Hudson Valley Transmission Line Plan: Updated Analysis of Need & Alternatives

Gidon Eshel

www.environmentalCalculations.com,

Research Professor of Environmental Physics, Bard College

Email ✉

Disclaimer: *The views and analyses presented in this document are solely those of the author. The author's long term and/or ad hoc affiliations and appointments do not imply endorsement by any party of this document or the scientific findings it conveys. The author is the sole contributor to this document.*

Copyright ©April 2015 Gidon Eshel.

Executive Summary: The main findings of this report are: (1) No additional transmission capacity into the downstate region is needed. In fact, an updated model built, presented and tested below, shows even *lower* likelihood of peak loads exceeding capacity than previously estimated. (2) NYISO's critiques of the Eshel (2014) report fall into two categories. The first comprises points that are emphatically wrong on technical, conceptual or—most often—technical *and* conceptual grounds, as shown in detail below. The second class comprises points that *may* be true, but make no difference when included in the calculation at face value. (3) NYISO projections systematically overestimate future downstate peak load, for three independent reasons. First, NYISO's reliance on questionable GDP projections is unwarranted. Second, NYISO failure to distinguish the role of trends and covarying fluctuations about them in evaluating their model skill at fitting past observations is erroneous. Third, NYISO's treating that skill as indicative of the model skill to forecast future peak loads is incorrect. Each of these errors is individually serious. When serialized, they discredit the forecasted peak loads. Comparison with a carefully constructed, tested and validated model shows that these forecasts are skewed consistently up.

1 Background

The Cuomo administration has indicated its wish to add one GW of north–south power transmission capacity along the Hudson Valley. The NY PSC followed up, soliciting proposals from interested developers. In response to these developments, I spent nine weeks during the summer of 2014 analyzing the proposal.

I embarked on this mostly uncompensated volunteer effort for several reasons. **First**, because I am a concerned citizen of New York who wishes his State to become a national leader in devising a new model for a reliable, efficient and environmentally sound power distribution network that takes nimble advantage of modern technologies. **Second**, because at the heart of the issue are the aspects of my academic profession I am most passionate about: analysis of resource use efficiency, and statistical forecasting; I am always on the lookout for good, environmentally important, interesting problems in those realms that lend themselves well to being solved using the tools in which I am an expert and which I partly developed. Both are true about this problem, as will become clear soon. **Third**, my family's home in Milan, NY is situated on a land parcel transected by the western branch of the existing Right of Way. As such, we may be impacted by possible eminent domain widening of the ROW. While it may be tempting to some to conclude that this latter point is my primary motivation, note that we bought the house just a few years ago, fully mindful of the ROW, the easement on a portion of our land, and the fact that our views may be undermined if the ROW corridor is expanded to westward. In all sincerity, my interests in the public good, and efficiency, far exceed any personal motivation.

I chose to focus exclusively on the asserted but, astonishingly, never previously quantified need for the project. These efforts resulted in a detailed technical document entitled *Hudson Valley Transmission Line Plan: Assessing Need & Alternatives*, and a non-technical summary of the main document. In keeping with standard academic publishing practices, once completed, the technical report was sent by a third party to two anonymous reviewers. While I know nothing about the reviewers' identity, background, or academic affiliation, I was explicitly assured by the third party handling the review process that both are respected professionals in academic–scientific disciplines in which statistical forecasting is widely used and in which rigorous standards are upheld. The work was revised in complete accordance with the reviews received anonymously, addressing every single point raised by the reviewer.The Eshel (2014) report thus meets the gold standard of academic publishing and scientific advances: having been vetted anonymously by colleagues in the context of peer review. The Eshel (2014) report, and its summary, are publicly available here.

The crux of Eshel (2014) findings was that the proposed project is unnecessary; no shortage of transmission capacity is evident in the best available data.

On December 29th 2014, NYISO published a document entitled *Comments of the New York Independent System Operator, Inc.* (addressed to Secretary Burgess of the NY PSC; hereafter "NYISO Comments" or "the Comments") which strove to (1) rebut the Eshel (2014) results and discredit the Report; (2) call into question the author's suitability for carrying out the reported work; and—apparently in anticipation of efforts 1–2 falling short—(3) advance alternative justifications for the project despite lack of need.

In the current document, I respond to NYISO's critique and place it in context.

2 On The Need for Technical Resolution

Unfortunately, parts of this report will likely prove prohibitively technical to some. This is not meant in any way to alienate or exclude readers. Rather, it is dictated by the nature of the problem, whether or not current assets will likely meet future peak loads, or fall short. This is a very technical problem, as the following sections explain, and its solution lies squarely within statistical forecasting, modeling and related, inevitably technical branches of applied mathematics.

Because of the disparity between the technical level of the problem and the comfort level of many readers of this report, to make up his or her mind, the nontechnical reader will likely wish to employ indirect judgment. In many such cases, the favored point of view is often chosen based on trust in the experts of one side vs. the other[1]. I sincerely hope this report will instill readers' confidence in the technical foundations of the conclusion that **no additional assets are necessary to meet probable future downstate peak loads**.

3 The Nature of the Problem Whose Solution is Disputed

The problem has two unequal parts.

First and foremost is devising an answer to the most elemental question: **Is the proposed project necessary?** If it is not, logically no further steps are necessary or warranted. This question is answered by forecasting future peak electricity loads—those occurring during a few hot, muggy summer afternoons of each year—in the downstate area, comprising New York City, Long Island's two counties, and Westchester and Rockland counties.

An optional secondary element may address **environmental alternatives**. Importantly, this element is only required if one continues to deem the project necessary despite the fact that its proponents never demonstrated any need, and the aforementioned Eshel (2014) report, the only quantitatively authoritative analysis of the problem to date, shows *no* need.

4 On the Absence of a Clear Statistical–Methodological Statement of NYISO's Peak Load Calculations

A key obstacle to a productive dialog with NYISO is the lack of transparency. Most important for the purposes of this report in particular and for settling cogently the dispute over the need for the proposed project in general is the lack of a detailed statement, with equations, of the model on which NYISO's peak load projections are based.

Modeling not accompanied by unambiguous statement of the methods used is the antithesis of established scientific practices. Yet forecasting future peak loads is an applied scientific problem. No scientific problem is ever considered adequately solved without future reproducibility facilitated by untempered access to all materials used in deriving the asserted scientific conclusions. NYISO's failure to comply with these basic standards and their repeated referencing of proprietary data products and models are deeply troubling.

Absent detailed information about their forecasts, the best yet imperfect window into NYISO's work is provided by their rebuttal Comments. It is these comments, therefore, that the remainder of this document addresses.

5 Specific NYISO Technical Points

From here on, technical arguments assume center stage. To help the less technically inclined readers, in the Appendix, section 9, I provide a short primer on forecasting.

5.1 R^2 is Suggestive, Not Definitive Quantification of Statistical Forecasting Model Skill; Cross Validation is Crucial

KEY POINTS: **One of the key challenges of successful, meaningful statistical forecasting involves combining physical intuition with statistical considerations to choose the right predictors and discard inadequate ones. Throughout the Comments, NYISO employs an inappropriate criterion for choosing model predictors, thus reaching seriously erroneous conclusions.**

Throughout the Comments, NYISO strives to bolster various of their points by showing higher R^2 or lower error of models they favor relative to the model Eshel (2014) used (the Comments' Figs. 3 & 4; here those measures are each other's redundant mirror image).

This is an often serious error in which forecasting—which is about making statements about the *future*—is willy nilly confused with fitting the observed past. As Niels Bohr once noted "prediction is very difficult, especially about the future."[1] NYISO's model choices based only on the models' skill in fitting the past is never sufficient, and often very misleading.

Using hindcast and cross validation forecast skills synonymously is common in exploratory data analysis. In such contexts, regression is used to devise an empirical mechanistic explanation of the variability of the studied phenomenon in terms of other explanatory variables that are better observed or understood in some situation. For example, the rather tight relationship between a person's height and weight allows us to predict the weights of members of a group of people whose weights are unknown but whose heights happened to be known.

This does not apply to forecasting, in which one is not trying to explain the variability of a studied phenomenon in terms of explanatory variables, but is rather striving to predict future values. Even if R^2 is the favored measure of model skill (and not often more revealing alternatives such as *Akaike Information Criterion, AIC*), it must be evaluated by cross validation, not simple hindcast as in NYISO's Comments.

5.1.1 Cross Validation Protocol Introduced

KEY POINTS: **In cross validation, one splits the available predictor and predictand data into a part used for model optimization [denoted by an (r) superscript below] and another part used for the all important model validation [denoted by a (w) superscript below]. Cross validating a statistical model is not always enough, but its absence is never acceptable.**

To better understand this point, and for clarity of later parts, let me now introduce cross validation in rather general terms. Let $\mathbf{A}$ denote a matrix whose columns are time series of predictors, and let the vector $\mathbf{e}$ similarly hold the time series of the phenomenon being forecasted. In cross validation, $\mathbf{A}$ and $\mathbf{e}$ are split into two parts, the retained and withheld ones,

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}^{(w)} \\ \mathbf{A}^{(r)} \end{pmatrix} \qquad \text{and} \qquad \mathbf{e} = \begin{pmatrix} \mathbf{e}^{(w)} \\ \mathbf{e}^{(r)} \end{pmatrix}, \tag{1}$$

with which the forecast for the withheld part of $\mathbf{e}$ is[2]

$$\hat{\mathbf{e}}^{(w)} = \mathbf{A}^{(w)} \left[\left(\mathbf{A}^{(r)T} \mathbf{A}^{(r)} \right)^{-1} \mathbf{A}^{(r)T} \mathbf{e}^{(r)} \right]. \tag{2}$$

That is, the model is developed and optimized (done by the bracketed part of the right hand side above) with *no* knowledge of the withheld values whatsoever, and then the optimized solution thus obtained is used to calculate the expected e values for the withheld points, the elements of $\hat{\mathbf{e}}^{(w)}$.

5.1.2 A Very Simple Example

KEY POINTS: **Using a simple idealized example, this section shows how terribly wrong forecasting can sometimes be even when using a model whose ability to fit past**

observations is unquestionably excellent. While Fig. 1a shows an increasing model skill with increasing number of predictors, Fig. 1f shows that those competing models' forecasting skills become rapidly catastrophic with increasing model complexity.

With this introduction, we are now ready to demonstrate the danger in confusing hindcast and forecast skills. Let's start with a simple case, and then proceed to more realistic situations pertinent to the problem of forecasting downstate peak loads. Fig. 1 presents a very simple case based on analyzing a specific realization of the general generating process

$$b(t) = 20 + \mathcal{N}(0, 10^2) \tag{3}$$

where to avoid confusion with subsequent realistic problems, in the special case of this simple example,

$$\mathbf{b} = \begin{pmatrix} b(t_1) & b(t_2) & \cdots & b(t_{20}) \end{pmatrix}^T, \tag{4}$$

plays the role of the general predictand $\mathbf{e}$. Also in Eq.3, the right hand side depicts a time dependent predictor given by random draws from a normal distribution with zero mean and variance $= 10^2$ units. The specific case being analyzed in Fig. 1 is generated for $t = [1, 20]$ and split into two equal parts. The first, $t = [1, 10]$ (shown by red dots in Fig. 1b–e), corresponds to $\mathbf{A}^{(r)}$ and $\mathbf{e}^{(r)}$ and is used to optimize the model. The latter part, $t = [11, 20]$ (blue dots in Fig. 1g–j), corresponds to $\mathbf{A}^{(w)}$ and $\mathbf{e}^{(w)}$ ($\mathbf{b}^{(w)}$ in the simple example) and is used for forecast validation.

