE and RMSE between two subsamples types (Table 2) is informative. Recalling the seed of 1 in Table 1, a GAM using the 369 support vector subsample ($+\ n^{\text{svm}}$; Table 2) has larger NSE and smaller RMSE relative to a GAM using the 431 nonsupport vectors ($-\ n^{\text{svm}}$). So although the input subsample of the nonsupport vectors is about 1.17 times larger (431/369), that larger subsample actually produces a somewhat inferior model with respect to both NSE and RMSE. The support vectors thus are conceptualized to represent a type of super subsample conveying more information than the nonsupport vectors.

5.4. Simulation experiments with SVMs

Successive or looping iterations of SVMs that are fed the same data produce results from slightly different models and hence support vectors. As an example, this study uses $n = 369$ (the number of support vectors for SVM) as a representative number of support vectors. Fixing to this sampling size, an experiment using 20 simulations was made. The experiment

Table 2
Nash-Sutcliffe model efficiency and root-mean-square error statistics for a GAM using the support vectors as one sample and the nonsupport vectors as another based on results of the SVM given a seed of 1.

Basis	NSE	RMSE	Sample size
Entire grid ($+\ n^{\text{svm}}$)	0.963	5.002	369 (Tables 1 and 3)
Entire grid ($-\ n^{\text{svm}}$)	0.953	5.593	$431 = 800 - 369$

[NSE, Nash-Sutcliffe model efficiency; RMSE, root-mean-square error; Sample size, if $+\ n^{\text{svm}}$, the subset from the whole sample of the support vectors, or if $-\ n^{\text{svm}}$, the complement sample of $800 - n$. The "entire grid" means the 200×200 grid of the known surface was used to compute the listed statistics.]

was based on generating random subsamples of size $n = 369$ from the population of 800 locations, fitting a GAM, and making predictions throughout the 200×200 grid (again these are out-of-sample locations). The results are shown by the $+\ n^{\text{rand}}$ listed in Table 3. The repeated SVMs (similar to Fig. 4 for clarity) are the $+\ n^{\text{svm}}$ listed in Table 3.

Summary statistics of NSEs and RMSEs (Table 3) show that a range of these statistics occur. In this experiment, there are differences between the central tendency of these to those of the single GAM (first row of Table 2). The NSE for the GAM in the first row of Table 2 is about 0.963, which is greater than the third quartile of NSE for the $+\ n^{\text{rand}}$ (first row of Table 3). The RMSE for the GAM in the first row of Table 2 is about 5.002, which is below the first quartile of RMSE for the $+\ n^{\text{rand}}$ (third row of Table 3). These two comparisons show that a random sample of size n produces a GAM that performs worse than the GAM that used the 369 support vectors as one sample.

Using another 20 simulations, the $+\ n^{\text{rand}} = 369$ NSE and RMSE statistics listed in Table 3 represent simulations in which random samples of size $n = 369$ were generated and a GAM fit after each sample generation. It again is seen that the NSEs are generally smaller and that the RMSEs are generally larger for random samples than they are for the support vectors.

Simulations also can document a potential range in support vector counts. Using 20 simulations, successive SVMs were fit to all of the 800 locations and the number of support vectors recorded. The number of support vectors listed in Table 4 ranges between about 30 to 60 percent of the original $n = 800$ dataset. The mean is about 351, which implies that on average about 45 percent of the original dataset is identified as support vectors for at least the $Z(X, Y)$ data in this study (computed using defaults of the `ksvm()` function). These percentages are termed "support vector percentages" and abbreviated as $\mathscr{SV}_{\%}$, and the distribution of these is shown in Fig. 7.

5.5. Discussion

Support vectors are locations that reside away from the prediction tube and form a type of covariate hull. The support vectors represent a nonrandom super subsample, which is more informative than a random sample of equal size. The Karatzoglou et al. (2018) SVM implementation has some probabilistic feature as its algorithms progress. The result is that some observations are identified more frequently than others as support vectors, and $\mathscr{SV}_{\%}$ for some locations are as high as 100 percent and other locations have $\mathscr{SV}_{\%}$ near zero. The $\mathscr{SV}_{\%}$ distribution is inherently affected by the complexity of the surface, the errors associated with it (e.g. $\mathscr{N}(0, \sigma)$ measurement errors), and somewhat by the number of observation locations. Identification of the most and least informative sites using the $\mathscr{SV}_{\%}$ could be useful to stakeholders and provide decision support in terms of the number of monitoring sites, upgrading data collection instrumentation suites or recording intervals, and rebalancing other data collection resources.

