Australian Government
Department of Defence
Science and Technology

Predicting Trends in Peer-Reviewed Publications

Dmitri Kamenetsky

Joint and Operations Analysis Division
Defence Science and Technology Group

DST-Group-TN-1881

ABSTRACT

This work investigated a method for predicting future trends in peer-reviewed publication data related to science and technology. The developed method use non-linear regression to fit a particular type of s-shaped curve. The method's prediction accuracy is measured on historical publication data of mature technologies. It was found that the prediction accuracy is acceptable for short-term (5-10 years) predictions, but declines for long-term (10+ years) predictions. The method was also used to predict the future publication trends of novel technologies. It was found that the publication rate of various upcoming technologies is expected to grow. In particular, CRISPR and deep learning are expected to grow rapidly. Some technologies however, are already at the peak of their trend and are expected to decline in the future.

RELEASE LIMITATION

Approved for public release

Published by

Joint and Operations Analysis Division
Defence Science and Technology Group
Fairbairn Business Park Department of Defence
Canberra ACT 2600 Australia

May 2019

Conditions of Release and Disposal

Predicting Trends in Peer-Reviewed Publications

Executive Summary

The Concepts and Futures (C&F) group, within Joint and Operations Analysis Division (JOAD) is a collaborative research facility for the study of emerging and disruptive technologies. The team aims to identify areas of threat and opportunity in developing technologies and provides foresight to policy, strategy and capability development for the Australian Defence Organisation (ADO) and its strategic partners.

In support of the technology futures research, this work investigated a method for predicting future trends in peer-reviewed publication data related to science and technology. The developed method uses non-linear regression to fit a particular type of s-shaped curve, called the Gompertz function. The method's prediction accuracy was measured on historical publication data of mature technologies. It was found that the prediction accuracy is acceptable for short-term (5-10 years) predictions, but declines for long-term (10+ years) predictions. The method was also used to predict the future publication trends of novel technologies. It was found that the publication rate of various upcoming technologies is expected to grow. In particular, clustered regularly interspaced short palindromic repeats (CRISPR) and deep learning are expected to grow rapidly. However some technologies, such as metal foam and organic light-emitting diode (OLED), have already reached the peak of their trend and are expected to decline in the future.

Author

Dmitri Kamenetsky
Joint and Operations Analysis Division

In 2005 Dmitri obtained a Bachelor of Science with first class Honours in Computer Science from the University of Tasmania. He then completed a PhD in statistical machine learning at the Australian National University and NICTA in 2009. His PhD thesis investigated methods for inference and parameter estimation in graphical models, in particular the Ising model. In 2010, he joined the NSI Division of the Defence Science and Technology Group. Since then he has worked on object detection, multi-camera pedestrian tracking, atmospheric turbulence mitigation and score-level fusion. Dmitri was seconded to the Technology Concepts and Futures Group in 2016-17.

Contents

This page is intentionally blank.

Glossary

ADO	Australian Defence Organisation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DST Group	Defence Science and Technology Group
JOAD	Joint and Operations Analysis Division
OLED	Organic Light-Emitting Diode
TIPS	Theory of Inventive Problem Solving

This page is intentionally blank.

1. Introduction

Concepts and Futures group, within Joint and Operations Analysis Division (JOAD) is a collaborative research facility for the study of emerging and disruptive technologies. The team aims to identify areas of threat and opportunity in developing technologies and provides foresight to policy, strategy and capability development for the Australian Defence Organisation (ADO) and its strategic partners.

Technology forecasting can be approached via a number of forecasting methods [1, 2], such as: expert opinion obtained from surveys and workshops, bibliometric analysis, statistical methods such as correlation and risk analysis, scenario simulations, back-casting, game theory, trend analysis, Delphi or the theory of inventive problem solving (TIPS). These methods can be classified as exploratory or normative [3]. Exploratory methods predict future events based on what has happened in the past. Normative methods begin with a possible future state and work backwards to determine the steps necessary to reach that state. Forecasting methods can also be classified as quantitative or qualitative [3]. Quantitative methods rely on mathematical models to examine the rates of change, but are limited in their understanding of social and political variables. Qualitative methods rely on opinion of experts to make their predictions. They are capable of describing complex processes, but can be highly subjective.

In this work we performed technology forecasting using trend analysis, which is an exploratory quantitative method. In particular, we performed trend prediction on historical publications of certain technologies obtained from the Scopus database [4].

2. Trend Prediction

2.1 Investigation

Initial investigations focused on historical publication trends for a set of mature technologies from the Scopus database. A technology was deemed mature if it had been developed and widely used by the general public for at least 20 years. The technologies analysed included the following: CDROM, Ethernet, GPS, HTML, HTTP, Internet, Kevlar®, laser, LCD, microchip, microprocessor, modem, MRI, penicillin, Prozac, radar, sonar, USB, Valium and WWW. The analysis of publication trends for these mature technologies can be seen in Figure 1. This figure shows that trends of these mature technologies can be classified into one of three broad categories: bell-shaped, exponential and multi-modal. Bell shaped trends have the classic “bell” shape. These trends grow until they reach a peak, after which they decrease down to their former level (see row 1 of Figure 2). Exponential trends are always increasing and their growth rate is accelerating (see row 2 of Figure 2). Multi-modal curves have more than one peak (see row 3 of Figure 2). These curves have multiple stages of growth and decline.

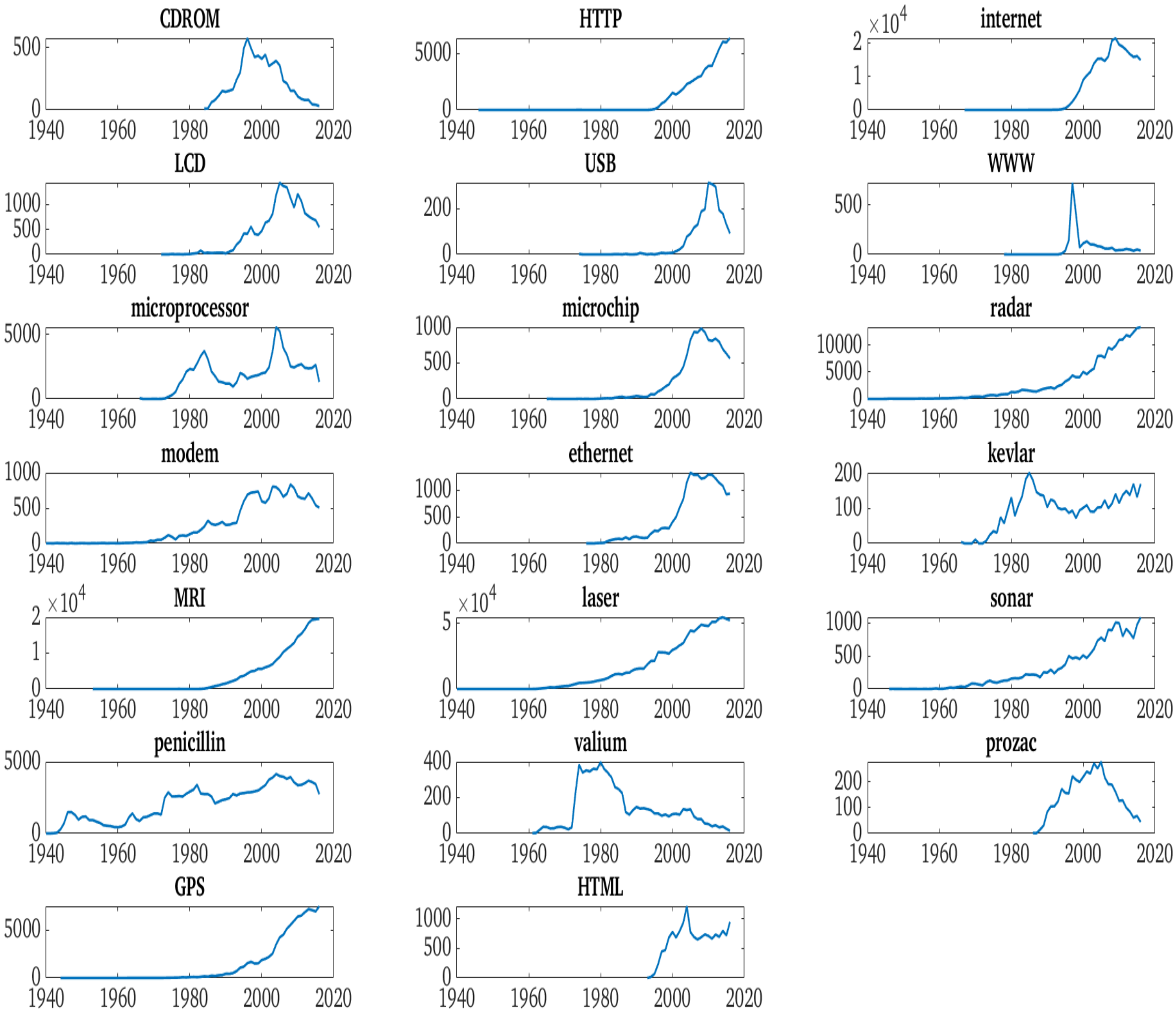

Figure 1. Publication trends of some mature technologies starting from 1940 until now.

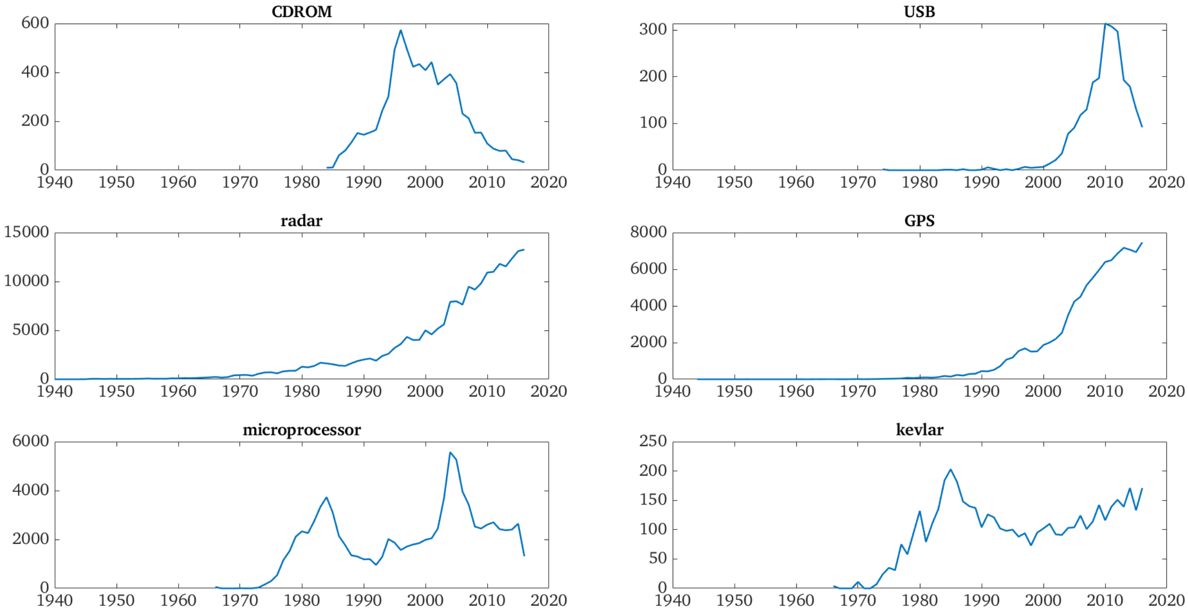

Figure 2. Examples of each trend shape category: bell-shaped (first row), exponential (second row), and multi-modal (third row).