Let's now use the general cross validation formalism to demonstrate why NYISO's choice of the preferred model based on hindcast skill of past observations is misguided and leads to fatally wrong conclusions. In Fig. 1, I strive to both hindcast and forecast the above $b(t)$ values (the time series generated by Eq. 3, held in the vector $\mathbf{b}$ given in Eq. 4, and plotted as solid circles in Fig. 1) in terms of polynomials of increasing degrees. That is, the ith column of $\mathbf{A}$ is a $p_{ji} = t_j^{i-1}$ so that, taking $n_p = 2$ as an example,

$$\mathbf{A} = \begin{pmatrix} t_1^0 & t_1^1 & t_1^2 \\ t_2^0 & t_2^1 & t_2^2 \\ & \vdots & \\ t_{20}^0 & t_{20}^1 & t_{20}^2 \end{pmatrix} = \begin{pmatrix} 1 & t_1^1 & t_1^2 \\ 1 & t_2^1 & t_2^2 \\ & \vdots & \\ 1 & t_{20}^1 & t_{20}^2 \end{pmatrix} \in \mathbb{R}^{20\times 3}. \tag{5}$$

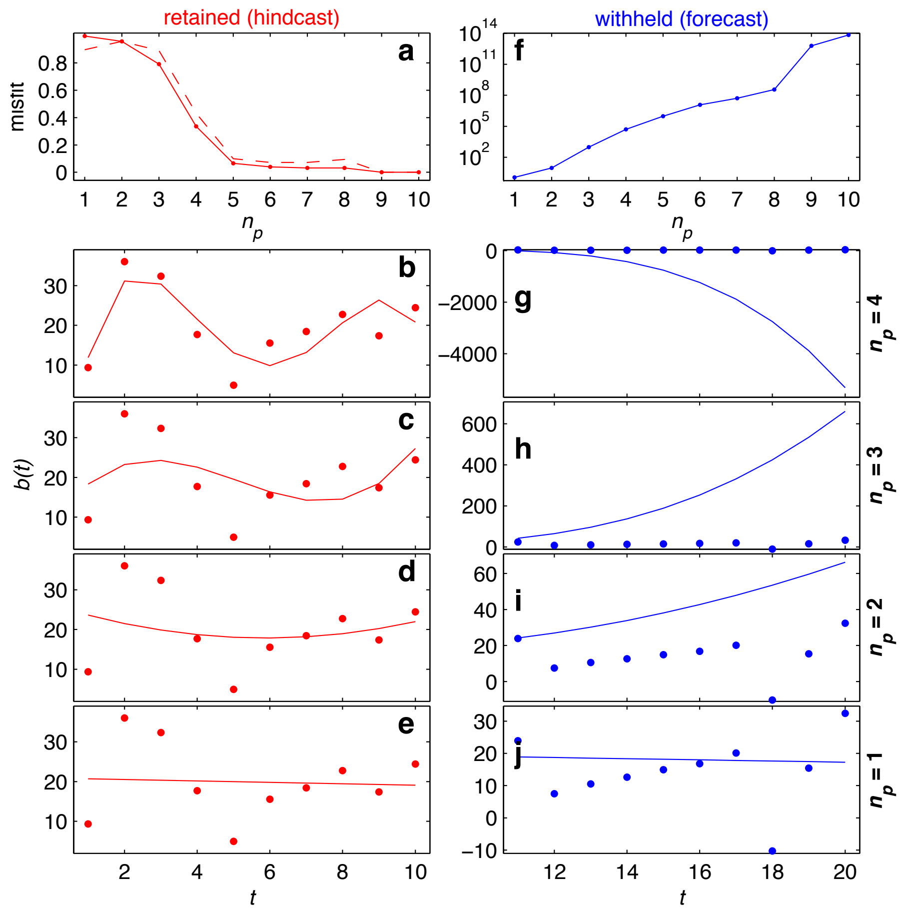

Figure 1: A demonstration of the severe limitations of hindcast R^2 as a basis for choosing a sound model. See text for model details. Panel a shows the hindcast misfit as a function of the highest degree of the polynomial fit, while panel f shows the same for the forecast error. Under each, the actual time series (filled dots) of the retained (blue) and withheld (red) portions of the timeseries and various polynomial estimates of it, with panels e and j show degree 1, panels d and i show degree 2, and so on, as indicated on the right, where the polynomial's highest degree is denoted by n_p.

The reason here I choose a power series representation of the predictand (polynomial rows for $\mathbf{A}$) is that with this choice, the hindcast fit is guaranteed to steadily improve, up until the limiting special case of $n_p = 10$, at which point the problem switches from the least squares problem is has been for $1 \leq n_p \leq 9$ to an interpolation problem (with an exact solution and thus $R^2 = 1$ because with $n_p = 10$ $\mathbf{A}$'s row dimension is equal to the number of data points being fitted).

This expected behavior is indeed seen in Fig. 1a, where the error steadily declines with increasing polynomial degree (number of predictor columns in $\mathbf{A}$). Fig. 1b–e shows the declining error for $1 \leq n_p \leq 4$.

So far, it seems, so good: our model skill steadily improves. But the colossal mistake this conclusion constitutes is immediately apparent upon examining Fig. 1f, in which I test the utility of the derived model not to hindcast values used during its development, but new ones the model has not been cognizant of before. In stark contrast to Fig. 1a, in which the error declines with increasing n_p, the reverse happens in Fig. 1f. Not only does the error increase, it increases astronomically fast, spanning 12 order of magnitude over $1 \leq n_p \leq 10$. **Clearly, confusing hindcast model skill with forecasting power is an unacceptably egregious error.**

5.1.3 Application to Downstate Peak Load Forecasting

Key Points: **While the preceding section shows that confusing hindcast with forecast skills *can* be dangerous, this section shows that this danger is unfortunately encountered in the problem of forecasting future downstate peak loads.**

To demonstrate next the extreme importance of distinguishing hindcast from forecast skill for the realistic issue of downstate peak load, Fig. 2 compares hindcast and cross validated forecast skills of five single predictor models. If the predictor (e.g., total population in the leftmost panel) at year i is denoted generically p_i, and downstate peak load of the same year is denoted e_i, the model is

$$\mathbf{A}\mathbf{x} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & p_1 \\ 1 & p_2 \\ & \vdots \\ 1 & p_N \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{e} \tag{6}$$

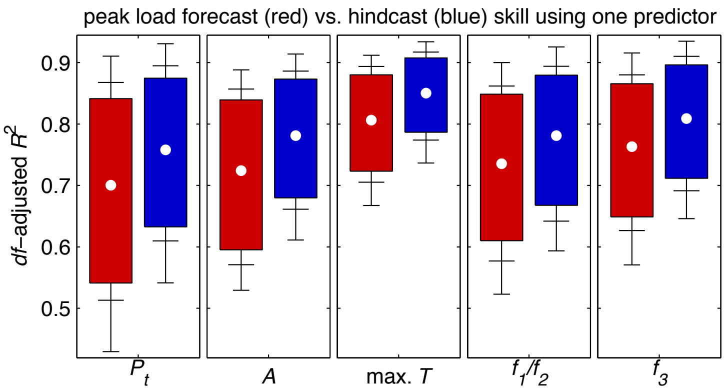

Figure 2: Skill of predicting downstate peak load using each one of the five considered predictors individually (from left to right: total downstate population, affluence, maximum temperature during peak load, the ratio of people ages 45–70 to those 20–45, and the fraction of people older than 70). Red is cross validated forecast, while blue is hindcast, as explained in the text. Colored rectangles encompass the central 90% of the cross validation populations, with wide and narrow horizontal whiskers encompassing 95% and 99%. Means are indicated by white circles. The figure shows that the forecasting skill (red) of each one of the models is appreciably lower than the hindcasting skill (blue). While the entire distribution degrades, much of the degradation occurs due to large increases in low skill forecasts absent from or rare in the hindcast (see the larger downward displacement of the bottom of the red bars than the same displacement of the top).

where the explicit left and right hand sides define $\mathbf{A}$, $\mathbf{x}$ and $\mathbf{e}$ respectively. For hindcast [and assuming $\mathbf{e}$ is not a uniform vector parallel to $\mathbf{A}$'s left column, so that $\text{rank}(\mathbf{A}) = 2$], this is solved in the canonical least squares form (Eshel 2012[3]),

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{e}. \tag{7}$$

In the current problem—with annual resolution data for 1993–2013, so that there are 21 data points and thus $\mathbf{A} \in \mathbb{R}^{21\times 2}$—for most applications it is sufficient to withhold three points at a time (i.e., $\mathbf{A}^{(w)} \in \mathbb{R}^{3\times 2}$ and $\mathbf{A}^{(r)} \in \mathbb{R}^{18\times 2}$), with the withheld years being $(1, 2, 3)$, $(1, 2, 4)$, $\cdots (1, 2, 21)$, $(1, 3, 4)$, $(1, 3, 5)$, $\cdots (1, 3, 21)$, $\cdots \cdots (19, 20, 21)$. Using the appropriate binomial coefficient, there are

$$\begin{pmatrix} 21 \\ 18 \end{pmatrix} = 1330 \tag{8}$$

such combinations, and most cross validation populations in this write-up are thus 1330 elements large.

In cross validation, then, we evaluate any considered model not based on its weakly relevant ability to fit past observations, but based on its highly relevant skill to predict values completely absent from the model development and optimization stages. This is the gold standard of suitably validated forecast models, and it is unfortunately entirely absent from NYISO's forecasting efforts.

Consistent with this knowledge, Fig. 2 shows that, as expected, in each one of the five single predictor models considered, forecast skill is discernibly lower than the hindcast skill. Importantly, the skill deterioration is not symmetrical, but instead skewed toward likelier low skill forecasts, with the bottom of the red bars and corresponding whiskers significantly more strongly degraded than the upper end of the population. Thus all skills decrease in forecasting relative to hindcasting, but unusually low skill forecasts arise more frequently.

5.2 Distinguishing the Roles of Trends and Fluctuations About Them

KEY POINTS: **Successful, meaningful statistical forecasting requires covariability of predictor and predictand. Covariability is about matching fluctuations: rise or fall of the predictor are more often than not accompanied by a corresponding, coherent rise or fall in the predictand. Trends—slow, steady highly structured changes in time—are antithetical to such predictor–predictand covariability, and do not constitute acceptable foundation for successful forecasting. Yet much of NYISO's modeling is trend based, leading to seriously erroneous conclusions.**

Another fundamental misunderstanding of statistical forecasting the NYISO Comments reveal involves identifying the role of trends in the presumed predictive power of individual predictors. Despite being a rudimentary element of the trade, mishandling trends is typically fatal, thoroughly invalidating any following element of the analysis. Such mishandling is unfortunately the hallmark of the NYISO prediction process. While thoroughly illuminating the issue is beyond the scope of this response (see Eshel 2012[4] for richer details), let me summarize it briefly below.

In the time series analysis literature, a "trend" typically describes any coherently structured non-cyclical element of the data. Regression or correlation analyses are about revealing covariability: how random looking point by point fluctuations in the explanatory variable *covary* coherently with corresponding fluctuations in the explained variable. Trends are the diametric opposite. First, in general they are unique attributes of individual variables, not a predictor-to-predictand bridge. Most importantly, they are time invariant or vary coherently and slowly in time, and are visually and analytically the antithesis of the above "random looking point by point fluctuations".

There are several risks associated with relying on trends for statistical forecasting. Among those are (1) trends tend to eventually run out or change quantitatively or qualitatively; (2) even if trends persist, predictor and predictand trends in general march to different, unrelated drums, and thus the nature of the trend based correspondence between predictor and predictand is likely to collapse; (3) trends likely reflect dynamics of other processes, not governing dynamics of the addressed variables; and (4) they tend to have far fewer degrees of freedom than the modeled series, and can thus yield highly inflated apparent model skill.

Let's examine some examples that illuminate the above shortcomings. Fig. 3 presents both observed U.S wide civilian workforce employment and projections thereof issued by the U.S. Dept. of Labor, Bureau of Labor Statistics[5]. While the methodology of the projections is beyond the current scope, they are clearly strongly influenced by preceding employment trends. The figure shows that while the 1992 and 2000 forecasts proved reasonably accurate (although, importantly, systematically *over*estimated employment), subsequent forecasts gave no hint of the 2008 recession in either 2006 or even, incredibly, in 2008. Consequently, the prediction error of the unduly optimistic forecasts roughly quadrupled from Fig. 3b to Fig. 3d. This makes it clear that since nothing persists indefinitely, it is only a matter of time before trend based forecasting reaches hopelessly erroneous conclusions.

A simple example of points (2) and (3) above addresses upper oceans dissolved CO_2 and temperature. In recent decades, throughout most of the world ocean both increased in concert. This makes it seductive to treat temperature as a skillful predictor of oceanic CO_2, or vice versa. Yet major volcanic eruptions, with no known predictability, will disrupt this lockstep evolution, cooling the earth by reflecting solar radiation to space while increasing atmospheric concentrations of CO_2, and thus rates of atmosphere-to-ocean CO_2 invasion. The result is a hiatus in the warming, accompanied by accelerated upper ocean dissolved CO_2 increases, entirely at odds with recent decades history. No, temperature and CO_2 are *not* competent predictors of each other. Finally, point (4) is trivial, as a straight line slope, e.g., has two degrees of freedom (mean and slope), and more elaborate trends have just a few more.

