

Improved Prediction of Managed Water Flow into Everglades National Park Using Empirical Dynamic Modeling

Erik Saberski¹; Joseph Park²; Troy Hill³; Erik Stabenau⁴; and George Sugihara⁵

Abstract: Alteration of natural surface flow paths across South Florida has been detrimental to the environmental health and sustainability of the Everglades and surrounding ecosystems. As part of the Comprehensive Everglades Restoration Plan (CERP), predicting flows into Everglades National Park (ENP) is a central concern of effective management strategies. Management efforts have established weekly target flows into Everglades National Park through optimization of numerically intensive hydrological models. These target flows are focused specifically on flows across US Highway 41, also known as the Tamiami Trail. To aide in timely management assessments in response to current or predicted hydrologic conditions, the Tamiami Trail Flow Formula (TTFF) was developed previously to predict weekly target flows based on linear regression of six theorized flow drivers. It is known that these drivers exhibit nonlinear dynamics, suggesting that there is room for improvement in relation to the strictly linear TTFF. We used empirical dynamic modeling (EDM), a nonparametric modeling paradigm for forecasting and analyzing nonlinear time series, to show that prediction accuracy is improved when nonlinearity is accounted for. This method relies on weighted linear regressions that depend on specific environmental conditions or system states, i.e., in which the regression gives greater weight to input variables that have values that are more similar to the current state. Surprisingly, we found that only two of the six standard TTFF variables are required in the nonlinear weekly forecast model (upstream and downstream water levels), and thus the EDM model not only improves accuracy but also greatly simplifies hydrologic forecasting. **DOI: 10.1061/(ASCE)WR.1943-5452.0001598.** *This work is made available under the terms of the Creative Commons Attribution 4.0 International license, <https://creativecommons.org/licenses/by/4.0/>.*

Introduction

Many analytic approaches use equation-based models as approximations of real-world systems to test hypothesized mechanisms or to predict future outcomes. However, real-world systems often are nonlinear and multidimensional, which can render explicit parametric approaches intractable. Empirical approaches, which extract information from the data instead of relying on hypothesized equations, represent a natural and flexible approach to modeling complex, nonlinear systems such as managed water resources.

Empirical dynamic modeling (EDM) is a nonparametric framework for modeling nonlinear systems based on the mathematical theory of reconstructing attractors (vector fields that can show how variables interact though time) from time-series data (Takens 1981). EDM initially was intended to address problems in ecology (Sugihara and May 1990; Sugihara 1994; Dixon et al. 1999;

Sugihara et al. 2012; Deyle et al. 2016; Ye and Sugihara 2016). However, its applications have extended to many areas such as climate change (van Nes et al. 2015), atmospheric sciences (Sugihara et al. 1999), neuroscience (Segundo et al. 1998), studying the dynamics of infant heart rhythms (Sugihara et al. 1996), identifying the drivers of influenza outbreaks (Deyle et al. 2016), and classifying complex behaviors in the nematode *C. elegans* (Lorimer et al. 2021; Saberski et al. 2021). To our knowledge, EDM has not yet been used specifically to map hydrologic dynamics. Here, we introduce the use of EDM as a tool for forecasting managed water flows in Everglades National Park (ENP) as a component of the Comprehensive Everglades Restoration Plan (CERP). A lucid and accessible introduction to EDM was provided by Chang et al. (2017).

Managed Flows: Everglades Restoration

The Florida Everglades originally consisted of 3 million acres of marsh draining the Kissimmee River Basin and Lake Okeechobee southward into Florida Bay. Starting in the late nineteenth century, ambitious plans to drain the Everglades to produce arable and habitable lands were initiated, eventually coalescing in 1948 under the Congressionally authorized Central and Southern Florida (C&SF) Project under auspices of the USACE. Design goals were to provide flood control and agricultural sustainability, and major features included the Herbert Hoover dike impounding Lake Okeechobee, creation of a large agricultural area along the southern lake border, a levee along the eastern boundary of the Everglades, and impoundment of three water conservation areas (WCAs) linking Lake Okeechobee to Everglades National Park and the southern coast (National Research Council 2008).

The result of these water control efforts was a fundamental alteration of the natural flow paths and hydroperiods (Fig. 1), which eventually was recognized as being detrimental to the environmental

¹Scripps Institution of Oceanography, Univ. of California San Diego, La Jolla, CA 92037 (corresponding author). ORCID: <https://orcid.org/0000-0002-6475-6187>. Email: eriksaberski@gmail.com

²US Department of the Interior, National Park Service, South Florida Natural Resources Center, Homestead, FL 33031. ORCID: <https://orcid.org/0000-0001-5411-1409>

³US Department of the Interior, National Park Service, South Florida Natural Resources Center, Homestead, FL 33031.

⁴US Department of the Interior, National Park Service, South Florida Natural Resources Center, Homestead, FL 33031. ORCID: <https://orcid.org/0000-0002-6574-9317>

⁵Professor, Scripps Institution of Oceanography, Univ. of California San Diego, La Jolla, CA 92037.

Note. This manuscript was submitted on August 19, 2021; approved on June 7, 2022; published online on September 16, 2022. Discussion period open until February 16, 2023; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Water Resources Planning and Management*, © ASCE, ISSN 0733-9496.