Relevance vector machines (RVMs) (Bishop, 2006; Tipping, 2000) have an "identical functional form" to SVMs (Tipping, 2001), but a defining characteristic of RVMs is that they can be considerably more

Table 3
Summary statistics from 20 simulations generating random samples and the NSE and RMSE computed from a 200 × 200 grid from GAMs using 2-dimensional smooths and GAMs based only on SVM support vectors.

Basis and sample size[a]	Statistic type	Minimum	First quartile	Median	Mean	Third quartile	Maximum
+ $n^{rand} = 369$	NSE	0.923	0.952	0.957	0.954	0.960	0.971
+ $n^{svm} = 369$	NSE	0.958	0.960	0.964	0.966	0.972	0.981
+ $n^{rand} = 369$	RMSE	4.407	5.155	5.388	5.498	5.696	7.196
+ $n^{svm} = 369$	RMSE	3.590	4.323	4.915	4.711	5.159	5.279

[NSE, Nash-Sutcliffe model efficiency; RMSE, root-mean-square error in units of length; + n^{rand}, a random subsample of indicated sample size; + n^{svm}, the number of support vectors in the support vector machine (SVM) were used to make each generalized additive model (GAM).]
[a] The integer seed on the random number generator set ahead of the SVM.

Table 4
Summary statistics for distribution of the number of support vectors of SVM of the 800 locations of the synthetic dataset with 20 simulations of leave-one-out.

Modeling approach	Minimum	First quartile	Median	Mean	Third quartile	Maximum
SVM	248	307	356	351	399	476

[SVM, support vector machine.]

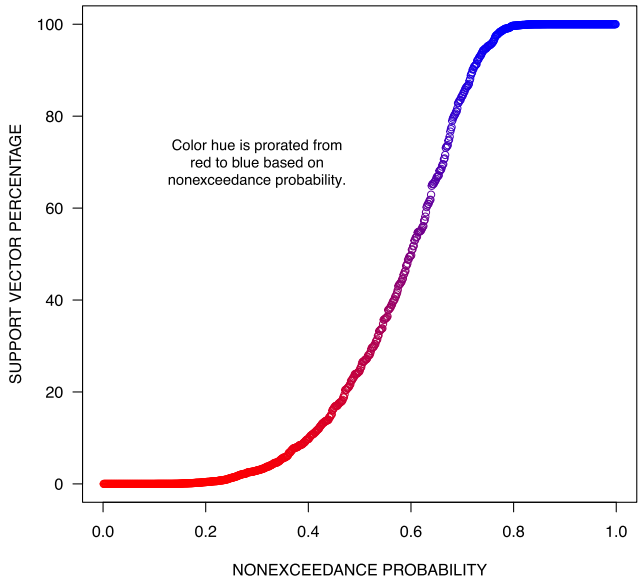

Fig. 7. Distribution of the support vector percentage ($\mathscr{SV}_{\%}$) as nonexceedance probability of each of 800 locations occurring as support vectors of SVMs with 20 simulations of leave-one-out with red to blue color ramp prorated by $\mathscr{SV}_{\%}$.

sparse in "relevance vectors" than an SVM is in its support vectors. The vectors of an RVM can be thought of as prototypical samples, whereas the support vectors of an SVM are outside the prediction tube. RVMs are computationally intensive, difficult to run for sample sizes larger than a couple of thousand, and are susceptible to local minima as opposed to SVMs that guarantee global minimization.

Some RVM simulations (data not shown), suggest that isolation of only the relevance vectors is too sparse a subsample for reliable regionalization by alternative methods, such as GAMs. For example and for the 800 locations of the dataset, the number of relevance vectors is often between 30 and 40. (The value 0.05 = 40/800 means less than 5 percent of the observations are needed to define the RVM.) Combining the relevance vectors and the support vectors would seem to identify effective subsamples; however, it is the author's opinion that RVM use to augment SVMs for the purposes here does not outweigh the algorithmic, computational, and documentation overhead. The author formed this conclusion using the synthetic data and the defaults of the `rvm()` function from Karatzoglou et al. (2018).