It is interesting to note that the publication trend for a given technology in some aspects also mirrors its technological maturity. In particular, as the number of publications increases so does the maturity of the technology. Once a technology becomes fully mature, the number of publications about it begins to plateau. The maturity of a technology is sometimes described with a technology maturity curve [5]. This is an s-curve (see Figure 3) that models how specific technologies develop, mature and become adopted over time. The technology maturity curve contains four distinct phases of development:

- **New technology** has not reached the first tipping point in the curve.
- **Improving or emerging technology** is within the exponential development stage of the curve between its two tipping points.
- **Mature technology** follows the second tipping point before the curve begins to decline.
- **Aging technology** is on the downward tail of the curve.

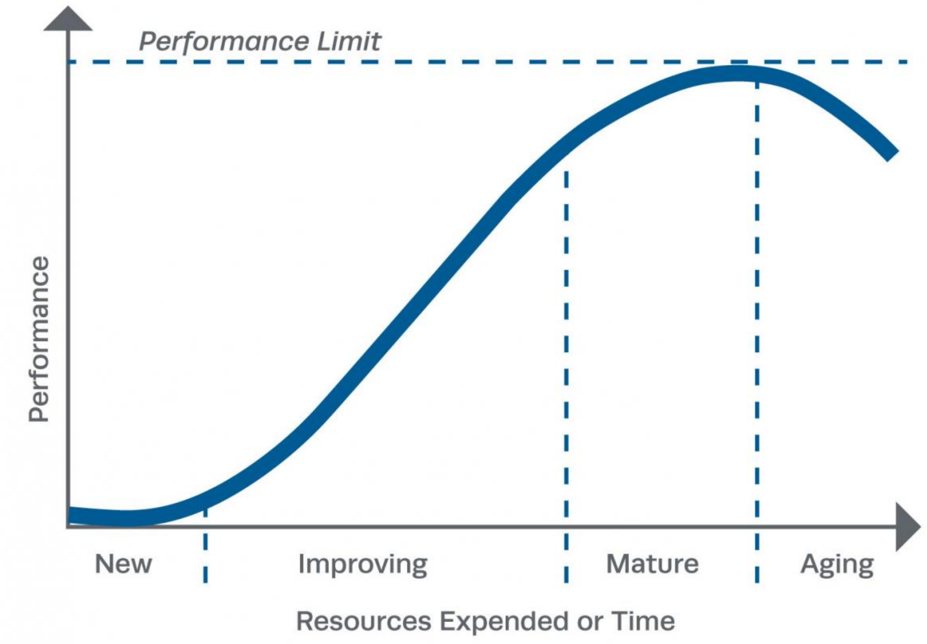

Figure 3. Technology maturity curve (image from [5]).

2.2 Method

The approach to trend prediction presented in this paper uses non-linear regression to fit a mathematical function to the publication trend data. It was important to choose an s-shaped function that closely models the mature technology publication trend data (see Figure 1 and Figure 2). A variety of mathematical functions were considered for this purpose. However, the Gompertz function [6] was selected, as it has been used to model the growth of items in the past, such as: mobile phone uptake [7], population [8], medical tumours [9] and financial market impact [10]. The Gompertz function $G(t)$ uses three independent parameters (a, b, c) that allow it to take various shapes:

$$G(t) = ae^{-be^{-ct}},$$

where a is the vertical asymptote as $t \to \infty$, b controls the horizontal displacement, c controls the growth rate and e is Euler's constant ($e = 2.71828 \ldots$).

Figure 4 shows an example fit of a Gompertz function to the publication data on "CDROM". This case demonstrates that the publication data on technology can be represented by a mathematical model, in this case the Gompertz function, and it is able to model both the rise and the decline in publications. Furthermore, this mathematical model permits the use of the Gompertz function for predictive purposes. In this case it predicts that the number of publications on "CDROM" will continue to decline. The Gompertz function can also be used for making predictions on a variety of other technologies (see Appendix A). The next section will investigate the predictive accuracy of this method.

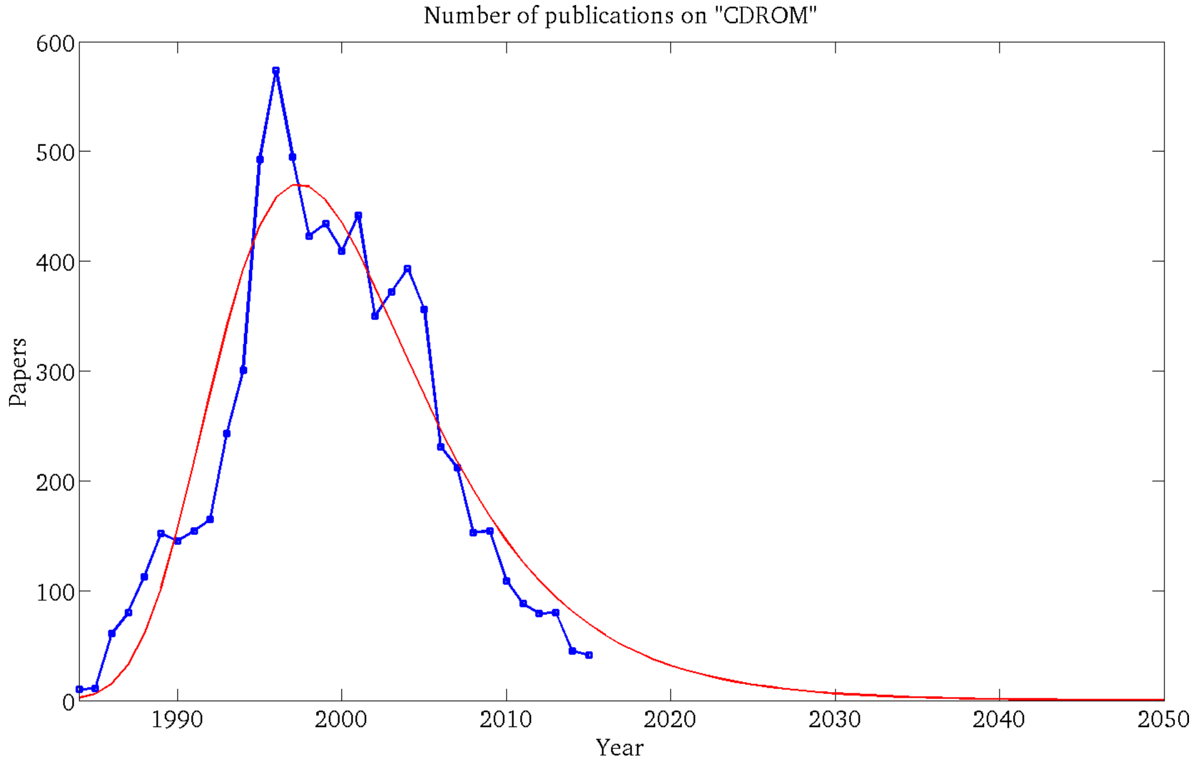

Figure 4. Fitting a Gompertz function to the publication data of "CDROM". Blue curve shows the historical publication data, while the red curve shows the predicted trend.

2.3 Prediction Accuracy on Mature Technologies

To measure the predictive accuracy of this method, we examined the outputs of the Gompertz function using data available at specific points in time (1995, 2000, 2005 and 2010) and compared that to the current technology publication data (in 2016).

Figure 5 shows the prediction accuracy of the Gompertz function given the "CDROM" publication data from Section 2.2. It can be seen that predictions made in 2005 and 2010 are reasonably accurate. For this study, predictions were classified as "close", "far under" or "far over" depending on their accuracy relative to the current publication data (in 2016). The predictive curves produced using the Gompertz function for the complete set of mature technologies used in this study can be found in Appendix A and are summarised in Table 1.

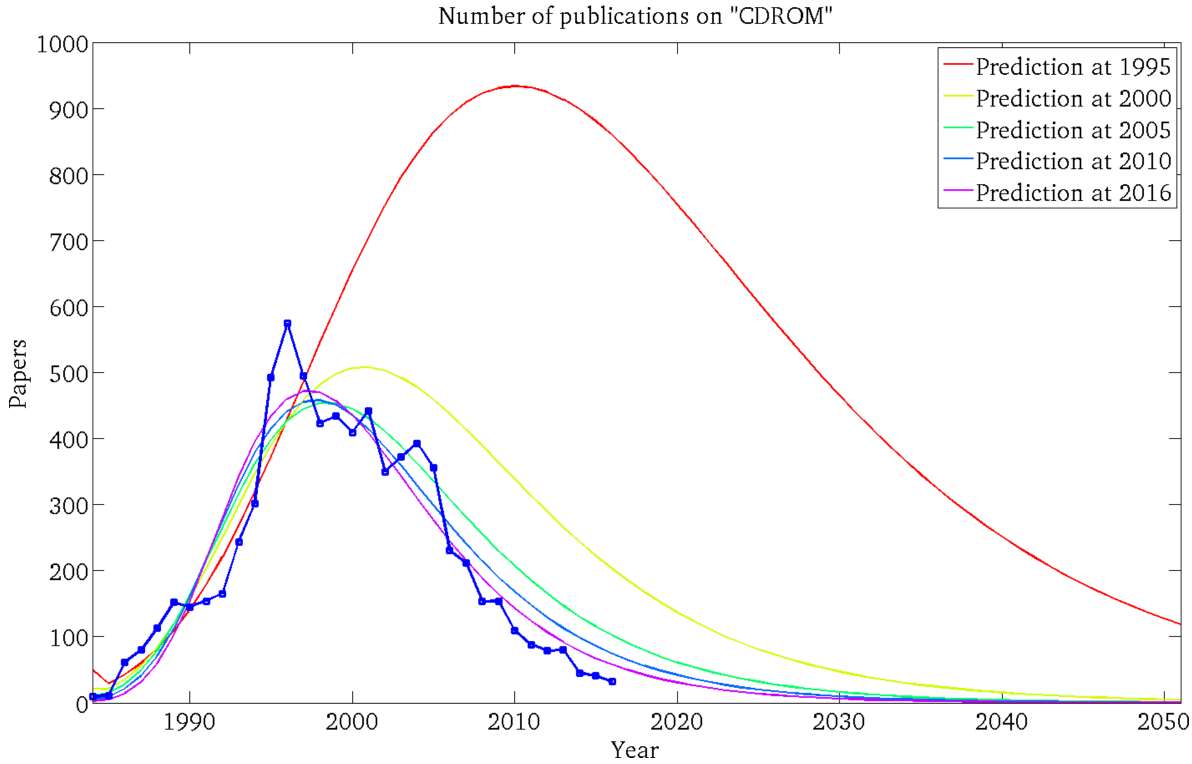

Figure 5. Predictions made at different points in time on the publication data of "CDROM".

Table 1. Prediction accuracy for mature technologies.