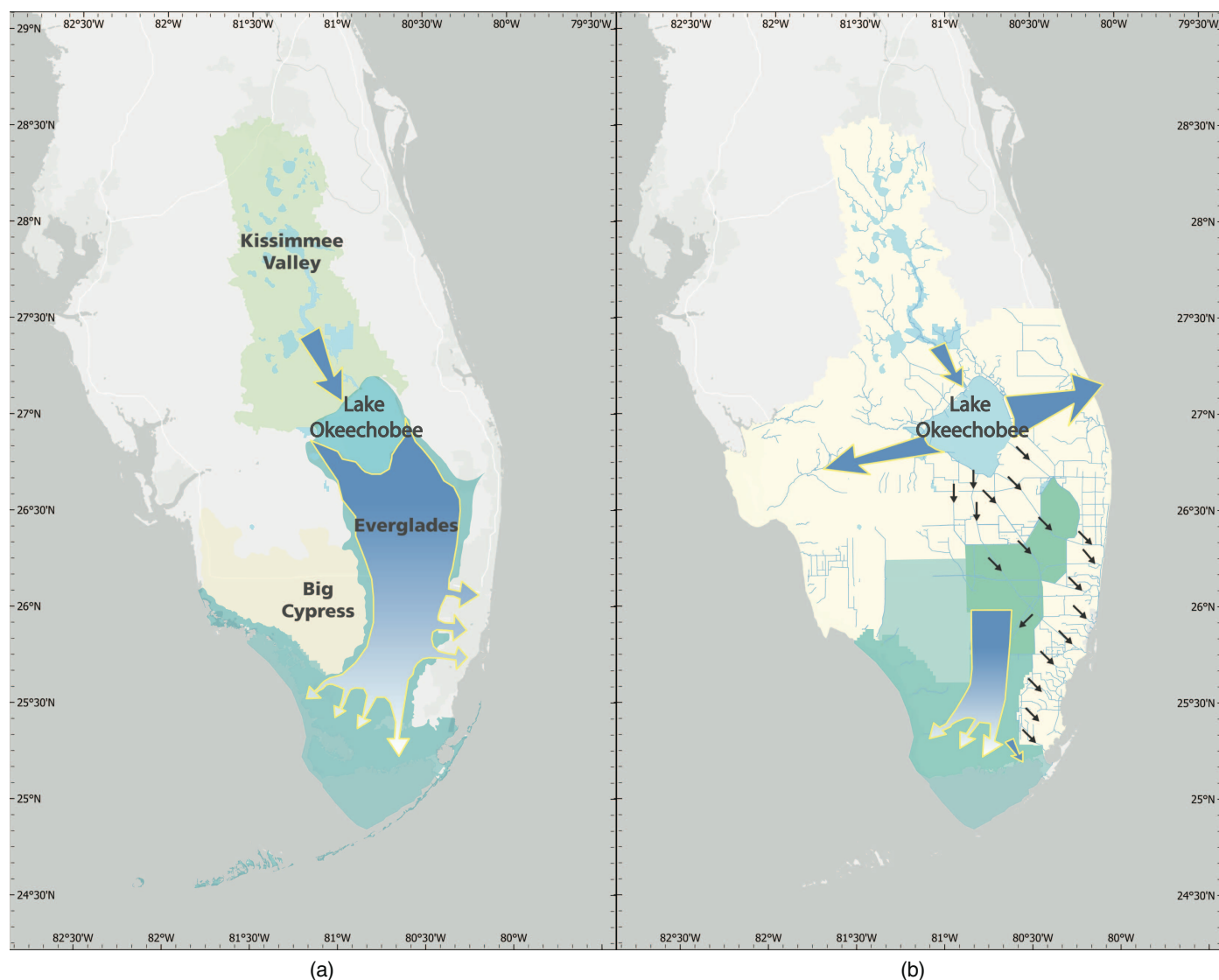


Fig. 1. (Color) Schematic of flow paths in South Florida: (a) predevelopment; and (b) modern. In the predevelopment era, the Kissimmee Valley floodplain drained into Lake Okeechobee, which then overflowed its southern rim in a river of grass to the southern peninsula. Postdevelopment, flow paths were channelized, represented by arrows, and the remaining segments of the Everglades were impounded with levees and canals. (Base map courtesy of South Florida Natural Resources Center.)

health and sustainability of the Everglades and its ecosystem services. Recognition of these changes led to the Congressionally mandated Comprehensive Everglades Restoration Plan in 2000, a framework for restoring, preserving, and protecting the South Florida ecosystem. The CERP originally was designed with 68 project components expected to take 30 years at an estimated cost of \$8 billion. Over the last 2 decades, it has been recognized that the CERP and state restoration efforts must encompass an adaptive management approach. Therefore, the restoration today is a complex, adaptive collaboration that is continuing to evolve (*National Academies of Sciences* 2018).

A central tenet of the CERP is to increase water flows and hydroperiods within Everglades National Park. A fundamental barrier to this was construction of the Tamiami Trail (US Highway 41) in the early twentieth century. The trail acts as a levee preventing natural flow from the upstream WCAs and natural areas. Flows from the upstream WCAs are managed primarily through the S-12 gated weirs (Fig. 2), with both upstream and downstream regulation limits.

A recent adaptation of restoration water management is the redevelopment of flow targets for releases into Everglades National Park as part of the Combined Operational Plan (COP) (*USACE* 2017). These targets serve as goals for maintaining healthy water levels throughout the greater Everglades system based on multiple environmental variables such as weekly rainfall and estimated evaporation, and are recalculated on a weekly basis. Because these flow targets are derived from up-to-date environmental conditions, future flow targets cannot be determined exactly without knowing the future environmental state. In order to best prepare for these weekly targets, the Tamiami Trail Flow Formula (TTFF) (*SFWMD* 2020b) was developed recently to forecast future flow targets. A diagram of the inputs and outputs to these models is presented in Fig. 3. The goal of the TTFF is not to predict the following week's flow into the ENP, but to forecast the following week's target flow. This subtle difference has huge implications: although the following week's target will guide what the managed flow into the system will be, predicting the target is fundamentally different than predicting the raw flow.

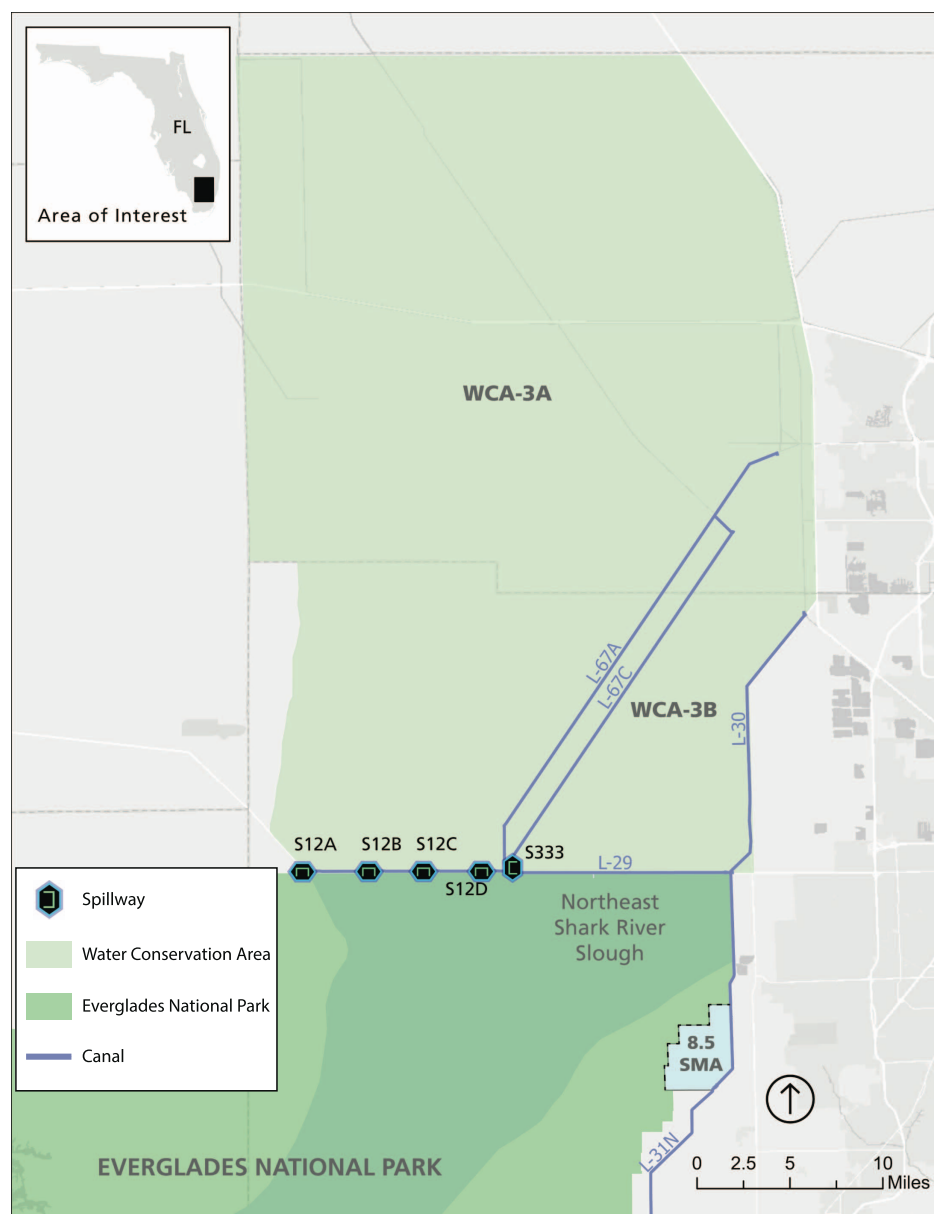


Fig. 2. (Color) Schematic of Everglades water control structures and projects. The Tamiami Trail, S-12 and S-333 structures separate upstream water conservation areas (WCAs) from Everglades National Park. (Base map courtesy of South Florida Natural Resources Center.)