6. Example application of support vector percentages

Using SVMs, a covariate space in five dimensions was studied based on retrospective analysis of U.S. Geological Survey (USGS) peak streamflows from an existing streamgage network in Texas (Asquith and Roussel, 2009). These experimental results are restricted to the study of a single statistic—the 10-year (0.1 annual exceedance probability) streamflow ($Q_{0.1}$). The $Q_{0.1}$ was chosen in part because of its highest adjusted R-squared of several regional regression equations (Asquith and Roussel, 2009). The use of other streamflow statistics could lead to different insights.

The objectives for the example application are (1) to identify the least informative streamflow monitoring locations (streamgages), and (2) to identify the most informative streamgages that also have not operated since at least 2000 to the present (2019). The year 2000 is an arbitrary choice and represents a premise that a streamgage not operated in the 21st century is a firmly discontinued streamgage but yet shown as informatively critical by the SVM.

Asquith and Roussel (2009) used about 640 streamgages to develop equations for estimation of peak-streamflow frequency for selected annual exceedance probabilities in and near Texas. For the analysis in this study, which list 536 streamgages with watershed properties and the corresponding $Q_{0.1}$ values, 536 streamgages in Texas were aggregated from the original 638 streamgage network. For each streamgage, $Q_{0.1}$ and the watershed properties of contributing drainage area (A), main-channel slope (S), and mean annual precipitations (P) also were assembled from the Asquith and Roussel (2009) data files. The latitudes (N) and longitudes (E) (U.S. Geological Survey, 2019) of the streamgages were converted to an Albers equal area projection system for North American (horizontal) Datum of 1983.

Using base-10 logarithmic transformations on $Q_{0.1}$, A, S, and P and no transformation on N and E, an SVM of the following form was used:

$$Q_{0.1} \sim A + S + P + N + E. \quad (8)$$

Further, a leave-one-out for each of the 536 streamgages was used along with a wrapping iteration (counted as a single simulation) to repeat the leave-one-out 20 times. Thus, 10, 740 SVMs were constructed. For each SVM, the support vectors (actually the streamgage identification numbers) were recorded, and then the percent of the time each streamgage was a support vector ($\mathscr{SV}_{\%}$, support vector percentage) was computed.

Approximately half of the 536 streamgages in the Texas network are support vectors about 100 percent of the time (Fig. 8). These streamgages are branded as always support vector sites (ASV). About 25 percent of the streamgages (100 × 134/536) occur as support vectors less than about 5 percent of the time for the 10, 740 SVM constructions, and these are branded as nonsupport vector sites (NSV).

The NSV are effectively always within the prediction tube, which means that all other streamgages including the ASV in the greater network could provide sufficient information to build a regional model having the capacity to make reliable estimates at the NSV locations. The ASV define a type of generalized hyper-dimensional covariate hull of watershed properties (inclusive of spatial location in Texas) conditional

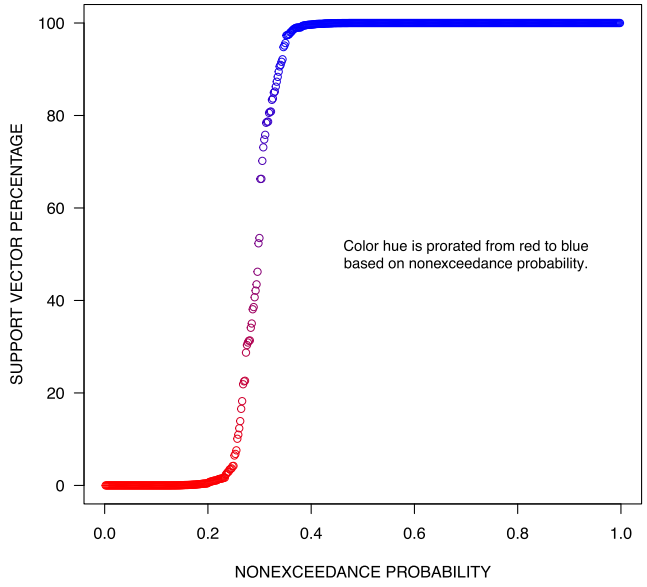

Fig. 8. Distribution of the support vector percentage ($\mathscr{SV}_{\%}$) as nonexceedance probability of the 536 streamgages in Texas based on leave-one-out SVM operation with 20 simulations of 536 leave-out-out SVM constructions with prorated color hue from red to blue based on nonexceedance probability.

on $Q_{0.1}$. These ASV sites are data points deemed by the SVM as critical during its machine-based training process; the NSV sites almost always have zero weights in the SVM. So it can be stated that by their very definition, all ASV sites are more important than NSV sites.