Technology	Shape	1995	2000	2005	2010
CDROM	Bell	Far over	Far over	Close	Close
Ethernet	Bell	Far under	Close	Far over	Far over
GPS	Exponential	Far over	Close	Close	Close
HTML	Multimodal	Far under	Far under	Far under	Far under
Internet	Bell	Far under	Far over	Far under	Close
Kevlar®	Multimodal	Far under	Far under	Far under	Far under
Laser	Exponential	Far under	Close	Close	Close
LCD	Bell	Close	Far over	Close	Close
Microchip	Bell	Far under	Far under	Far under	Far over
Microprocessor	Multimodal	Far under	Far under	Close	Far over
Modem	Bell	Close	Far over	Far over	Close
MRI	Exponential	Far under	Far under	Far under	Close
Penicillin	Multimodal	Close	Close	Close	Close
Prozac	Bell	Far under	Close	Close	Close
Radar	Exponential	Far under	Close	Close	Close
Sonar	Exponential	Far under	Far under	Close	Close
USB	Bell	Far under	Far under	Far over	Far over
Valium	Bell	Close	Close	Close	Close
WWW	Bell	Close	Close	Close	Close

Figure 6 highlights how prediction accuracy changes based on the date of the prediction. Not surprisingly, the Gompertz function has higher predictive accuracy for events that occurred closer to the year 2016. In fact predictions made in 2010 (6 years ahead) are more than 70% likely to be "close", while predictions made in (15 years ahead) were only 28% likely to be "close".

Interestingly when the prediction is "far" from accurate it is often "under" (not "over") the actual value (except for 2010 predictions for which the "far under" and "far over" count only differs by one and therefore is not statistically different). This result is consistent with Amara's Law [11, 12], which states:

> We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.

Figure 7 shows how prediction accuracy varies based on the trend shape. We can see that the Gompertz function performs better for bell-shaped trends (50% "close") and exponential trends (60% "close"). The Gompertz function however, is poorly suited for the multimodal trends with just 31% being "close" and hence a different model is needed for those cases.

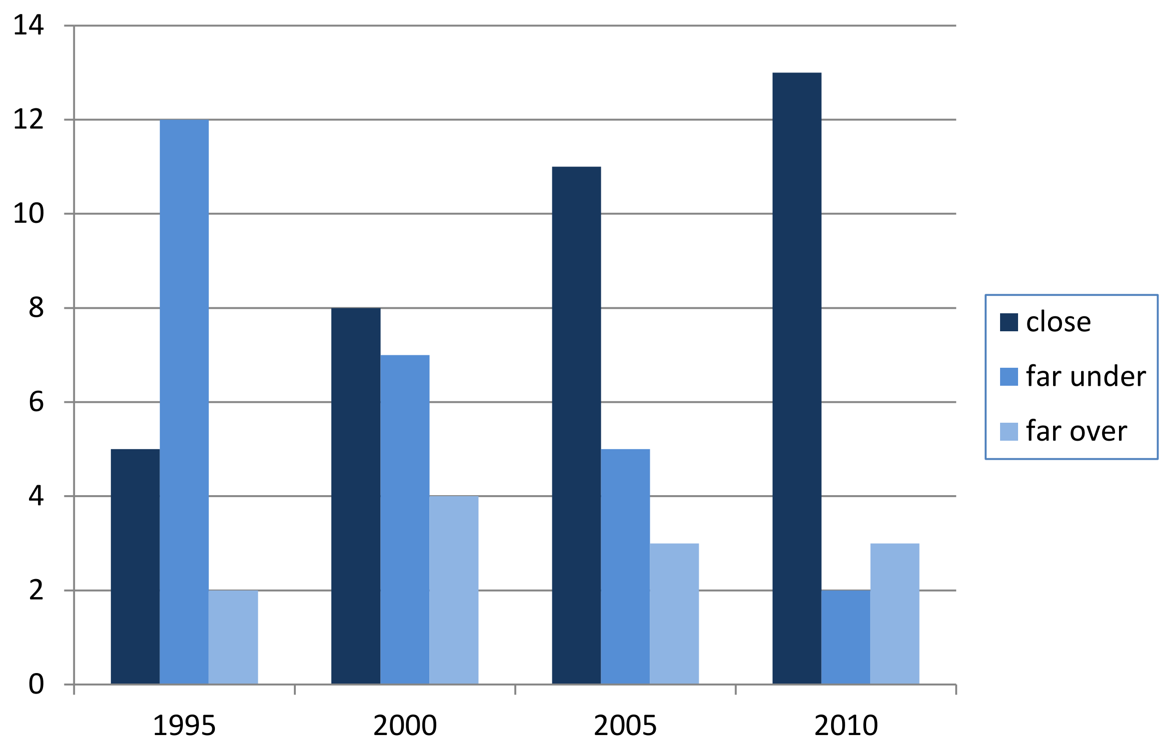

Figure 6. Prediction accuracy grouped by year of prediction.

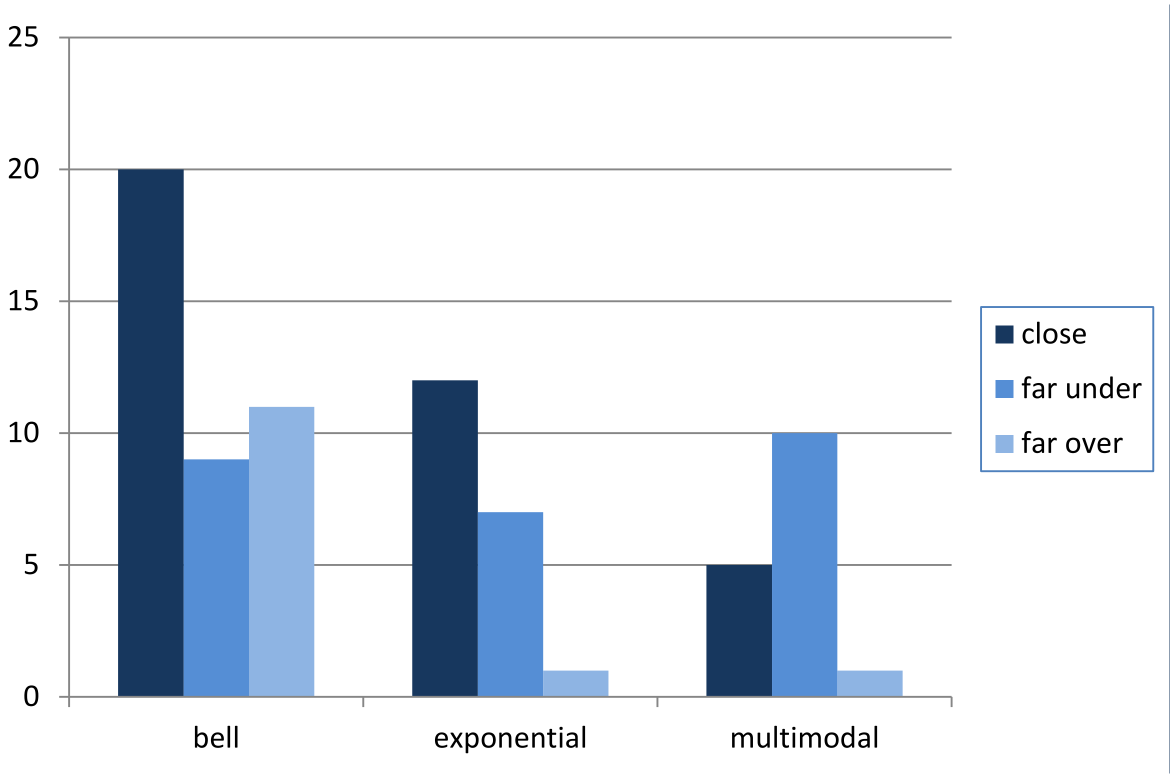

Figure 7. Prediction accuracy grouped by trend shape.

2.4 Predictions for Novel Technologies

Publication trend prediction is difficult due to the large variability in future events. For example, if a technology has a sudden breakthrough then we can expect its publication rate to increase. On the other hand, if a different technology is found to be significantly better, then the current technology publication rate will decrease. To account for these types of unpredictable events, we fitted multiple (100) independent Gompertz functions to the publication data on novel technologies. Each fitted curve can be considered as a possible path that the publication trend could take. The density of the curves represents the likelihood of any given path (outcome) occurring – areas that are more dense have a higher likelihood.

We created each independent prediction curve by using a random instantiation of the existing publication data. Instantiations were created by multiplying each y-value by a random number in the range [0.5, 1.5]. Figure 8 shows an example of this for "cloaking" technologies. The current publication trend is in blue, while the predicted curves are in red. This figure shows that there is some uncertainty in "cloaking" publication trends since there is a high variance between the red curves. Based on the majority of the red curves in Figure 8, we can predict that the publications on "cloaking" technologies will likely peak around 2020 and then decline for another 20-30 years. Figure 9 summarises all the predicted curves by showing the expected (mean) curve in solid red and ±1 standard deviations in dashed red (upper/lower bounds).

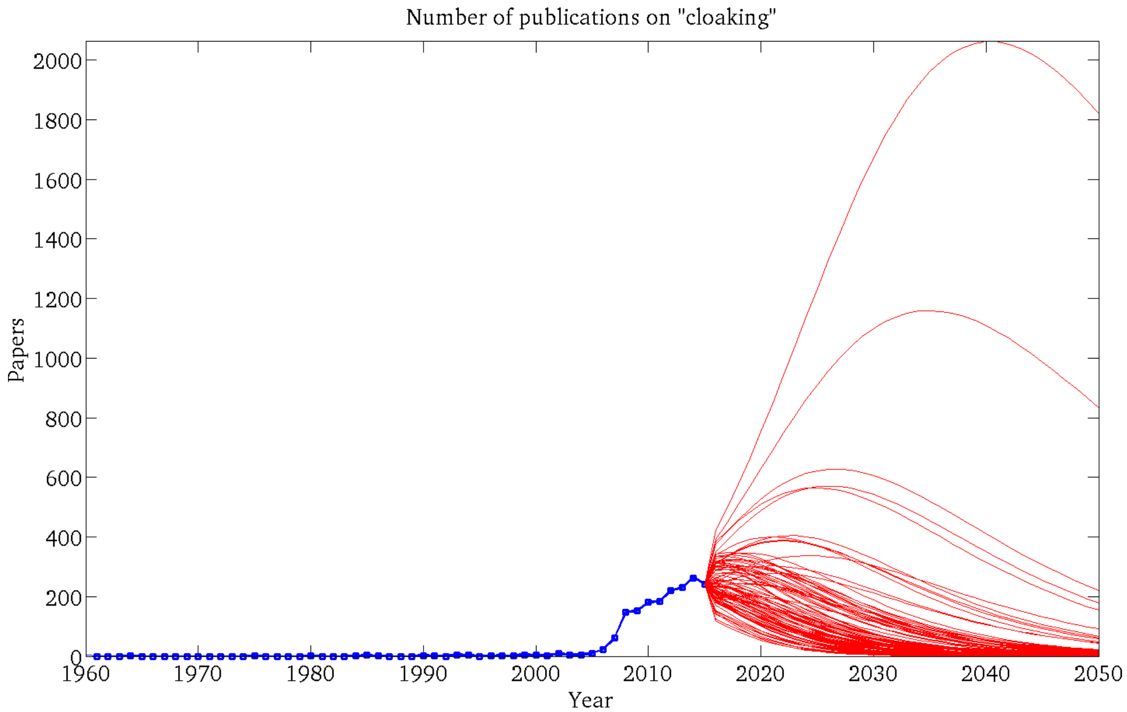

Figure 8. Number of publications on "cloaking" technologies as a function of time. Blue curve shows the current data, while red curves show possible future trends.