The quality of the forecasts made by the TTFF are particularly critical because the system must be managed as a whole, taking into account both the desired downstream conditions and the upstream storage capacity. Upstream basins are large and respond slowly to changes in operational efforts. Thus, considerable lead time is needed to adjust basin water levels. Forecasting next week's target flow is extremely important because it can provide valuable lead time for managers to prepare the upstream basin level appropriately to meet the future desired target effectively.

Setting the incoming flow volumes appropriately is critical for achieving projected deliveries through Tamiami Trail and not creating adverse impacts due to flooding. Improvements in forecast skill will reduce the likelihood of ecologically adverse conditions within the WCAs or, simply put, having too much or too little water to manage the system properly. A primary aim of the present work was to examine the TTFF to ascertain the completeness of its

information content and to compare it with and determine the potential benefits of forecasts made using EDM.

Target Flows and the Tamiami Trail Flow Formula

Target flows were determined over the 1965–2005 period using the Regional Simulation Model (RSM) (SFWMD 2020a) and an inverse modeling tool, identifying optimal flows in response to hydrologic constraints (SFWMD 2020b). The resultant time series is referred to as $Q^{\text{sum}}(t)$, representing cumulative weekly target flows across Tamiami Trail into Everglades National Park. To model these target flows in response to current or future conditions, the Tamiami Trail Flow Formula, a linear model, was developed (SFWMD 2020b). TTFF developers recognized the nonlinear nature of the problem but decided that a linear formulation performed adequately and was simpler and easier to understand than

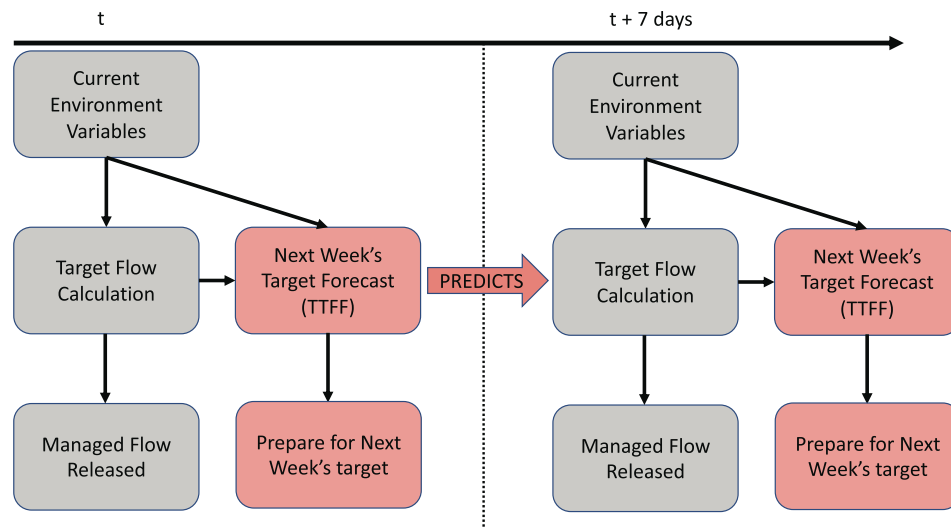


Fig. 3. (Color) Inputs and outputs to TTF. Each week, the current week's environmental conditions are used to generate a target flow that dictates management for the following week. This data is also used to forecast next week's target flow using the TTF, a forecast that implicitly accounts for next week's environmental conditions. Overall, this produces both a target to guide management for the upcoming week and a forecast to predict next week's new target, giving time for managers to prepare.

a nonlinear or machine learning model. The TTF presumes that precedent values of rain, evapotranspiration, upstream and downstream water levels, and flow are required to best predict target flows.

The TTF is

$$\hat{Q}^{\text{sum}}(t) = \beta_1 S^{\text{WCA}}(t) + \beta_2 S^{\text{ENP}}(t) + \beta_3 Q^{\text{sum}}(t-1) + \beta_4 R(t) + \beta_5 \text{PET}(t) + \beta_6 \text{ZA}(t) \quad (1)$$

where $\hat{Q}^{\text{sum}}(t)$ = predicted target flow release for coming week, and is the sum of S-12A, S-12B, S-12C, S-12D, and S-333 (Fig. 2); $S^{\text{WCA}}(t)$ = spatial average of observed water levels in WCA-3A at start of current week (the start of a week is Sunday and the end of a week is Saturday); $S^{\text{ENP}}(t)$ = observed water level in Everglades National Park, Northeast Shark River slough (NESRS), for current week; $Q^{\text{sum}}(t-1)$ = target flow releases for previous week; $R(t)$ = areal average of total weekly rainfall for WCA-3A; PET = total weekly potential evapotranspiration at 3AS3WX station; ZA = Zone A regulation water level of current week in WCA-3A (when water levels in WCA-3A are above ZA, flood control water releases are authorized across Tamiami Trail); and β = linear regression fit coefficients.

Plotting the raw variables against target flows (Fig. 4) revealed that the highest correlation among the variables is the previous week's target flow (autocorrelation). This makes sense because water flows relatively slowly through the Everglades, giving the system large inertia. Therefore we expected flows to have relatively little change from the prior week and to exhibit temporal autocorrelation. Upstream and downstream water levels are correlated noticeably with future flows. The other two variables that have positive correlations with flow are upstream (WCA-3A) and downstream (NESRS) water-levels. Despite the positive correlations, the data suggest that a nonlinear fit may be more suitable for these variables (e.g., exponential relationship between flow and NESRS level). The remaining variables, rain, PET, and ZA, have no clear indication of linear relationship or covariation (correlation) of any kind, which often occurs with nonlinear dynamics and overlapping effects from explanatory variables (Sugihara et al. 2012). These variables also likely are coupled with each other (e.g., upstream

water level influences downstream water level, and rainfall influences water levels), creating a complex web of dynamics that may be difficult to define with parametric models. Taken together, this suggests that predictions might be improved when the system is viewed through a nonlinear, nonparametric lens.

Despite the strictly linear nature of the TTF, it seemingly has impressive short-term predictive accuracy, achieving a correlation between observed and predicted weekly values of $\rho = 0.90$. However, this largely is due to the significant amount of autocorrelation in the data on a weekly time scale: a constant predictor (predicting that the value next week will be the same as the current) achieves a predictive accuracy of $\rho = 0.88$. Other metrics for predictive accuracy show that the formula has much room for improvement: it correctly predicts the directional change in flow only 60% of the time, and the prediction accuracy of changes in flow ($\Delta Q = Q_{t+1} - Q_t$) is $\rho = 0.45$.

Nonlinear, Nonparametric Approaches

As noted previously, the TTF was generated through a generalized linear model of six variables hypothesized to be influential to flow using data collected from 1965 to 2005. Because the TTF was generated from a single best-fit solution of the entire data record, the model is implicitly stationary: resolved coefficients of the TTF are fixed constants reflecting the global nature of the statistical regression. This is fundamentally distinct from dynamic nonlinear models, in which relationships among variables can change. In fact, nonlinear models can be constructed piecewise from segmented linear models to address how relationships among variables change as the system state evolves.

