Statistical Basis for Predicting Technological Progress

Béla Nagy
J. Doyne Farmer
Quan M. Bui
Jessika E. Trancik

SFI WORKING PAPER: 2012-07-008

SANTA FE INSTITUTE

Statistical Basis for Predicting Technological Progress

Béla Nagy[1,*], J. Doyne Farmer[1], Quan M. Bui[1,2] Jessika E Trancik[1,3,*]
1 Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
2 St. John's College,1160 Camino Cruz Blanca, Santa Fe, NM, 87505, USA
3 Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
∗ **E-mail: Corresponding bn@santafe.edu, trancik@mit.edu**

Abstract

Forecasting technological progress is of great interest to engineers, policy makers, and private investors. Several models have been proposed for predicting technological improvement, but how well do these models perform? An early hypothesis made by Theodore Wright in 1936 is that cost decreases as a power law of cumulative production. An alternative hypothesis is Moore's law, which can be generalized to say that technologies improve exponentially with time. Other alternatives were proposed by Goddard, Sinclair et al., and Nordhaus. These hypotheses have not previously been rigorously tested. Using a new database on the cost and production of 62 different technologies, which is the most expansive of its kind, we test the ability of six different postulated laws to predict future costs. Our approach involves hindcasting and developing a statistical model to rank the performance of the postulated laws. Wright's law produces the best forecasts, but Moore's law is not far behind. We discover a previously unobserved regularity that production tends to increase exponentially. A combination of an exponential decrease in cost and an exponential increase in production would make Moore's law and Wright's law indistinguishable, as originally pointed out by Sahal. We show for the first time that these regularities are observed in data to such a degree that the performance of these two laws is nearly tied. Our results show that technological progress is forecastable, with the square root of the logarithmic error growing linearly with the forecasting horizon at a typical rate of 2.5% per year. These results have implications for theories of technological change, and assessments of candidate technologies and policies for climate change mitigation.

Introduction

Innovation is by definition new and unexpected, and might therefore seem inherently unpredictable. But if there is a degree of predictability in technological innovation, understanding it could have profound implications. Such knowledge could result in better theories of economic growth, and enable more effective strategies for engineering design, public policy design, and private investment. In the area of climate change mitigation, the estimated cost of achieving a given greenhouse gas concentration stabilization target is highly sensitive to assumptions about future technological progress [1].

There are many hypotheses about technological progress, but are they any good? Which, if any, hypothesis provides good forecasts? In this paper, we present the first statistically rigorous comparison of competing proposals.

When we think about progress in technologies, the first product that comes to mind for many is a computer, or more generally, information technologies. The following quote by Bill Gates captures a commonly held view: "Exponential improvement – that is rare – we've all been spoiled and deeply confused by the IT model" [2]. But as we demonstrate here, information technologies are not special in terms of the functional form that describes their improvement over time. Information technologies show rapid rates of improvement, but many technologies show exponential improvement. In fact, all the technologies we study here behave roughly similarly: Information technologies closely follow patterns of improvement originally postulated by Wright for airplanes [3–8], and technologies such as beer production

or offshore gas pipelines follow Moore's law [9, 10], but with a slower rate of improvement [8, 11–15].

It is not possible to quantify the performance of a technology with a single number [16]. A computer, for example, is characterized by speed, storage capacity, size and cost, as well as other intangible characteristics such as aesthetics. One automobile may be faster while another is less expensive. For this study we focus on one common measure of performance: the inflation-adjusted cost of one "unit". This metric is suitable in that it can be used to describe many different technologies. However, the nature of a unit may change over time. For example, a transistor in a modern integrated circuit today may have quite different performance characteristics than its discrete counterpart in the past. Furthermore, the degree to which cost is emphasized over other performance measures may change with time [17]. We nonetheless use the changes in the unit cost as our measure of progress, in order to compare competing models using a sizable dataset. The crudeness of this approach only increases the difficulty of forecasting and makes it particularly surprising that we nonetheless observe common trends.

Analysis

We test six different hypotheses that have appeared in the literature [3, 9, 18–20], corresponding to the following six functional forms:

$$\begin{aligned}
\text{Moore} \quad \log y_t &= at + b + n(t) \qquad (1)\\
\text{Wright} \quad \log y_t &= a \log x_t + b + n(t)\\
\text{lagged Wright} \quad \log y_t &= a \log(x_t - q_t) + b + n(t)\\
\text{Goddard} \quad \log y_t &= a \log q_t + b + n(t)\\
\text{SKC} \quad \log y_t &= a \log q_t + c \log(x_t - q_t) + b + n(t)\\
\text{Nordhaus} \quad \log y_t &= at + c \log x_t + b + n(t)
\end{aligned}$$

The dependent variable y_t is the unit cost of the technology measured in inflation adjusted dollars. The independent variables are the time t (measured in years), the annual production q_t, and the cumulative production $x_t = \sum_{i=1}^{t} q_i$. The noise term $n(t)$, the constants a, b and c and the predictor variables differ for each hypothesis.

Moore's law here refers to the generalized statement that the cost y of a given technology decreases exponentially with time, i.e.

$$y_t = B \exp(-mt), \qquad (2)$$

where $m > 0$ and $B > 0$ are constants [9, 12]. (We assume throughout that $t > 0$, and we have renamed $a = -m$ and $b = \log B$ in Eq. (1)). Moore's law postulates that technological progress is inexorable, i.e. it depends on time rather than controllable factors such as research and development.

Wright's law, in contrast, postulates that cost decreases at a rate that depends on cumulative production, i.e.

$$y_t = B x_t^{-w}, \qquad (3)$$

where $w > 0$ and $B > 0$ are constants, and we have renamed $a = -w$ and $b = \log B$ in Eq. (1). Wright's law is often interpreted to imply "learning by doing" [5, 21]. The basic idea is that cumulative production is a proxy for level of effort, so that the more we make the more we learn, and knowledge accumulates without loss.

Another hypothesis is due to Goddard [18], who argues that progress is driven purely by economies of scale, and postulates that:

$$y_t = B q_t^{-s}, \qquad (4)$$

where $s > 0$ and $B > 0$ are constants, and we have renamed $a = -s$ and $b = \log B$ in Eq. (1).

We also consider the three multi-variable hypotheses in Eq. (1): Nordhaus [20] combines Wright's law and Moore's law[1], and Sinclair, Klepper, and Cohen (SKC) [19] combine Wright's law and Goddard's law. For completeness we also test Wright's law lagged by one year. Note that these methods forecast different things: Moore's law forecasts the cost at a given time, Wright's law at a given cumulative production, and Goddard's law at a given annual production.

We test these hypotheses on historical data consisting of 62 different technologies that can be broadly grouped into four categories: Chemical, Hardware, Energy, and Other[2]. The data are sampled at annual intervals with timespans ranging from 10 to 39 years. The choice of these particular technologies was driven by availability – we included all of the data that we could find to assemble the largest database of its kind. For a detailed description see the Supporting Information.

To compare the performance of each hypothesis we use hindcasting, which is a form of cross-validation. We pretend to be at time i and make a forecast $\hat{y}_j^{(f,d,i)}$ for time j using hypothesis (functional form) f and data set d, where $j > i$. The parameters for each functional form are fit using ordinary least squares based on all data prior to time i, and forecasts are made based on the resulting regression[3]. We score the quality of forecasts based on the logarithmic forecasting error

$$e_{fdij} = \log y_j^{(d)} - \log \hat{y}_j^{(f,d,i)} \tag{5}$$

The quality of forecasts is examined for all datasets and all hypotheses (and visualized as a three-dimensional error mountain, as shown in the Supporting Information). For Wright's law an illustration of the growth of forecasting errors as a function of the forecasting horizon is given in Fig. **??**.

Developing a statistical model to compare the competing hypotheses is complicated by the fact that errors observed at longer horizons tend to be larger than those at shorter horizons, and errors are correlated across time and across functional forms. After comparing many different possibilities (as discussed in detail in the Supporting Information), we settled on the following approach. Based on a search of the family of power transformations, which is known for its ability to accommodate a range of variance structures, we take as a response the square root transformation of the logarithmic error. This response was chosen to maximize likelihood when modeled as a linear function of the hindcasting horizon = target − origin $= j - i$, using a linear mixed effects model.

Specifically, we use the following functional form to model the response.

$$r_{fdij} \equiv |e_{fdij}|^{0.5} = \alpha_f + a_d + (\beta_f + b_d)(j - i) + \epsilon_{fdij}, \tag{6}$$

where r_{fdij} is the expected root error. The parameters α_f and β_f depend on the functional form and are called *fixed effects* because they are the same for all datasets. α_f is the intercept and β_f is the slope parameter.

The parameters a_d and b_d depend on the dataset, and are called *random effects* because they are not fitted independently, but are instead treated as dataset-specific random fluctuations from the pooled data. The quantities a_d and b_d are additive adjustments to the average intercept and slope parameters α_f and β_f, respectively, to take into account the peculiarities of each dataset d.

In order to avoid adding 62 a_d parameters plus 62 b_d parameters, we treated the $\begin{pmatrix} a_d \\ b_d \end{pmatrix}$ pair as a two-dimensional random vector having a bivariate normal distribution with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and variance-

[1] Note that the conclusions presented do not work against Nordhaus' point about the difficulty in separating learning from exogenous sources of change [20].

[2] All data can be found in the online Performance Curve Database at pcdb.santafe.edu.

[3] An alternative is to adjust the intercepts to match the last point, which produces better short term forecasts. For example, for Moore's law this corresponds to using a log random walk of the form $\log y_{t+1} = \log y_t - \mu + n(t)$, where $n(t)$ is an IID noise term. We have not done this here to be consistent with the way these hypotheses have been historically presented. The method we have used here also results in more stable errors. Our purpose here is not to propose an optimal forecasting method, but rather to compare existing hypotheses.

covariance matrix $\begin{pmatrix} \psi_a^2 & \psi_{ab} \\ \psi_{ab} & \psi_b^2 \end{pmatrix}$. This approach dramatically reduces the number of parameters. We parameterize the dataset-specific adjustments as random deviations from the average $\begin{pmatrix} \alpha_f \\ \beta_f \end{pmatrix}$ at a cost of only 3 additional parameters instead of $2 \times 62 = 124$. This parsimonious approach makes maximum likelihood estimation possible by keeping the number of parameters in check.

Finally, we add an ε_{fdij} random field term to take into account the deviations from the trend. This is assumed to be a Gaussian stochastic process independent of the $\begin{pmatrix} a_d \\ b_d \end{pmatrix}$ random vector, having mean 0, and given a_d and b_d, having variance equal to a positive σ^2 times the fitted values:

$$\mathrm{Var}\left(\varepsilon_{fdij} \middle| a_d, b_d\right) \;=\; \sigma^2 \,\mathrm{E}\left(r_{fdij} \middle| a_d, b_d\right) \tag{7}$$

We also define an exponential correlation structure within each error mountain (corresponding to each combination of dataset and hypothesis, see the Supporting Information), as a function of the differences of the two time coordinates with a positive range parameter ρ and another small positive nugget parameter η quantifying the extent of these correlations:

$$\mathrm{Corr}(\varepsilon_{fdij},\, \varepsilon_{f'd'i'j'}) \;=\; \delta_{ff'}\delta_{dd'}(1-\eta)\exp\left\{-(|i-i'|+|j-j'|)/\rho\right\}, \tag{8}$$

where the two Kronecker δ functions ensure that each error mountain is treated as a separate entity.

Equations (7) and (8) were chosen to deal with the observed heteroscedasticity (increasing variance with increasing logarithmic forecasting error) and the serial correlations along the time coordinates i (hindcasting origin) and j (hindcasting target). Based on the likelihood, an exponential correlation function provided the best fit. Note that instead of a Euclidean distance (root sum of the squares of differences), the Manhattan measure was used (the sum of the absolute differences), because it provided a better fit in terms of the likelihood.

Using this statistical model, we compared five different hypotheses. (We removed the Nordhaus model from the sample because of poor forecasting performance. This model gave good in-sample fits but generated large and inconsistent errors when predicting out-of-sample, a signature of over-fitting.) Rather than the $62 \times 5 \times 2 = 620$ parameters needed to fit each of the 62 datasets separately for each of the five functional forms, there are only 16 free parameters: $5 \times 2 = 10$ parameters α_f and β_f, three parameters for the covariance matrix of the bivariate random vector (a_d, b_d), and three parameters for the variance and autocorrelation of the residuals ϵ_{fdij}.

Results and Discussion

We fit the error model to the 37,745 different r_{fdij} data points using the method of maximum likelihood. In Fig. 2 we plot the expected root error $r_{fij} = \alpha_f + \beta_f(j-i)$ for the five hypotheses as a function of the hindcasting horizon. While there are differences in the performance of these five hypotheses, they are not dramatic. The intercept is tightly clustered in a range $0.16 < \alpha_f < 0.19$ and the slope $0.024 < \beta_f < 0.028$. Thus all the hypotheses show a large initial error, followed by a growth in the root error of roughly 2.5% per year[4].

The error model allows us to compare each hypothesis pairwise to determine whether it is possible to reject one in favor of another at statistically significant levels. The comparisons are based on the intercept and slope of the error model of Eq. (6). The parameter estimates are listed in Tables S1 and S3 and the corresponding p-values in Tables S2 and S4. For example, at the 5% level, the intercept of Goddard is significantly higher than any of the others and the slope of SKC is significantly greater than that of Wright, lagged Wright and Goddard. With respect to slope, Moore is at the boundary of being rejected

[4]This is a central tendency for the pooled data.

in favor of Wright. Fig. 2 makes the basic pattern clear: Goddard does a poorer job of forecasting at short times, whereas SKC and to a lesser extent Moore do a poorer job at long times.

We thus have the surprising result that most of the methods are quite similar in their performance. Although the difference is not large, the fact that we can eliminate Goddard for short term forecasts indicates that there is information in the cumulative production not contained in the annual production, and suggests that there is a learning effect in addition to economies of scale. But the fact that Goddard is not that much worse indicates that much of the predictability comes from annual production, suggesting that economies of scale are important. (In our database technologies rarely decrease significantly in annual production; examples of this would provide a better test of Goddard's theory.) We believe the SKC model performs worse at long times because it has an extra parameter, making it prone to overfitting.

Although Moore performs slightly worse than Wright, given the clear difference in their economic interpretation, it is surprising that their performance is so similar. A simple explanation for Wright's law in terms of Moore's law was originally put forward by Sahal [22]. He noted that if cumulative production grows exponentially[5], i.e.

$$x_t = A\exp(gt), \tag{9}$$

then eliminating t between Eqs. (2) and (9) results in Wright's law, Eq. (3), with $w = m/g$. Indeed, when we look at production vs. time we find that in almost every case the cumulative production increases roughly exponentially with time. This is illustrated in Fig. 3, where we show three representative examples for production and cost plotted as a function of time. Fig. 3 also shows histograms of R^2 values for fitting g and m for the 62 datasets. The agreement with exponential behavior ranges from very good to rather poor, but of course these are short time series and some of them are very noisy.