Let's now look at trends exhibited by the downstate peak load time series and two of its

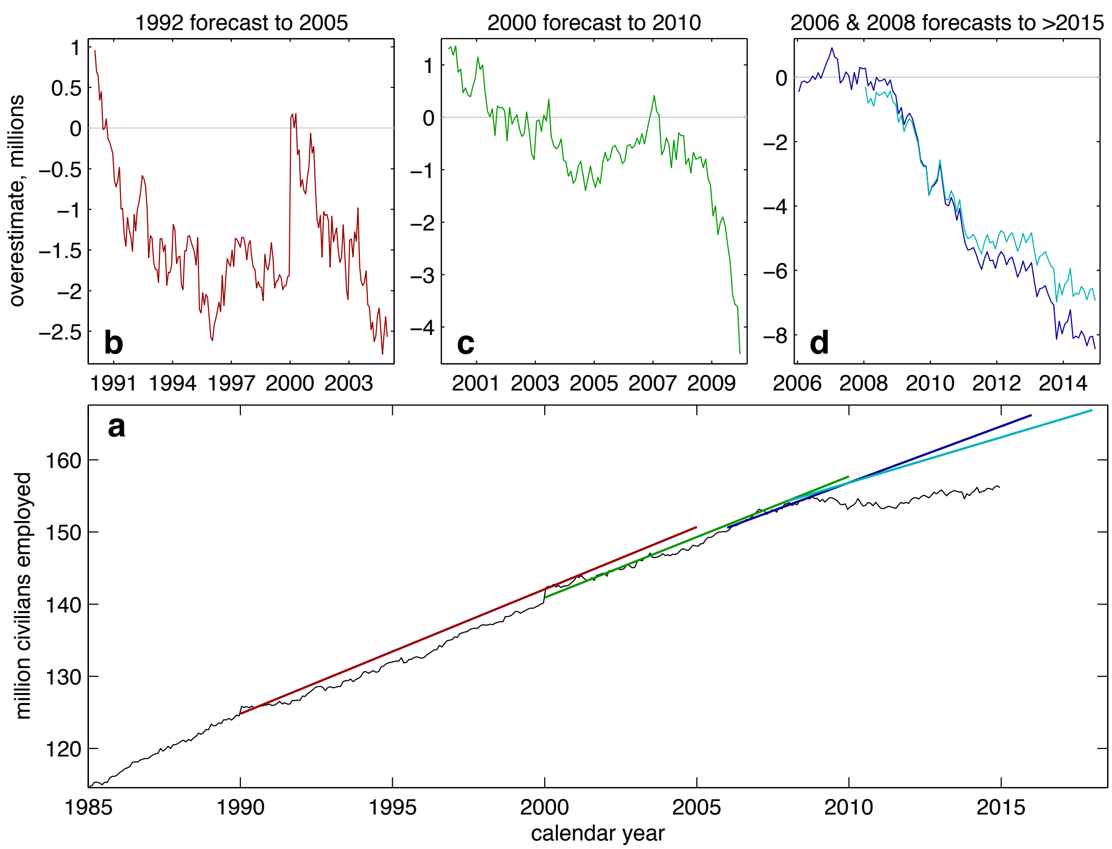

Figure 3: Some risks associated with trend based forecasting. Panel a shows in solid black the observed employment among the civilian U.S. workforce, along with three forecasts, for 1990–2005, 2000–2010, 2006–2016, and 2008–2018. Both data and archived forecasts are from the U.S. Dept. of Labor, Bureau of Labor Statistics. To better appreciate the nature of forecast errors, panels b–d zoom in on the forecast overestimation for the forecast horizons, in the same colors. Note that the errors are not random (fluctuating incoherently with equal likelihood of over- and underestimation) but are heavily skewed toward overestimation, a direct telltale consequence of over-reliance on past trends.

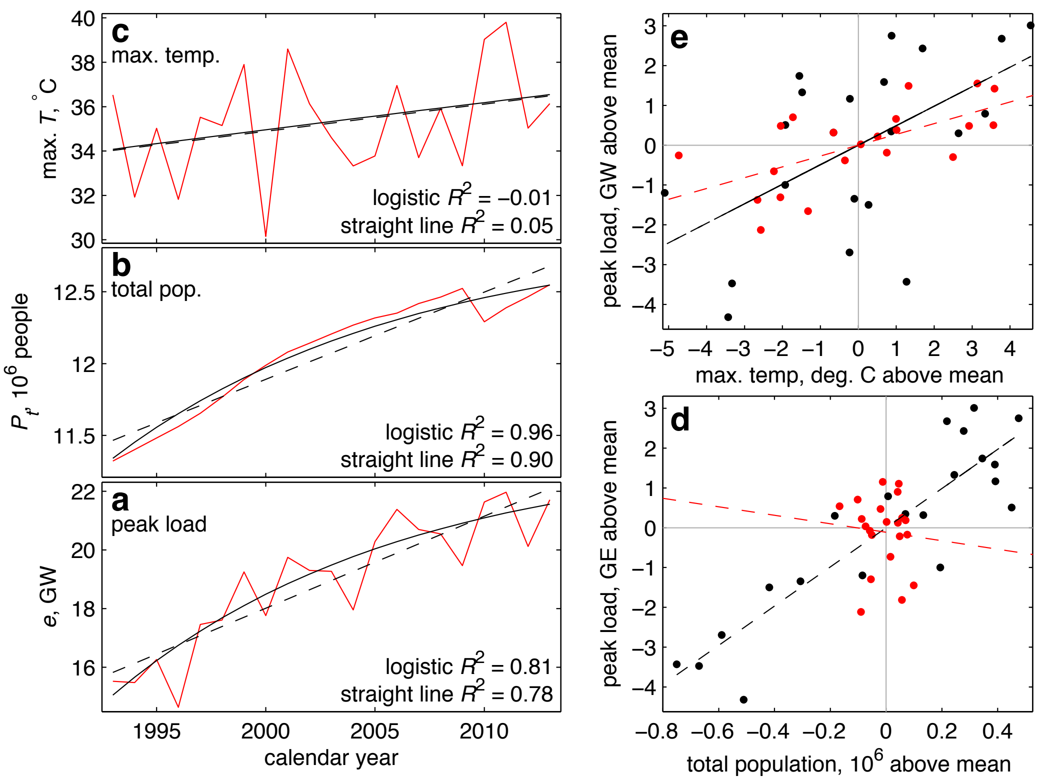

Figure 4: Illustration of the necessary nature of trend analysis, using the relationship of total population and maximum temperature to downstate peak load. In a–c, the observed peak load (a), population (b) and maximum temperature (c) are plotted in red, along with two best fit least squares trends, a straight line trend in dashed black and the logistic in solid black. The R^2 of the fits to the data are reported in the lower right. The superiority of the logistic trend for peak load and total population, and of the uniform slope trend for maximum temperatures are also supported by AIC, not shown here. Panels d–e show the impact of retaining or discarding the trend, and the choice of the appropriate trend, on the relationship of population (d) or maximum temperature (e) with downstate peak load (the vertical axis in d and e). See text for the details of the trend calculations.

explanatory variables, or predictors, with the aid of Fig. 4. Panels a–c examine the trends exhibited by these three variables (a being the predictand, and b–c addressing two of its predictors). For each of the three time series, I fit two trends. The first is a uniform slope straight line (dashed in Fig. 4a–c), given for any variable v by the choice of $\hat{\mathbf{y}} = (\mu,\ \psi)^T$ that minimizes the misfit in

$$\mathbf{G}\mathbf{y} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ & \vdots \\ 1 & t_N \end{pmatrix} \begin{pmatrix} \mu \\ \psi \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{v} \tag{9}$$

(i.e., the $\mathbf{y}$ that renders φ, the sum of squared errors

$$\varphi = (\mathbf{G}\mathbf{y} - \mathbf{v})^T (\mathbf{G}\mathbf{y} - \mathbf{v}) \tag{10}$$

the global minimum).

The second trend explored in Fig. 4 is logistic, given for variable v by

$$v_i = v^{(0)} \left[1 - e^{-(\zeta+\eta i)}\right] \tag{11}$$

where i is the year index, with $t_1 = 1993$ and $t_{21} = 2013$, and with general shape best visualized by the solid black curves in Fig. 4a,b. For choosing the better fitting trend, here I use the *adjusted* R^2 (i.e., the R^2 corrected for the unequal number of degrees of freedom of the trend and the sum of squared residual, see, e.g., Ryan 1997[6], p. 224). Similar conclusions are also reached by relying on the Akaike Information Criterion AIC (see, e.g., Burnham and Anderson 2013[7]) values for the same models.

Because Eq. 11 is nonlinear in time, I optimize its parameters $(v^{(0)}, \zeta, \eta)$ using direct nested numerical search over $\max_i(v_i) + [10^{-6}, 20]$ at 10^3 equally spaced increments, followed by directed adaptive refinement, for $v^{(0)}$, along with direct linear minimization for (ζ, η) following a rearrangement of Eq. 11 as

$$v_i/v^{(0)} = 1 - e^{-(\zeta+\eta i)}, \tag{12}$$

$$v_i/v^{(0)} - 1 = -e^{-(\zeta+\eta i)}, \tag{13}$$

$$1 - v_i/v^{(0)} = e^{-(\zeta+\eta i)}, \tag{14}$$

and finally

$$\log\left(1 - v_i/v^{(0)}\right) = -\left(\zeta + \eta i\right). \tag{15}$$

In vector form, this becomes

$$\begin{pmatrix} \log\left(1 - v_1/v^{(0)}\right) \\ \log\left(1 - v_2/v^{(0)}\right) \\ \vdots \\ \log\left(1 - v_N/v^{(0)}\right) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ & \vdots \\ 1 & N \end{pmatrix} \begin{pmatrix} -\zeta \\ -\eta \end{pmatrix}, \tag{16}$$

which readily lends itself to direct linear optimization for ζ and η.

Fig. 4a,b show that—based on adjusted R^2 or AIC—the trends of both downstate peak load and total population are highly significant, and are better described by the logistic equation, Eq. 11 above, than by a uniform slope trend. Conversely, maximum temperatures exhibit a weak uniform slope trend, and a wholly useless logistic trend (the reason R^2 is lower for the identical looking logistic trend is the added parameter, so that the residual about the logistic curve has one degree of freedom less than the residual about the uniform trend).

Finally, Fig. 4e,d show the all important relationship between downstate peak load as the predictand, and its two would-be predictors, both with the best trend retained (black dots) or removed (red dots). In Fig. 4e, the predictor–predictand relationship changes quantitatively but not qualitatively: the slopes of the solid black and dashed red trends differ in magnitude, but not sign. Conversely, Fig. 4d depicts the opposite situation , in which a positive predictor–predictand relationship with the trend included (black dots) reverses to a negative relationship once the trend is removed. This latter situation is a definitive cautionary tale about the risk one accepts—knowingly or not—when one indulges in trend based forecasting without analytically establishing first a clear understanding of the role the trend, and fluctuations about it, play in the overall past predictor–predictand relationship.

As a compelling example of this risk being realized, ruining the forecast, consider forecasting downstate peak loads e based on total population P_t while not clearly separating the role of the trends from that of actual predictor–predictand covariability. Fig. 4d shows that in this situation, while one expects a positive relationship between P_t and e (the black dashed line), reality conspires instead to exhibit *no* relationship to speak of, and certainly none that is positively actionable, between the suitably detrended e and P_t.

This reversal of P_t–e relation is readily understood physically. While the slow and steady rise

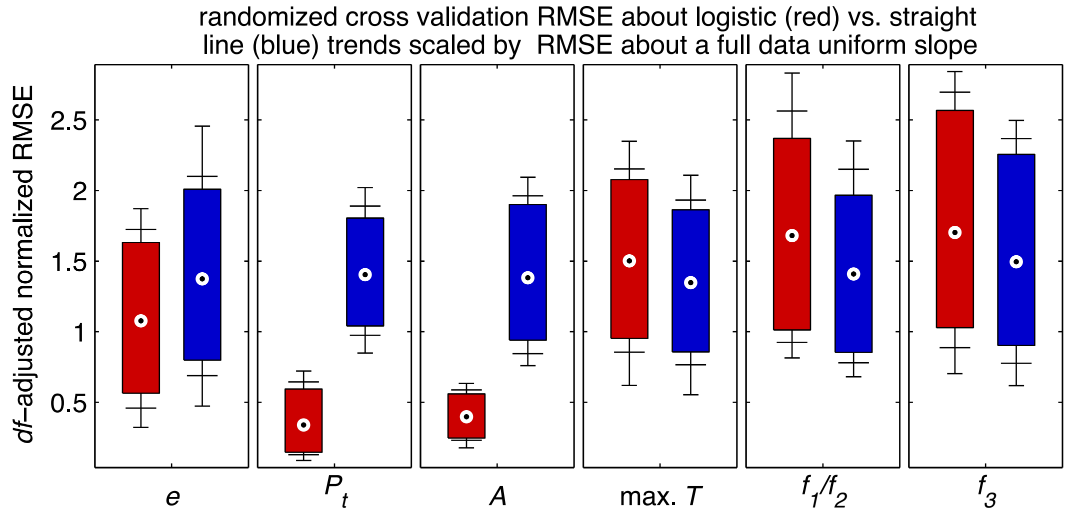

Figure 5: Cross validation evaluation of logistic vs. uniform slope trends for each of the considered variables (left to right): downstate peak load, total population, affluence (defined precisely in the following section), maximum temperature, the ratio of people ages 45–70 to those 20–45, and the fraction of people older than 70. The valuation metric is root means squared error about the trends, adjusted for loss of degrees of freedom. The solid colored rectangles present the central 90% of the cross validation population, and the whiskers the 95% and 99%. The means are shown by the dots.

in downstate population does lead to increasing peak loads, the minute year to year fluctuations in total populations do *not* result in coherent changes in peak loads because other factors that impact individual years' peak loads—notably weather—impact peak loads much more strongly, completely dwarfing the P_t contributions to e variability. Unfortunately, the role of trends is carelessly ignored in the Comments and thus, by extension, in NYISO's peak load projection efforts.