For example, if one assumes that the dynamics are changing slowly over time, the linear solution can be recalculated every few years to find a new set of coefficients specific to recent data. Similarly, if dynamics are theorized to change seasonally, one may calculate coefficients for each month of the year. Such partitions are known as "similar states," in which a state refers to a set of conditions associated with a specific set of dynamics. Similar states exhibit similar dynamics.

Typically, the state of a natural system depends on multiple factors (so-called "state variables"). For example, in the seasonal

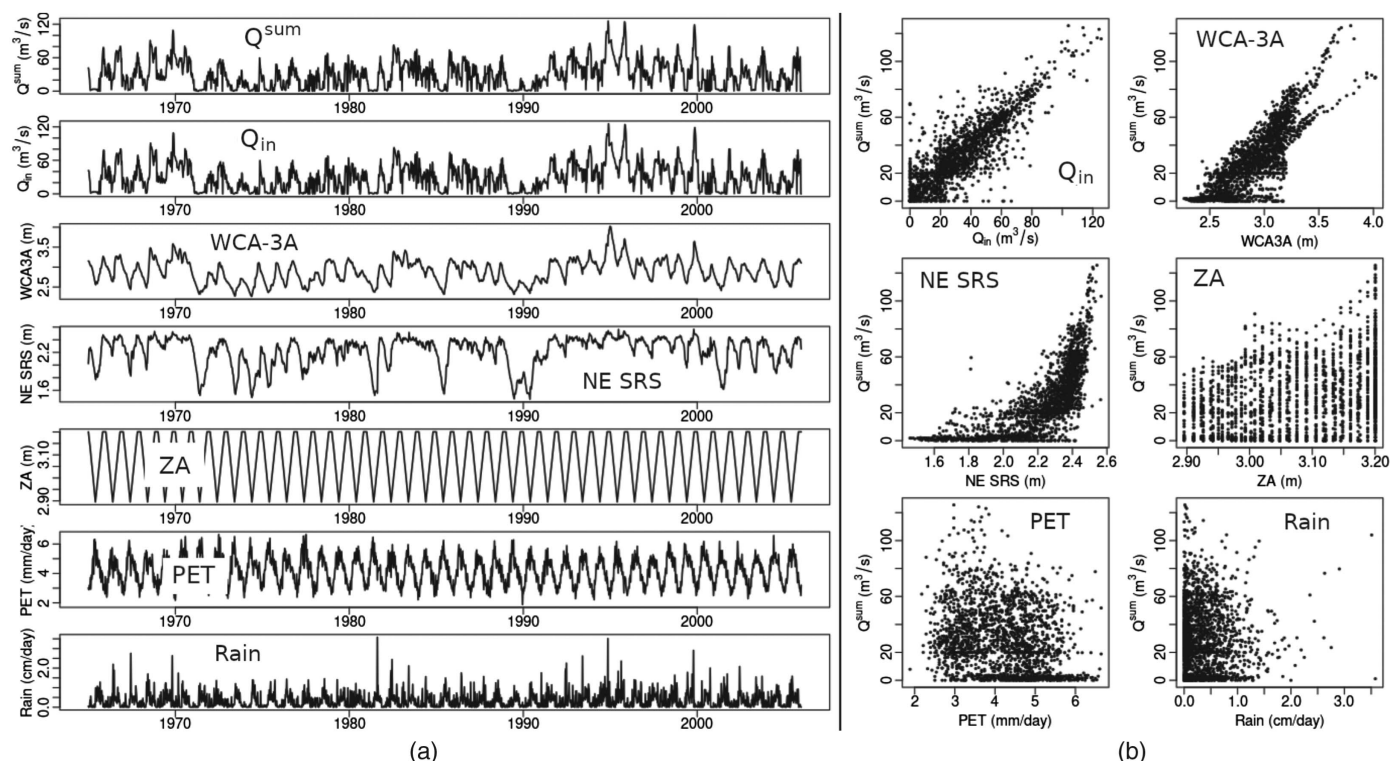


Fig. 4. (a) Time series of Q^{sum} and the presumed causal variables; and (b) scatter plots of the variables versus Q^{sum} .

model which calculated coefficients for each month (described previously), the time of year would be considered to be one state variable. Just as the seasonal model recalculates coefficients depending on the month, similar nonlinear models can be built to account for other state variables. For example, in the case of the TTFF it may be sensible to rederive coefficients by partitioning the data into subsets with similar flow rates, because high flow rates may have different dynamics than low flow rates.

EDM

Empirical dynamic modeling focuses on reconstructing a system's state space: a multidimensional representation of system variables as a function of time (Sugihara and May 1990; Sugihara 1994; Sugihara et al. 2012). Forecasting leverages the fact that points localized in state space (nearest neighbors) exhibit similar dynamics (Sugihara and May 1990; Sugihara 1994). Whereas the preceding examples describe ways to define the state space based on one variable (e.g., month or flow regime), EDM considers multiple variables together to identify similar states without presuming specific relationships (Deyle and Sugihara 2011; Ye and Sugihara 2016); instead, dynamics are derived directly from the data.

EDM can be used to screen the available time-series data and identify which variables are relevant and usefully can be included in a nonlinear forecasting model. Additionally, EDM involves the use of a causality test, convergent cross-mapping (CCM, Sugihara et al. 2012), which identifies nonlinear coupling between variables directly from time-series data (Fig. S1). This contrasts the common modeling procedure of the TTFF in which the specific variables used are asserted or hypothesized to be relevant.

We utilized a state-space forecasting technique within the EDM framework called sequential locally weighted global linear maps (S-Maps, Sugihara 1994). At each point in time, coefficients are recalculated based on a linear fit that maps state variables onto a target variable, similar to the TTFF formulation. However, each

fit at time t is weighted toward states similar to that of time t . This is analogous to the nonlinear methods described previously; however, instead of rigid cut-offs defined by the partition (e.g., partitioning data strictly by month, with weights of either 1 or 0), weights are applied smoothly based on proximity, with an exponential kernel applied to all points in the state space. A nonlinearity parameter, θ , can be adjusted to change how state-specific forecasts are: $\theta = 0$ gives all states equal weight regardless of state similarity, equating to a global autoregressive model. However, higher values of θ localize the forecast to more state-specific conditions, accounting for how system dynamics change over time. Typically, when using S-Map (or other EDM methods such as those described in the following section), the value being predicted from is left out of local linear regressions in order to avoid in-sample fitting and to obtain unbiased predictions (i.e., leave-one-out cross-validation).

The S-Map method adds only one additional step into the process of the original TTFF. Both the TTFF and S-Map forecasts utilize regressive maps from driving variables onto flow targets; S-Maps also apply weights to these regressions such that nearest neighbors are weighted more heavily in each prediction.

Models

The system state can be defined in many ways. For example, the state can be defined simply using the value of one variable (e.g., a high-flow state versus a low-flow state), or combinations of multiple (e.g., high and low flow in the winter versus high and low flow in the summer). We analyzed the performance of four different ways to define system state

1. Interannual predictor: recalculates linear regressions of the six variables from Eq. (1) using the previous 5 years of data [time points $(t-260), t]$ to predict flow at time $t + 1$.
2. Seasonal predictor: recalculates coefficients using historical data within 6 weeks of the current year-day. For example, if a forecast is predicting flow in the first week of March, all historical data

between mid-January and mid-April are used to generate linear coefficients.