A map is shown (Fig. 9) that separately identifies the ASV and NSV sites. The ASV are plotted as upright triangles and NSV are plotted as inverted triangles. The remaining sites are plotted as rotated squares with a red to blue color ramp prorated by $\mathscr{SV}_{\%}$ between the ASV and NSV as was similarly done for Fig. 8. There are 283 streamgages that are support vectors 100 percent of the time for which 152 have not operated since at least the 2000 calendar year. These 152 streamgages are identified by the concentric circles shown (Fig. 9).

The 152 streamgages are interpreted as especially important for statistical regionalization of peak streamflow, but the corresponding watersheds are currently (2019) not monitored. Most of these sites are located in smaller drainage areas than the greater network. The overall network median contributing drainage area is about 300 square kilometers (km^2), whereas the median contributing drainage area for the 152 streamgages is about 25 km^2 (Table 5).

These results show that many discontinued streamgages potentially are critical for regionalization of the $Q_{0.1}$ peak streamflow. The corresponding watersheds are quite a bit smaller in area than the overall network (Table 5). So by being classified as ASV, these streamgages generally reside near the boundaries of 5-dimensional covariate hull conditioned on the $Q_{0.1}$ peak streamflows themselves.

A need for peak streamflow on small watersheds in Texas is known and documented (Harwell and Asquith, 2011), and since about 2000, efforts have been made with stakeholders in Texas to collect additional peak streamflow data on about 50 small watersheds in western Texas (Asquith et al., 2018). This example application affirms an intuitive understanding of Texas streamgage history—there potentially has been a data-collection imbalance on peak streamflow data between small watersheds and the overall streamgage network.

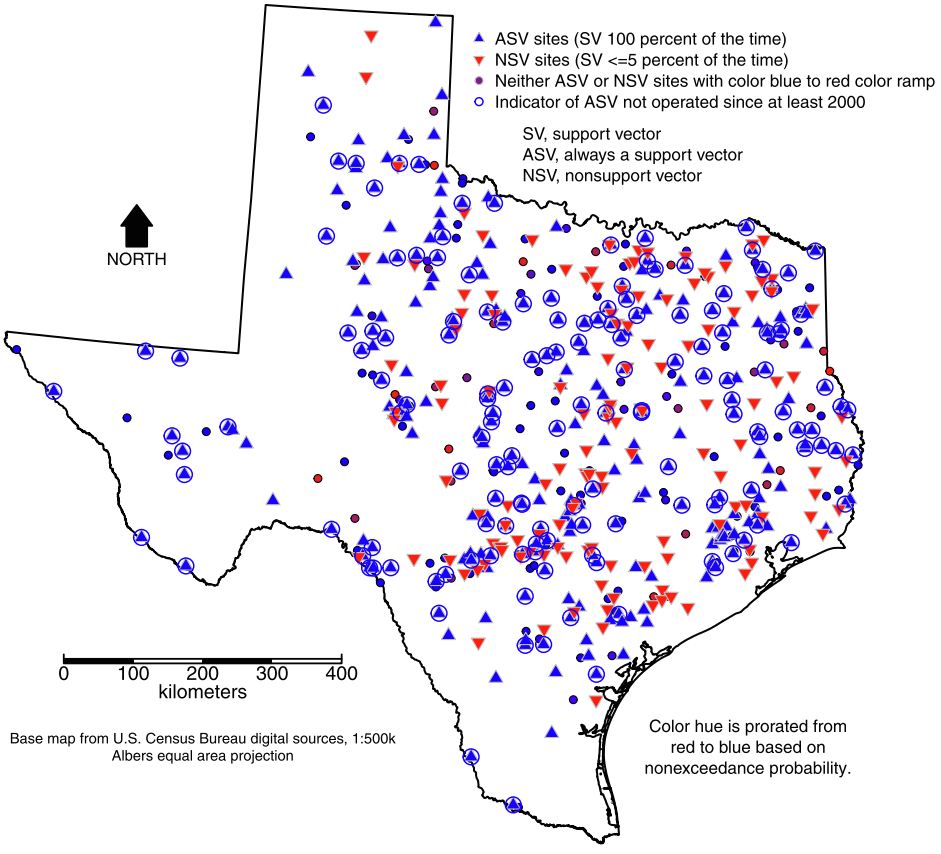

Fig. 9. Locations of 536 streamgages in Texas from Asquith and Roussel (2009) used to create SVMs for estimation of 10-year (0.1 annual exceedance probability) peak streamflow using watershed properties of contributing drainage area, main-channel slope, mean annual precipitation, and spatial coordinates. Symbology summarizes distribution of support vector percentages from 20 iterations of 536 leave-one-out SVM constructions.