We test this in Fig. 4 by plotting the measured value of w_d against the derived value $\hat{w}_d = m/g$ for each data set d. The values cluster tightly along the identity line, indicating that Sahal's conjecture is correct.

The differences in the data sets can be visualized by plotting a_d and b_d as shown in Fig. 5. All but one data set is inside the 95% confidence ellipsoid, indicating that the estimated distribution of (a_d, b_d) is consistent with the bivariate normal assumption. The intercepts vary in a range roughly $-0.10 < a_d < 0.17$ and the slopes $-0.018 < b_d < 0.015$. Thus the variation in the corresponding logarithmic forecasting error for the different datasets is comparable to the average error for all datasets (Fig. 5) and about an order of magnitude larger than the difference between the hypothesized laws (Fig. 2).

To illustrate the practical usefulness of our approach we make a forecast of the cost of electricity for residential scale photovoltaic solar systems (PV). Fig. 6 shows the best forecast (solid line) as well as the expected error (dashed lines)[6]. The expected cost in 2020 is 6 cents/kWh, with a range (3, 12), and in 2030 it is 2 cents, with a range (0.4, 11)[7]. The current cost of the cheapest alternative, coal-fired electricity, is roughly 5 cents/kWh[8]. In contrast to PV, coal-fired electricity is not expected to decrease in cost, and will likely increase if there are future penalties for CO_2 emissions [23].

The key postulate that we have made here is that the processes generating the costs of technologies through time are generic except for technology-specific differences in parameters. This hypothesis is powerful in allowing us to to view any given technology as being drawn from an ensemble. This means that we can pool data from different technologies to make better forecasts, and most importantly, make

[5] Note that if production grows exponentially, cumulative production also grows exponentially with the same exponent.

[6] Note that these are not confidence limits, but rather projected absolute log deviations from the best forecast, calculated from Eq. (6) using α_{Moore}, β_{Moore}, $a_{Photovoltaics2}$, and $b_{Photovoltaics2}$. The sharp drop in the one year forecast relative to the last observed data point comes from the fact that forecasts are based on the average trend line, and because this data series is particularly long. PV costs have risen recently due to increased material costs and other effects, but many industry experts expect this to be a short-lived aberration from the long-term cost trend. See footnote 4 and Section 5 of the Supporting Information.

[7] This does not include the additional cost of energy storage technologies. Note also that this is for residential scale PV. Industrial scale PV is typically about two-thirds the cost of electricity from residential scale systems.

[8] This is the wholesale cost at the plant (busbar), which may be most directly comparable to industrial scale PV.

error estimates. This is particularly useful in studying technology trends, where available data is limited. Of course we must add the usual caveats about making forecasts – as Neils Bohr reputedly said, prediction is very difficult, especially of the future.

Our analysis reveals that decreasing costs and increasing production appear to be closely related, and that the hypotheses of Wright and Moore are more similar than they might appear. We should stress, though, that they are not the same. For example, consider a scenario in which the exponential rate of growth of PV production suddenly increased, which would decrease the current production doubling time of roughly 3 years. In this case, Wright predicts that the rate at which costs fall would increase, whereas Moore predicts that it would be unaffected. Distinguishing between the two hypotheses requires a sufficient number of examples where production does not increase exponentially, which our current database does not contain. The historical data shows a strong tendency, across different types of technologies, toward constant exponential growth rates. Recent work has, however, demonstrated super-exponential improvement for information technologies [24] over long time spans, suggesting that Moore's law is only a reasonable approximation over short spans of time. This evidence from information technology, and the results presented here, suggest that Moore may perform significantly worse than Wright over longer time horizons.

Acknowledgments

This work was funded under National Science Foundation grant NSF-0738187. All the conclusions and opinions are those of the authors and do not necessarily reflect those of the NSF. Partial funding for this work was provided by The Boeing Company (Contact: Scott H. Mathews, TF). We thank all contributors to the performance curve database (pcdb.santafe.edu).

References

1. Gillingham K, Newell RG, Pizer WA (2008) Modeling endogenous technological change for climate policy analysis. Energy Economics 30: 2734 - 2753.

2. Fried I (2010) Gates: We've been spoiled by Moore's Law. CNET News .

3. Wright TP (1936) Factors affecting the costs of airplanes. Journal of Aeronautical Sciences 10: 302-328.

4. Argote L, Epple D (1990) Learning curves in manufacturing. Science 247: 920-924.

5. Dutton JM, Thomas A (1984) Treating progress functions as a managerial opportunity. Acadamy of Management Review 9: 235-247.

6. IEA (2000) Experience Curves for Energy Technology Policy. OECD/IEA.

7. McDonald A, Schrattenholzer L (2001) Learning rates for energy technologies. Energy Policy 29: 255-261.

8. Alberth S (2008) Forecasting technology costs via the experience curve - myth or magic? Technological Forecasting and Social Change 75: 952-983.

9. Moore GE (1965) Cramming more components onto integrated circuits. Electronics Magazine 38.

10. Moore GE (1975) Progress in digital integrated electronics. IEEE International Electron Devices Meeting : 11-13.

11. Kurzweil R (2005) The Singularity Is Near: When Humans Transcend Biology. Viking Penguin. http://singularity.com/.

12. Koh H, Magee CL (2006) A functional approach for studying technological progress: Application to information technology. Technological Forecasting and Social Change 73: 1061-1083.

13. Nordhaus WD (2007) Two centuries of productivity growth in computing. The Journal of Economic History 67: 128-159.

14. Koh H, Magee CL (2008) A functional approach for studying technological progress: Extension to energy technology. Technological Forecasting and Social Change 75: 735-758.

15. Amaya MA, Magee CL (2008) The progress in wireless data transport and its role in the evolving internet. Working paper, Massachusetts Institute of Technology.

16. Dosi G, Nelson RR (2010) Technical change and industrial dynamics as evolutionary processes. In: Hall B, Rosenberg N, editors, Handbook of the Economics of Innovation, Elsevier/North Holland.

17. Abernathy WJ, Wayne K (1974) Limits of the learning curve. Harvard Business Review : 109-119.

18. Goddard C (1982) Debunking the learning curve. IEEE Transactions on Components, Hybrids, and Manufacturing Technology 5: 328-335.

19. Sinclair G, Klepper S, Cohen W (2000) What's experience got to do with it? sources of cost reduction in a large specialty chemical producer. Management Science 46: 28-45.

20. Nordhaus WD (2009) The perils of the learning model for modeling endogenous technological change. Technical report, National Bureau of Economic Research.

21. Thompson P (2010) Learning by doing. In: Hall B, Rosenberg N, editors, Handbook of the Economics of Innovation, Elsevier/North Holland.

22. Sahal D (1979) A theory of progress functions. AIIE Transactions 11: 23-29.

23. McNerney J, Farmer JD, Trancik JE (2011) Historical costs of coal-fired electricity and implications for the future. Energy Policy 39: 3042-3054.

24. Nagy B, Farmer JD, Trancik JE, Gonzales JP (2011) Superexponential long-term trends in information technology. Technological Forecasting and Social Change 78: 1356 - 1364.

Figure Legends

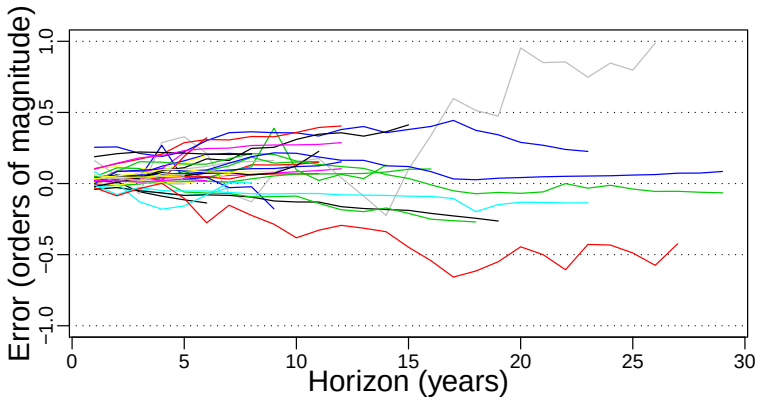

Figure 1. An illustration of the growth of errors with time using the Wright model. The mean value of the logarithmic hindcasting error for each dataset is plotted against the hindcasting horizon $j - i$, in years. An error of $10^{0.5} \approx 3$, for example, indicates that the predicted value is three times as big as the actual value. The longest data-sets are: PrimaryAluminum (green), PrimaryMagnesium (dark blue), DRAM (grey), and Transistor (red).

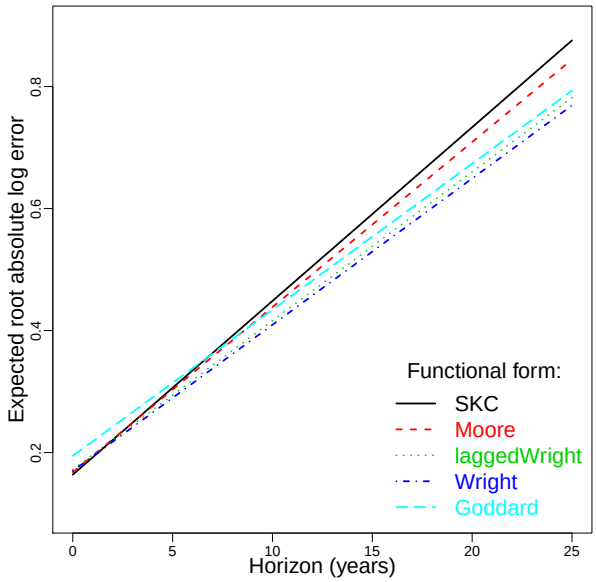

Figure 2. An illustration of the growth of errors of each hypothesized law vs. time. The plot shows the predicted root absolute log error r_{fij} vs. forecasting horizon $(j - i)$ using each of the functional forms (see Eq. (6)). The performance of the five hypotheses shown is fairly similar, though Goddard is worse at short horizons and SKC and Moore are worse at long horizons.

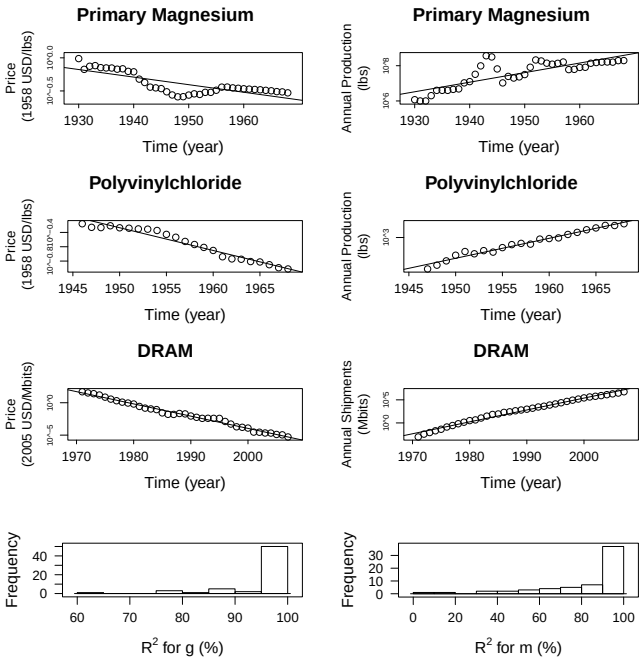

Figure 3. Three examples showing the logarithm of price as a function of time in the left column and the logarithm of production as a function of time in the right column, based on industry-wide data. We have chosen these examples to be representative: The top row contains an example with one of the worst fits, the second row an example with an intermediate goodness of fit, and the third row one of the best examples. The fourth row of the figure shows histograms of R^2 values for fitting g and m for the 62 datasets.

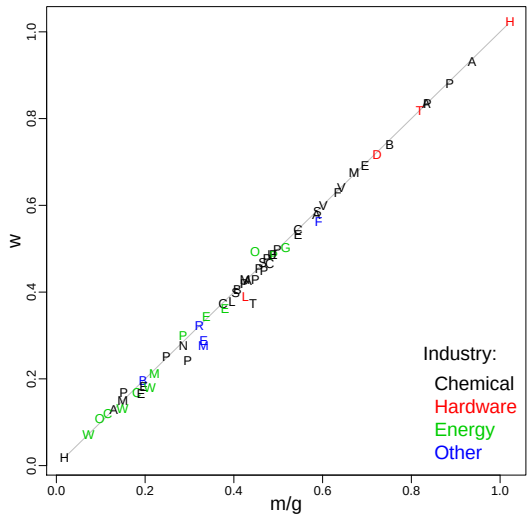

Figure 4. An illustration that the combination of exponentially increasing production and exponentially decreasing cost are equivalent to Wright's law. The value of the Wright parameter w is plotted against the prediction m/g based on the Sahal formula, where m is the exponent of cost reduction and g the exponent of the increase in cumulative production.

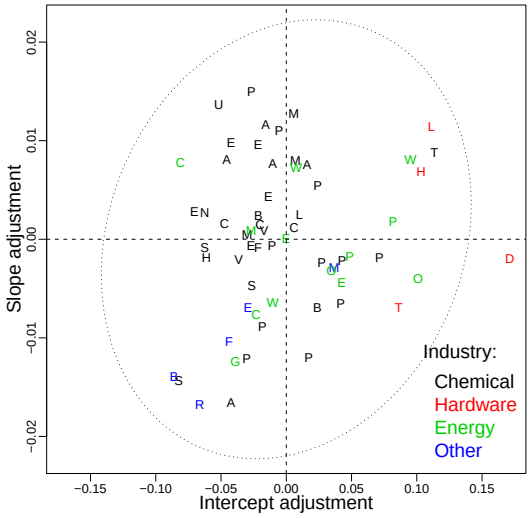

Figure 5. An illustration of how individual datasets deviate from the pooled data. The data-specific contribution to the slope, b_d, is plotted against the data specific contribution to the intercept, a_d, and compared to the ellipse of two standard deviation errors. The best forecasts are obtained for those found in the lower left quadrant, such as Beer, Sodium, RefinedCaneSugar, and Aluminum.