3. Variable-specific predictor: recalculates coefficients at each time (t) using the 10% of values in the data set that have the values closest to that of the given variable (v), i.e., the timepoints (t^*) with the smallest values $|v_t - v_{t^*}|$. This was performed on the five theorized drivers of flow.
4. EDM S-Maps: in contrast to the variable-specific model in which data are partitioned based on the value of a single variable, S-Maps use all input variables to define the state space. At each point in time, the nearest neighbors are defined as the other points in time that have similar sets of all of the variables (as measured by Euclidean distance in state space), not just one variable. Coefficients are recalculated at each point in time, just as in the preceding models; however, regressions are weighted toward state-space coordinates that have similar states (i.e., similarly valued state-variables at a particular time). The variables included in these S-Map forecasts are the six TTFF variables as well as a sine and cosine term each with a 1-year period to represent the time of year.

Results and Discussion

Model Coefficients

When the TTFF was formulated, the static coefficients were associated with physical processes. For example, rain had a positive coefficient that was interpreted as more rainfall should increase overall flow (SFWMD 2020b). However, depending on the data used in the linear regression to formulate the TTFF, one can obtain either a negative coefficient, a positive coefficient, or a coefficient near 0 (Fig. 5). For example, in Model 1 (interannual predictor), we found that the linear regression calculated using only data from 1985 to 1990 yielded a negative coefficient for rain (Fig. 5). This highlights that it can be dangerous to make physical interpretations based on linear coefficients if the results change depending on the data used. In this case, we suggest that the influence of rain on flow targets can in fact change from positive to negative depending on the state of the system. For example, certain states may cause

rainfall to increase the downstream water level more so than upstream, which in turn may reduce the overall target flow.

Each model resulted in coefficients that change over time. For example, coefficients change for the interannual predictor [Fig. 5(a)] and seasonal predictor [Fig. 5(b)]. The interannual predictor reveals coefficients with significant temporal variation, exhibiting dynamics reflective of external interactions. There was a large excursion in the NESRS, ZA, and PET coefficients from the mid-1990s to 2000. This was a period of high water levels in the upstream WCAs, with accordingly negative influence of NESRS (downstream water levels) and positive forcing associated by upstream water availability (ZA).

The seasonal predictor recovers dynamics reflective of the South Florida summer monsoon, with a dry season from November through April, and a wet season from May to October. Here, WCA-3A and ZA (upstream water supply) closely reflect these monsoon patterns with a distinct shift from positive to negative coefficients in April and November. Furthermore, the downstream NESRS exhibits a delayed response consistent with water management releases.

Model Performance and Causal Inference

Fig. 6 compares the accuracy of all models tested. We used the mean absolute error (MAE) as our main metric for model accuracy because it is a meaningful value for managers implementing these predicted target values. The TTFF achieves a MAE of 7.2 m³/s. The interannual predictor performed the worst, with a MAE of flow forecasts of 7.3 m³/s. Three of the variable-specific predictors (using Rain, ZA, and PET to define the system state) also performed poorly (worse than the TTFF). The other three variable-specific predictors (using NESRS, WCA-3A, and flow to define state-space) outperformed the TTFF, achieving a MAE under 7.2 m³/s. The second-best model was the seasonal model, achieving a MAE of 6.9 m³/s. However, S-Map forecasts provided the highest fidelity, achieving a MAE of 6.5 m³/s.

The six variables selected as independent variables of the TTFF model make complete sense from the perspective of a hydrologic system. However, all variables may not provide significant information for improving forecasts.

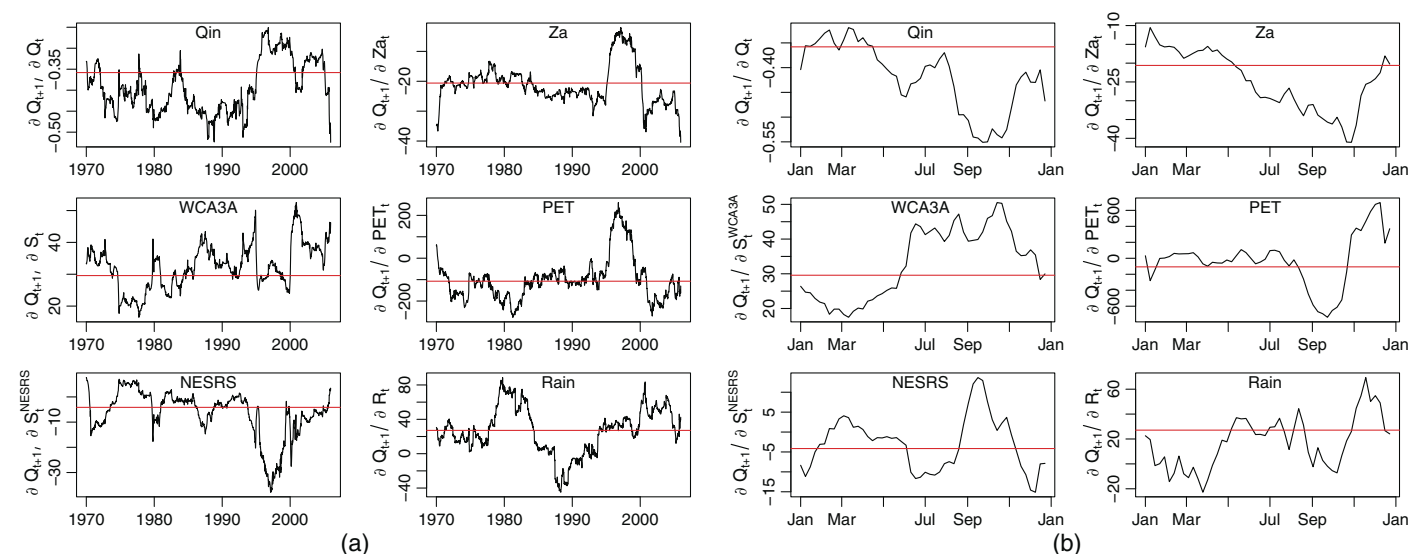


Fig. 5. (Color) (a) Recalculating the TTFF coefficients every 5 years yields different coefficients over time compared with the linear coefficients defined by the TTFF (red lines); and (b) recalculating the coefficients depending on the time of year reveals seasonal dynamics among the TTFF variables.

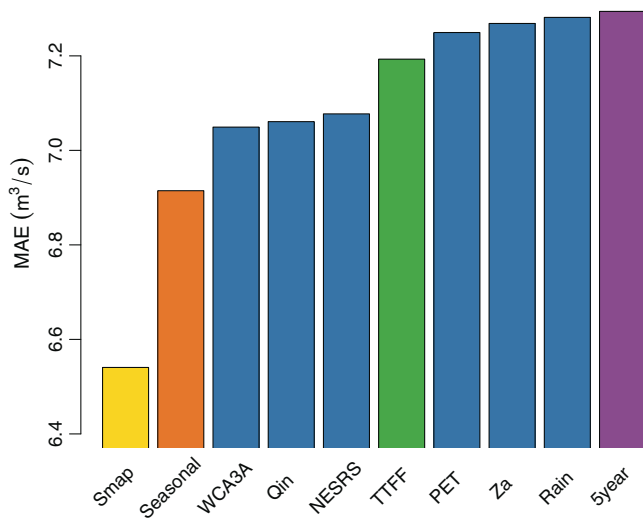


Fig. 6. (Color) Comparison of different predictive models' accuracy (mean absolute error between observed and predicted flow change). The TTFF (green) is a linear regression across all historical data. The 5-year predictor (purple) recalculates the TTFF coefficients in a 5-year moving window. The seasonal predictor (orange) recalculates the coefficients using historical data within 6 weeks of the given year date. Each variable (v) predictor (blue) makes forecasts from time t using only historical data with similar values of v_t . S-Maps (gold) account for all the nonlinearities among the variables to make a smooth state space that is not specific to the state of any one variable.