In summary, this section presented several of the many pitfalls associated with trend based forecasting, the underlying foundation of the NYISO forecasting procedure as portrayed by their Comments. Each of these individually, let alone their combination, is sufficiently potentially catastrophic to invalidate NYISO's forecasting.

5.2.1 Which Should be Favored, Uniform Slope or Logistic Trend?

Because the preceding section makes clear just how important trends can be in the following analysis and by extension for resolving the dispute that is the focus of this report, Fig. 5 presents the results of direct competition between linear and logistic trends for each of the considered variables. The performance measure used is *df*-adjusted cross validated RMSE (root mean squared error, where *df* stands for degrees of freedom) about the respective trends, scaled by each variable's RMS scatter about a uniform slope trend derived from the full data (i.e., *not* cross validated). A reading along the vertical axes of Fig. 5 of 0.5, e.g., means that the cross validated residual scatter about the considered trend is half as large as the same scatter about the full data uniform slope trend. The cross validation uses 3:18 withheld:retained data points, providing 1330 combinations, each yielding three cross validation points.

Fig. 5 indicates that total population and affluence (second and third from left, resp.) are both distinctly better described by a logistic trend, while all other variables are roughly equally well characterized by either logistic or uniform slope trends. Note that the leftmost panel, for downstate peak load, corroborates and expands the view of the trend exhibited by this variable previously introduced in Fig. 4a. In particular, while it is still true that the downstate peak load time series is better characterized by a logistic trend (that, importantly, asymptotes toward an upper bound in the next several decades), as Fig. 4a indicates, the leftmost panel of Fig. 5 shows that the superiority of the logistic trend is robust and reproducible (all elements of the distributions are systematically lower for the logistic fits) yet modest in magnitude. In the name of simplicity and parsimony, therefore, a uniform slope description of e may be adequate and even favorable for most purposes.

In summary of this section, of the considered predictors, only maximum temperature exhibits a fluctuation-based-covariability with peak loads. For all other predictors, the relationship is based exclusively on the trend of predictor and predictand alike. Thus apart from $T_{\max}$, which can be used as a fluctuation-based predictor, all other predictors can only enter any envisioned predictive model of future downstate peak loads represented by their trends, not the raw predictor as done in the NYISO Comments.

5.3 On the Inclusion of GDP

Key Points: **NYISO's modeling of future downstate peak load is based on poorly known, highly improbable and speculative rise in GDP. Yet GDP is not an acceptable**

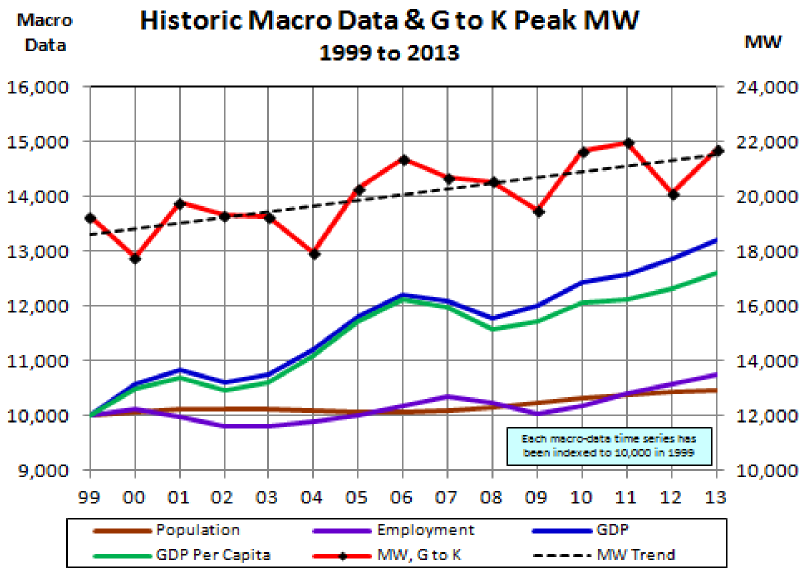

Figure 6: Fig. 3 of the NYISO Comments.

predictor of downstate peak loads, leading to seriously erroneous conclusions.

NYISO's main criticism, to which they devote the most attention, is that the discarding in the Eshel (2014) report—after carefully considering—of GDP as a peak load predictor is erroneous, and that including GDP would have resulted in a better model. This is categorically false, as I will show shortly, but more importantly betrays misunderstanding of several fundamental aspects of statistical forecasting. Specifically, contrary to the enthusiasm for GDP as powerfully predictive of future downstate peak loads expressed in the Comments, some basic analysis reveals unambiguously that GDP inclusion actually *degrades* forecasting skill of downstate peak loads. In this section I demonstrate this, and explain the reason for this behavior.

First, by way of house keeping: NYISO never made clear what exact data set they used for GDP. They mention the for-profit Moody's Analytics, to which I have no access, but they fail to specify precisely which of Moody's data sets they use. In general, it is very poor data analysis hygiene to rely on commercially purchased processed data from for profit third parties.

This general caveat notwithstanding, to complete the work discussed here I devised a novel per capita GDP timeseries, striving to reproduce as closely as I could the appearance of the per capita GDP timeseries NYISO used (green in the Comments' Fig. 3, reproduced here as Fig. 6) and its very high R^2 with the downstate peak load timeseries demonstrated in Fig. 4 of the NYISO

Comments.

To derive the per capita GDP time series used in this report, I rely on publicly available, free raw data from the Interactive Data web site of the U.S Dept. of Commerce, Bureau of Economic Analysis[8]. From the SIC portion of this data set, I obtain 1993–1997 per capita GDP for the states of CT, NJ and NY, in 1997 Dollars, which I convert to 2009 Dollars by multiplying by 1.3367 (this conversion factor is obtained from the U.S. Dept. of Labor, Bureau of Labor Statistics inflation calculator). I next augment these data with their 1998–2013 counterparts using the NAICS portion of the data set, already in 2009 Dollars. I then blend these three timeseries with per capita GDP data for the New York-Newark-Jersey City metropolitan area, (Metropolitan Statistical Area 35620, NY-NJ-PA, hereafter denoted $\text{pcGDP}_{\text{NYC}}$), which is only available for 2001-2013 (also already in 2009 Dollars). Based on trial and error (with the above stated objectives of reproducing as best I can the appearance of NYISO's per capita GDP timeseries and its high R^2 with the downstate peak load timeseries), I finally combine these sources into the affluence variable A using

$$A = \begin{cases} \frac{1}{3}\left(\text{pcGDP}_{\text{NYS}} + \text{pcGDP}_{\text{NJ}} + \text{pcGDP}_{\text{CT}}\right) & \text{for } 1993 \le \text{year} \le 2000 \\ \frac{1}{6}\left(\text{pcGDP}_{\text{NYS}} + \text{pcGDP}_{\text{NJ}} + \text{pcGDP}_{\text{CT}} + 3\text{pcGDP}_{\text{NYC}}\right) & \text{for } 2001 \le \text{year} \le 2013 \end{cases} \tag{17}$$

The resultant time series shows no discontinuity at 2000, and meets both criteria: it is visually extremely close to the per capita GDP time series the Comments present, and its temporal correlation with the downstate peak load time series is 0.92. With this A time series we are now ready to examine the validity of NYISO's claims.

Fig. 3 of the Comments, reproduced here as Fig. 6, presents downstate peak load (red), with a trend (dashed black; presumably uniform slope least squares best fit, although this is never explicated in the Comments), along with several economic measures. The NYISO Comments then present Fig. 4 (not reproduced here), in which they show higher a posteriori R^2 achieved by two models of peak load based on either GDP or per capita GDP relative to alternative models based on either population or employment. This is misleading, and the conclusion the NYISO Comments draw from this—that per capita GDP must be included as a predictor of future downstate peak loads—is false. The reasons for this follow.

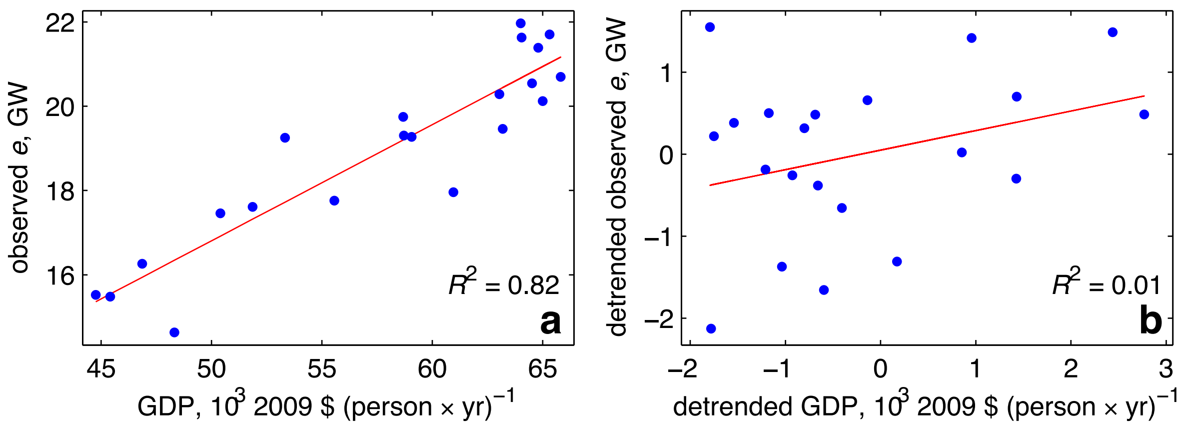

Figure 7: Affluence (A, downstate mean per capita GDP) as a predictor of downstate peak power load e, with and without the optimal time trends of the two variables. In panel a, both variables still contain their respective trends, and the hindcast prediction of e in terms of A yields the reported very respectable $R^2 \approx 0.82$. In panel b, however, when both variables are suitably detrended (e as a uniform slope and A logistically), A's predictive power of e collapses and disappears altogether.

5.3.1 The Historic Relationship Between GDP and Peak Load relies Exclusively on Time Trends

Key Points: **This section describes time trends, explains why they form an unacceptably poor foundation for statistical forecasting, and shows that unfortunately the historic relationship of downstate peak loads and GDP are entirely trend based, and thus unusable for forecasting.**

As Fig. 7 unambiguously shows, once time trends are properly removed, fluctuations in per capita GDP have absolutely no utility in explaining fluctuations in peak load. The entire relationship between affluence and peak load is carried by the two variables' trends. If we had any reason to believe we can expect both variables' trends to continue indefinitely, the above finding would not have been a problem. In reality, however, there is absolutely no reason to expect such stasis. In fact, given that A is far better described by a logistic trends, there is every reason to expect A's trend to flatten with time. **In summary of this point**, affluence is only useful for explaining downstate peak loads because of its time trend, and this trend is much more likely to change markedly in the future than it is to maintain its current value. In other words, using A to explain future e is exceptionally unsound.

5.3.2 Future GDP Values Are Poorly Known, Constituting an Unacceptably Poor Foundation for Forecasting Future Peak Loads

Key Points: **Even if GDP proved useful for forecasting future downstate peak loads (which the preceding section shows it is not), our knowledge of future GDP is too poor for actually employing this putative but unrealized potential.**

Independently of and in addition to the above, while there exist reasonable quality past GDP data, future GDP projections—needed if GDP is to be used as a predictor of future downstate peak loads—are highly speculative.

All statistical forecasting must carefully consider not only the strength of past predictor–predictand links, and the robustness in time of this link, but equally importantly the availability and fidelity of future values of the predictor. At the risk of stating the obvious, even a predictor that is perfectly correlated with the predictand during the available past is utterly useless if its future values are unknown, or are known with such large uncertainty as to produce a resultant forecast devoid of practical utility. While this is self evidently true, its broader context and practical implications are discussed by DelSole and Shukla (2009[9]).

As a general highly pertinent digression, note that fundamentally, observed phenomena can be classified based on whether or not the fundamental, first principle equations governing their behavior are known or unknown. The reason we can design bridges or power towers is that the governing equations of the interactions of these structures with such stress inducers as traffic, wind gusts or ice buildup are largely known, and can be solved uniquely for each considered configuration. Conversely, e.g., the reason the economic outlook of economist and New York *Times* columnist Paul Krugman so rarely mirrors that of fellow economist and former Harvard President Lawrence Summers is that economics addresses aggregate human behavior, and human behavior has no known governing equations.