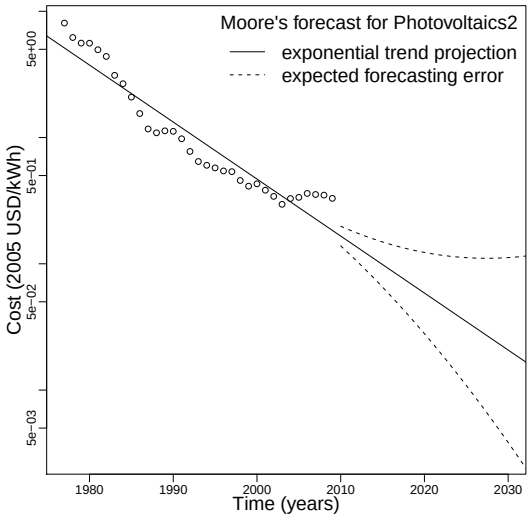

Figure 6. A projection of future PV electricity costs from the Photovoltaics2 historical data set (1977 – 2009) using Moore's exponential functional form. The solid line is the expected forecast and the dashed line is the expected error.

Statistical Basis for Predicting Technological Progress Supporting Information

Béla Nagy, J. Doyne Farmer, Quan Minh Bui, Jessika E. Trancik

March 27, 2012

Contents

1 Data set

This section contains four tables for the 62 technologies used in the paper, divided into four industry groups: Chemical, Hardware, Energy, and Other. In each row, after the name of a particular product, a contiguous time period is specified for which data was available, followed by the number of data points in the resulting yearly time series. Then the parameter estimates g, m, and w are given, followed by the corresponding cumulative production volume doubling times, unit price halving times, and progress ratios, respectively.

References for all the data sources and all the data for the 62 technologies used in this paper are available for online visualization and free download in the Performance Curve Database at `http://pcdb.santafe.edu/`. Based on what was available in this database as of November

Chemical Industry	*time period*	*data points*	*g*	*m*	*w*	*doubling time*	*halving time*	*progress ratio*
AcrylicFiber	1960 - 1972	13	0.076	0.045	0.58	4.0	6.8	0.67
Acrylonitrile	1959 - 1972	14	0.077	0.033	0.43	3.9	9.1	0.74
Aluminum	1956 - 1972	17	0.035	0.004	0.13	8.7	67	0.91
Ammonia	1960 - 1972	13	0.047	0.039	0.83	6.4	7.7	0.56
Aniline	1961 - 1972	12	0.027	0.025	0.93	11	12	0.52
Benzene	1953 - 1968	16	0.036	0.027	0.74	8.4	11	0.60
BisphenolA	1959 - 1972	14	0.065	0.027	0.41	4.6	11	0.76
Caprolactam	1962 - 1972	11	0.092	0.050	0.55	3.3	6.0	0.69
CarbonDisulfide	1963 - 1972	10	0.019	0.009	0.47	16	32	0.72
Cyclohexane	1956 - 1972	17	0.060	0.023	0.37	5.0	13	0.77
Ethanolamine	1955 - 1972	18	0.049	0.027	0.53	6.1	11	0.69
EthylAlcohol	1958 - 1972	15	0.031	0.006	0.17	9.8	51	0.89
Ethylene	1954 - 1968	15	0.083	0.016	0.18	3.6	19	0.88
Ethylene2	1960 - 1972	13	0.058	0.028	0.49	5.2	11	0.71
EthyleneGlycol	1960 - 1972	13	0.041	0.029	0.69	7.3	10	0.62
Formaldehyde	1962 - 1972	11	0.041	0.026	0.63	7.4	12	0.65
HydrofluoricAcid	1962 - 1972	11	0.035	0.001	0.018	8.5	460	0.99
LowDensityPolyethylene	1953 - 1968	16	0.11	0.044	0.38	2.7	6.8	0.77
Magnesium	1954 - 1972	19	0.022	0.003	0.15	13	90	0.90
MaleicAnhydride	1959 - 1972	14	0.055	0.024	0.43	5.4	13	0.74
Methanol	1957 - 1972	16	0.038	0.025	0.68	8.0	12	0.63
NeopreneRubber	1960 - 1972	13	0.033	0.009	0.28	9.1	32	0.82
Paraxylene	1958 - 1968	11	0.10	0.043	0.42	3.0	7.0	0.75
Pentaerythritol	1952 - 1972	21	0.039	0.018	0.45	7.7	17	0.73
Phenol	1959 - 1972	14	0.042	0.035	0.83	7.1	8.5	0.56
PhthalicAnhydride	1955 - 1972	18	0.035	0.031	0.88	8.6	9.7	0.54
PolyesterFiber	1960 - 1972	13	0.12	0.059	0.48	2.4	5.1	0.72
PolyethyleneHD	1958 - 1972	15	0.093	0.042	0.46	3.2	7.1	0.73
PolyethyleneLD	1958 - 1972	15	0.077	0.038	0.50	3.9	7.8	0.71
Polystyrene	1944 - 1968	25	0.086	0.025	0.24	3.5	12	0.84
Polyvinylchloride	1947 - 1968	22	0.073	0.033	0.43	4.1	9.2	0.74
PrimaryAluminum	1930 - 1968	39	0.044	0.011	0.25	6.8	28	0.84
PrimaryMagnesium	1930 - 1968	39	0.075	0.011	0.17	4.0	26	0.89
Sodium	1957 - 1972	16	0.014	0.007	0.47	21	45	0.72
SodiumChlorate	1958 - 1972	15	0.043	0.017	0.40	7.0	17	0.76
Styrene	1958 - 1972	15	0.051	0.030	0.59	5.9	10	0.67
TitaniumSponge	1951 - 1968	18	0.12	0.051	0.38	2.6	5.9	0.77
Urea	1961 - 1972	12	0.065	0.032	0.49	4.6	9.5	0.71
VinylAcetate	1960 - 1972	13	0.055	0.033	0.60	5.5	9.1	0.66
VinylChloride	1962 - 1972	11	0.061	0.039	0.64	5.0	7.7	0.64

Hardware Industry	*time period*	*data points*	g	m	w	*doubling time*	*halving time*	*progress ratio*
DRAM	1972 - 2007	36	0.26	0.19	0.72	1.2	1.6	0.61
HardDiskDrive	1989 - 2007	19	0.28	0.28	1.0	1.1	1.1	0.49
LaserDiode	1983 - 1994	12	0.32	0.14	0.39	0.95	2.2	0.76
Transistor	1969 - 2005	37	0.26	0.21	0.82	1.2	1.4	0.57

Energy Industry	*time period*	*data points*	g	m	w	*doubling time*	*halving time*	*progress ratio*
CCGTElectricity	1987 - 1996	10	0.075	0.009	0.12	4.0	34	0.92
CrudeOil	1947 - 1968	22	0.025	0.004	0.17	12	68	0.89
ElectricPower	1940 - 1968	29	0.046	0.016	0.34	6.5	19	0.79
Ethanol	1981 - 2004	24	0.06	0.023	0.36	5.0	13	0.78
GeothermalElectricity	1980 - 2005	26	0.042	0.022	0.50	7.2	14	0.71
MotorGasoline	1947 - 1968	22	0.028	0.006	0.21	11	48	0.86
OffshoreGasPipeline	1985 - 1995	11	0.11	0.049	0.49	2.7	6.1	0.71
OnshoreGasPipeline	1980 - 1992	13	0.068	0.007	0.11	4.4	45	0.93
Photovoltaics	1976 - 2003	28	0.097	0.028	0.30	3.1	11	0.81
Photovoltaics2	1977 - 2009	33	0.092	0.045	0.48	3.3	6.7	0.71
WindElectricity	1984 - 2005	22	0.19	0.040	0.18	1.6	7.5	0.88
WindTurbine	1982 - 2000	19	0.12	0.018	0.13	2.5	17	0.91
WindTurbine2	1988 - 2000	13	0.23	0.017	0.073	1.3	18	0.95

Other Industry	*time period*	*data points*	g	m	w	*doubling time*	*halving time*	*progress ratio*
Beer	1952 - 1968	17	0.077	0.015	0.20	3.9	20	0.87
ElectricRange	1947 - 1967	21	0.029	0.010	0.29	10	31	0.82
FreeStandingGasRange	1947 - 1967	21	0.014	0.009	0.56	21	35	0.68
MonochromeTelevision	1948 - 1968	21	0.074	0.024	0.28	4.1	12	0.82
RefinedCaneSugar	1936 - 1968	33	0.006	0.002	0.32	47	150	0.80

Table 1: Statistics for the datasets used in this study. g is the exponent for the increase in production, m the exponent for the drop in cost, w the exponent for Wright's law, the doubling time refers to the increase in production, the halving time to the decrease in cost, and the progress ratio is 2^{-w}, interpreted as the drop in cost with a doubling of production. All times are in years.

2010 we selected the performance curves that had both price and production data for at least 10 years, with no missing values in between. The resulting 62 datasets are commodity-type products that retained a functional unit equivalence during their (at least decade-long) evolution.

For example, the Transistor dataset represents many different technologies that have been in production during the 37 years in its time period. Transistors have changed dramatically during that period in terms of factors such as speed and power requirements. Nonetheless, one could very crudely say that a transistor as a functional unit is equivalent to any other; hence, we make the crude approximation of comparing the price of such a unit from the sixties to one from the 2000s. Similarly, a bit can be viewed as a functional unit for the DRAM and HardDiskDrive data.

All unit prices are yearly averages after adjusting for inflation. The Transistor, DRAM, HardDiskDrive, and Photovoltaics2 data was converted to real 2005 U.S. dollars using the GDP deflator. The other data sets were published previously, and were converted to real values by their respective authors.

Since originally for each dataset either only yearly production or only total cumulative production was available, we obtained the missing variable from the other one by adding up yearly production or differencing cumulative production, respectively. In order to avoid missing values, this resulted in a shortening of the original time period by one year because for the first year either the previous experience measure was absent (assumed to be zero) or the yearly production was unknown.

2 Exponential increase of production

In this section we present additional evidence for the exponential increase in production, which is one of our new findings in this paper. One way to evaluate the assumption of the constant growth rates g and m and the constant learning rate determined by w is to assess goodness of fit by looking at the distribution of R^2 percentages for the regression lines used for estimating g, m, and w. High R^2 values in Supplementary Figure 1 indicate that in most cases the exponential approximation is accurate for g and m, and the power law fit is good for w.

3 Hindcasting results

We illustrate the hindcasting method using the Transistor dataset. This time series started in the year 1969 and ended in 2005. Thus the first hindcast was made in 1973 based on only five data points, targeting the rest of the time period from 1974 to 2005. The last hindcast was based on the data from 1969 to 2004 having a single target: the year 2005. The resulting projection lines over the actual data points are drawn in Supplementary Figure 2 for the six different functional forms. The most striking feature of Moore's projections is that they consistently underestimated prices for the latter half of the Transistor dataset. On the other hand, Nordhaus's projections strayed from the data in both directions. In contrast, we can see that the other four functional forms demonstrated a more satisfactory prediction performance on this particular dataset, having projection lines near the actual data points and thus avoiding huge deviations (that are measured on the log scale).

Another way to visualize the prediction errors of the hindcasting procedure is to plot them as a surface over two time coordinates: the origin and the target of the hindcasts. Since the target year is always after the origin, the result is a mountain of prediction errors over a triangular area. Figures 3 to 8 use topographical colors to indicate the magnitude of these errors on the log scale (base 10). Note that the coloring scheme only extends from the sea level in blue (meaning no or

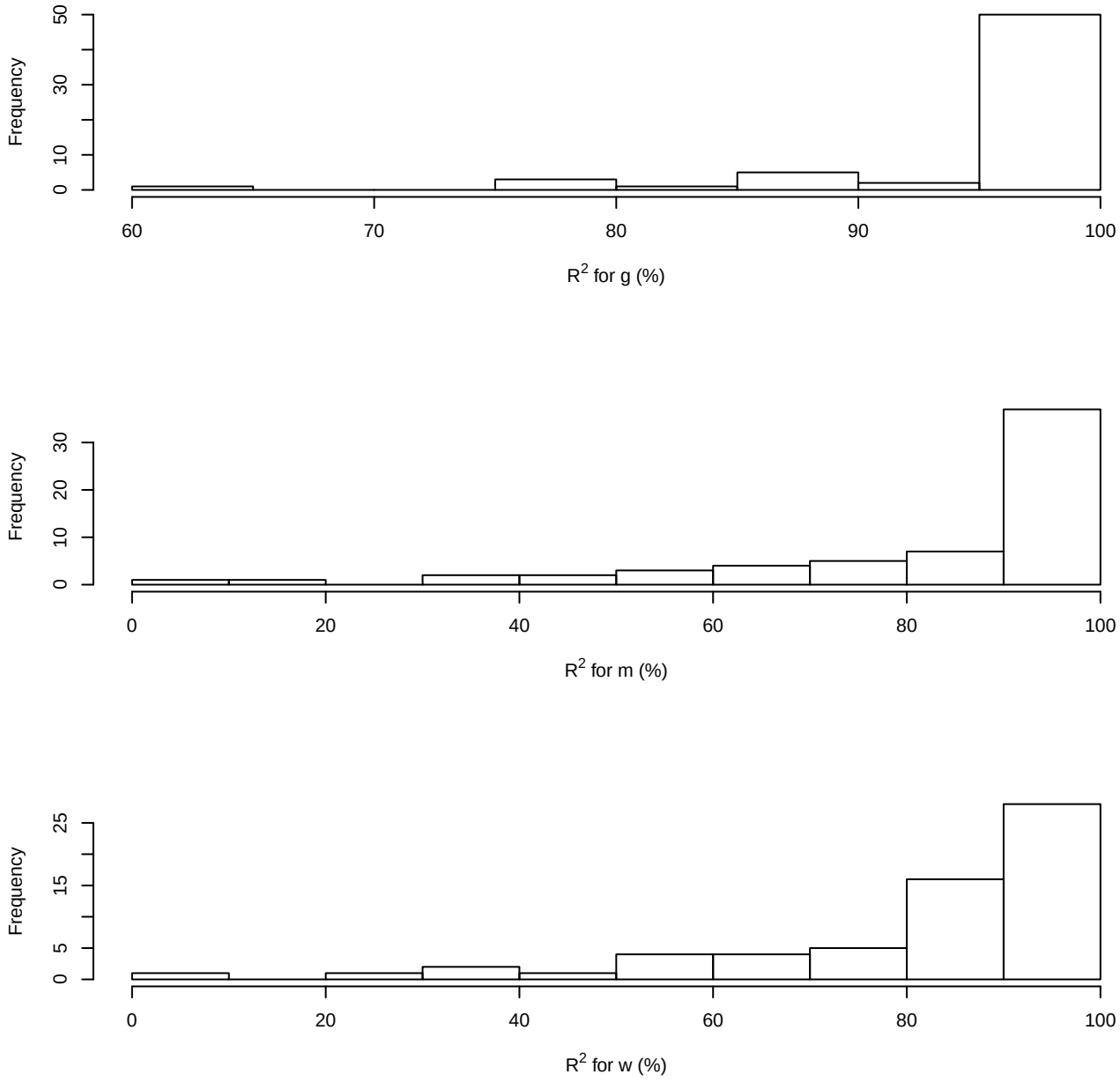

Figure 1: Histograms of R^2 values for fitting g, m, and w for the 62 datasets in percent. The majority of the cases have R^2 values in excess of 90%.