To verify that the five hypothesized variables are causal drivers of flow targets, we performed the EDM nonlinear causality test convergent cross-mapping (Sugihara et al. 2012). Despite the limited correlation between flow targets and these variables (Fig. 4), CCM revealed evidence for nonlinear coupling between all five variables and flow targets (Fig. S1). An introduction to CCM was presented by (Sugihara Lab 2015). Variables that have weak coupling (low CCM values, e.g., Rain and PET) do not necessarily provide useful information for improving predictions beyond the information provided by the strong drivers (e.g., upstream and downstream water levels). To further evaluate whether the five theorized driving variables of the TTFF are important for making predictions, we measured the performance of the S-Map predictor with variables removed one at a time (Fig. 7). Three of these variables (ZA, PET, and Rain) had little to no negative impact on overall predictions when removed. This suggests that these variables, although they were shown to be weak causal drivers with CCM, may not be important for defining the state space of the system; no matter what their values were, the dynamics of Q^{sum} did not change appreciably. As a further check, we performed an exhaustive assessment of state-space variable combinations using the EDM multiview algorithm (Fig. S2) (Ye and Sugihara 2016). The multiview approach tests the predictive accuracy of using different combinations of variables (with varying time delays) to reconstruct the state space. This gives a more complete measure of the importance of variables for making predictions (Fig. S2). Combined with the CCM results, these analyses confirmed that the variables ZA, PET, and Rain in the historical data, although they potentially are important, are not historically important for the overall goal of predicting integrated water flows on a weekly time scale across the Tamiami Trail.

We found that predictions were significantly hindered when WCA-3A and NESRS were removed (Fig. 6). Physically, this aligns

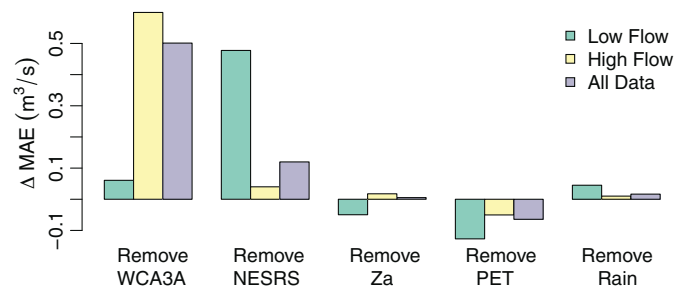


Fig. 7. (Color) Removing variables in S-Map forecasts to measure the impact on forecasts. The only two variables that had a significant negative impact on forecasts (increased MAE) when they were removed were water levels in the WC3A and NESRS regions.

with the fact that upstream (WCA-3A) and downstream (NESRS) water levels are the primary variables determining weir flow, whereas rain and PET accumulated over one week are integrated drivers of these upstream and downstream water levels. Furthermore, focusing on periods of high flow and low flow revealed that the WCA-3A water stage is more important for making forecasts during high flows, whereas the stage in the NESRS is more important when predicting low-flow regimes.

The accuracy of the S-Map forecasts was further improved to a MAE of 6.3 m³/s when ZA, PET, and Rain were excluded from the embedding. Fig. 8(a) shows this improvement compared with the performance of the TTFF on the test data set of weekly sampled data spanning 1965–2005. Furthermore, the prediction improvement varied depending on the flow: S-Maps outperformed the TTFF during all flow regimes, although periods of lower flow had the greatest improvement [Fig. 8(b)]. This also was true for a contemporary data set (weekly data from 2007 to 2020) [Figs. 8(c and d)].

The TTFF achieves a seemingly significant predictive accuracy with correlation between observed and predicted target flows of 0.90. However, upon inspection, it is apparent that such accuracy is not hard to achieve; simply predicting that next week's flow will be the same as this week's flow achieves a comparable correlation of 0.88. By removing variables one at a time from the TTFF, model performance stays essentially constant (Appendix). This suggests that relationships presumed by TTFF may not be as fully informative for forecasting dynamics of the system as one might presume.

Because the correlation between observed and predicted values is obscured by the high level of autocorrelation in the system, correlation is not the best metric to determine the significance of predictions. Here, we focused on mean absolute error as a standard for measuring predictive accuracy. Using S-Maps, we found an average improvement of 0.9 m³/s per weekly prediction (from a MAE of 7.2 to 6.3 m³/s). This translates to a predicted flow of over 500,000 m³ of water over the course of 1 week. Still, without a point of reference, the relative magnitude of this improvement is difficult to assess. We found that predicting the correct directional change (higher or lower next week than the current week) increased from 60% with the TTFF to 70% with S-Maps. We determined a null standard for this metric to be 55% by predicting that next week's change will be the same as that of the previous week (i.e., if the flow target increased last week, it will increase again next week). Thus, an improvement from 60% to 70% corresponds to an improvement from 5% above the null to 15% above the null. Furthermore, the correlation between predicted and observed changes in target flows from the prior week ($\Delta Q = Q_{t+1} - Q_t$) improved from $\rho = 0.45$ with the TTFF to $\rho = 0.58$ with S-Maps.