Similarly, neither does per capita GDP. Its future projections are thus based on various assumptions and conjectures and implicitly or explicitly assume that past behavior will reproduce itself in the future. While at times true, at other times this may lead to such forecasting catastrophes as the hopelessly wrong employment projections presented in Fig. 3a,d. NYISO's devotion to per capita GDP—whose governing equations are unknown—as a requisite predictor of future downstate peak loads is thus fundamentally questionable.

This general limitation notwithstanding, are Moody's GDP forecasts reliable and trustworthy, forming a solid foundation for using them to forecast future downstate peak loads? We don't know. There are no publicly available, detailed accounts of their methodology, there are no public (or indeed any well cataloged) archives of their past forecasts, and a *Google Scholar* search for "Moody Analytics GDP forecast" yields not a single peer reviewed or even gray literature technical paper in which Moody's historic GDP forecasts are rigorously validated.

This unsettling lack of quantitative accountability is a sharp deviation from established practices in other mainstream forecasting efforts, like those addressing weather. For example, ECMWF (the European Centre for Medium-Range Weather Forecasts, the agency that produces weather forecasts for much of the globe) has decades worth of archived forecasts on its web site, freely and readily downloadable by anybody. Various aspects of verifying these archived forecasts are the subject of thousands of peer reviewed scientific papers published annually. For example, a *Google Scholar* search for "ECMWF forecast verification" yields 2193 papers since 2014. This is handily eclipsed by ECMWF's U.S. counterpart, the U.S Dept. of Commerce, National Oceanic and Atmospheric Administration's National Weather Service, whose archived forecasts have been the subject of 4192 peer reviewed papers since 2014 (obtained by *Google Scholar* searching for "NWS forecast verification").

While there is no directly relevant published work on Moody's GDP forecasting, I did find one quasi-technical paper that does address Moody's technical integrity and forecasting skills in other arenas. This paper joins the above observations, raising serious concerns as well. The paper, Joffe (2012[10]), cites a U.S. Senate investigation report that describes systemic staff shortages that seriously undermined Moody's ability to carry out its ratings line of business.

Because basic standards of democratic transparency and accessibility of knowledge are not only not met, but are indeed fundamentally rejected by the proprietary Moody's Analytics, forecasting future downstate peak loads based on their unverified and unverifiable GDP forecast is frivolously speculative.

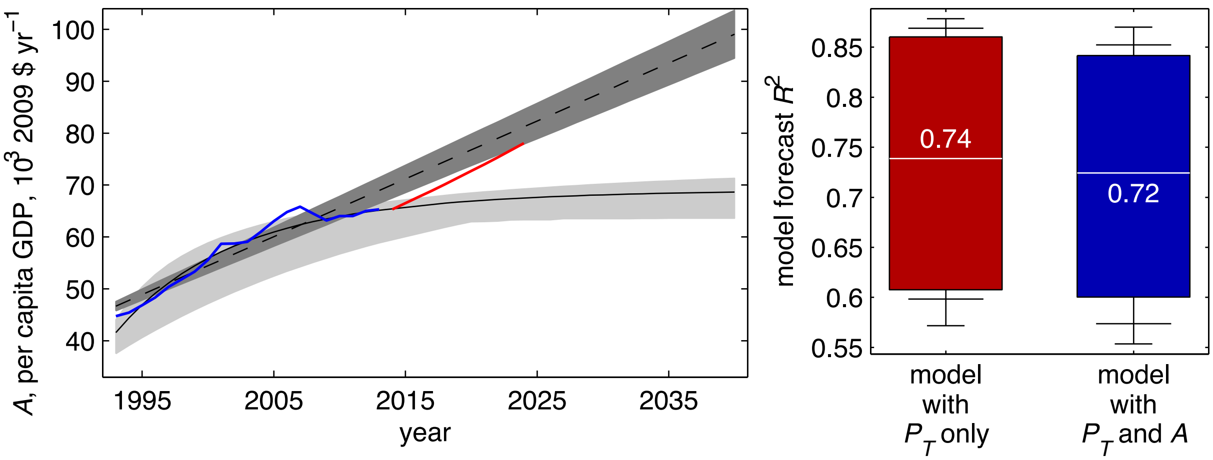

Figure 8: Observed downstate per capita GDP (left panel, blue curve), and two least squares best fit trends to these data (gray shading showing the central 95% probability with the best estimate in solid black, and dashed black). In red is a slightly less optimistic version of Moody's prediction. See text for more details and discussion.

5.3.3 So Moody's Downstate Per-Capita GDP Are Unverified; Do They At Least Make Intuitive Sense?

On page 21, the NYISO Comments quote Moody Analytics' prediction of a 22% per capita downstate GDP decadal growth over 2014–2024. If we assume this growth is compounded annually, the annual geometric growth rate γ satisfies

$$(1+\gamma)^{11} = 1.22 \qquad \text{or} \qquad \gamma = \sqrt[11]{1.22} - 1 \approx 0.018 \tag{18}$$

or an increase rate of per capita GDP of 1.8% annually. This is shown in red in the left panel of Fig. 8, along with the recently observed downstate per capita GDP history in blue. The panel also shows the logistic (light shading with solid black) and uniform slope (dark shading with dashed black) trends that best fit the historic data (blue), where the black curves are the best estimates and the shadings present the central 95% of the cross validation population.

Fig. 8's left panel thus shows that Moody's projections are (1) more optimistic (have a more positive and increasing slope) than even the clearly inferior uniform slope trend (recall that this inferiority is an inescapable corollary of Fig. 5, third panel from left); and (2) are completely at odds with recent downturn, cheerily ignoring the greatest recession in American history since the Great Depression.

Based on those, it appears that no, Moody's projections are not only unverified and unverifiable;

they are in all likelihood also unrealistically optimistic.

5.3.4 So is GDP a Good Predictor of Peak Loads?

No. It is not even an acceptable or marginally useful predictor. It is not a predictor at all. To see this numerically, and unambiguously, consider the right panel of Fig. 8. The calculations whose results that panel reports are as follows.

Employing physical intuition—a key, often underused, tool in the forecaster's toolbox—it is clear that total population must be included in any model of peak loads. Short of complete rewrite of the basic rules of human energy use, one can envision no circumstance under which larger populations will not result in larger peak loads. Thus P_t is definitely in the predictive model. Given this, the question is: "Does the inclusion of affluence improve the performance of a model of downstate peak load that already includes total population size as a predictor?"

An age old practice in science (epitomized popularly as Occam's Razor) dictates favoring the most parsimonious, least conjectural of all competing equal skill models of a given phenomenon. Given that a key measure of model simplicity is its number of parameters, and that each added predictor adds an additional parameter, adding predictors to an existing model must only be done after showing convincingly that the addition not only enhances skill, but that the skill addition remains positive after correcting for lost degrees of freedom (reduced parsimony). Thus the skill of the more complex of two competing models must be significantly larger than the simpler model's.

The right panel of Fig. 8 shows the reverse for the addition of A (per capita GDP) as a second predictor to a model of downstate peak load already containing total population, as discussed above. Not only does the addition of A not increase skill, incredibly, it actually decreases it. So the more complex (less parsimonious) model of peak downstate loads that contains both P_t and A as predictors is *less* skillful. Since more complex *and* less skillful is a clearly inferior combination, A must be logically excluded as a predictor of downstate peak load. **That is, NYISO's assertion that per capita GDP is a skillful and important predictor of future downstate peak loads that must be included in a model of those peak loads is based on a suite of serialized rather rudimentary errors, and is false.**

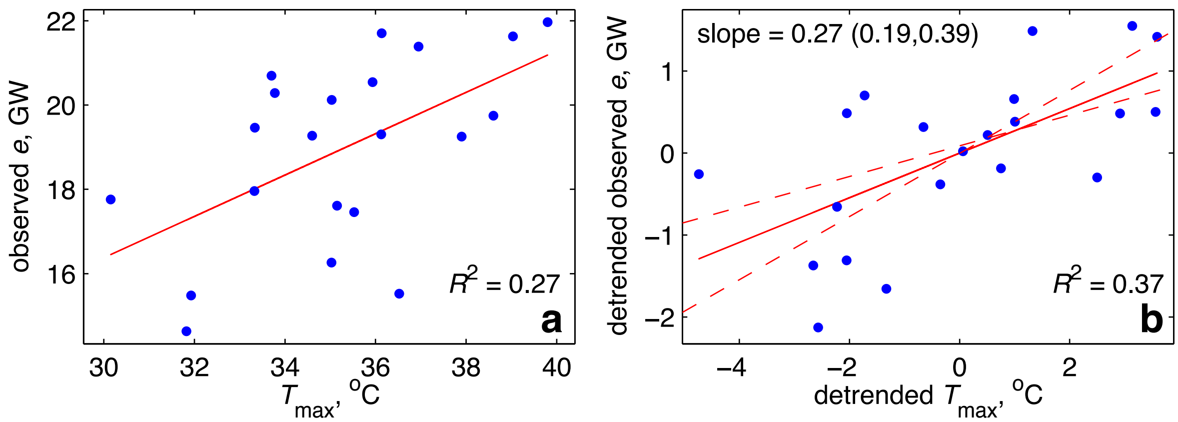

Figure 9: Observed downstate peak load vs. $T_{\max}$ (blue dots). On the left, both variables are plotted with their trends retained, and the model $e = \alpha + \beta T_{\max}$ yields the shown hindcast $R^2 \approx 0.27$. On the right, the blue dots show peak load vs. $T_{\max}$ after both variables' uniform slope trends (denoted by the "(τ)" superscript) are removed, now yielding the model $(e - e^{(\tau)}) = \alpha + \beta(T_{\max} - T_{\max}^{(\tau)})$ with the shown clearly superior hindcast $R^2 \approx 0.37$. Carrying out all 1330 possible withholding experiments with unique combinations of three withheld years yields 1330 estimates of the slope β, the central 99th percentile of the distribution of which is bounded by the dashed red lines. The numerical value of the optimal slope, along with the 0.5 and 99.5 percentiles of the slope distribution, are indicated in the upper left corner.

6 An Updated Model of Future Downstate Peak Loads

6.1 Predictors Used

While the population fraction of people over age 70 is expected to rise in coming decades according to the Cornell future downstate population projections data I use [see the Eshel (2014) report for discussion], that fraction is still quite small, probably too small to be reasonably expected to exert significant control over future peak loads. In addition, its time series is nearly the mirror image of the f_2/f_1, thus being mostly redundant. I thus exclude f_3 from the model, using only $r = f_2/f_1$ and P_t as population based predictors. As vectors holding all annual values, these are denoted, as in the Eshel (2014) report, $\mathbf{r}$ and $\mathbf{P}_t$, respectively.

The third and final possible predictor is $T_{\max}$ denoted, conforming with the Eshel (2014) report, by C_i as a scalar showing the value of year i or by $\mathbf{C}$ as a vector holding all annual values. The decision of whether or not $\mathbf{C}$ should enter the model is partly based on the already discussed Fig. 2 (central panel) and Fig. 4c,e, and partly on Fig. 9. Fig. 9 shows that (1) $T_{\max}$ is far more predictive of peak load when both variables' trends are removed, and that (2) while the relationship is of limited power (the R^2 is modest), it is very robust (the slope is tightly clustered around the optimal value of 0.27 GW per degree C, with all 1330 slope estimates falling inside [0.18,0.40]). Including $T_{\max}$ in the model is also important on independently verified physical grounds because it connects the forecast to future climate change. Because, as expected, the slope is positive—downstate peak loads rise with rising $T_{\max}$, as all three red lines in Fig. 9b show—inclusion of $T_{\max}$ in the model is conservative in that it can only raise expected future peak loads (because future $T_{\max}$ values increase with time). Finally, on algebraic grounds, the scaled projections of $\mathbf{C}$ (the vector holding the $T_{\max}$ time series) on the other two predictors are low, so that adding $\mathbf{C}$ will not undermine the conditioning of the data matrix, as is required from an additional predictor in a stepwise regression formalism [see Ryan (1997[1]) pp. 222-223, or Eshel (2012[2]) pp. 184-185].

Thus $T_{\max}$ should enter the model. The three predictors used here are, as in the Eshel (2014) report, $\mathbf{r}$, $\mathbf{P}_t$, and $\mathbf{C}$.

6.2 The Forecasting Setup

The model presented below is based on the premise—discussed and demonstrated earlier in this report—that the relationship of $\mathbf{r}$ and $\mathbf{P}_t$ with downstate peak loads $\mathbf{e}$ are mostly due to trends, not fluctuations about the trends. Conversely, the $\mathbf{C}$–$\mathbf{e}$ relationship is nearly entirely due to fluctuations about the trends. This is crucially important not only because of the need to cleanly separate the contributions of trends and point-by-point fluctuations about them in statistical forecasting, as I have emphasized several times earlier in this report, but also because we simply know, or even *can* know, nothing about individual future fluctuations.