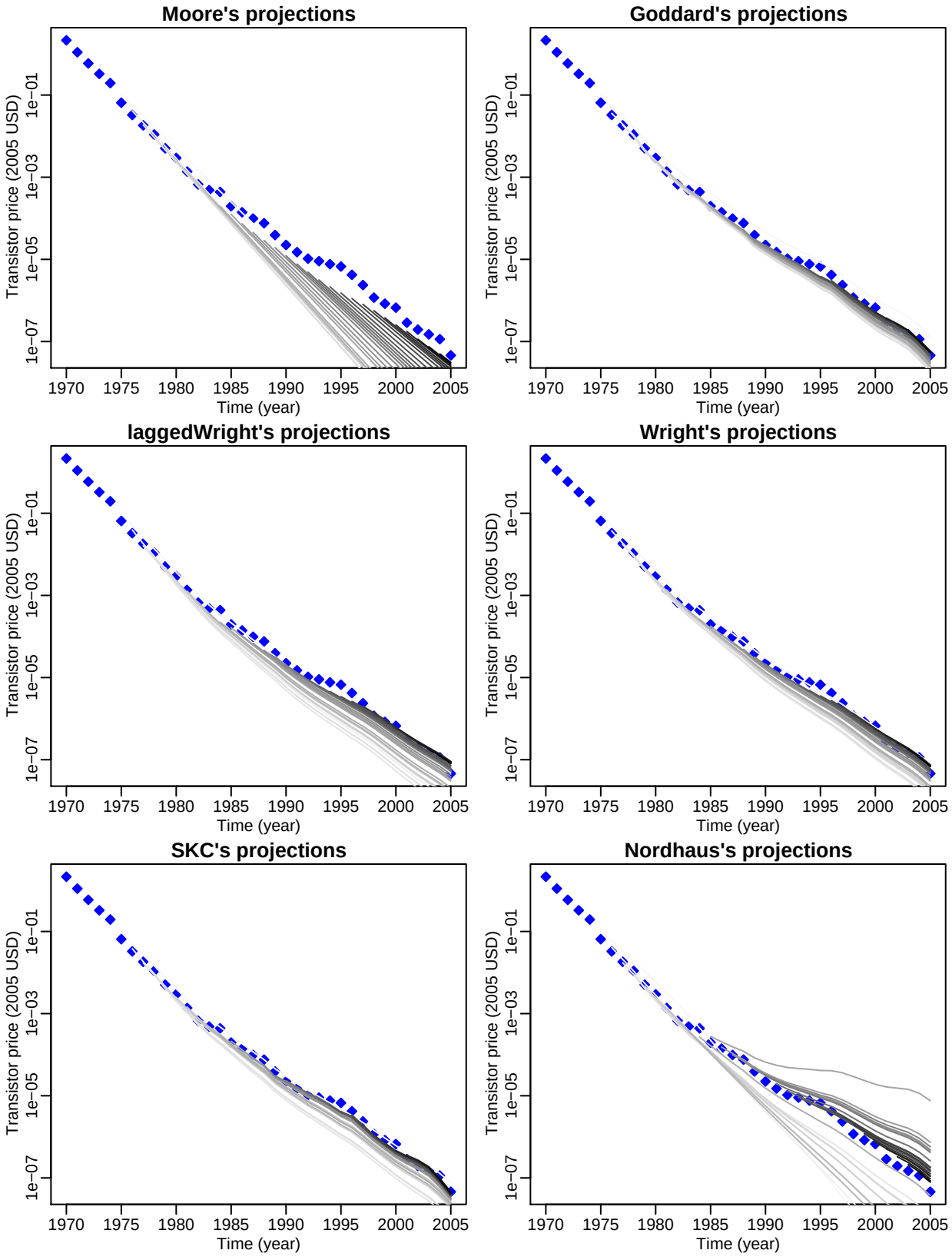

Figure 2: Predictions of six functional forms for the transistor dataset. The hindcasting origin varies from 1973 to 2004 (and thus the target year varies from 1974 to 2005). Forecasts are plotted in shades of gray; the real data is shown in blue diamonds, and ranges from 1969 to 2005. The forecast for 1974 is based on five years of data from 1969 - 1973.

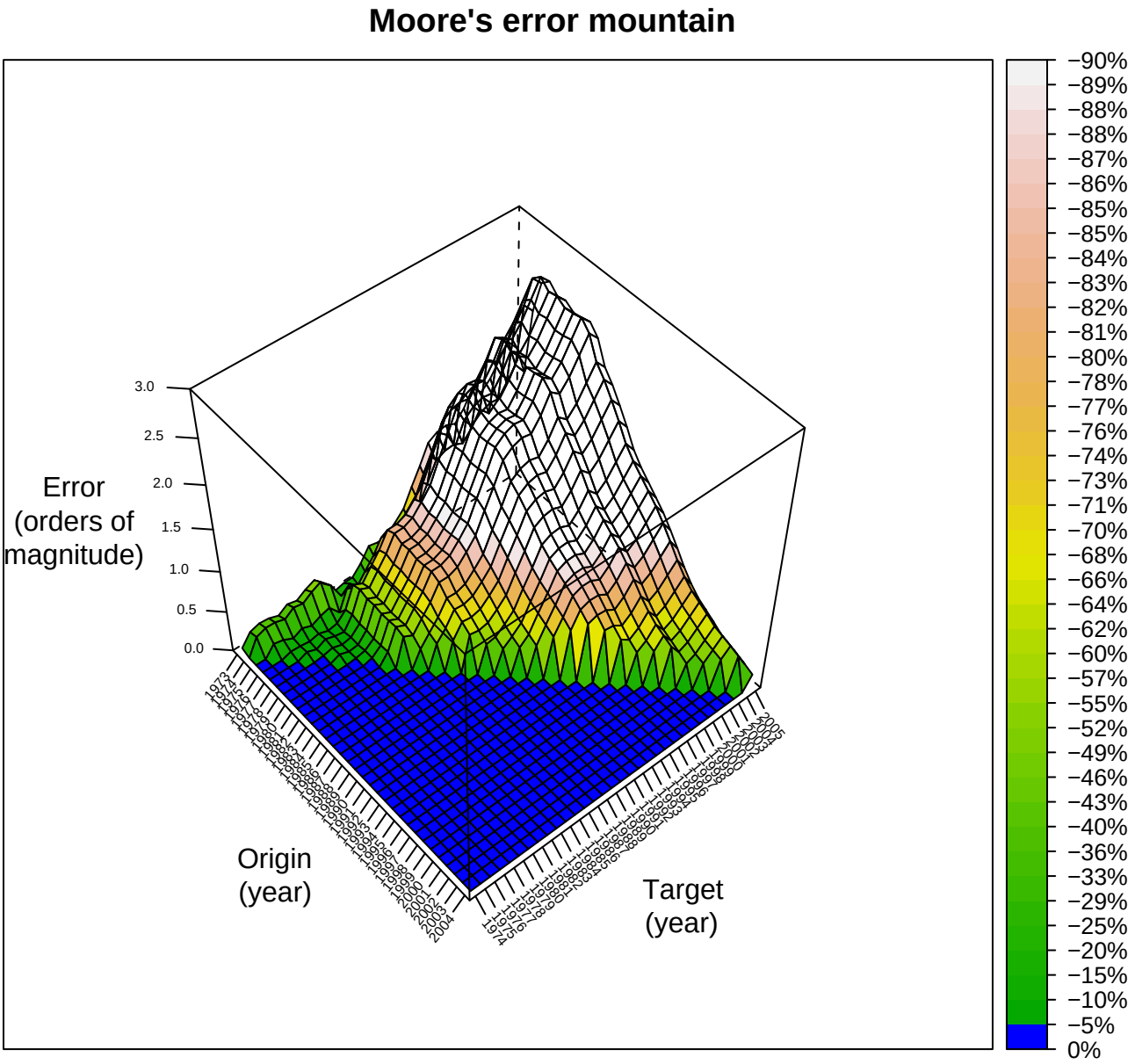

Figure 3: Moore's prediction errors for the Transistor dataset as a function of the hindcasting origin from 1973 to 2004 and the target year from 1974 to 2005.

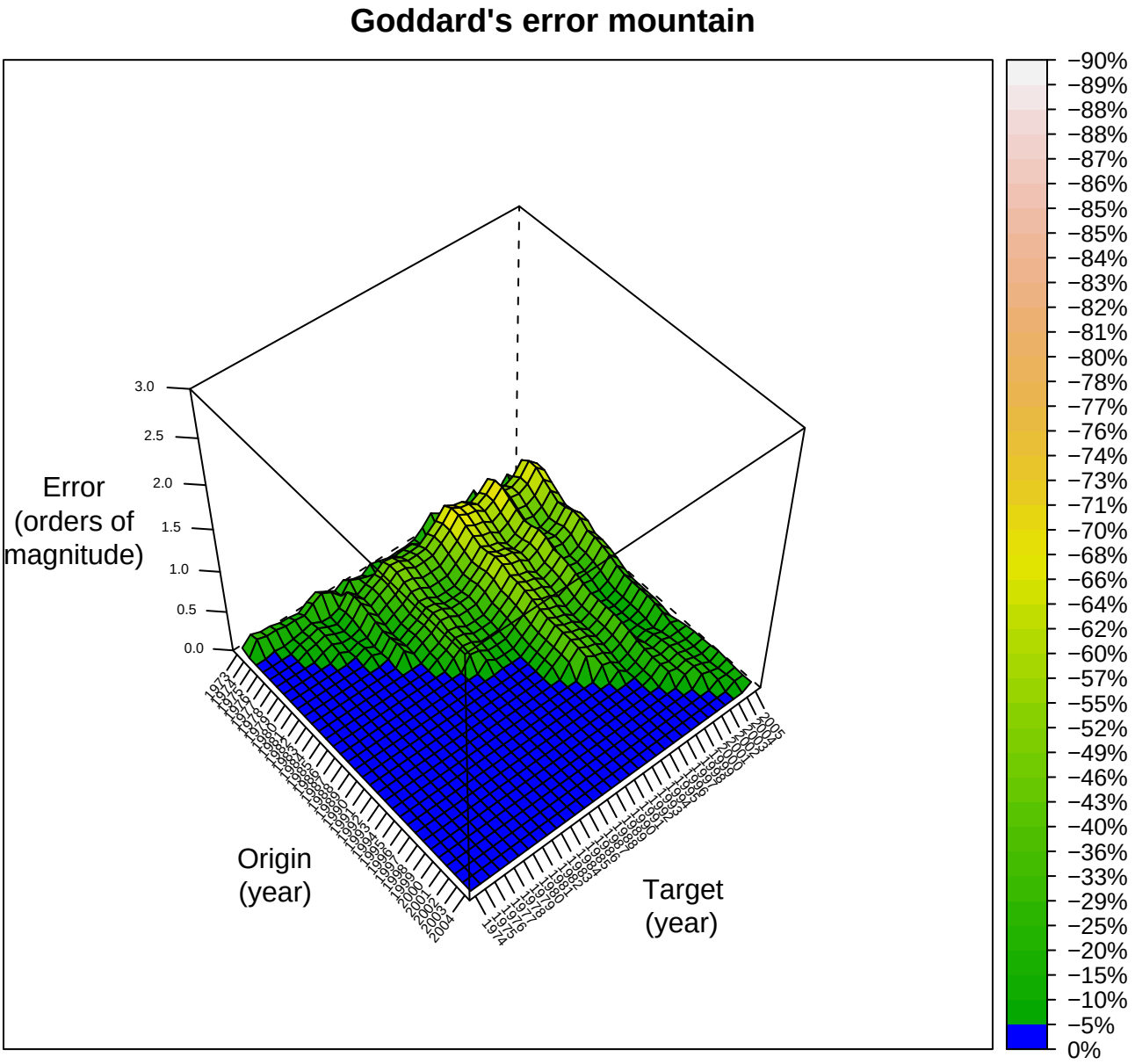

Figure 4: Goddard's prediction errors for the Transistor dataset as a function of the hindcasting origin from 1973 to 2004 and the target year from 1974 to 2005.

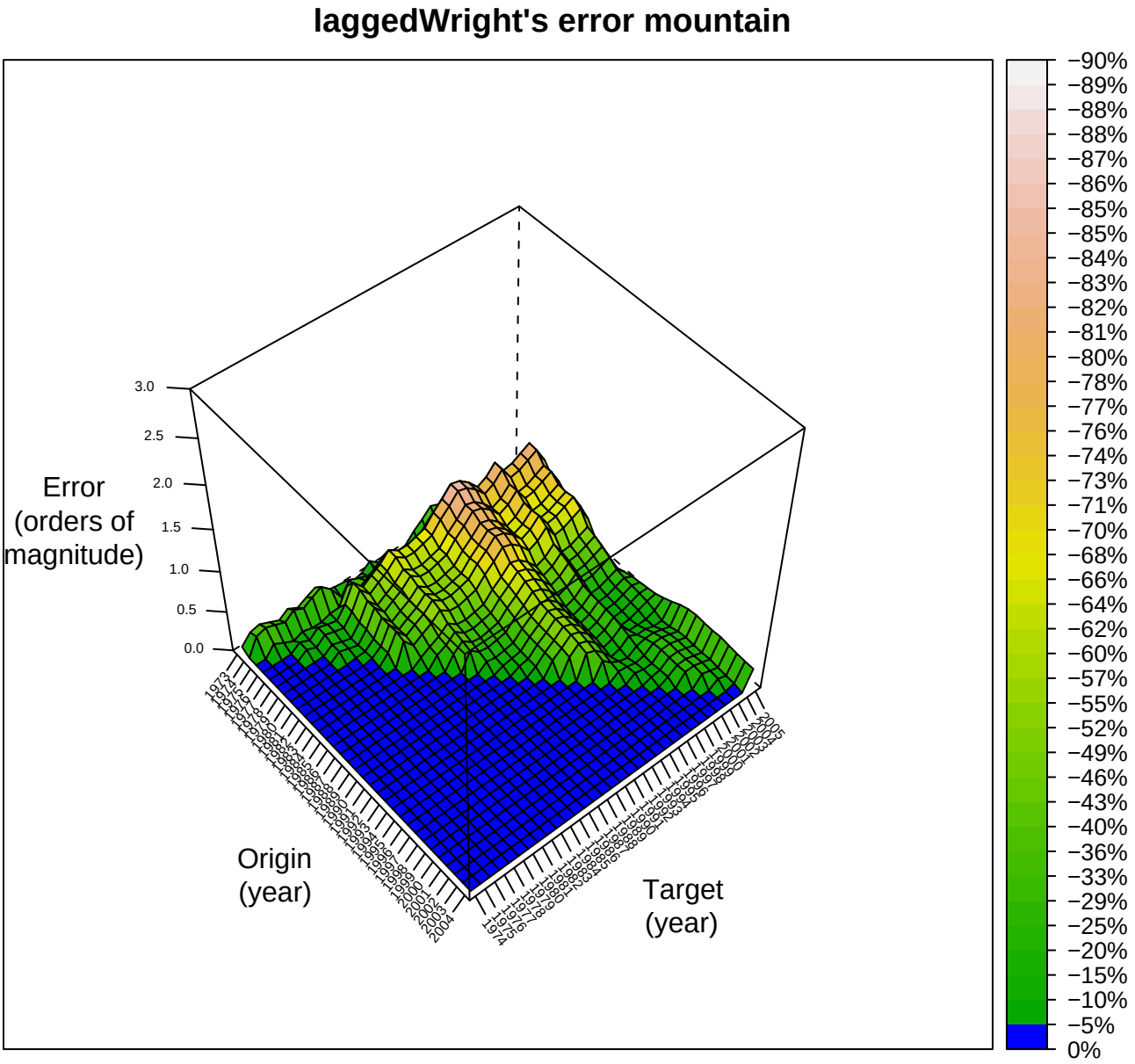

Figure 5: laggedWright's prediction errors for the Transistor dataset as a function of the hindcasting origin from 1973 to 2004 and the target year from 1974 to 2005.