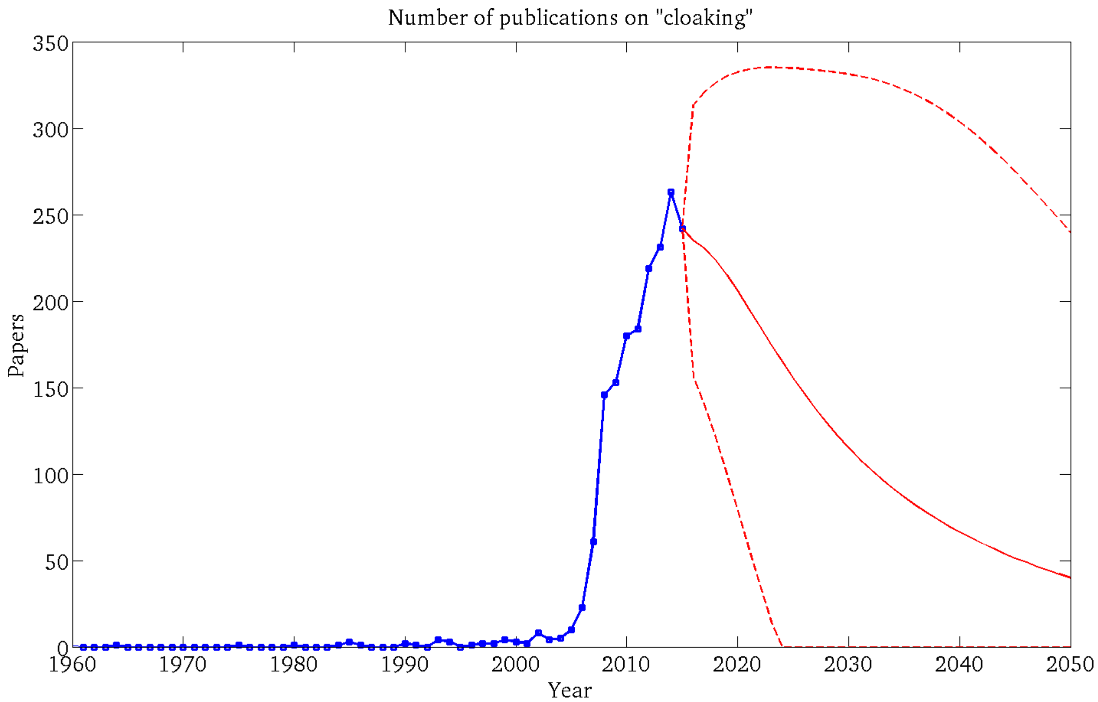

Figure 9. Number of publications on "cloaking" technologies as a function of time. Blue curve shows the current data, solid red curve shows the expected trend, while the dashed red curves show the upper/lower bounds on the expected trend.

This study investigated the use of the Gompertz function for predicting publication trends for a number of novel and emerging technologies, which are presented in Appendix B. These predictions are also summarised in Table 2. Most technologies are expected to experience some growth. In particular, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and Deep Learning are expected to grow rapidly. Other technologies however, are already at their peak and are expected to decline, such as: Nanodots, OLED, Optical Computing and Virtual Reality. A number of technologies like Bioplastics, Metal Foam and Nanomedicine still need 10-15 years to develop before they have the potential to grow.

Table 2. Predictions for novel technologies.

Technology	Prediction
Aerogel	Gradual growth
Augmented reality	Gradual growth
Biofuels	Growth until 2020 followed by a decline
Bioplastics	Steady until 2025, followed by possible growth
Biosensors	Gradual growth
CRISPR	Rapid growth
Deep learning	Rapid growth
Directed energy	Gradual growth
Energy harvesting	Gradual growth
Fusion power	Gradual growth
Graphene	Gradual growth
Metal foam	Steady until 2030, followed by possible growth
Nanodots	Decline
Nanomaterial	Gradual growth
Nanomedicine	Steady until 2025, followed by rapid growth
OLED	Decline
Optical computing	Decline
Quantum computing	Decline
Small satellites	Gradual growth
Virotherapy	Decline
Virtual reality	Decline
Wireless energy	Gradual growth

3. Conclusion and Future Work

This study investigated the use of the Gompertz function as a mathematical model for predicting future trends in publication data. The prediction accuracy of the Gompertz function was measured on historical peer-reviewed publication data of mature technologies. It was found that the prediction accuracy is acceptable for short-term predictions (5-10 years), but declines for long-term predictions (10+ years). The predictions were more accurate for bell-shaped and exponential publication trends. This study also used the Gompertz function to predict the future publication trends of novel and emerging technologies. It predicts that many of the examined novel technologies are expected to grow. In particular, CRISPR and deep learning are expected to grow rapidly. Other technologies like optical computing are already at their peak and will decline in the future.

There are a number of avenues for future work on trend prediction:

- This study only used number of publications as the source of data. The use of additional sources, such as patents and Google trends, could help to increase the accuracy of trend predictions. These sources contain additional information that is not provided by the publication data. We expect novel technologies to appear much earlier in patent data, compared to publication data, indicating that patent data could be a better predictor of new technology. Google trends tend to become active once the technology has reached the public, indicating that this is a good source of data for predicting technology adoption rates. This creates the opportunity to fuse information from multiple sources to obtain a more accurate prediction.

- It would be interesting to compare and contrast the publication trends of various technologies. There are likely commonalities in trends for technologies in similar domains, which may help in emerging technology prediction. Furthermore, it may be possible to match trends for different technologies in different domains. For example, if an emerging novel technology follows a similar trend pattern as a mature technology then it is likely to continue to follow it. This would allow the use of trends from older, mature technologies to help in predicting future emerging technology trends.

- In this study prediction accuracy was only measured subjectively. Future work should focus on using and developing quantitative measures of accuracy. One possibility is to use the absolute difference between the predicted value and the actual value. To make this measure robust, this difference should be normalised relative to the maximum value in the trend.

- This study focused only on the use of the Gompertz function. Investigation into the use of other s-shaped functions for non-linear regression should be considered. These could be logistic curves, polynomial functions or a combination of functions where each one models a separate trend shape.

- This study did not consider the link between a technology's publication trend and the uptake or commercial use of that technology. Investigation into the correlation between publication trends and commercial acceptance (uptake) of technology may assist in improving the quality of emerging technology prediction.

4. Acknowledgements

We are grateful to Jayson Priest, Ryan Hales and Dale Quinn for their useful comments and review of this report.

5. References

1. Miller, P. E. and Swinehart, K. D. (2011) *Technological Forecasting: a Strategic Imperative*. College of Business and Technology, East Tennessee State University, USA
2. Glenn, J. C. (2009) *Introduction to the Futures Research Methods Series*. The Millennium Project
3. Keenan, M. (2006) Overview of Foresight Methods. *Technology Foresight for Organizers*
4. *Scopus: abstract and citation database of peer-reviewed literature*. Available from: https://www.elsevier.com/solutions/scopus.
5. The MITRE Corporation (2014) *The MITRE Systems Engineering Guide*
6. Jukić, D., Kralik, G. and Scitovski, R. (2004) Least-squares fitting Gompertz curve. *Journal of Computational and Applied Mathematics* **169** (2) 359-375
7. Islam, T., Fiebig, D. G. and Meade, N. (2002) Modelling multinational telecommunications demand with limited data. *International Journal of Forecasting* **18** (4) 2002/10/01/ 605-624
8. Zwietering, M. H., et al. (1990) Modeling of the Bacterial Growth Curve. *Applied and Environmental Microbiology* **56** (6) 1875-1881
9. Sottoriva, A., et al. (2010) Cancer Stem Cell Tumor Model Reveals Invasive Morphology and Increased Phenotypical Heterogeneity. *Cancer Research* **70** (1) 46
10. Caravelli, F., et al. (2016) Optimal growth trajectories with finite carrying capacity. *arXiv* (1510.05123)
11. *Amara's Law*. Available from: https://spotlessdata.com/blog/amaras-law.
12. Ridley, M. *Amara's Law*. (2017) Available from: http://www.rationaloptimist.com/blog/amaras-law/.

Appendix A Predictions for Mature Technologies

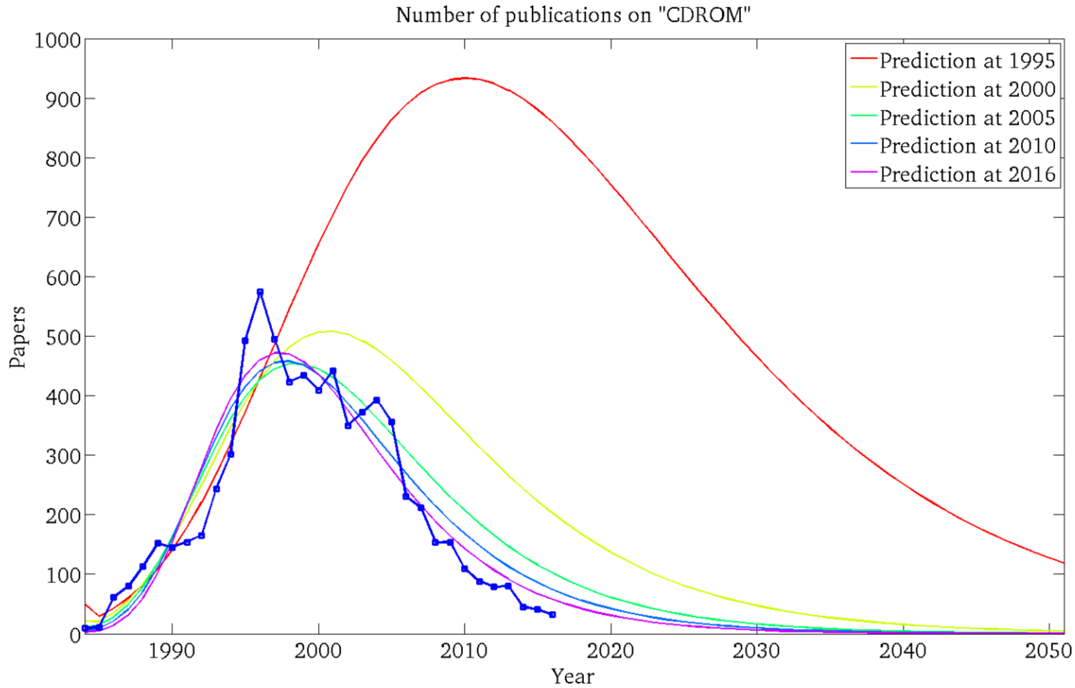

Figure 10. Prediction for "CDROM".