Table 5
Summary statistics for distribution of contributing drainage area of selected watersheds in Texas from 20 iterations of 536 leave-one-out SVM constructions.

Dataset type	Mini-mum (km^2)	First quartile (km^2)	Median (km^2)	Mean (km^2)	Third quartile (km^2)	Maxi-mum (km^2)
All:536, $n = 536$	0.259	14.89	297.84	168.73	1,474.34	24,162
$ASV_{(<2000)}$, n = 152	.337	3.25	25.02	45.95	545.65	22,173

[km^2, square kilometers; All:536, all 536 streamgages in the network; $ASV_{(<2000)}$, always support vector streamgages that have not operated since at least 2000 for which the number of streamgages is $n = 152$. Note, the $ASV_{(<2000)}$ could slightly change when re-using the code in the Supplementary materials accompanying this paper.]

7. Conclusions

The distinction between support and nonsupport vectors by SVMs measures site importance for hydrometeorologic monitoring network analyses. Support vectors are those observations outside the prediction tube during SVM training, and importantly, those observations close to the SVM predictions have no impact on the SVM itself because their weights are zero.

Simulations of a synthetic surface (2-dimensional) show that support vectors are especially important and contribute relatively more information than a random sample of equal size. Though SVMs have a nonlinear regression operational mode, the use of GAMs was really the core estimation method. Using the known 200 × 200 grid of the surface as a cross-validation tool, it was found that a GAM based on a random sample performs on average worse than a GAM using support vector subsamples of the data.

The example application demonstrated $\mathscr{SV}_{\%}$ interpretation to identify discontinued streamgages in Texas that are especially informative for a regional model of peak streamflow frequency, particularly data from small watersheds. Scientists or stakeholders interested in enhancing such regional models could potentially find these results informative.

In the context of network analysis, four major insights are made. First, the $\mathscr{SV}_{\%}$ values for sites are immediately useful for assigning quantitative relative importance to active and discontinued sites. Second, large differences between predictions from an SVM to a GAM (or simpler multilinear regression) at monitored and unmonitored sites could represent data gaps. Third, SVMs require limited data setup and are computationally fast. Fourth, SVMs are complex but in simple one-dimension use can be geometrically described to many audience types including lay stakeholders.

Being that they contain more information than a random sample, sites that always or very nearly so are support vectors can be interpreted as more important than sites that are infrequently to never are support vectors. This is key to the use of support vectors and SVMs within network analyses. A final remark is that the analyses here are not to be confused that the author is advocating that statistical predictions from nonSVMs (say from a GAM) be based on just the data identified as support vectors by an SVM for any given statistical prediction method. The use of either the whole sample with allowance for analysts to also consider leave-one-out or other cross-validation schemes remains a general standard of practice.

CRediT authorship contribution statement

W.H. Asquith: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work reported in this paper is a result of discussions with USGS colleagues and several stakeholders over many years related to how to efficiently provide useful decision-support data to stakeholders on groundwater and surface-water monitoring networks. The work received some support from (1) a joint USGS and U.S. Environmental Protection Agency study funded by the Gulf Coast Ecosystem Restoration Council, (2) a regional water-availability study of the Mississippi River alluvial plain through the USGS Water Availability and Use Science Program, and (3) a small watershed, peak-streamflow gaging project between the Texas Department of Transportation (TxDOT) (Contract No. 0000009037) and the USGS Oklahoma-Texas Water Science Center. Early draft comments by Saul Nuccitelli (TxDOT) are greatly appreciated, and Christopher Konrad (USGS) provided internal-agency peer review. The work also greatly benefitted from the four peer reviewers functioning on behalf of the Journal. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jhydrol.2019.124522.

References

Andricevic, Roko, 1989. Groundwater Monitoring Network—Analysis and Design. St. Anthony Falls Hydraulic Laboratory, University of Minnesota Digital Conservancy, accessed on April 12, 2019 athttps://hdl.handle.net/11299/114153.