Based on Fig. 5 I favor a more complex (three parameter) logistic trend only for $\mathbf{P}_t$, and a simpler (two parameter) uniform slope trend for the age-based population ratio $\mathbf{r}$ and maximum temperature $\mathbf{C}$. Denoting a given variable's optimal, either logistic or uniform slope, trend by a (τ) superscript (see caption of Fig. 9), the model predictors are $\mathbf{r}^{(\tau)}$, $\mathbf{P}_t^{(\tau)}$, and $\tilde{\mathbf{C}} \stackrel{\text{def}}{=} \langle \mathbf{C} \rangle + (\mathbf{C} - \mathbf{C}^{(\tau)})$ (where angled brackets denote long-term time mean). That is, along with the mean term (a column of ones, denoted $\mathbf{o}$), the basic data matrix of the model is thus

$$\mathbf{A} = \begin{pmatrix} \mathbf{o} & \mathbf{r}^{(\tau)} & \mathbf{P}_t^{(\tau)} & \tilde{\mathbf{C}}, \end{pmatrix} \tag{19}$$

visually similar to but conceptually more finely resolved than the earlier model presented in Eshel (2014).

The subtle yet important difference is based on the careful analysis of the role trends presented in earlier sections. The population based predictors $\mathbf{r}$ and $\mathbf{P}_t$, whose relations with downstate peak load are exclusively due to their respective trends, enter the model only as trends, $\mathbf{r}^{(\tau)}$ and $\mathbf{P}_t^{(\tau)}$. This cleanly highlights the real nature of predictor–predictand relation, eliminating the distractions (uncorrelated year-to-year fluctuations). On the other extreme is $T_{\max}$ or $\mathbf{C}$, whose relation to downstate peak load is nearly entirely due to year-to-year fluctuations, which the above $\tilde{\mathbf{C}}$ cleanly represents.

The overall forecast is derived from statistics of all 1330 unique combinations of three withheld years among the 21 available past years (see Eq. 8 and the discussion in which it is embedded). For each of those 1330 withholding forecasts, I modify Eq. 2 into

$$\hat{\mathbf{e}} = \mathbf{A}\left[\left(\mathbf{A}^{(r)T}\mathbf{A}^{(r)}\right)^{-1}\mathbf{A}^{(r)T}\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)}\right] + \left(\mathbf{e}^{(\tau)}\right)^{(w)} \tag{20}$$

where, recall, the superscripts (τ), (w), and (r) denote the optimal (uniform slope or logistic, as appropriate) time trend, withheld data, and retained data, respectively.

Note the all important treatment of the time trend of $\mathbf{e}$ in Eq. 20: the vector with which $\mathbf{A}^{(r)}$'s columns are dotted is $\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)}$, the detrended peak load vector with the parameters of the trend chosen based on retained data only, and the removed trend—now evaluated for the withheld years—is then added to the solution, done by the addition of the rightmost term.

Also note that in Eq. 20 neither $\hat{\mathbf{e}}$ nor $\mathbf{A}$ have a (w) superscript because they hold lines related to withheld, retained and future years. This is so because Eq. 2's $\mathbf{A}^{(w)}$ is replaced here with the unadorned

$$\mathbf{A} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{A}^{1993-2013} \\ \mathbf{A}^{2018-2040} \end{pmatrix} \in \mathbb{R}^{44\times 4} \tag{21}$$

whose 44 rows correspond to the combined number of years in the two indicated discontinuous periods, and whose four columns reflect the mean term column of ones plus the three physical predictors $\mathbf{r}^{(\tau)}$, $\mathbf{P}_t^{(\tau)}$, and $\tilde{\mathbf{C}}$. Also in Eq. 20, $\mathbf{A}^{(r)} \in \mathbb{R}^{18\times 4}$ holds the past observations of the 18 retained past years (the full 21 minus the withheld three). With these definitions, $\hat{\mathbf{e}}$ and $\mathbf{A}$ in

Eq. 20 can be rearranged with no loss of generality so that

$$\hat{\mathbf{e}} \stackrel{\text{def}}{=} \begin{pmatrix} \hat{\mathbf{e}}^{(w)} \\ \mathbf{e}^{(r)} \\ \hat{\mathbf{e}}^{2018-2040} \end{pmatrix} \in \mathbb{R}^{(3+18+23)\times 1} \quad \text{and} \quad \mathbf{A}^{1993-2013} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{A}^{(w)} \\ \mathbf{A}^{(r)} \end{pmatrix} \in \mathbb{R}^{(3+18)\times 4} \tag{22}$$

in which $\hat{\mathbf{e}}^{(w)} \in \mathbb{R}^{3\times 1}$ holds the forecast for the three withheld years (whose predictor data are held in $\mathbf{A}^{(w)} \in \mathbb{R}^{3\times 4}$), $\mathbf{e}^{(r)} \in \mathbb{R}^{18\times 1}$ holds the hindcast for the 18 retained years (whose predictor data are held in $\mathbf{A}^{(r)} \in \mathbb{R}^{18\times 4}$), and $\mathbf{e}^{2018-2040} \in \mathbb{R}^{23\times 1}$ holds the future forecasts (corresponding to $\mathbf{A}$'s lower 23 rows).

Because we know nothing about future individual fluctuations, and because there is no reason to attach any significance to the ones that were randomly realized in the past (since the trends dominate predictor–predictand relationship), in each of the 1330 realizations, I perturb randomly both past observations and expected future predictor values. That is,

$$\mathbf{A}^{(r)} = \mathbf{A}_0^{(r)} + \mathbf{A}_p^{(r)}, \tag{23}$$

where

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{o} & \mathbf{r}^{(\tau)} & \mathbf{P}_t^{(\tau)} & \tilde{\mathbf{C}} \end{pmatrix} \in \mathbb{R}^{21\times 4} \tag{24}$$

is the basic unperturbed data matrix from which all 1330 $\mathbf{A}_0^{(r)} \in \mathbb{R}^{18\times 4}$ are drawn, and, consistent with the Eshel (2014) report, $\mathbf{o}$ is an 21-vector of ones, and the τ superscript indicates the optimal time trend of the respective variable, calculated anew for each withholding experiment over the 18 retained years only.

The second matrix in Eq. 23, unique to each of the 1330 solutions, is

$$\mathbf{A}_p^{(r)} = \begin{pmatrix} \mathbf{0} & \mathbf{n}(m_r, s_r) & \mathbf{n}(m_{P_t}, s_{P_t}) & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{18\times 4} \tag{25}$$

in which $\mathbf{0}$ is an 18-vector of zeros, and each $\mathbf{n}(m_v, s_v)$ is a unique 18-vector of random draws from a normal distribution with mean and variance given by variable v's sample mean and variance calculated over all past years 1993–2013. This guarantees that our solution is robust and independent of the irreproducible vagaries of past noise.

In a similar manner, I also perturb for each of the 1330 solutions the matrix $\mathbf{A}^{2018-2040}$ of future expected predictor values calculated based on the trends alone, viz.

$$\mathbf{A}^{2018-2040} = \mathbf{A}_0^{2018-2040} + \mathbf{A}_p^{2018-2040}, \tag{26}$$

where all notation, and means and variances, are as defined above.

6.3 A Crucial Technical Point: On Inverting Eq. 20

6.3.1 The Problem

Eq. 20 requires inverting ${\mathbf{A}^{(r)}}^T\mathbf{A}^{(r)}$. While deceptively trivial to the uninitiated, this is a make-or-break step in which disastrously erroneous results can be readily obtained if great care is not paid where needed. The issue is the potential ill-posedness ${\mathbf{A}^{(r)}}^T\mathbf{A}^{(r)}$ may sometime exhibit.

First, to clarify unambiguously, in the full, 21 year long, matrix of basic predictors, the columns are mutually linearly independent and thus yield a formally full rank $\mathbf{A}$, $\text{rank}(\mathbf{A}) = 4$. Yet the conditioning is somewhat tenuous, with the 1st and 4th singular values satisfying $\sigma_4 \lessapprox 0.01\sigma_1$. Because of this modest stability, some specific withholding experiments—in which each time series comprises only 18, deliberately noise contaminated, data points—can become numerically unstable. The issue is best understood with the aid of the SVD (singular value decomposition) representation of the matrix,

$$\mathbf{A}^{(r)} = \mathbf{U}\mathbf{D}\mathbf{V}^T. \tag{27}$$

Unfortunately, the scope of this fundamentally important issue far exceeds the scope of this report, but the interested reader can consult chapter 9 (subsection 9.4 in particular) in Eshel (2012[3]).

With this representation of $\mathbf{A}^{(r)}$,

$${\mathbf{A}^{(r)}}^T\mathbf{A}^{(r)} = \left(\mathbf{U}\mathbf{D}\mathbf{V}^T\right)^T\mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{V}\mathbf{D}^T\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T. \tag{28}$$

Because $\mathbf{U}$ is orthonormal, $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, and thus

$${\mathbf{A}^{(r)}}^T\mathbf{A}^{(r)} = \mathbf{V}\mathbf{D}^T\mathbf{D}\mathbf{V}^T. \tag{29}$$

Therefore,

$$\left({\mathbf{A}^{(r)}}^T\mathbf{A}^{(r)}\right)^{-1} = \left(\mathbf{V}\mathbf{D}^T\mathbf{D}\mathbf{V}^T\right)^{-1}. \tag{30}$$

Because $\mathbf{V}$ is also orthonormal, $\mathbf{V}^{-1} = \mathbf{V}^T$, and thus

$$\left({\mathbf{A}^{(r)}}^T\mathbf{A}^{(r)}\right)^{-1} = \left(\mathbf{V}^T\right)^{-1}\left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{V}^{-1} = \mathbf{V}\left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{V}^T. \tag{31}$$

Consistently,

$$\left(\mathbf{A}^{(r)T}\mathbf{A}^{(r)}\right)^{-1}\mathbf{A}^{(r)T} = \mathbf{V}\left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{V}^T\left(\mathbf{U}\mathbf{D}\mathbf{V}^T\right)^T \tag{32}$$

$$= \mathbf{V}\left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{V}^T\mathbf{V}\mathbf{D}^T\mathbf{U}^T \tag{33}$$

$$= \mathbf{V}\left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{D}^T\mathbf{U}^T. \tag{34}$$

Let's use this to recast the bracketed term of Eq. 20 (which determines the least squares fit) as

$$\left(\mathbf{A}^{(r)T}\mathbf{A}^{(r)}\right)^{-1}\mathbf{A}^{(r)T}\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)} = \mathbf{V}\left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{D}^T\mathbf{U}^T\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)}. \tag{35}$$

This can be simplified by visualizing its right hand side,

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}\begin{pmatrix} 1/\sigma_1 & & & & \\ & 1/\sigma_2 & & & \vdots \\ & & 1/\sigma_3 & & \cdots\ \mathbf{0}\ \cdots \\ & & & 1/\sigma_4 & \vdots \end{pmatrix}\begin{pmatrix} \mathbf{u}_1^T\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)} \\ \vdots \\ \mathbf{u}_4^T\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)} \\ \mathbf{u}_5^T\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)} \\ \vdots \\ \mathbf{u}_{18}^T\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)} \end{pmatrix} = \tag{36}$$

$$= \sum_{i=1}^{4}\left(\frac{\mathbf{u}_i^T\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)}}{\sigma_i}\right)\mathbf{v}_i. \tag{37}$$

This emphasizes that for poorly conditioned withholding experiments, the trailing (4th) term can ruin the solution because its scalar coefficient—while having a well defined numerator, the projection of the detrended peak load on $\mathbf{u}_4$—has a very small yet nonzero denominator. In a handful of cases, this applies to the 3rd term as well.

While it may appear that the denominator being nonzero, formally permitting the division of the projection numerator by σ_i for all is, is a virtue, it is in fact a major liability. If the poorly determined terms corresponded to a $\sigma_i = 0$ exactly, of course the operation would not have been permitted, and the program execution would have stopped with an error message. As things stand, however, this does not happen in any of the 1330 experiments. Instead, the operation is permitted even in cases in which σ_4 is many orders of magnitude smaller than σ_1. Yet in such cases, the poorly determined $\mathbf{v}_4$ appears in the series solution with a very large scalar coefficient, ruining the solution.

6.3.2 The Cure

To deal with this potential dominance of the essentially ill-defined $\mathbf{v}_{3,4} \in \mathbb{R}^{4\times 1}$ over some solutions, I impose a stringent truncation cutoff c satisfying

$$\sigma_{i=[1,c]} > 0.001\sigma_1. \tag{38}$$

With this criterion, evaluated uniquely for each of the 1330 withholding realizations,

$$\begin{aligned}
\hat{\mathbf{e}} &= \mathbf{A}\left[\left(\mathbf{A}^{(r)T}\mathbf{A}^{(r)}\right)^{-1}\mathbf{A}^{(r)T}\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)}\right] + \left(\mathbf{e}^{(\tau)}\right)^{(w)} \qquad (39)\\
&= \mathbf{A}\left[\sum_{i=1}^{c\le 4}\left(\frac{\mathbf{u}_i^T\left(\mathbf{e}-\mathbf{e}^{(\tau)}\right)^{(r)}}{\sigma_i}\right)\mathbf{v}_i\right] + \left(\mathbf{e}^{(\tau)}\right)^{(w)} \qquad (40)
\end{aligned}$$

where $\mathbf{A}$, recall, is *not* subject to the SVD representation because it is a different matrix than the matrix being decomposed, $\mathbf{A}^{(r)}$, with 44 instead of 18 rows. This method leads to stable, robust solutions, whose statistics are presented in Fig. 10.