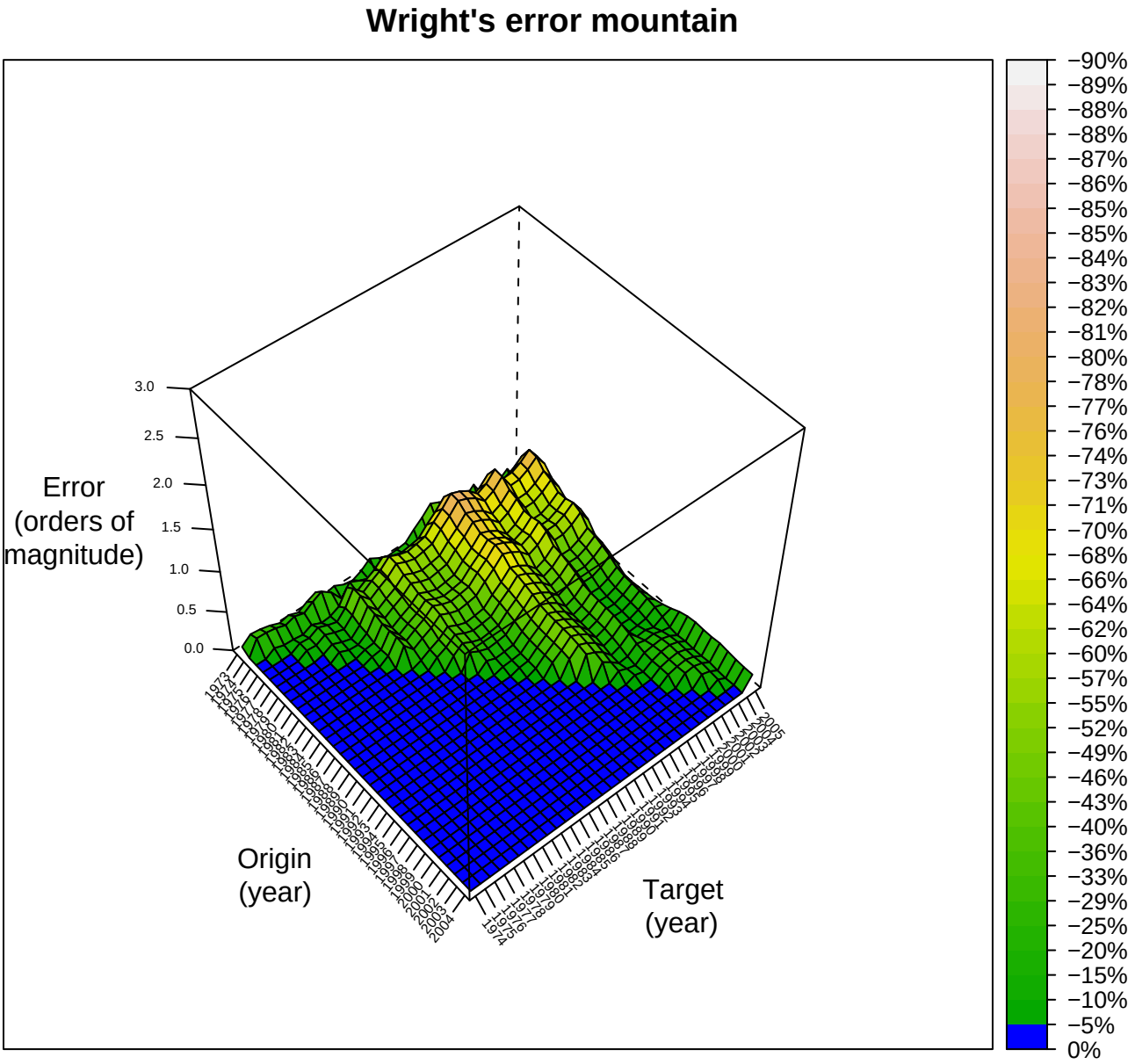

Figure 6: Wright's prediction errors for the Transistor dataset as a function of the hindcasting origin from 1973 to 2004 and the target year from 1974 to 2005.

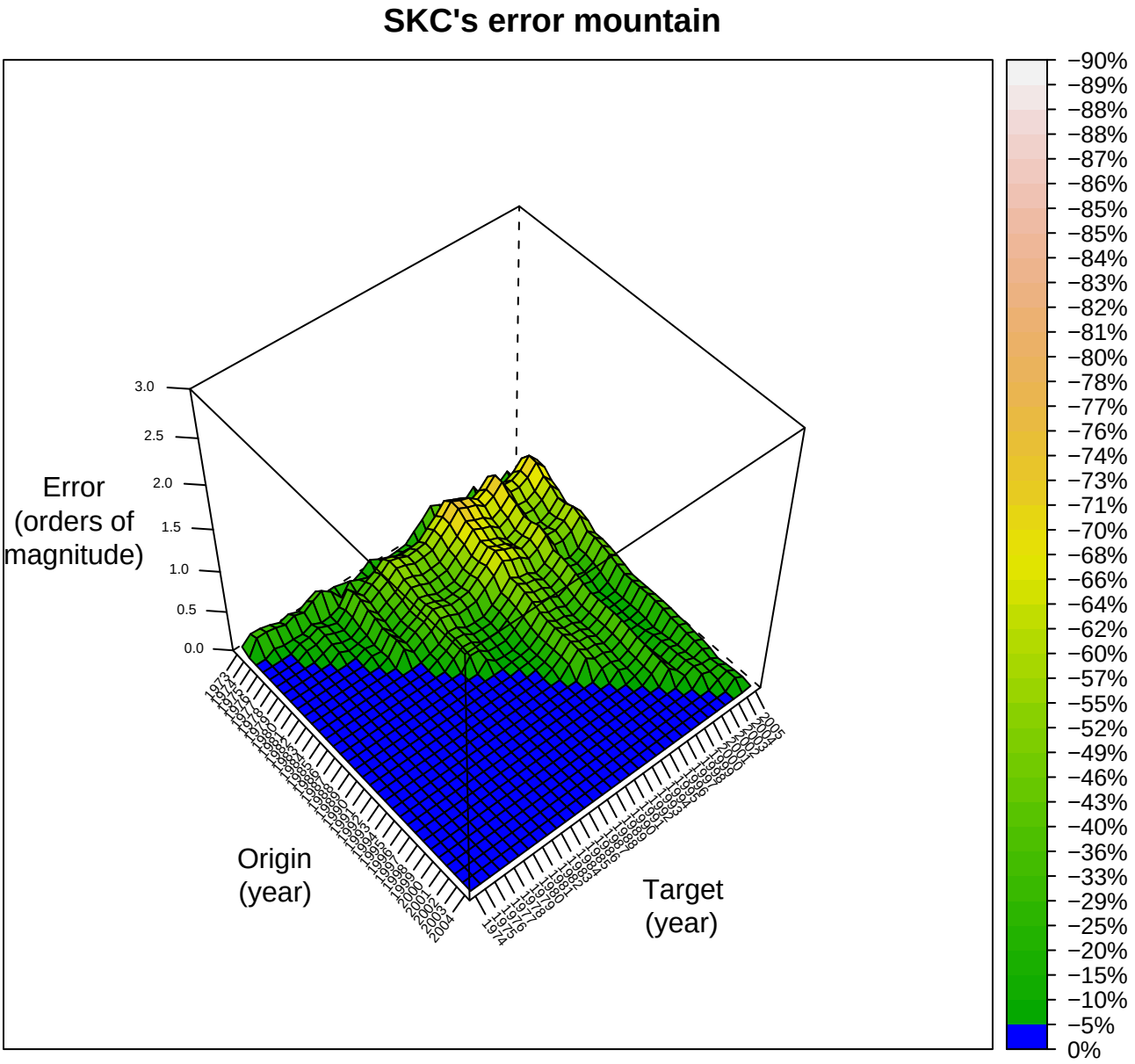

Figure 7: SKC's prediction errors for the Transistor dataset as a function of the hindcasting origin from 1973 to 2004 and the target year from 1974 to 2005.

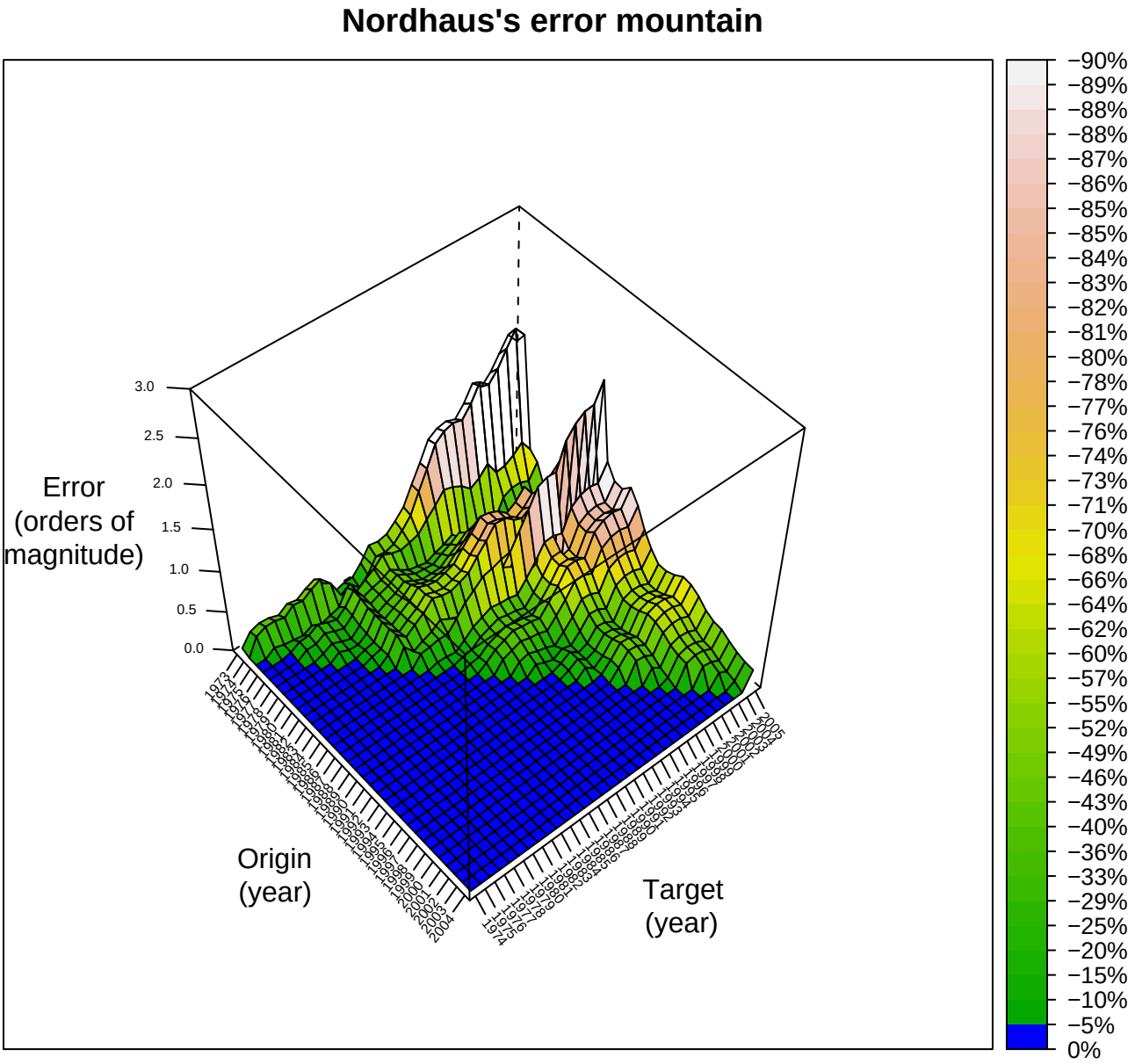

Figure 8: Nordhaus's prediction errors for the Transistor dataset as a function of the hindcasting origin from 1973 to 2004 and the target year from 1974 to 2005.