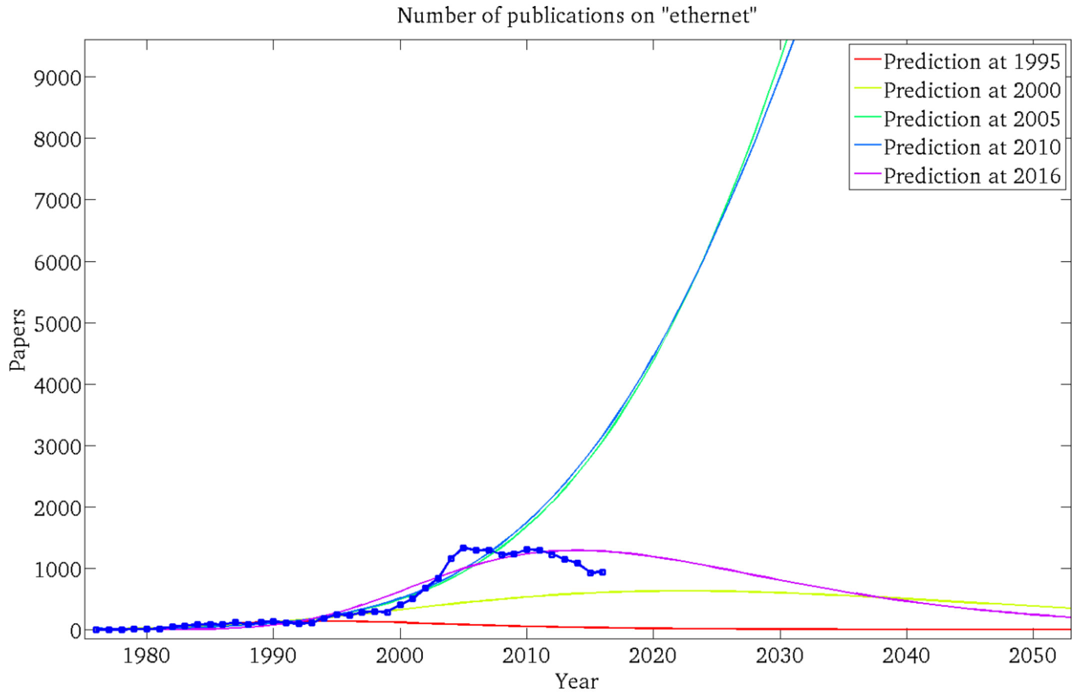

Figure 11. Prediction for "Ethernet".

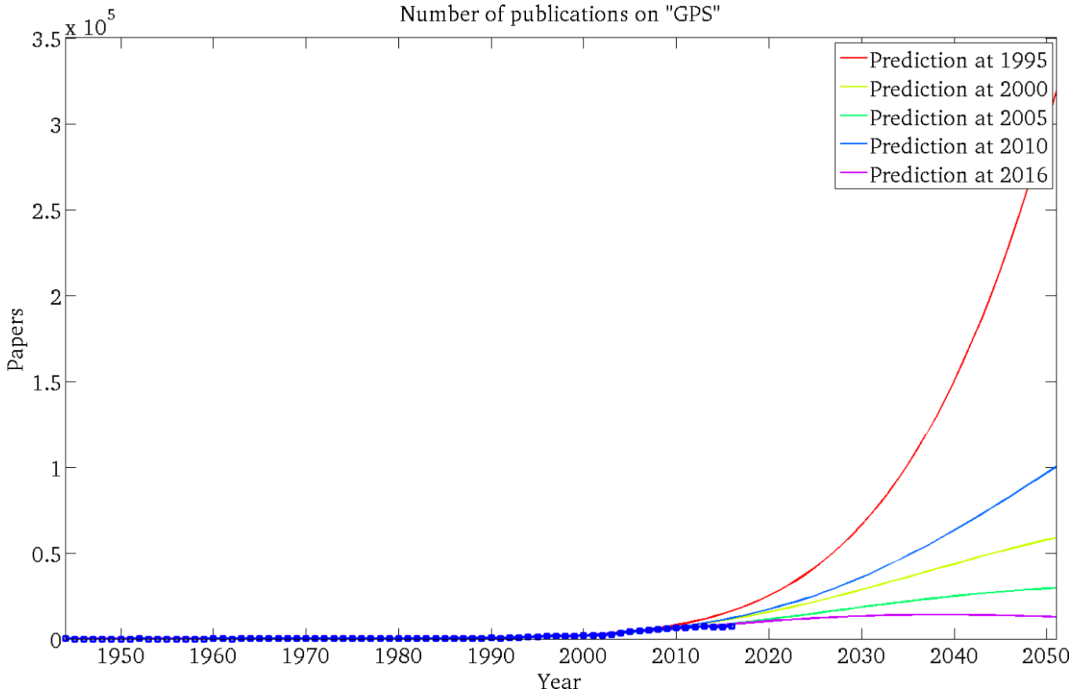

Figure 12. Prediction for "GPS".

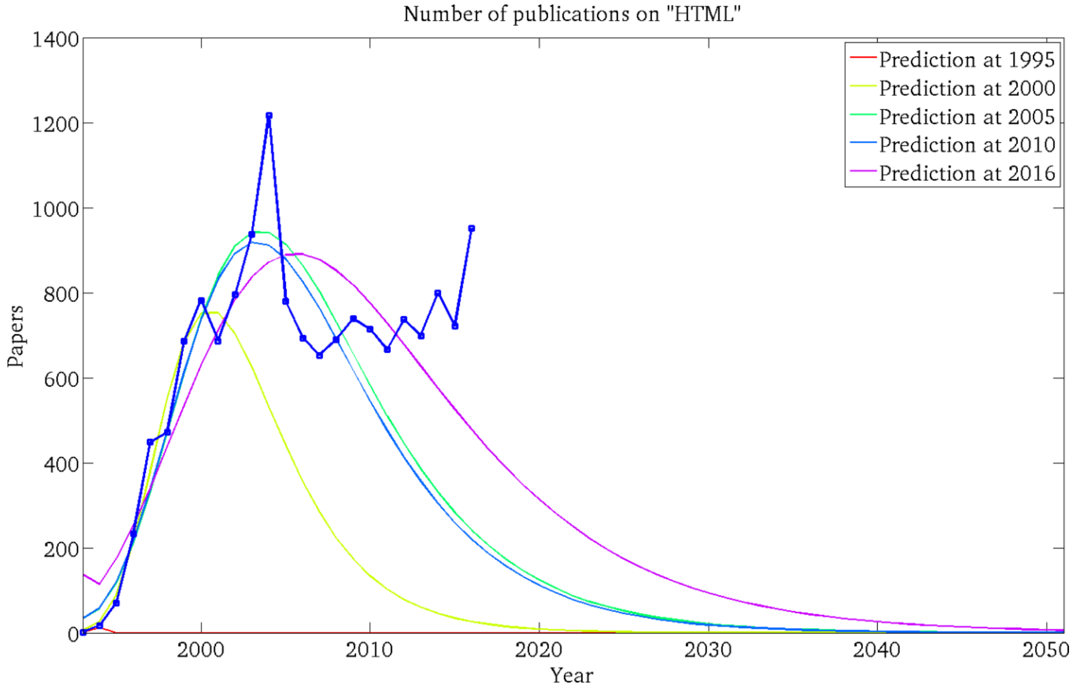

Figure 13. Prediction for "HTML".

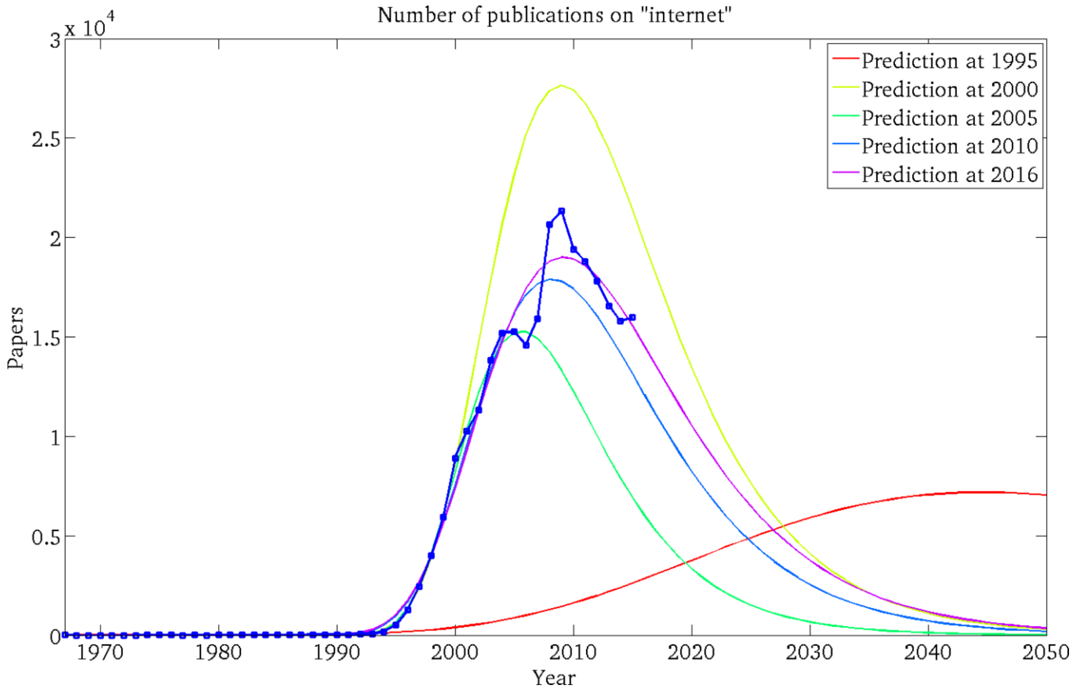

Figure 14. Prediction for "internet".

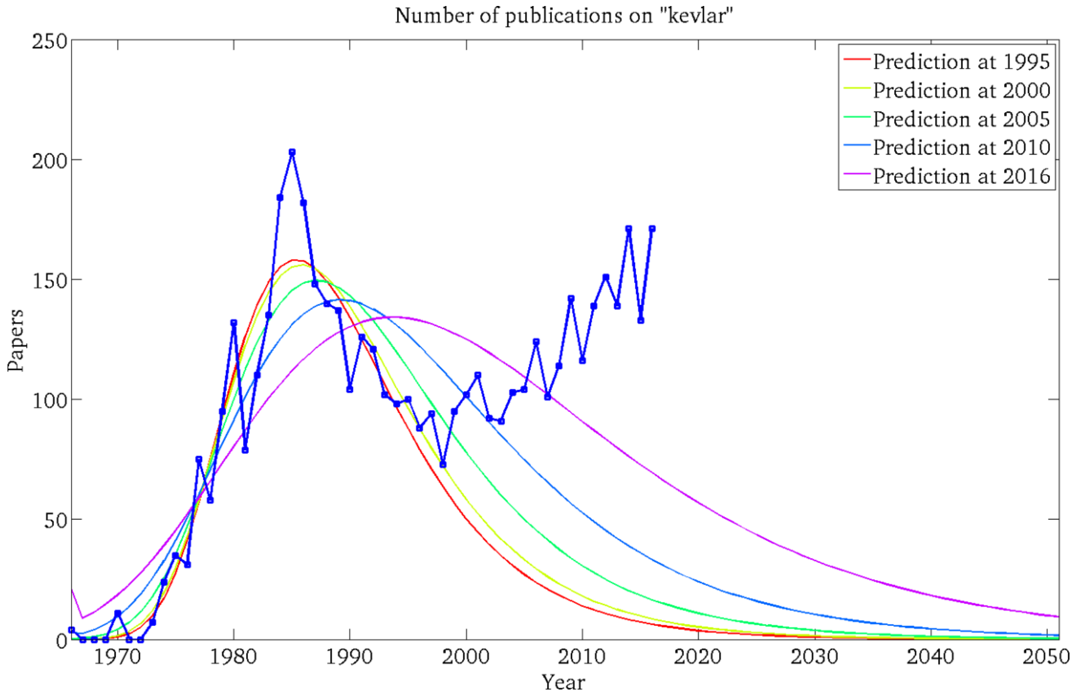

Figure 15. Prediction for "Kevlar".