Andricevic, Roko, 1990. Cost-effective network design for groundwater flow monitoring. Stoch. Hydrol. Hydraul. 4 (1), 27–41. https://doi.org/10.1007/BF01547730.

Asquith, W.H., Roussel, M.C., 2009, Regression equations for estimation of annual peak-streamflow frequency for undeveloped watersheds in Texas using an L-moment-based, PRESS-minimized, residual-adjusted approach: U.S. Geological Survey Scientific Investigations Report 2009–5087, 48p.https://pubs.usgs.gov/sir/2009/5087/.

Asquith, W.H., Harwell, G.R., Winters, K.E., 2018. Annual and approximately quarterly series peak streamflow derived from interpretations of indirect measurements for a crest-stage gage network in Texas through water year 2015. U.S. Geological Survey Scientific Investigations Report 2018–5107, 24p.https://doi.org/10.3133/sir20185107.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York978-0-387-31073-2.

Brunsdon, Chris, Comber, Lex, 2015. An introduction to R for spatial.

Carlisle, D.M., Wolock, D.M., Howard, J.K., Grantham, T.E., Fesenmyer, Kurt, Wieczorek, Michael, 2016. Estimating natural monthly streamflows in California and the likelihood of anthropogenic modification: U.S. Geological Survey Open-File Report 2016-1189, 27p.https://doi.org/10.3133/ofr20161189.

Chebbi, Afef, Bargaoui, Z.K., Abid, Nesrine, Cunha, M.C., 2017. Optimization of a hydrometric network extension using specific flow, kriging and simulated annealing. J. Hydrol. 555, 971–982. https://doi.org/10.1016/j.jhydrol.2017.10.076.

Chen, Yen-Chang, Chiang, Wei, Yeh, Hui-Chung, 2007. Rainfall network design using kriging and entropy. Hydrol. Process. 22 (3), 340–346. https://doi.org/10.1002/hyp.6292.

Clifton, Lei, Clifton, D.A., Zhang, Yang, Watkinson, Peter, Tarassenko, Lionel, Yin, Hujun, 2014. Probabilistic novelty detection with support vector machines. IEEE Trans. Reliab. 63 (2), 455–467. https://doi.org/10.1109/TR.2014.2315911.

Davis, J.C., 2002. Statistics and Data Analysis in Geology. John Wiley, New York 987-0-471117-275-8.

Esquivel, J.M., Morales, G.P., Esteller, M.V., 2015. Groundwater monitoring network design using GIS and multicriteria analysis. Water Resour. Manage. 29, 3175–3194. https://doi.org/10.1007/s11269-015-0989-8.