6.4 Additional Assumptions Consistent with NYISO's Critique

In the calculations leading to Fig. 10, I have implemented several of NYISO's critiques of the Eshel (2014) model. Note that this inclusive policy does not necessarily suggest agreement with those points. Rather, the choice of treating NYISO's comments as a given indicates the wish to promote a dialog, and to point out that they change nothing in the final conclusion, that **no need for additional transmission assets is present**.

The implemented changes are:

1. NYISO deemed too optimistic the queue considered in Eshel (2014). The plot is now based on NYISO's latest published queue.

2. NYISO suggested that the Eshel (2014) assumed overly optimistic existing assets. These assets are now lowered from the maximum handled to date, the 2011 maximum peak, to the mean peak of the last five available years, 2009-2013, see the horizontal green line on top of the red observation lines, $\sim$1 GW lower than before.

3. NYISO viewed as overly optimistic the Eshel (2014) assumed completion rate for assets already on the queue. They are now lowered, considering both 45% and 48% completion rates in addition to the earlier 50% rate.

4. NYISO also viewed as overly optimistic the Eshel (2014) assumed annual energy saving rate. Here, this rate is now lowered to 0.4%, 0.5% or 0.6% annually. As a yardstick, recall that the Eshel (2104) report cited a recent NYSERDA study[4] that estimated an achievable electricity savings potential in the State of 18% by 2032, equivalent to a mean savings rate of 0.9% per year. Another estimate cited, that of Woolf et al.[5] is 1.5% savings per year, which those authors deemed "clearly feasible" for New York State in the decade following their 2011 publication.

I now assume Indian Point is already closed. For further robustness, in the model forecasts, whose statistics are condensed into Fig. 10's shading encompassing the blue forecast curve, I now include as upper bound not the 95 percentile, as before, but the 99.5 percentile (the uppermost of the shading).

6.5 Statistics of Expected Future Downstate Peak Loads

Combining the low completion rate of queue assets (45% instead of the earlier 50%) with the lowest energy savings rate (0.4% yr^{-1}) yields the lower of the three green curves on the upper right.

The very conservatively calculated future peak loads never exceed, or even approach, available assets. This holds even when assuming the most pessimistic availability while at the same time assuming the worst, five chances out of a thousand, peak loads. Note that because the current forecasts are based on a conceptually different model than the one the Eshel (2014) report used, and are derived using new code reflecting the new foundation, the two models are independent, each lending support to the other. Clearly, **the earlier conclusions still stands. There is no need for additional transmission assets**.

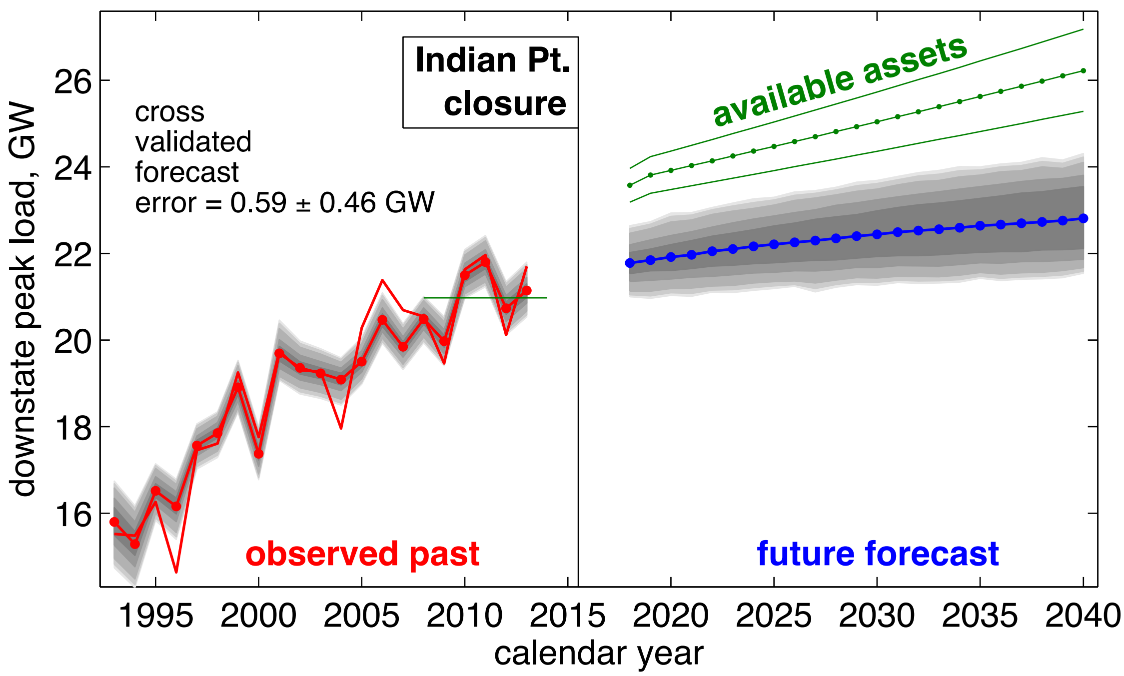

Figure 10: Observed (red) and predicted (blue) downstate peak loads, along with existing (horizontal green) or expected availability (considering the three asset expansion scenarios discussed in the text, given as green curves). Shades of gray indicate (from the center outwardly, from dark to light gray) percentiles 30–70, 15–85, 5–95, 2.5–97.5 and 0.5–99.5. Indian Point is assumed closed already. The mean and standard deviation of $3 \times 1330 = 3990$ cross validated annual forecast errors are, as indicated, approximately 0.6 ± 0.5 GW. For comparison, note that the overall rise over the considered period, roughly 6 GW, is tenfold larger than the mean error. Similarly, the standard deviation of the *optimally detrended* observed downstate peak load time series (i.e., the characteristic magnitude of the year-to-year fluctuations about the trend), 0.93 GW, is 60% larger than the forecast error.

7 Non-Technical Points

7.1 Deviation from Established Energy Sector Practices

One key critique the NYISO Comments level against the Eshel (2014) report is that it "deviates from established NYISO modeling practices". For example, p. 18 of the Comments states that *"Figure 4 illustrates that the NYISO found that the model using population has the worst* R^2 *coefficient whereas the affluence variable has the best. This is consistent with more than a decade of NYISO modeling experience which, in keeping with established utility industry practice, considers GDP to be a more significant variable than either population or employment in most cases."*

Apart from the technical errors discussed earlier—cross validated skill, not hindcast R^2, evaluates model performance, and reliance on trends is unacceptable regardless of whether the forecaster is mindful of this or not—this point is fundamentally correct. My approach is indeed a marked deviation from established NYISO practices, in most cases representing a 180°, about face reversal. Where this critique errs is in viewing this deviation as a liability. In fact, it is a key asset.

Statistical forecasting can take many non-unique forms, and diversity of forecasting techniques is essential to the integrity of any forecasting enterprise. Yet any approach is ultimately either technically defensible or erroneous. No forecasting effort that falls into the latter category is acceptable or worthy of attention or credence.

As the above sections of this report technically demonstrate unimpeachably, NYISO's "established practices" are in fact technically naive and riddled with errors, and decisive deviation from them is thus urgently needed.

7.2 Alleviating Congestion Costs as a Rationale for the Project

Clearly recognizing the peculiarity of claiming the necessity of a project with no demonstrated necessity, the Comments go to considerable length to advance other reasons the project may be needed despite being unnecessary. The Comments mostly do so by quoting extensively a Brattle Group report (e.g., pp. 7–10 of the Comments).

7.2.1 Balancing Societal Needs and Prudent Resource Use

A few of the Brattle Group's points may be relevant. For example, "mitigation of weather and load uncertainty" or of "extreme events" are worthy societal objectives. I would advance a simple, and decisive, counter argument. All else being equal, it is always better to be able to handle a perturbation of a given magnitude than it is to be unable to handle it. This is, however, a question of sensible allocation of resources. If the Brattle Group logic carries the day, e.g., Manhattan would be connected to the mainland with bridges and tunnels whose joint throughput exceeds Thanksgiving traffic, let alone any weekday's rush hour. Obviously, this fantasy will require monetary and land use investments that are too enormous to merit even a casual consideration. Yet this is precisely the logic the Brattle Group will have us apply to power delivery to the downstate.

Similarly, committing for decades large tracts of New York land assets for mitigating minor, temporary load events whose frequency is very low and stands to drastically further decline in coming decades is illogical

7.2.2 How Wise is Reducing Congestion Losses?

The rest of the Brattle Group arguments revolve around reducing congestion losses. This argument puts logic on its head, completely missing the key fundamental point: congestion is an asset, not a liability.

As section 6 above shows, there is no shortage of transmission capacity at peak loads, but a nontrivial and growing surplus. The issue of congestion is purely economic; congestion raises power prices for a few hours on a few afternoons a year.

Yet why wouldn't it?! For example, while Vail ski slopes are freely accessible in August, it is the presence of snow that makes them desirable. And for that, we all think nothing of paying a massive premium.

Similarly, want power when power is at a premium? Not a problem, it can be readily delivered using current infrastructure, but it will cost. Want to save money? By all means; that latter choice is what Demand Response is all about. Lower your demand when prices are high, and save.

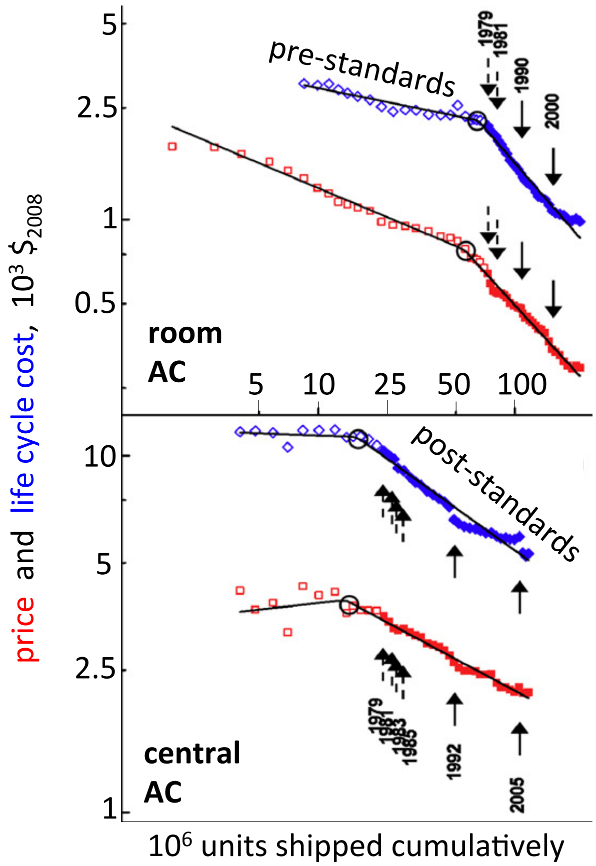

Figure 11: Price (red squares) and life cycle cost (blue diamonds) of room and central air conditioning units (top and bottom panels, respectively, with all axes given in $\log_{10}$) as a function of cumulative purchases (a proxy for time elapsing and experience gained). Introduction of California and Federal energy efficiency standards are indicated by vertical dashed and solid arrows, respectively. Pre-standard data are plotted with empty symbols, with post-standards ones solid. Modified from Fig. 1 (p. 5) of Van Buskirk et al. (2014).

The customer's call.

Nobody is thinking of "relieving" the price premium on skiing in Vail in January. Nobody is losing sleep over how much more expensive a Porsche is relative to a Camry. We take it as a given that that which is highly coveted - will cost. Why should consuming power when the grid is overloaded be any different?! It should not, and "relieving" congestion therefore disregards the simple market logic of supply and demand, and actively and adversely promotes market failures.

Congestion "relief" thus constitutes a deliberate choice to artificially interfere with the marketplace so as to actively discourage energy efficiency using inherently anti-capitalist policy tools.

This is a particularly puzzling shortsightedness given that increasing energy efficiency is unquestionably one of the key societal objectives of our time, and that cost minimization and gain maximization are among the most powerful human motivators and thus effective policy tools.

Energy price is an extremely powerful tool for promoting energy conservation. New York cannot profess to favor energy savings while at the same time enacting policies that reduce power costs, especially during peak loads. To achieve peak load reductions, or to slow their increase, one must allow costs during peak loads to increase roughly following market dictates. Doing the opposite by "congestion relief" is to prevent market forces from enhancing energy efficiency.