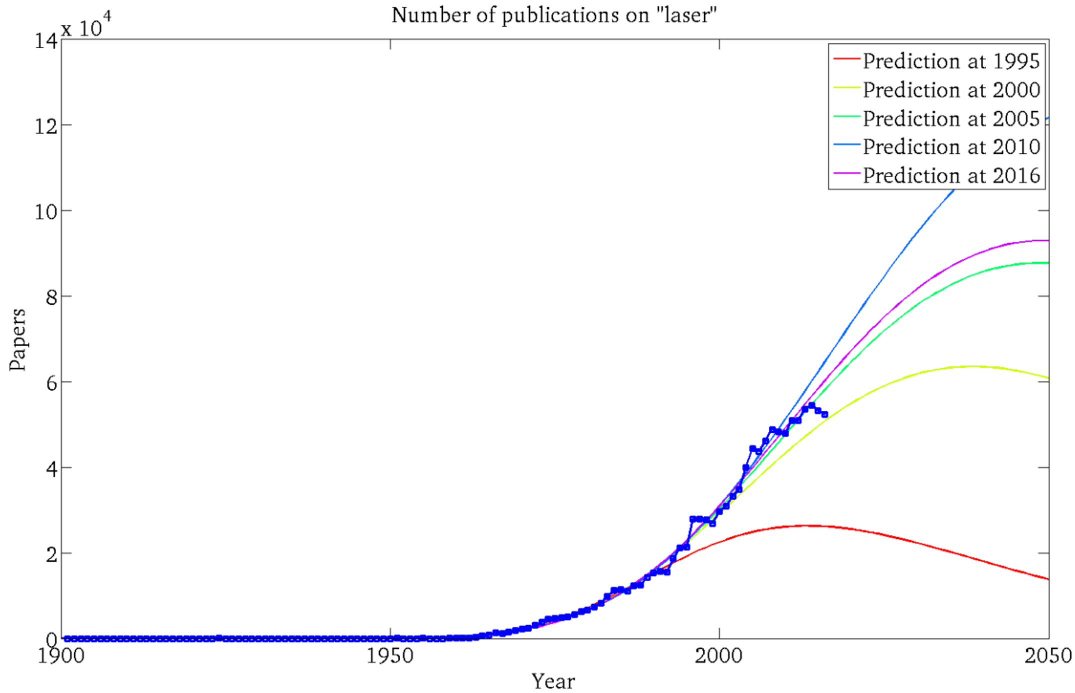

Figure 16. Prediction for "laser".

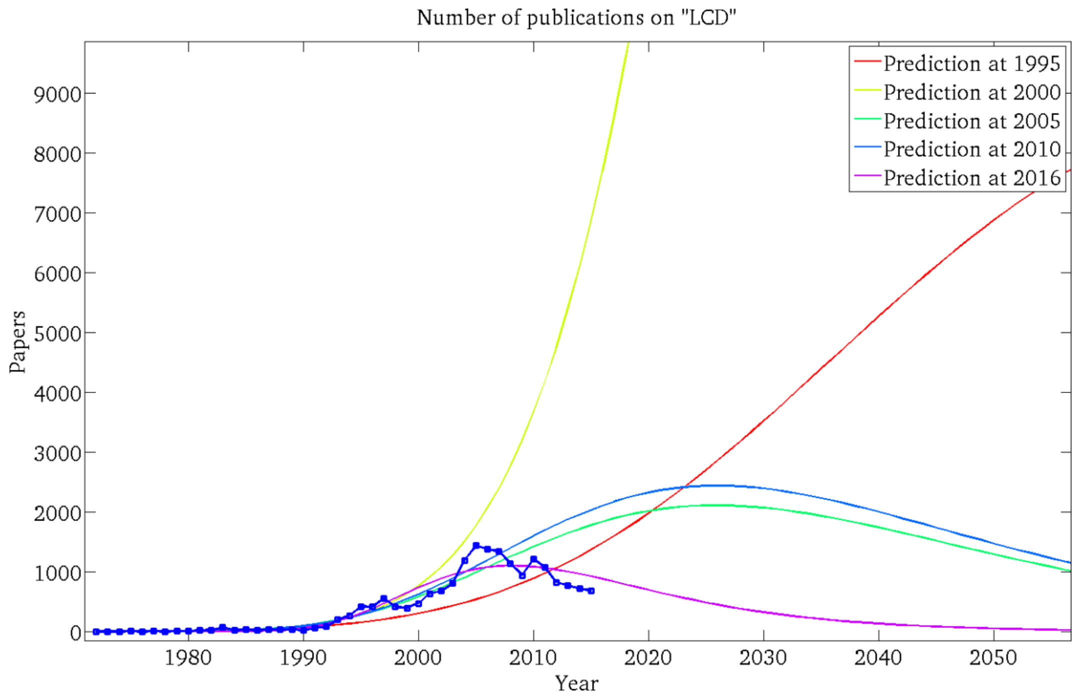

Figure 17. Prediction for "LCD".

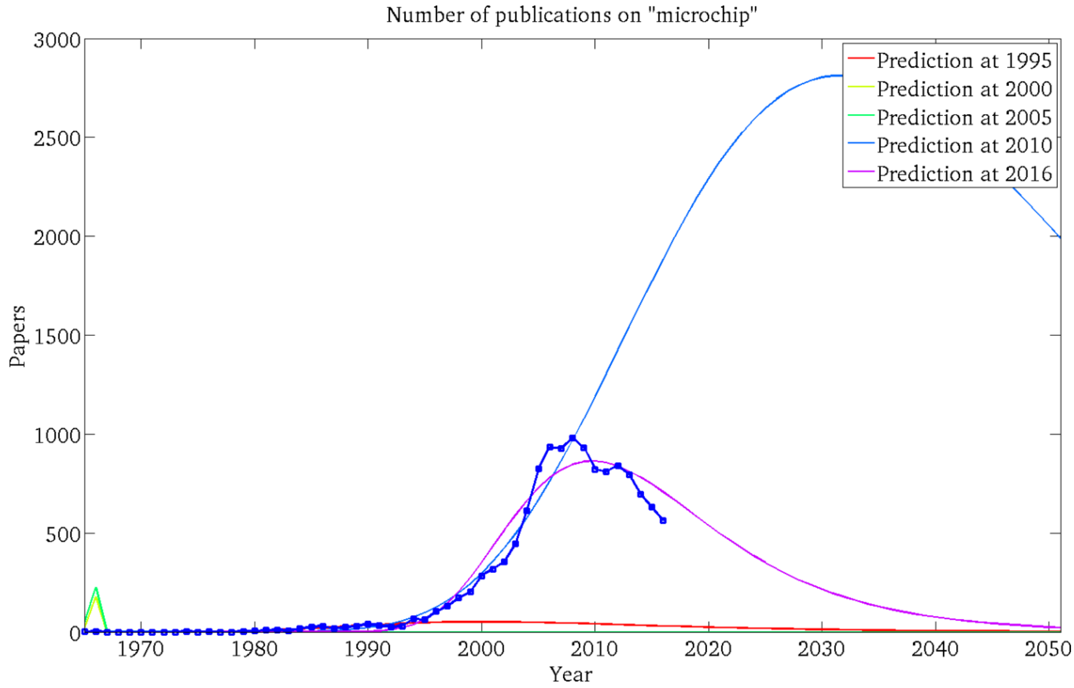

Figure 18. Prediction for "microchip".

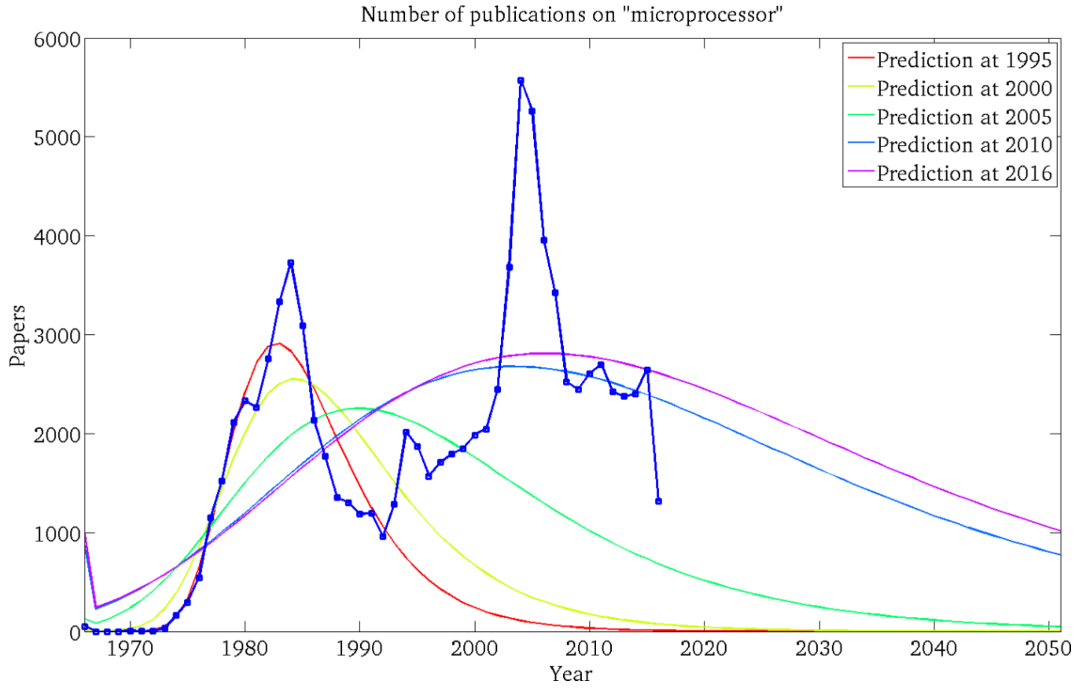

Figure 19. Prediction for "microprocessor".