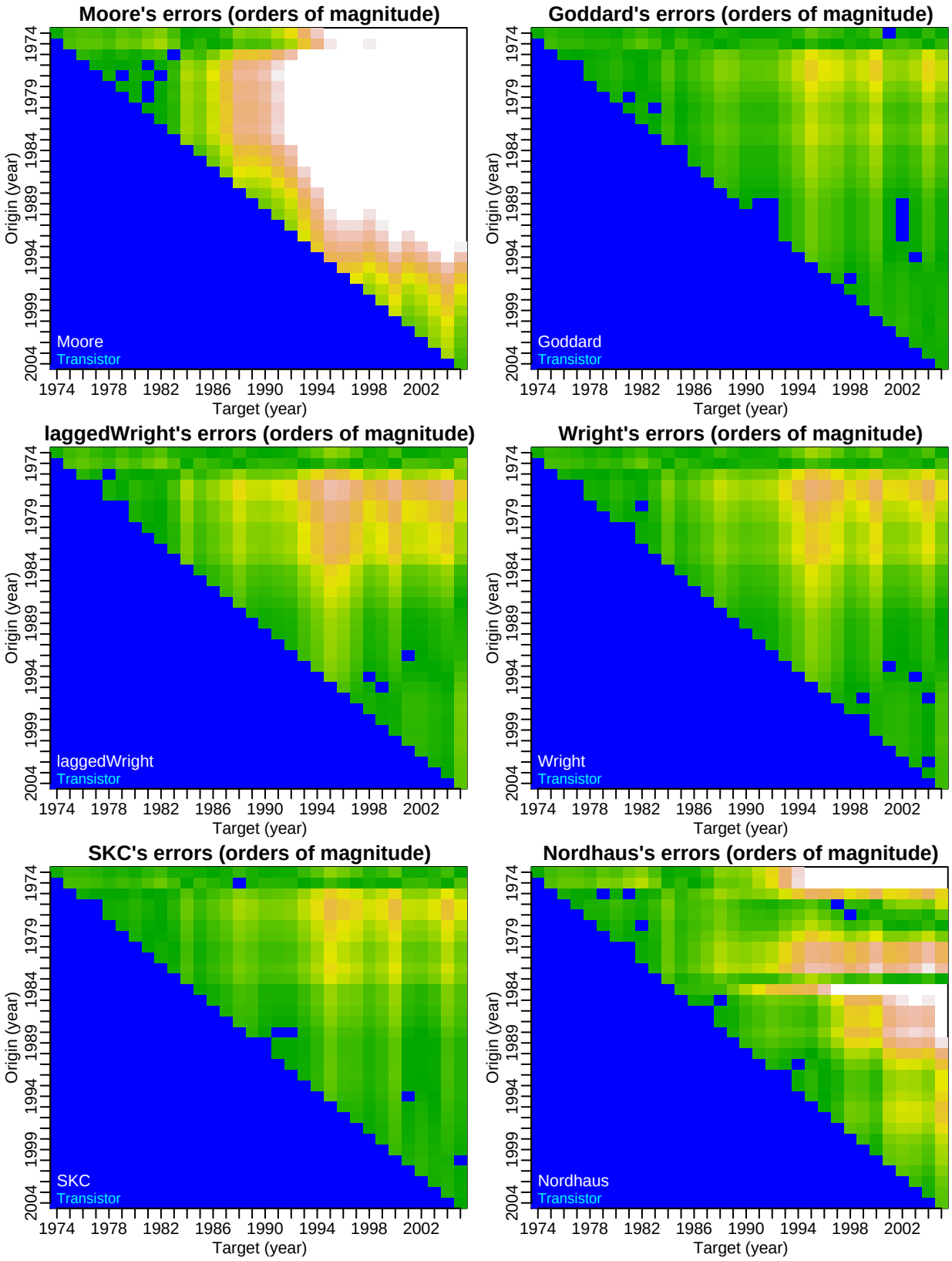

Figure 9: Summary of the prediction errors of six functional forms for the Transistor dataset as a function of the hindcasting origin from 1973 to 2004 and the target year from 1974 to 2005.

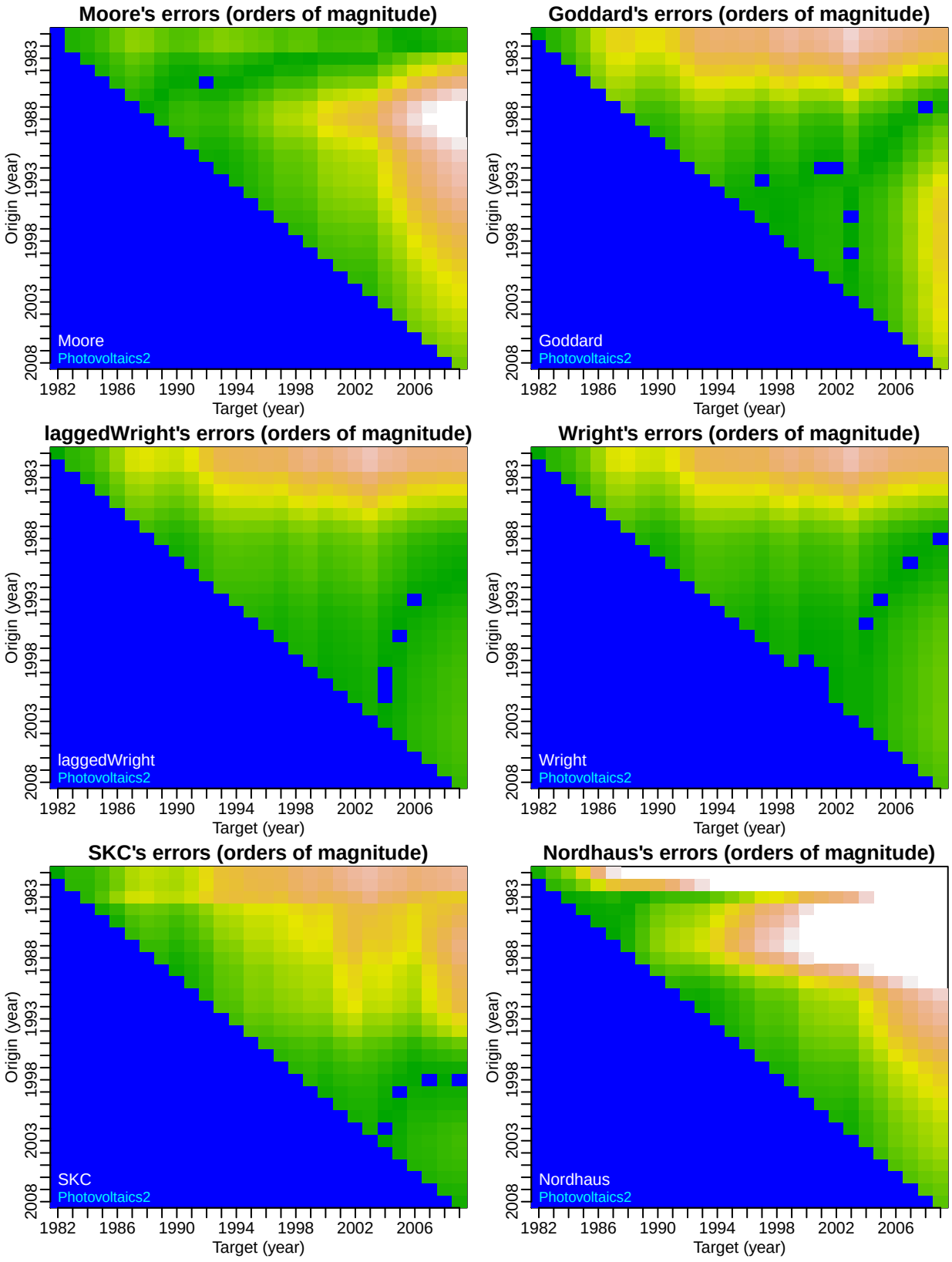

Figure 10: Summary of the prediction errors of six functional forms for the Photovoltaics2 dataset as a function of the hindcasting origin from 1981 to 2008 and the target year from 1982 to 2009.

negligible error between -6% and +6%), up to the snow line at 1.0 (above which the rest of the mountain is left uncolored). This is where the absolute error on the log scale reaches 1.0, meaning that the predicted price is either 10 times greater than the actual ($+900\%$ hindcasting error) or 10 times less (-90% error). To illustrate how the errors on the log scale translate to percentages, the color key on the right of these three-dimensional plots indicate the negative error range on the original scale from 0% to -90%. (However, we should keep in mind that because the logarithm of the absolute difference is taken, a 1.0 on the log scale can mean either $+900\%$ or -90%).

The functional form inspired by Moore does not predict transistors very well. This is ironic since transistors are the technology that Moore's Law was originally formulated to describe. In fact, among all the 62 performance curves, Moore's functional form was the least accurate for the Transistor dataset. We can see that a large portion of the error mountain in Figure 3 rises not only above the 1.0 mark at the snow line, but also exceeds the 2.0 mark, i.e. hindcasting that the price would be less than one-hundredth of what it actually turned out to be. The summit is at 2.67, meaning that in this case the actual price is underestimated by a factor of 467.

For a side-by-side comparison of all the hindcasts on the Transistor dataset, Supplementary Figure 9 is a bird's-eye view from the top of how the six competing forms fared against one another. Supplementary Figure 10 is a similar plot for the Photovoltaics2 dataset. Supplementary Figures 11 to 16 are top view supergraphics for the error mountains generated by the six functional forms for the other 60 datasets.

4 Error model

Visualizing the error mountains is a quick and intuitive way to screen out inadequate functional forms like Nordhaus that show erratic behavior. This was especially easy to do with multiple-variable forms (with or without interaction terms) that were prone to overfitting, which is manifested by the fact they gave good in-sample fits but generated large and inconsistent errors when trying to predict out-of-sample. Hence, Nordhaus and all the other multiple-variable forms that failed to generate a relatively consistent prediction error surface (without huge jumps) were not included in the subsequent formal analysis. The statistical comparison we made here in constructing our error model is only between the "finalists".

After ruling out all multiple-variable forms but SKC's by visual inspections of the error mountains, we are left with five candidates competing for the hindcasting champion title: Moore, Goddard, laggedWright, Wright, and SKC. For example, does Moore's comparably weaker performance on the Transistor dataset make it an inferior functional form? How do the others compare? Is any one of them significantly better than any of the others?

There are no obvious ways to answer these questions, but one way is to build a suitable statistical model for the errors generated by the remaining five functional forms, based on the data displayed for those functional forms in Supplementary Figures 9 to 16. The statistical model employed here is an extended linear mixed-effects model, fitted by maximum likelihood, using the *lme* function in the *nlme* package in R. The mixed-effects designation here refers to the presence of both fixed and random effects (as explained later in this section). The basic linear mixed-effects model needed extension because the hindcasting data was both heteroscedastic (with unequal variances) and correlated (not independent).

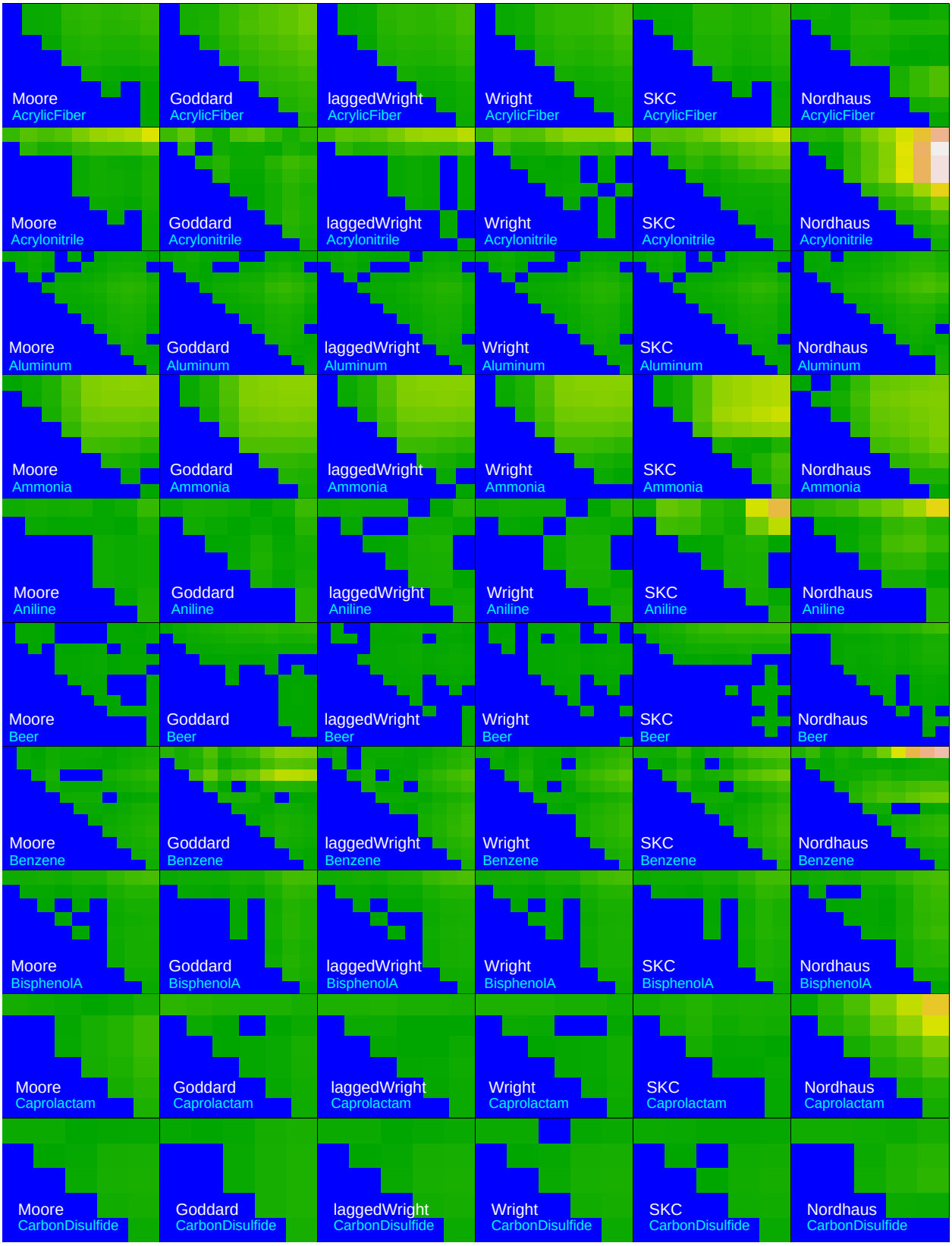

Figure 11: Summary of the prediction errors generated by six functional forms as a function of the hindcasting origin (vertical axis) and the target year (horizontal axis).