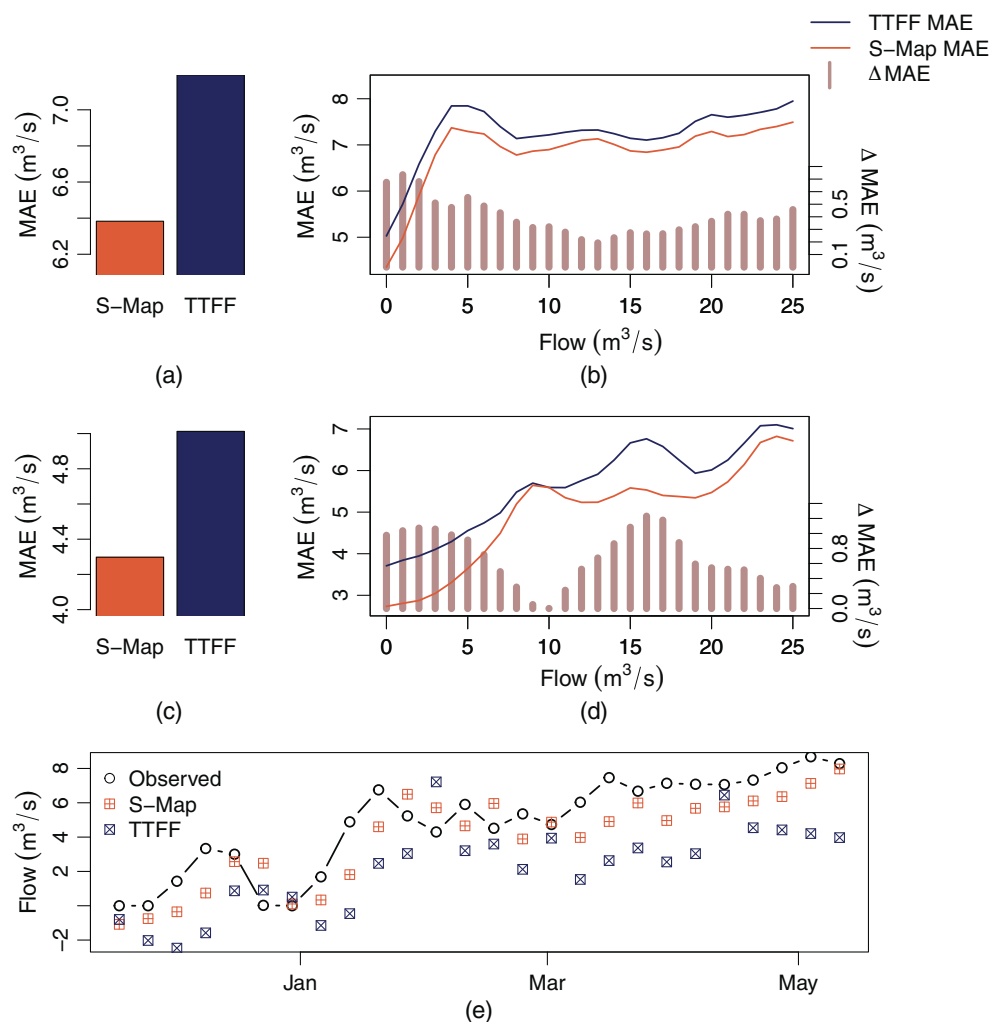


Fig. 8. (Color) (a) performance of S-Maps and the TTFF on the original data set (1965–2005); (b) average error for both forecasting algorithms as a function of flow for original data set (1965–2005), and the difference between the errors (maroon); (c) performance of S-Maps and the TTFF on contemporary data (2007–2020); (d) average error for both forecasting algorithms as a function of flow for contemporary data (2007–2020), and the difference between the errors (maroon); and (e) predictions made by the TTFF (blue) and S-Maps (gold) for a 6-month period with relatively low flow. S-Maps here do not utilize Rain, Za, or PET. S-Map forecasts significantly outperformed the TTFF during low-flow regimes.

State-Space (Nonlinear) Relationships

If a real-world system exhibits purely linear dynamics, reducing the amount of data used in the best-fit solution should hinder predictive accuracy because it reduces the signal-to-noise ratio (assuming equal amounts of noise throughout the time series). However, if partitioning the data into state-dependent subsets leads to improved predictions, the system dynamics are in fact different within each partition (i.e., the system is nonlinear). We found that certain partitions led to increased predictability compared with the general linear solution (TTFF), suggesting that this system is indeed nonlinear. Specifically, we found that the seasonal partitions performed best (aside from S-Maps), suggesting that the dynamics of this system are highly dependent on seasonal forcing.

A 5-year moving window performed the worst of the models tested, obtaining a MAE of $7.3 \text{ m}^3/\text{s}$ (Fig. 6). This suggests that the system does not change significantly on a year-to-year basis. However, that is not to say that dynamics do not change interannually at all; rather, the potential nonlinearity accounted for does not improve predictions more than the negative impact of using fewer data points, which reduces the signal-to-noise ratio. We did find that a 5-year window outperformed all other window sizes tested [ranging

from 2 to 20 years (Fig. S3)]. This suggests that window sizes that are too small have a large signal-to-noise ratio, whereas window sizes that are too large obscure the nonlinear processes.

The Zone A regulation potentially has a strong causal influence on flow targets; it had the third highest CCM value (Fig. S1) and was the third most important variable in multiview embeddings (Fig. S2). This likely was due to the strong seasonal forcing in this system; the Zone A regulation is a waveform with constant annual periodicity. Although this value may influence managed flows in the region, it more likely contributes to predictive models as a variable that helps define the time of year (season). This is affirmed by the predictive accuracy barely diminishing when it was removed from S-Map embeddings (Fig. 7), which already include sine and cosine terms to provide seasonal information.

When partitioning these results specifically into periods of high and low target flows, we found that water levels in WCA-3A are more important when making predictions during high-flow periods, whereas water levels in the NESRS are relatively more important during low-flow periods (Fig. 7). This may be explained by water management operations in this region: when upstream water levels (WCA-3A) are high, water is available for release into ENP.

Conversely, when downstream (NESRS) water levels are low, there is a need to release water to mitigate drought, likely at lower flow values.

The TTFF was formulated specifically using weekly data collected from 1965 to 2005. It is instructive to measure whether the predictive improvement obtained with S-Maps constructed using data from the same 1965–2005 period is still exists in recent data. S-Maps still outperformed the TTFF using data spanning 2007–2020 [Figs. 8(c and d)].

Although EDM outperformed the TTFF over the course of the entire time series on average, it still may be possible that the TTFF outperforms S-Maps during specific flow regimes. An important management concern of flows from the WCAs into ENP are low-flow regimes during dry season and drought conditions. S-Maps outperformed the TTFF during low flow (0–25 m³/s), especially during flows close to 0 m³/s [Figs. 8(b and d)]. Fig. 8(e) shows an example of the significant improvement gained using S-Maps.

The S-Map forecasts are fundamentally similar to those of the TTFF. The main differences are that (1) S-Maps utilize fewer variables (they do not include Rain, ZA, or PET), and, most importantly, and (2) S-Maps solve for linear fits only on similar states rather than on all historical data. However, just these two changes significantly improve forecasts. This demonstrates that nonlinear forecasting does not need to be complex; rather, it can be implemented almost as easily as linear formulations, and provides insight into nonlinear relationships.

Conclusion

A guiding principle of the Comprehensive Everglades Restoration Plan is to get the water right. This refers to restoring the quantity, quality, timing, and distribution of water throughout the greater Everglades system. This work focuses on the quantity aspect of this plan. A core component of this objective is management of water delivered across the Taimiami Trail from the upstream water conservation areas into Everglades National Park. This management is highly constrained by competing interests of hydroperiod and water depths for ecologic benefit, flood control for agricultural and urban interests, and water quality. These issues become particularly acute during the seasonal dry periods and droughts. Although efforts continue to remove barriers to natural sheetflow across the Trail, the active management of this complex, nonlinear objective is a fundamental lever in the water managers toolbox toward Everglades restoration.

This work highlighted the importance of model selection when dealing with real-world systems. In cases in which the system is multidimensional and dynamic, it is ambitious to assume that a single linear equation can describe the dynamics of a system. Despite this, such linear models often are favored due to their simplicity. However, significantly improved nonlinear approaches do not necessitate significantly complicated models. Here, we used the same linear regressive approach as was used to formulate the TTFF; however, we added a nonlinear perspective by partitioning the data into similar states. This effectively changed the focus of the model from determining the single set of rules that defines the system to determining the rules of this system when it appears at a specific point in time. This nonlinear perspective significantly improved predictions of weekly integrated flows from the WCAs into ENP, while also revealing dynamical truths about the system. Because nonlinear dynamics are ubiquitous in nature, such nonlinear approaches should also be ubiquitous in management efforts.