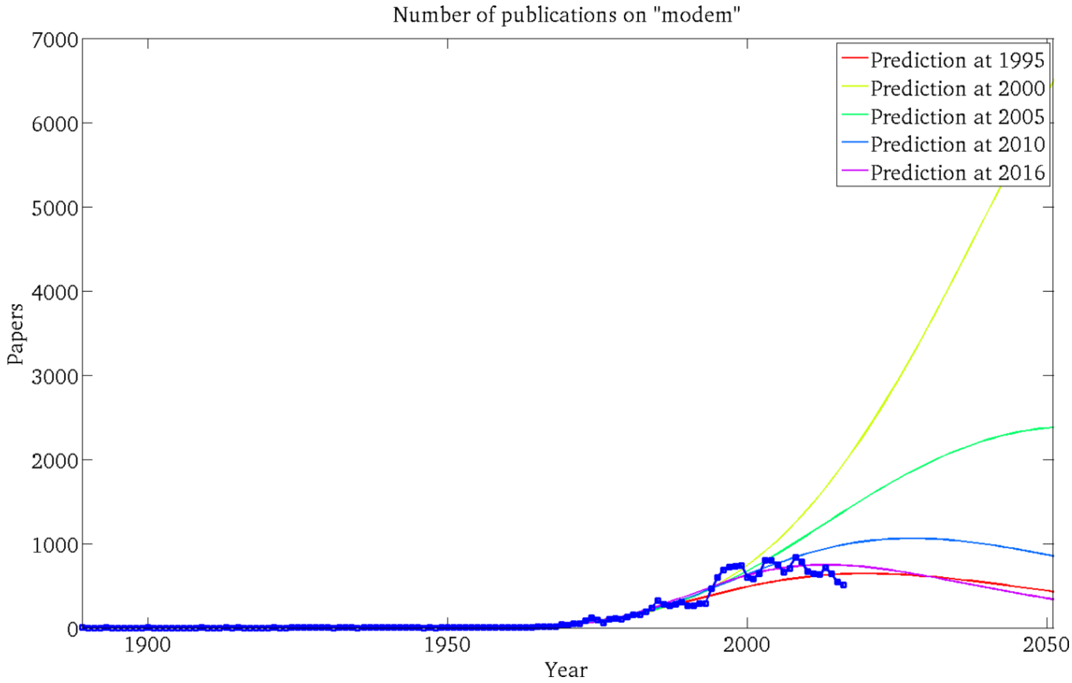

Figure 20. Prediction for "modem".

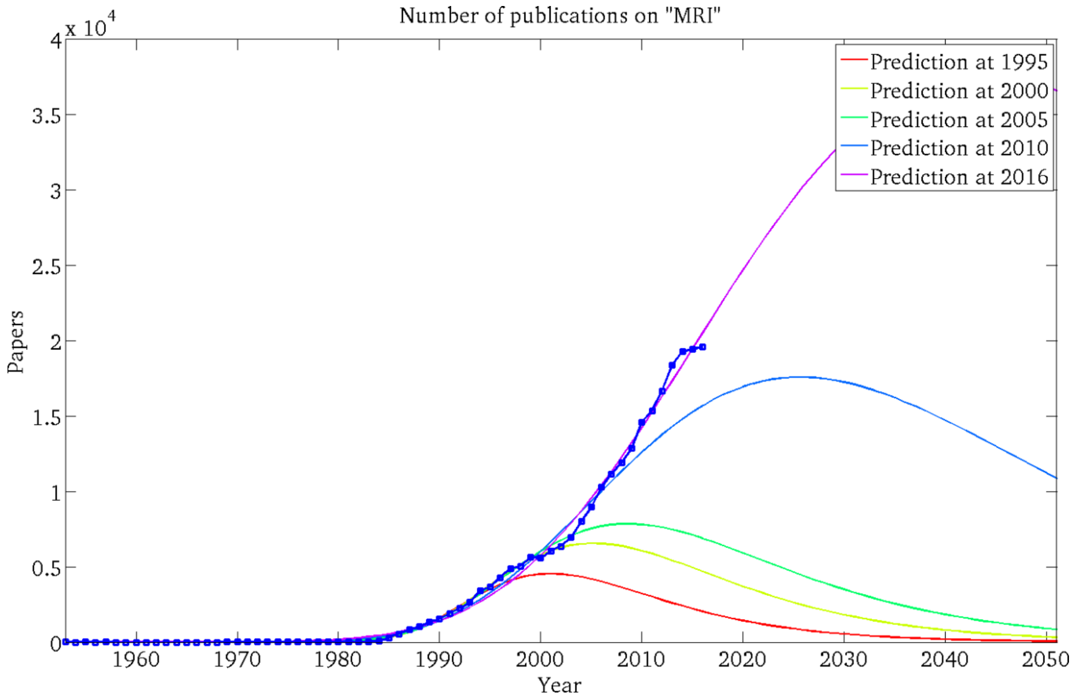

Figure 21. Prediction for "MRI".

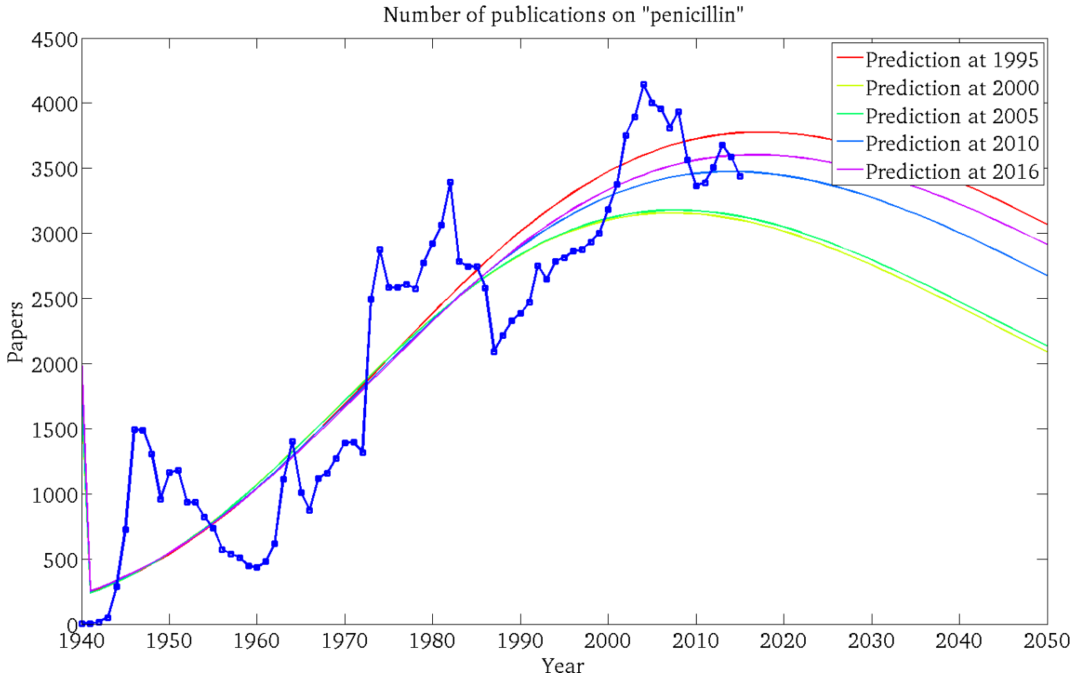

Figure 22. Prediction for "penicillin".

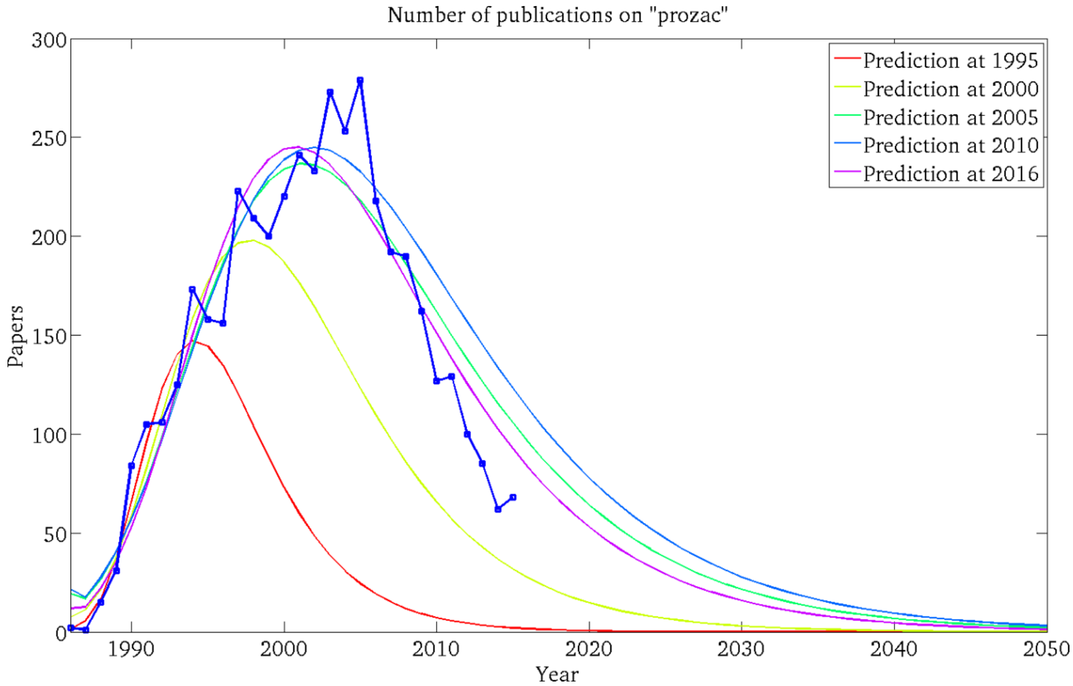

Figure 23. Prediction for "Prozac".

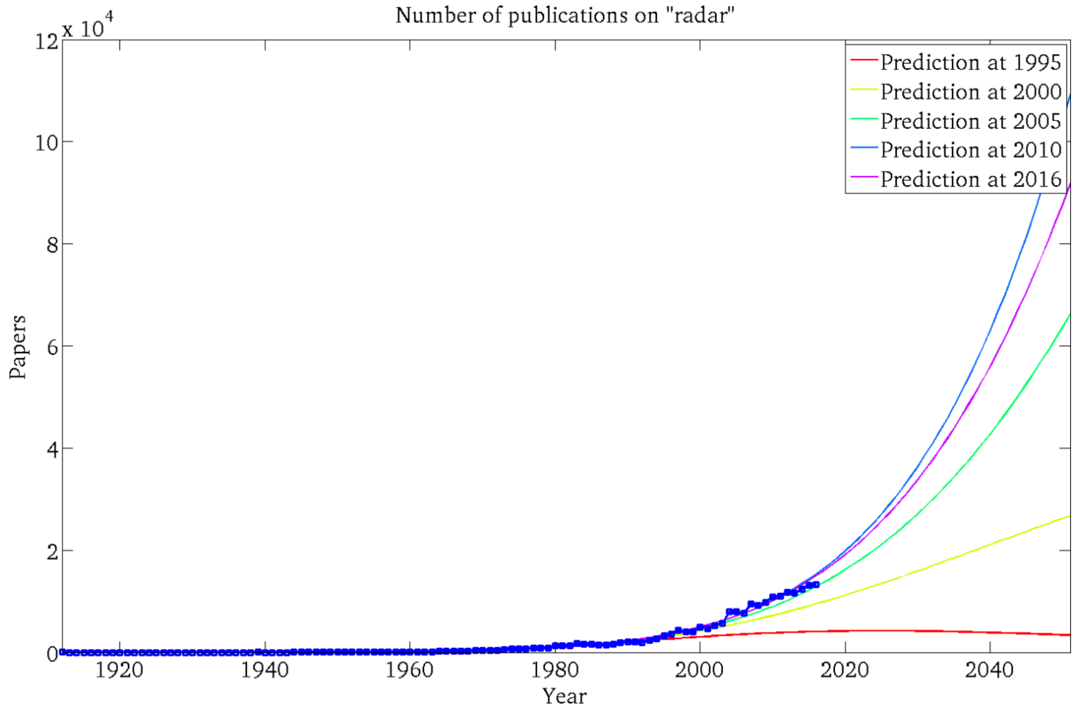

Figure 24. Prediction for "radar".