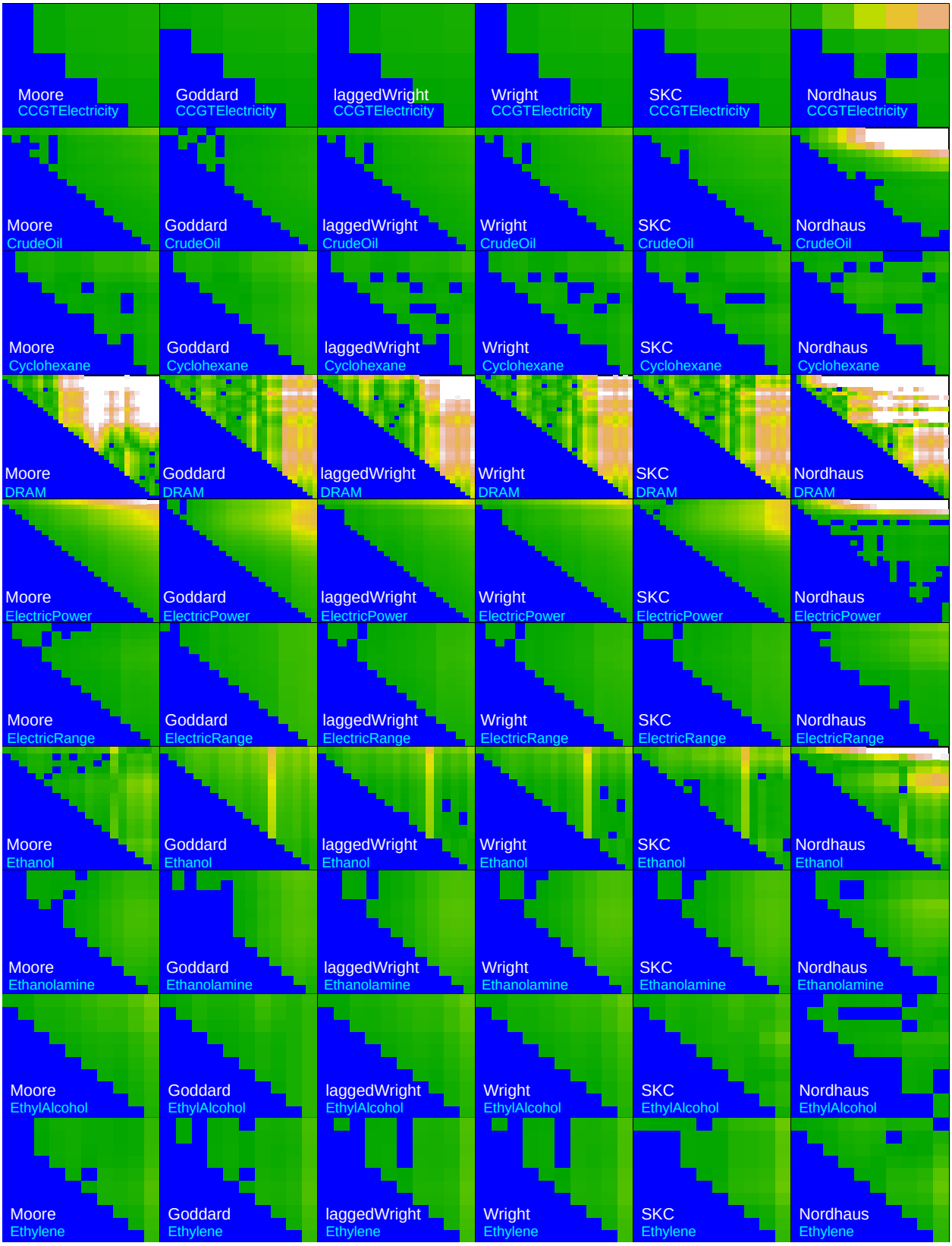

Figure 12: Summary of the prediction errors generated by six functional forms as a function of the hindcasting origin (vertical axis) and the target year (horizontal axis).

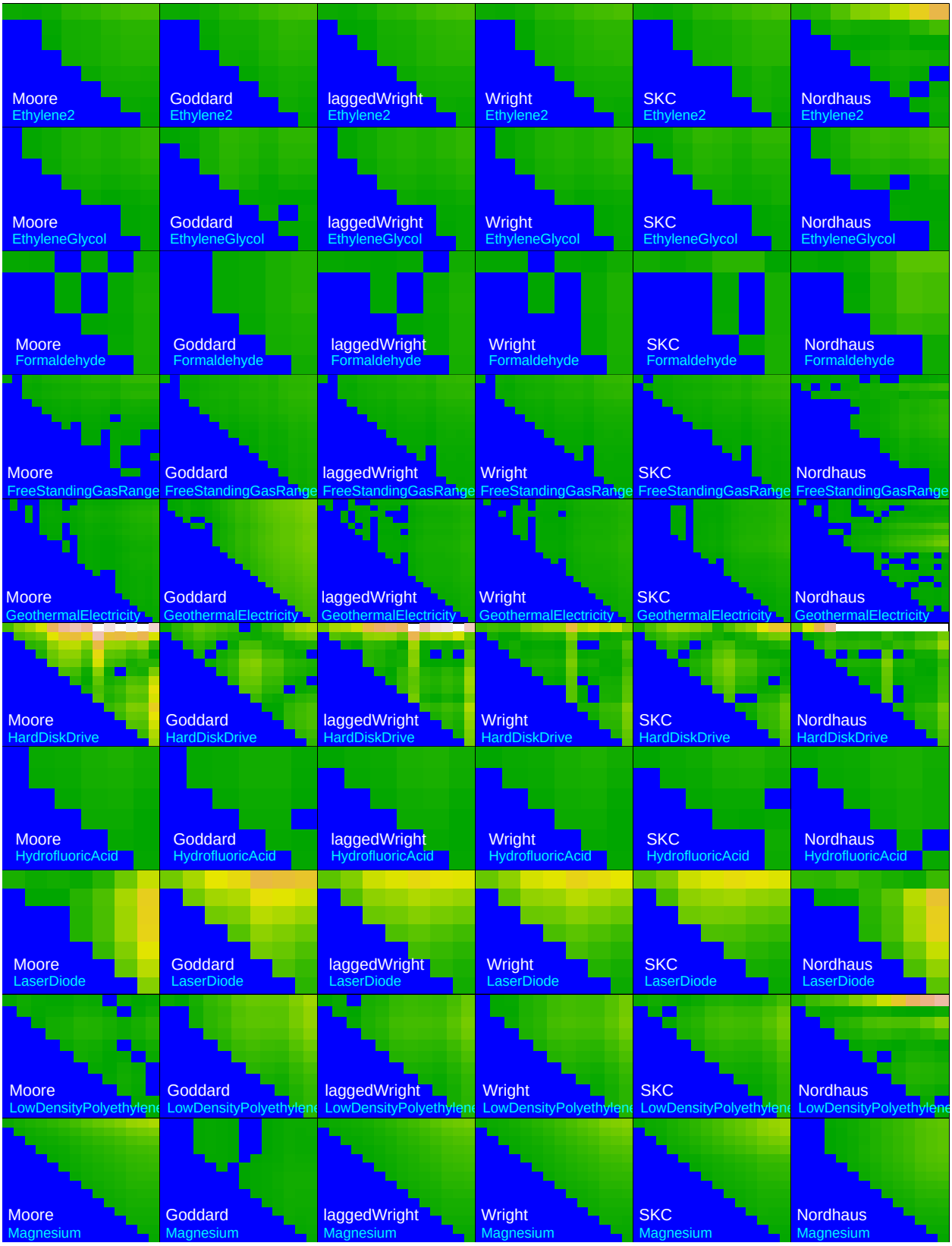

Figure 13: Summary of the prediction errors generated by six functional forms as a function of the hindcasting origin (vertical axis) and the target year (horizontal axis).

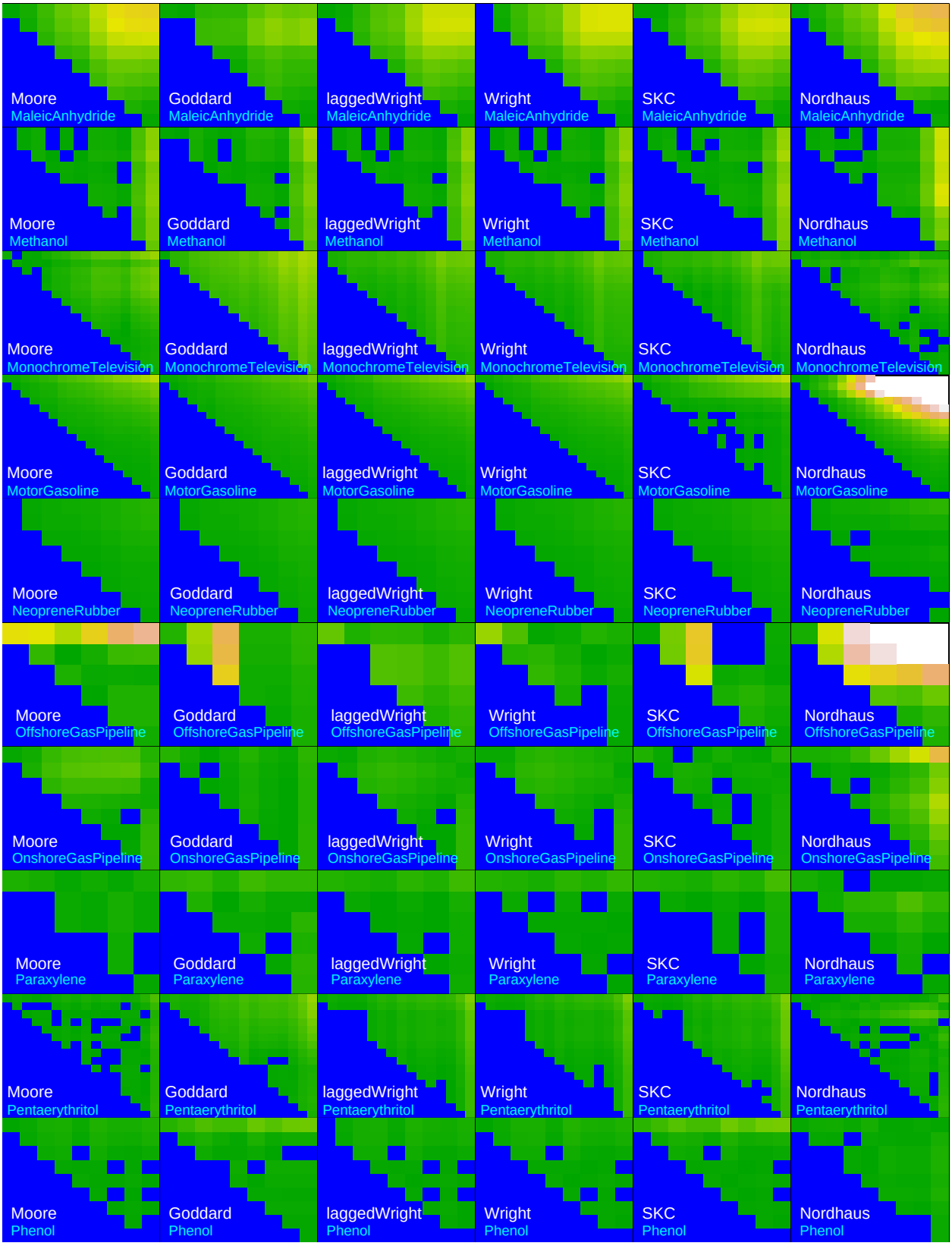

Figure 14: Summary of the prediction errors generated by six functional forms as a function of the hindcasting origin (vertical axis) and the target year (horizontal axis).

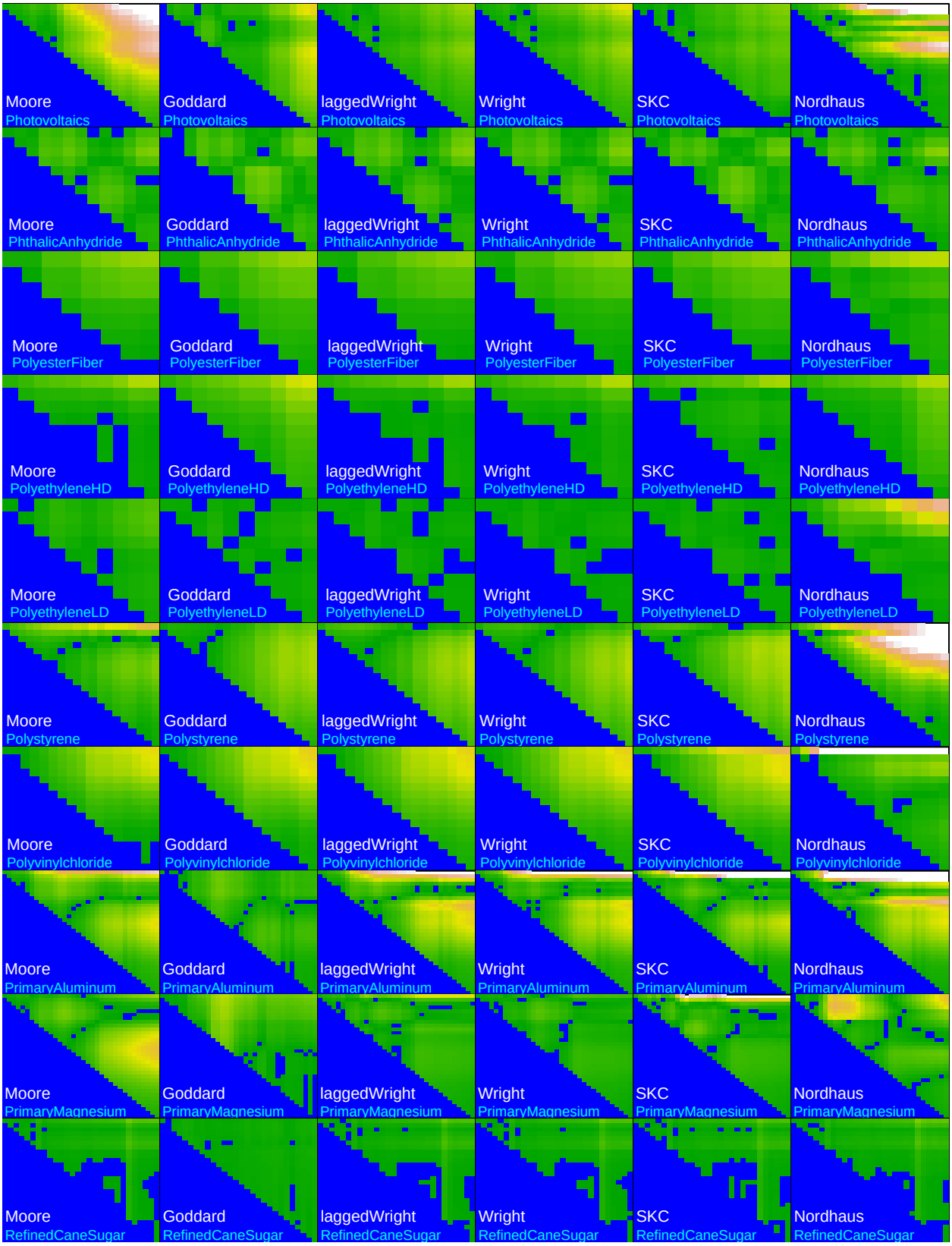

Figure 15: Summary of the prediction errors generated by six functional forms as a function of the hindcasting origin (vertical axis) and the target year (horizontal axis).