Faraway, J.J., 2005. Linear Models with R. Chapman & Hall/CRC, Boca Raton, Florida1-58488-425-8.
Faraway, J.J., 2006. Extending the Linear Model with R—Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman & Hall/CRC, Boca Raton, Florida978-1-58488-424-8.
Fisher, J.C., 2013. Optimization of water-level monitoring networks in the eastern Snake River Plain aquifer using a kriging-based genetic algorithm method. U.S. Geological Survey Scientific Investigations Report 2013–5120 (DOE/ID-22224), 74p.https://pubs.usgs.gov/sir/2013/5120/.
Harwell, G.R., Asquith, W.H., 2011. Annual peak streamflow and ancillary data for small watersheds in central and western Texas. U.S. Geological Survey Fact Sheet 2011–3082, 4p.https://pubs.usgs.gov/fs/2011/3082/.
Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall/CRC, Boca Raton, Florida978-0-41234-390-2.
Hastie, T.J., Tibshirani, R.J., Friedman, Jerome, 2008. The Elements of Statistical Learning—Data Mining, and Inference and Prediction, second ed. Springer, New York978-0-38784-857-0.
Karatzoglou, Alexandros, Meyer, David, Hornik, Kurt, 2006. Support vector machines in R.J. Stat. Softw. 15 (9), 1–28. https://doi.org/10.18637/jss.v015.i09.
Karatzoglou, Alexandros, Smola, Alex, Hornik, Kurt, 2018. kernlab—Kernel-based machine learning lab. R package version 0.9-27, dated October 10, 2018.https://CRAN.R-project.org/package=kernlab.
Kiang, J.E., Stewart, D.W., Archfield, S.A., Osborne, E.B., Eng, Ken, 2013. A national streamflow network gap analysis. U.S. Geological Survey Scientific Investigations Report 2013–5013, 79p.https://pubs.usgs.gov/sir/2013/5013/.
Kollat, J.B., Reed, P.M., Maxwell, R.M., 2011. Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics. Water Resour. Res. 47, W02529. https://doi.org/10.1029/2010WR009194.
Kuhn, Max, Johnson, Kjell, 2016. Applied Predictive Modeling. Springer, New York978-1-4614-6848-6.
Markus, Momcilo, Knapp, H.V., Tasker, G.D., 2003. Entropy and generalized least square methods in assessment of the regional value of streamgages. J. Hydrol. 283, 107–121. https://doi.org/10.1016/S0022-1694(03)00244-0.
Medina, K.D., 1987. Analysis of surface-water data network in Kansas for effectiveness in providing regional streamflow information. U.S. Geological Survey Water-Supply Paper 2303, 28p.https://pubs.usgs.gov/wsp/2303/report.pdf.
National Research Council, 2004. Assessing the National Streamflow Information Program—Chapter 4 Streamflow Network Design. The National Academies Press, Washington, DC978-0-309-09210-4https://doi.org/10.17226/10967. 176p.
Olea, R.A., 1999. Geostatistics for Engineers and Earth Scientists. Kluwer Academic ISBN 978-1-4615-5001-3.
Pimentel, M.A.F., Clifton, D.A., Clifton, Lei, Tarassenko, Lionel, 2014. A review of novelty detection. Signal Process. 99, 215–249. https://doi.org/10.1016/j.sigpro.2013.12.026.
Putthividhya, Aksara, Tanaka, Kenji, 2012. Optimal rain gauge network design and spatial precipitation mapping based on geostatistical analysis from colocated elevation and humidity data. Int. J. Environ. Sci. Dev. 3 (2), 124–129. https://doi.org/10.7763/IJESD.2012.V3.201.
R Development Core Team, 2019. R—A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, version 3.6.1. https://www.R-project.org.
Stasinopoulos, M.D., Rigby, R.A., Heller, G.Z., Voudouris, Vlasios, Bastiani, De, 2017. Fernanda Flexible Regression and Smoothing using GAMLSS in R. Chapman & Hall/CRC, Boca Raton, Florida978-1-138-19790-9.
Steinwart, Ingo, Christmann, Andreas, 2008. Support Vector Machines. Springer, New York978-0-387-77241-7 601p.
Tipping, M.E., 2000. The relevance vector machine. In: S.A. Solla, T.K. Leen, K.Müller (Eds.). Advances in Neural Information Processing Systems 12, Proceedings, MIT Press. pp. 652–658.
Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244.
Tonkin, M.J., Larson, S.P., 2002. Kriging water levels with a regional-linear and point-logarithmic drift. Groundwater 40 (2), 185–193. https://doi.org/10.1111/j.1745-6584.2002.tb02503.x.
U.S. Geological Survey, 2019. U.S. Geological Survey National Water Information System—Web interface, accessed January 20, 2019.https://doi.org/10.5066/F7P55KJN.
Van Greer, F.C., Te Stroet, C.B.M., Yangxiao, Zhou, 1991. Using Kalman filtering to improve and quantify the uncertainty of numerical groundwater simulations—1. The role of system noise and its calibration. Water Resour. Res. 27 (8), 1987–1994. https://doi.org/10.1029/91WR00509.
White, J.B., 2017. Forecast first—an argument for groundwater modeling in reverse. Groundwater 55 (5), 660–664. https://doi.org/10.1111/gwat.12558.
Williams–Sether, Tara, 2015. Regional regression equations to estimate peak-flow frequency at sites in North Dakota using data through 2009. U.S. Geological Survey Scientific Investigations Report 2015–5096, 12p.https://doi.org/10.3133/sir20155096.
Wood, S.N., 2017. Generalized Additive Models—An Introduction with R, second ed. Chapman & Hall/CRC, Boca Raton, Florida978-1-4987-2833-1 476p.
Wood, S.N., 2019. mgcv—Mixed GAM computation vehicle with automatic smoothness estimation. R package version 1.8-29, dated September 20, 2019.https://CRAN.R-project.org/package=mgcv.
Wu, Y., 2004. Optimal design of a groundwater monitoring network in Daqing, China. Environ. Geol. 45 (4), 527–535. https://doi.org/10.1007/s00254-003-0907-x.