Unless the champions of the proposed project as a means of congestion cost "relief" have found a superior alternative to market based capitalism, their stated goals, if realized, will reduce energy efficiency in New York, setting us back instead of moving us forward.

This failing of congestion "relief" becomes all the clearer when applied to a specific human activity. Because congestion is a summer phenomenon, with much of the added power demand due to air conditioning, it makes sense to demonstrate this using air conditioning efficiency vs. cost. Van Buskirk et al. (2014[1]) have considered the problem, and their relevant findings are summarized in their Fig. 1, partially reproduced with some modification here as Fig. 11.

As expected, Van Buskirk et al. (2014) found that appliances of a given quality and range of features become progressively more affordable. Most importantly, however, Van Buskirk et al. (2014) also showed that this cost reduction is accelerated appreciably in response to the introduction of energy efficiency standards. Because the expected impacts on individual consumers of such efficiency standards and higher power costs are similar (but not identical), these rate acceleration—the key to mitigating future downstate peak loads—are an example of the positive societal role congestion costs can play. Eliminating or even reducing these costs is therefore societally counterproductive.

Finally, congestion "relief" has another built in fundamental flaw, its regressive nature. That is, "relieving" congestion-related price spikes will help large commercial power users far more than it will help the bulk of New York's population, because the relief is per kWh consumed. And in the downstate region, large power users tend to be various financial and banking industry megaplayers who need no financial assistance. If the congestion "relief" objective in rooted in the concern for undue hardship congestion costs impose on low income communities, a very reasonable motivation, it would be far more sensible and impactful to set up a program that deals directly with this very real problem rather than indirectly and regressively alleviate those costs to all consumers, thus spending the bulk of the cost on "helping" those who need no help at all.

8 Conclusions

This report examined quantitatively the key arguments in the NYISO rebuttal of the Eshel (2014) report and its main conclusion, that no additional transmission capacity into the downstate area is needed. The report demonstrates that none of these arguments is meritorious. The report examines closely the assertion that GDP is an essential predictor of future downstate peak loads, and shows that once several errors in the NYISO Comments are corrected, GDP loses all of its allure. The report then shows that much of the NYISO forecasting is based on erroneous treatment of temporal trends. Once those are corrected, it is further shown, the skill of the NYISO forecasting largely disappears. Finally, a revised model of downstate peak loads through 2040 is introduced and carefully validated, achieving the very low cross validated error of 0.59 ± 0.46 GW. Employing the model for future projections, finally, only strengthens the original conclusion, that **peak loads exceeding expected assets are highly improbable**. The proposed transmission project that is at the heart of this dispute thus remains with no defensible justification.

9 Appendix: A Primer on Forecasting

As stated below, the resolution of this dispute lies with professional grade forecasting of future downstate peak loads. To help readers reach their own conclusions as to whose argument is more persuasive, below I provide a short introduction to the forecasting problem.

There are two fundamental ways to forecast future events, neither perfect. The first is deterministic, preferable when known rigid rules govern the behavior of the phenomenon being forecasted. By "known rigid rules", I mean such fundamental governing laws of physics as conservation of energy or momentum. The "rigid" is crucial here, indicating that no flexibility is allowed, each of those invariants *must* be exactly conserved. Your weather forecast is likely the example of deterministic modeling with which you are most familiar. Known physical laws govern the evolution through time of atmospheric motions, and the distribution of heat and humidity in the atmosphere and upper ocean, and this knowledge is embodied in computer codes known as numerical weather prediction models. Our ever refining weather forecast is the output (solution) of these models obtained by running them forward in time roughly ten days out several times daily, each time with the most up to date information about the current state of the atmosphere.

The other type of forecasting is statistical, used when no known *exact* or *rigid* physics govern the addressed phenomenon. Think, e.g., about election outcome predictions. There are of course no rules about what will make people come to the polls or, if they do, what candidate they will vote for. As a result, one must resort to *predictors*, readily measurable attributes of the population at hand that experience shows have something to say about the vote aggregate such a population is likely to yield. Note the vague wording, which is the point; "something to say" is the imperfect diametric opposite of rigid or exact. Because of this slop, none of those predictions are exact, and they certainly cannot possibly apply to an individual; we have no way of knowing what an individual will prefer unless we study him or her personally. But for the population as a whole, those predictors *can* prove useful. For example, still referring to the election outcome example, if it's a low income minority precinct, the Democratic candidate is likelier to prevail, but in an affluent, white Long Island precinct, the GOP candidate is likelier to carry the day. Of course neither of those is more than a rule of thumb; they simply help make better guesstimates.

Forecasting peak electricity loads for tomorrow, or five days hence, can be reasonably done

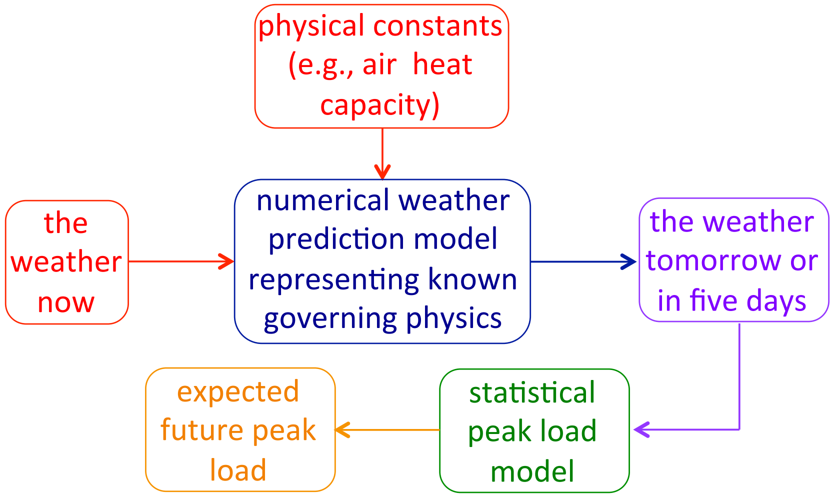

Figure 12: A schematic of the hybrid deterministic–statistical model of future peak loads.

by combining deterministic and statistical forecasting, because the daily fluctuations are mostly weather related. You would use standard numerical weather prediction for getting the expected weather, and then use a statistical model to convert the expected weather into expected electricity loads. To be unambiguous and leave no reader behind, Fig. 12 presents a schematic representation of the model in this example.

In the schematic, red indicates inputs the numerical weather prediction model needs to make its future weather predictions, navy is that model itself (again, the embodiment of our knowledge of the governing physics represented as a long set of computer code instructions), and purple is the output of the weather prediction model (the expected future weather). Most important for us, green is the statistical model. It takes in as input the expected future weather, and predicts from it what the peak load—the final, orange, box—will likely be.

A very simple example of a statistical peak load model may be

$$\begin{pmatrix} \text{expected} \\ \text{peak} \\ \text{load} \\ \text{on} \\ \text{Friday} \end{pmatrix} = \alpha \begin{pmatrix} \text{expected} \\ \text{maximum} \\ \text{temperature} \\ \text{on} \\ \text{Friday} \end{pmatrix} + \beta \begin{pmatrix} \text{expected} \\ \text{humidity} \\ \text{during} \\ \text{Friday's} \\ \text{temperature} \\ \text{maximum} \end{pmatrix}, \tag{41}$$

whose left hand side is what we are after, what we are trying to forecast (orange in the preceding

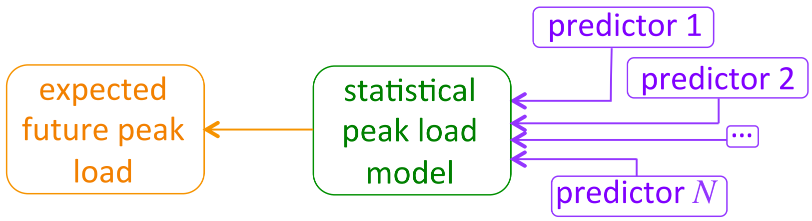

Figure 13: A Schematic of the purely statistical model of future peak loads.

schematic), sometimes called "the predictand" and denoted e. The two inputs on the right hand side (enclosed in parentheses) are predictions the numerical weather prediction model generated, and α and β are this model's parameters. Their values are chosen, using any number of techniques, based on past relationship between predictors—here Friday's expected humidity and maximum temperature—and the predictand, Friday's peak load.

However, beyond a season or so, and certainly for years ahead, peak loads can only be forecasted statistically; there exists no deterministic alternative. At the heart of this dispute, therefore, is statistical forecasting, presented schematically in Fig. 13.

The art and science of statistical forecasting comprises the following elements.

1. Choosing proper predictors (the purple boxes on the right of the above schematic) and *extremely carefully and rigorously* demonstrating that the model with a given predictor is superior to a simpler alternative model without it. Failure to follow this firmly established forecasting protocol is central to why key elements of NYISO's critique are utterly erroneous. More importantly, this and similar violations of basic bedrock tenets of statistical modeling offer windows into major sources of error in NYISO's analysis.
2. Choosing the representations of the chosen predictors in the model. For example, a predictor may be simply its value, or its value squared, or the cosine of its value, or whatever. The possibilities are almost countless. To give this choice an unambiguous precision, let's imagine we have two predictors, call them p_1 and p_2. With the notation introduced earlier, the model can be

$$e_i = \alpha\, p_1 + \beta\, p_2, \tag{42}$$

$$e_i = \alpha\, p_1^2 + \beta\, p_2^3, \tag{43}$$

$$e_i = \alpha \cos(p_1) + \beta\, p_2^{1/3}, \tag{44}$$

or any conceivable alternative.

3. Once 1 & 2 are properly accomplished (*not* a simple matter!!), choosing the best, most representative values of the coefficients, α and β above.

With the proliferation of off the shelf ready to use statistical software packages (such as Microsoft's *Excel* the NYISO analysts favor), any user of those "black boxes" can very easily proceed with each of these steps, cavalierly confusing this autopilot guided exercise with building a forecasting model. The two, alas, are extremely rarely the same; most forecasting efforts carried out by such means—including, notably, NYISO's prediction of future downstate peak loads—shed *no* light on the future.

10 Notes

Section 2

1. Allen, G. E., 1987: The role of experts in scientific controversy. In Engelhardt, H. T. Jr. and A. L. Caplan, *Scientific controversies: Case studies in the resolution and closure of Disputes in Science and Technology*, 169-202, ISBN: 0521275601.

Section 5

1. See p. 431. of Ellis, Arthur K., 1970: *Teaching and Learning Elementary Social Studies*, Allyn and Bacon, 444 pp., ISBN: 9780205086115.

2. Eshel, G., 2012: *Spatiotemporal Data Analysis*, Princeton Univ. Press, 368 pp., ISBN: 9780691128917.

3. Eshel, G., 2012: *Spatiotemporal Data Analysis*, Princeton Univ. Press, 368 pp., ISBN: 9780691128917.

4. Eshel, G., 2012: *Spatiotemporal Data Analysis*, Princeton Univ. Press, 368 pp., ISBN: 9780691128917.

5. http://data.bls.gov.

6. Ryan, Thomas P., 1997: *Modern Regression Methods*, Wiley, New York, 515 pp., ISBN: 0471529125.

7. Burnham, K. P. and D. R. Anderson, 2013: *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer, New York, 488 pp., ISBN: 1441929738.

8. The data are obtained from the Bureau of Economic Analysis Interactive Data site by selecting "GDP by State" and "per capita real GDP".

9. DelSole, T. and J. Shukla, 2009: Artificial skill due to predictor screening, *J. Clim.*, **22**, 331-345.

10. Joffe M. D., 2012: Rating government bonds: Can we raise our grade?, *Econ. J. Watch*, **9**(3),

350–365.

Section 6

1. Ryan, Thomas P., 1997: *Modern Regression Methods*, Wiley, New York, 515 pp., ISBN: 0471529125.

2. Eshel, G., 2012: *Spatiotemporal Data Analysis*, Princeton Univ. Press, 368 pp., ISBN: 9780691128917.

3. Eshel, G., 2012: *Spatiotemporal Data Analysis*, Princeton Univ. Press, 368 pp., ISBN: 9780691128917.

4. New York State Energy Research and Development Authority, 2014: *Energy Efficiency and Renewable Energy Potential Study of New York State*, Vol. 2, NYSERDA Report 14-19, April 2014, Fig. 3, p. 13.

5. Woolf, T., M. Wittenstein and R. Fagan, 2011:*Indian Point Energy Center Nuclear Plant Retirement Analysis: Replacement Options, Reliability Issues and Economic Effects*, Synapse Energy Economics, Inc., Cambridge, MA, October 17 2011, 33 pp.

Section 7

1. Van Buskirk, R. D., C. Kantner, B. F. Gerke and S. Chu, 2014: A retrospective investigation of energy efficiency standards: Policies may have accelerated long term declines in appliance costs, *Env. Res. Let.*, **9**(11), doi: 10.1088/1748-9326/9/11/114010.