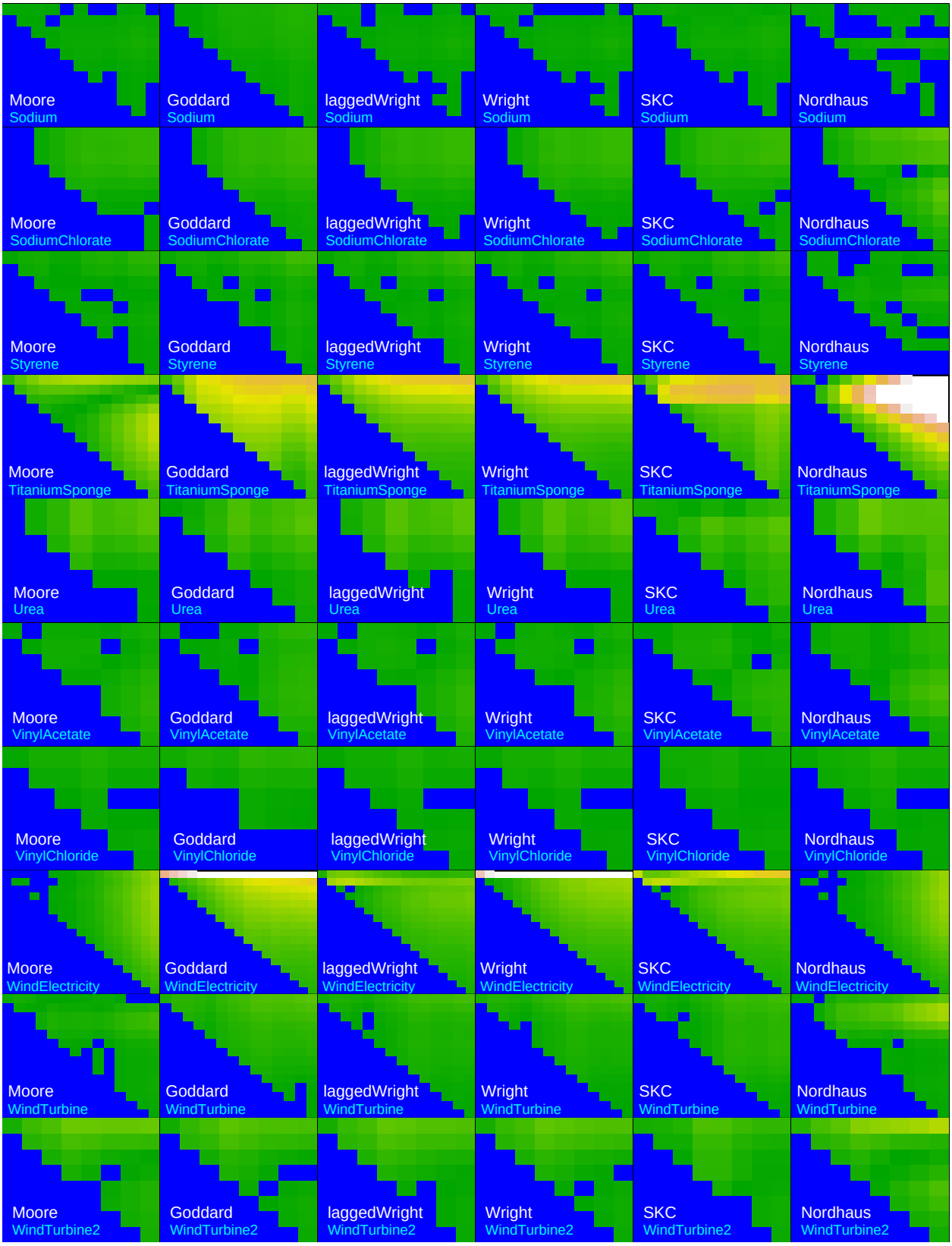

Figure 16: Summary of the prediction errors generated by six functional forms as a function of the hindcasting origin (vertical axis) and the target year (horizontal axis).

4.1 Choosing the response

When trying to build a model that captures the essence of a large dataset, one is faced with many theoretical and practical challenges, involving several subjective decisions, such as the various tradeoffs between simplicity and goodness of fit. One of the most important early choices is what to model in the first place. We started out by trying to model the prediction errors directly (i.e. the height of the error mountains), but that did not lead to acceptable diagnostics of the resulting fits.

Next we experimented with transforming these prediction errors in order to obtain a response that will allow a better fit. We searched the family of power transformations, which is known for its flexibility to accommodate a wide range of variance structures for the purposes of linear modeling. Exponents in the neighborhood of 0.5 provided the best fits in terms of regression diagnostics, so we fixed it at 0.5. This meant applying a square root transformation to the original errors.

Formally, the response r_{fdij} used in this model is the root-transformed absolute hindcasting error on the log scale (base 10) of the hindcast made in year i for year j using the functional form f for dataset d. In other words, this is the square root of the height of the error mountain of f for d at the time coordinates i and j:

$$r_{fdij} \;=\; \left|\log y_j^{(d)} - \log \hat{y}_j^{(f,d,i)}\right|^{0.5}, \tag{1}$$

where $y_j^{(d)}$ is the actual price in the dataset d in year j and $\hat{y}_j^{(f,d,i)}$ is the price estimated by the functional form f for year j, using all data in d available up to and including year i, where $i < j$.

4.2 Modeling the response

The main advantage of choosing the response given by equation (1) is that it can be modeled in a parsimonious manner as a linear function of the hindcasting horizon $=$ target $-$ origin $= j - i$. The effect of each functional form f can be characterized by two numbers: an intercept α_f and a slope parameter β_f that specifies this linear relationship. But the individual curves themselves can have large effects on the response and we need to take that into account, too. So, instead of modeling only the average effect of the functional form f by a linear function of the hindcasting horizon $(j-i)$ with the linear relationship $\alpha_f + \beta_f(j-i)$, we model the joint effect of the functional form f and the performance curve data d with the adjusted linear trend $(\alpha_f + a_d) + (\beta_f + b_d)(j-i)$, where the a_d and b_d quantities are additive adjustments to the average intercept and slope parameters α_f and β_f, respectively, to take into account the peculiarities of the dataset d.

In order to avoid adding 62 a_d parameters plus 62 b_d parameters, we treated the $\begin{pmatrix} a_d \\ b_d \end{pmatrix}$ pair as a two-dimensional random vector having a bivariate normal distribution with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and variance-covariance matrix $\begin{pmatrix} \psi_a^2 & \psi_{ab} \\ \psi_{ab} & \psi_b^2 \end{pmatrix}$. This way we can parameterize these adjustments as random deviations from the average $\begin{pmatrix} \alpha_f \\ \beta_f \end{pmatrix}$ at a cost of only 3 additional parameters instead of $2 \times 62 = 124$, resulting in a parameterization that is not only much more parsimonious but also makes maximum likelihood estimation possible by keeping the number of parameters in check.

In statistical terminology, we can say that the effects of the 62 performance curves are random (as opposed to the fixed effects of the five functional forms). The interpretation is that we view

each curve as a random draw from a hypothetical population of performance curves. Of course, the underlying assumption here is that it makes sense to talk about such an ensemble of performance curves that have enough characteristics in common to make this notion meaningful.

4.3 Statistical model

Now we are ready to define the extended linear mixed-effects model by the following equation:

$$r_{fdij} \;=\; (\alpha_f + a_d) + (\beta_f + b_d)(j - i) + \varepsilon_{fdij}, \tag{2}$$

where the response is modeled as the sum of a trend term (that is a linear function of the hindcasting horizon $(j - i)$ as before) plus an ε_{fdij} random field term to take into account the deviations from the trend. This is assumed to be a Gaussian stochastic process independent of the $\begin{pmatrix} a_d \\ b_d \end{pmatrix}$ random vector, having mean 0, and given a_d and b_d, having variance equal to a positive σ^2 times the fitted values:

$$\text{Var}\,(\,\varepsilon_{fdij}|\,a_d, b_d) \;=\; \sigma^2\,\text{E}\,(\,r_{fdij}|\,a_d, b_d) \tag{3}$$

and an exponential correlation structure within each mountain that is a function of the differences of the two time coordinates with a positive range parameter ρ and another small positive nugget parameter η quantifying the extent of these correlations:

$$\text{Corr}(\varepsilon_{fdij},\; \varepsilon_{f'd'i'j'}) \;=\; \delta_{ff'}\delta_{dd'}(1 - \eta)\exp\{-(|i - i'| + |j - j'|)/\rho\}, \tag{4}$$

where the two Kronecker δ functions ensure that each mountain surface is treated as a separate entity.

Equations (3) and (4) were chosen to deal with the fact that variances tend to increase with altitude on the error mountains and that there are serial correlations along the time coordinates i (hindcasting origin) and j (hindcasting target). The heteroscedasticity (increasing variance with increasing elevation) problem is handled by the variance function (3) and the time dependence is taken care of by the correlation function (4). Based on the likelihood, this exponential correlation function provided the best fit. Note that instead of the usual Euclidean distance (root sum of squares of differences), here the so-called "Manhattan" measure was used (the sum of the absolute differences), because it provided a much better fit in terms of the likelihood.

4.4 Intercept and slope parameter estimates

The maximum likelihood estimates for the five intercept and five slope parameters are listed in Tables 2 and 4, respectively. It is evident that all five functional forms perform similarly in terms of hindcasting accuracy, because most of these estimates are not significantly different from one another. The corresponding approximate p-values for all pairwise comparisons are listed in Tables 3 and 5. The highest intercept estimate for Goddard means that it does a relatively poor job of forecasting at short times, whereas the higher slope estimates for Moore and SKC mean that they are not as good at long times. Otherwise the models are roughly equivalent.

5 Extrapolation method

For the purposes of this paper we have chosen to fit the model to all the past data available at time i and use the resulting parameter estimates to make the forecasts. Thus the forecast corresponds to

Intercept	*estimate*
α_{SKC}	0.164
α_{Moore}	0.168
α_{Wright}	0.170
$\alpha_{\text{laggedWright}}$	0.172
α_{Goddard}	0.195

Table 2: Intercept estimates.

Intercept difference	*p-value*
$\alpha_{\text{Moore}} - \alpha_{\text{SKC}}$	0.555
$\alpha_{\text{laggedWright}} - \alpha_{\text{SKC}}$	0.224
$\alpha_{\text{laggedWright}} - \alpha_{\text{Moore}}$	0.531
$\alpha_{\text{Wright}} - \alpha_{\text{SKC}}$	0.389
$\alpha_{\text{Wright}} - \alpha_{\text{Moore}}$	0.786
$\alpha_{\text{Wright}} - \alpha_{\text{laggedWright}}$	0.722
$\alpha_{\text{Goddard}} - \alpha_{\text{SKC}}$	0.000
$\alpha_{\text{Goddard}} - \alpha_{\text{Moore}}$	0.001
$\alpha_{\text{Goddard}} - \alpha_{\text{laggedWright}}$	0.005
$\alpha_{\text{Goddard}} - \alpha_{\text{Wright}}$	0.002

Table 3: Testing whether the pairwise intercept differences are significantly different from zero.

the trend line for the entire data set, and points in the distant past get just as much weight as points in the present. We have chosen this method because it is the "vanilla" model, and probably the most widely used to apply these laws. This method is also well-suited to compare standard hypotheses, which is the main objective of the paper. We do not argue that this is the best possible forecasting method; developing more accurate forecasting methods will be a topic of future research. In the meantime we want to provide more background information on some of the anomalies commented on in the text, and in particular figure 5 and footnote 7.

In figure 5 there is an immediate drop in the one-year forecast relative to the last observed price. This is a direct consequence of the use of the vanilla method, which is the ideal model if the data are generated by independent fluctuations around a deterministic trend. To illustrate this with Moore's law, suppose the true random process generating the data is of the form:

$$\log y_t = at + n(t) \tag{5}$$

This is the most straightforward interpretation of Moore's law. If the noise terms $n(t)$ are uncorrelated in time, then the method we have used to make forecasts here is ideal. But if the noise terms are correlated in time this is no longer the case. Suppose, for example, that the process is better described by a random walk with drift, of the form

$$\log y_{t+1} = \log y_t - \mu + n(t), \tag{6}$$

where μ is a drift term, and where the noise fluctuations $n(t)$ are uncorrelated in time. In this case the best forecast for $\log y_{t-1}$ is $\log y_t - \mu$. There are many intermediate possibilities, for example if $\log y_t$ is a long-memory process.

Slope	*estimate*
β_{Goddard}	0.02396
β_{Wright}	0.02397
$\beta_{\text{laggedWright}}$	0.02438
β_{Moore}	0.02706
β_{SKC}	0.02848

Table 4: Slope estimates.

Slope difference	*p-value*
$\beta_{\text{Moore}} - \beta_{\text{SKC}}$	0.394
$\beta_{\text{laggedWright}} - \beta_{\text{SKC}}$	0.011
$\beta_{\text{laggedWright}} - \beta_{\text{Moore}}$	0.093
$\beta_{\text{Wright}} - \beta_{\text{SKC}}$	0.005
$\beta_{\text{Wright}} - \beta_{\text{Moore}}$	0.050
$\beta_{\text{Wright}} - \beta_{\text{laggedWright}}$	0.788
$\beta_{\text{Goddard}} - \beta_{\text{SKC}}$	0.007
$\beta_{\text{Goddard}} - \beta_{\text{Moore}}$	0.059
$\beta_{\text{Goddard}} - \beta_{\text{laggedWright}}$	0.788
$\beta_{\text{Goddard}} - \beta_{\text{Wright}}$	0.991

Table 5: Testing whether the pairwise slope differences are significantly different from zero.

We find clear evidence for memory in the noise terms. Taking into account correlations in the noise terms produces better forecasts for short time horizons. For longer time horizons, greater than 3 - 5 years, they are roughly equivalent.

The use of the vanilla model leads to some peculiar results. For example, we find that short term forecasts get worse as we add more historical data — in other words, recent data is more useful than data in the far past. This is not surprising if the true dynamics are closer to Eq. 6 than to Eq. 5 — more data systematically means that the most recent point will show larger deviations from the trend line. However, continually adjusting trend lines to take into account the most recent data compromised the goodness of fit of our error model by generating excessive noise (discontinuities in the error mountains). In other words, improving the short-term forecasts would have compromised our ability to compare standard hypotheses (functional forms commonly used to forecast technological improvement).

In Figure 5, we are using one of the longest time series in the data set, and the forecast is based on the entire series. Thus the errors for a time horizon of one are large compared with the typical series in the data set. This is not a problem at longer time horizons. The error estimates after the first five years become more trustworthy.

In a future paper we intend to work on constructing a “best model". In this paper, our goal to place the problem of forecasting technological change using past performance in a solid statistical context, and to quantify the quality of the forecasts.