Table 1. TTFF prediction accuracy with removal of variables

Equation	Number of variables	Accuracy (ρ)	Accuracy (MAE)
1	6	0.903	7.20
2	5	0.903	7.20
3	4	0.903	7.21
4	3	0.889	7.71
5	2	0.886	7.68
6	1	0.884	7.27

Appendix. Reductions of TTFF

Reductions of TTFF (Table 1)

$$\hat{Q}^{\text{sum}}(t) \approx \alpha_1 S^{\text{WCA}}(t) + \alpha_2 S^{\text{ENP}}(t) + \alpha_3 Q^{\text{sum}}(t-1) + \alpha_4 \text{PET}(t) + \alpha_5 \text{ZA}(t) \quad (2)$$

$$\hat{Q}^{\text{sum}}(t) \approx \gamma_1 S^{\text{WCA}}(t) + \gamma_2 S^{\text{ENP}}(t) + \gamma_3 Q^{\text{sum}}(t-1) + \gamma_4 \text{ZA}(t) \quad (3)$$

$$\hat{Q}^{\text{sum}}(t) \approx \zeta_1 S^{\text{WCA}}(t) + \zeta_2 S^{\text{ENP}}(t) + \zeta_3 Q^{\text{sum}}(t-1) \quad (4)$$

$$\hat{Q}^{\text{sum}}(t) \approx \eta_1 S^{\text{WCA}}(t) + \eta_2 Q^{\text{sum}}(t-1) \quad (5)$$

$$\hat{Q}^{\text{sum}}(t) \approx Q^{\text{sum}}(t-1) \quad (6)$$

where α , γ , ζ , and η = linear regression fit coefficients.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request. An accessible Python package for performing EDM analyses is available at <https://pypi.org/project/pyEDM/>.

Acknowledgments

This work was funded in collaboration of the US Department of the Interior, the National Park Service, Everglades National Park, and the University of California San Diego through the Cooperative Ecosystem Studies Units (CESU) Network (<http://www.cesu.psu.edu/>). This work was supported by DoD-Strategic Environmental Research and Development Program 15 RC-2509, NSF DEB-1655203, NSF ABI-1667584, DOI USDI-NPS P20AC00527, the McQuown Fund, and the McQuown Chair in Natural Sciences, University of California San Diego.

Supplemental Materials

Figs. S1–S3 are available online in the ASCE Library (www.ascelibrary.org).

References

- Chang, C.-W., M. Ushio, and C.-H. Hsieh. 2017. "Empirical dynamic modeling for beginners." *Ecol. Res.* 32 (6): 785–796. <https://doi.org/10.1007/s11284-017-1469-9>.
- Deyle, E. R., M. C. Maher, R. D. Hernandez, S. Basu, and G. Sugihara. 2016. "Global environmental drivers of influenza." *Proc. Natl. Acad.*

- Sci. U.S.A.* 113 (46): 13081–13086. <https://doi.org/10.1073/pnas.1607747113>.
- Deyle, E. R., and G. Sugihara. 2011. “Generalized theorems for nonlinear state space reconstruction.” *PLoS One* 6 (3): e18295. <https://doi.org/10.1371/journal.pone.0018295>.
- Dixon, P. A., M. J. Milicich, and G. Sugihara. 1999. “Episodic fluctuations in larval supply.” *Science* 283 (5407): 1528–1530. <https://doi.org/10.1126/science.283.5407.1528>.
- Lorimer, T., R. Goodridge, A. K. Bock, V. Agarwal, E. Saberski, G. Sugihara, and S. A. Rifkin. 2021. “Tracking changes in behavioural dynamics using prediction error.” *PLoS One* 16 (5): e0251053. <https://doi.org/10.1371/journal.pone.0251053>.
- National Academies of Sciences. 2018. *Progress toward restoring the everglades: The seventh biennial review: 2018*. Washington, DC: The National Academies Press.
- National Research Council. 2008. *Progress Toward Restoring the Everglades: The Second Biennial Review—2008*. Washington, DC: The National Academies Press.
- Saberski, E., A. K. Bock, R. Goodridge, V. Agarwal, T. Lorimer, S. A. Rifkin, and G. Sugihara. 2021. “Networks of causal linkage between eigenmodes characterize behavioral dynamics of *Caenorhabditis elegans*.” *PLoS Comput. Biol.* 17 (9): e1009329. <https://doi.org/10.1371/journal.pcbi.1009329>.
- Segundo, J. P., G. Sugihara, P. Dixon, M. Stiber, and L. F. Bersier. 1998. “The spike trains of inhibited pacemaker neurons seen through the magnifying glass of nonlinear analyses.” *Neuroscience* 87 (4): 741–766. [https://doi.org/10.1016/S0306-4522\(98\)00086-4](https://doi.org/10.1016/S0306-4522(98)00086-4).
- SFWMD (South Florida Water Management District). 2020a. “Combined operational plan for water deliveries from water conservation area 3A to Everglades National Park: Tamiami Trail Flow Formula.” Accessed August 1, 2021. <https://usace.contentdm.oclc.org/utis/getfile/collection/p16021coll7/id/15783>.
- SFWMD (South Florida Water Management District). 2020b. “Regional simulation model (RSM).” Accessed August 1, 2021. <https://www.sfwmd.gov/science-data/rsm-model>.
- Sugihara, G. 1994. “Nonlinear forecasting for the classification of natural time series.” *Philos. Trans. R. Soc. London, Ser. A* 348 (1688): 477–495. <https://doi.org/10.1098/rsta.1994.0106>.
- Sugihara, G., W. Allan, D. Sobel, and K. D. Allan. 1996. “Nonlinear control of Sugihara et al. 1996 rate variability in human infants.” *Proc. Natl. Acad. Sci. U.S.A.* 93 (6): 2608–2613. <https://doi.org/10.1073/pnas.93.6.2608>.
- Sugihara, G., M. Casdagli, E. Habjan, D. Hess, P. Dixon, and G. Holland. 1999. “Residual delay maps unveil global patterns of atmospheric nonlinearity and produce improved local forecasts.” *Proc. Natl. Acad. Sci. U.S.A.* 96 (25): 14210–14215. <https://doi.org/10.1073/pnas.96.25.14210>.
- Sugihara, G., R. May, H. Ye, C. H. Hsieh, E. Deyle, M. Fogarty, and S. Munch. 2012. “Detecting causality in complex ecosystems.” *Science* 338 (6106): 496–500. <https://doi.org/10.1126/science.1227079>.
- Sugihara, G., and R. M. May. 1990. “Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series.” *Nature* 344 (6268): 734–741. <https://doi.org/10.1038/344734a0>.
- Sugihara Lab. 2015. “Introduction to empirical dynamic modeling.” Posted August 29, 2015. YouTube video. . <https://www.youtube.com/watch?v=fevurdpiRYg&list=PL-SSmlAMhY3bnogGTe2tf7hpWpl508pZZ>.
- Takens, F., 1981. “Detecting strange attractors in turbulence.” In *Dynamical systems and turbulence, Warwick 1980*, 366–381. Berlin: Springer.
- USACE. 2017. “Combined operational plan (COP).” Accessed August 1, 2021. <https://usace.contentdm.oclc.org/digital/collection/p266001coll1/id/4300/>.
- van Nes, E. H., M. Scheffer, V. Brovkin, T. M. Lenton, H. Ye, E. Deyle, and G. Sugihara. 2015. “Causal feedbacks in climate change.” *Nat. Clim. Change* 5 (5): 445–448. <https://doi.org/10.1038/nclimate2568>.
- Ye, H., and G. Sugihara. 2016. “Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality.” *Science* 353 (6302): 922–925. <https://doi.org/10.1126/science.aag0863>.