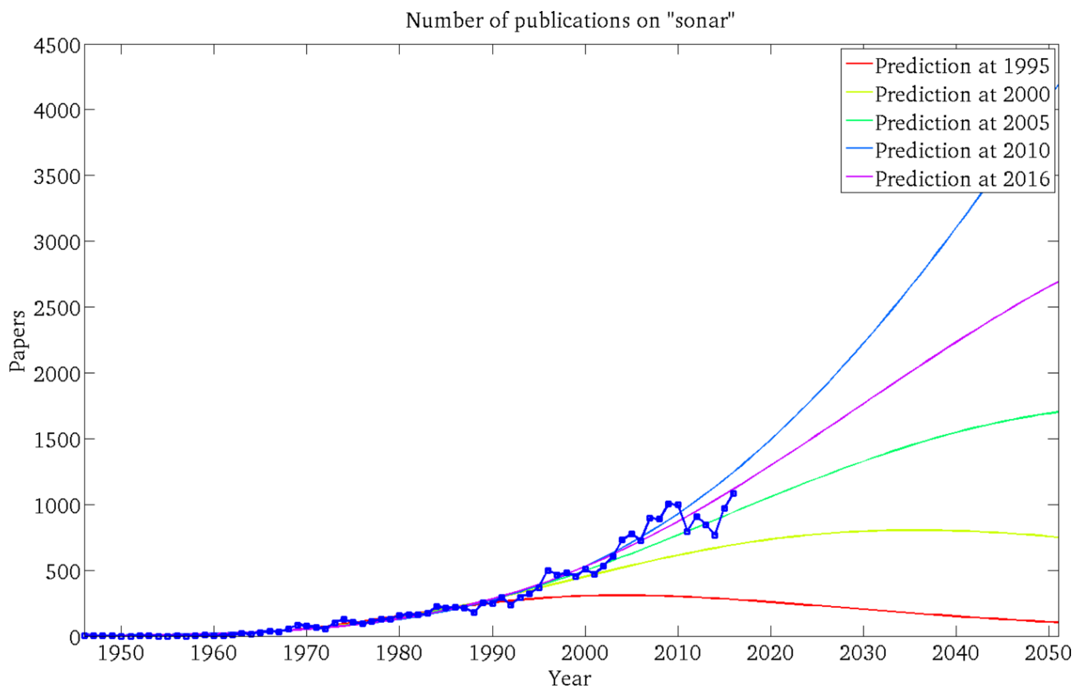

Figure 25. Prediction for "sonar".

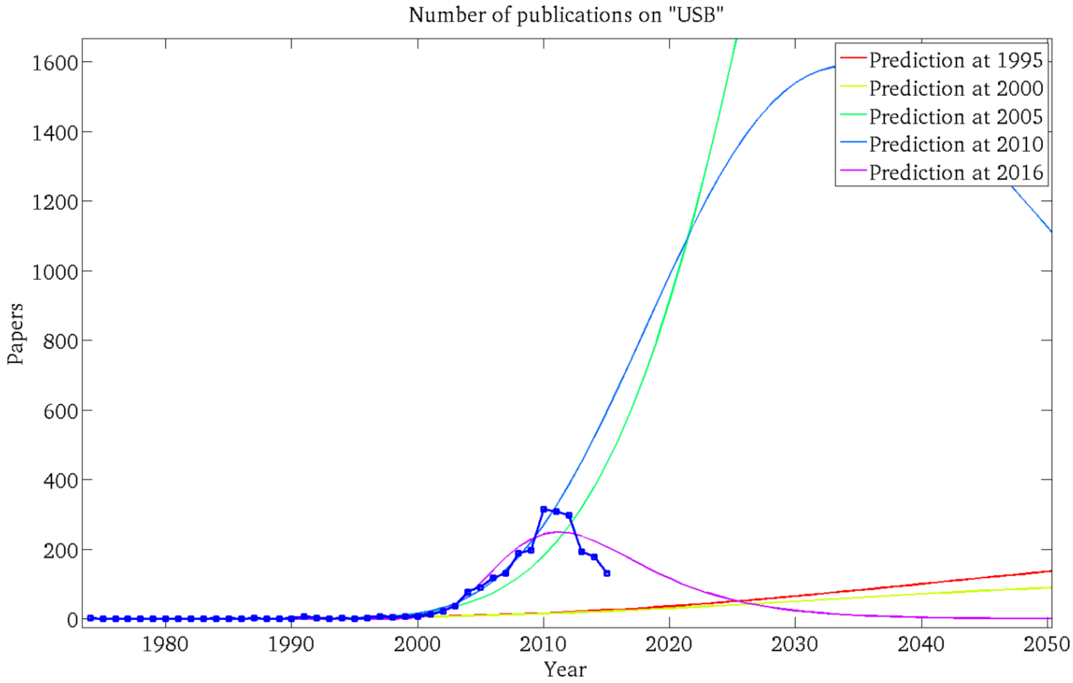

Figure 26. Prediction for "USB".

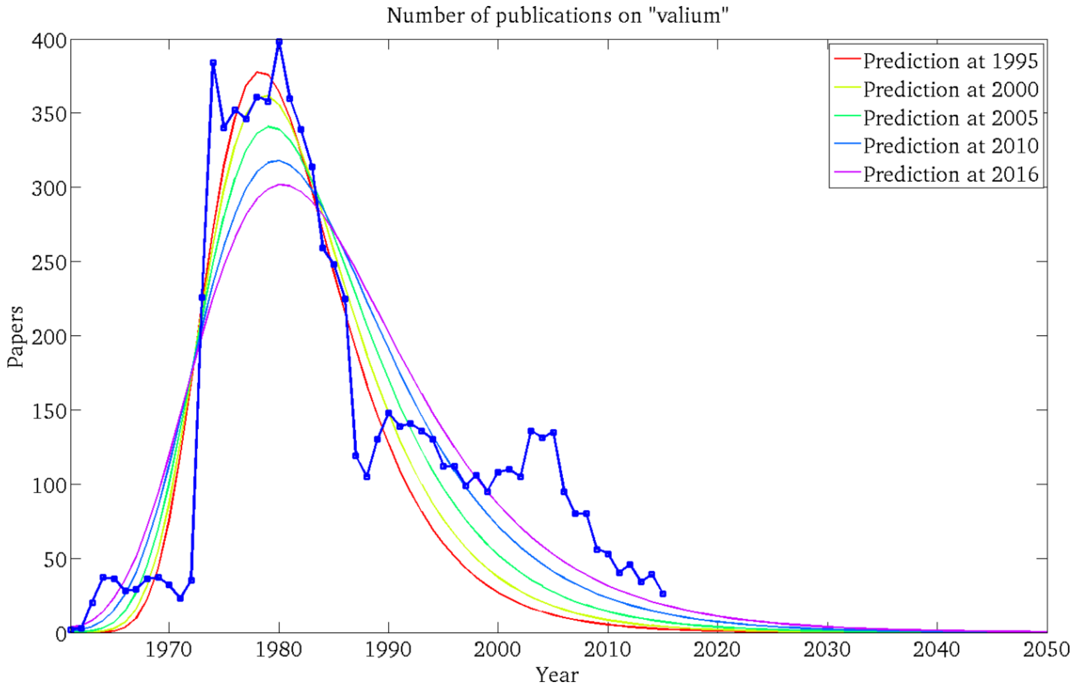

Figure 27. Prediction for "Valium".

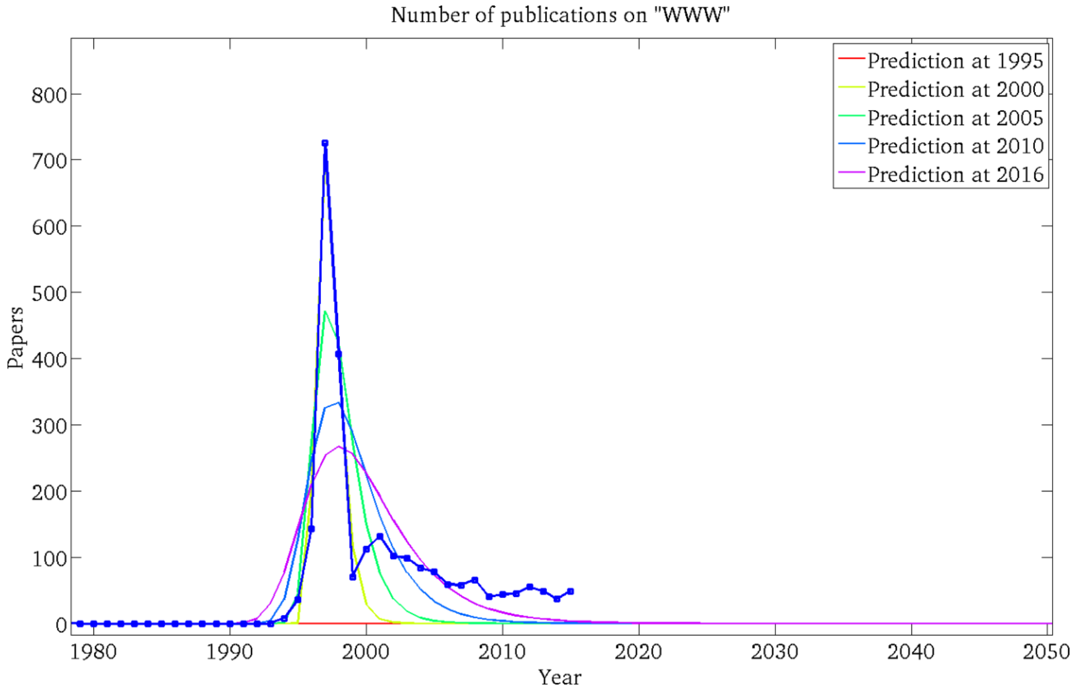

Figure 28. Prediction for "WWW".

Appendix B   Predictions for Novel Technologies

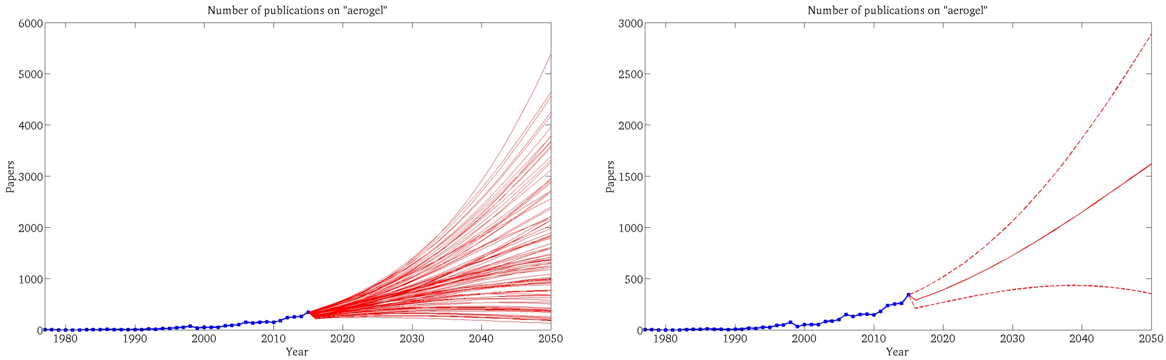

Figure 29. Prediction for "aerogel" technologies.

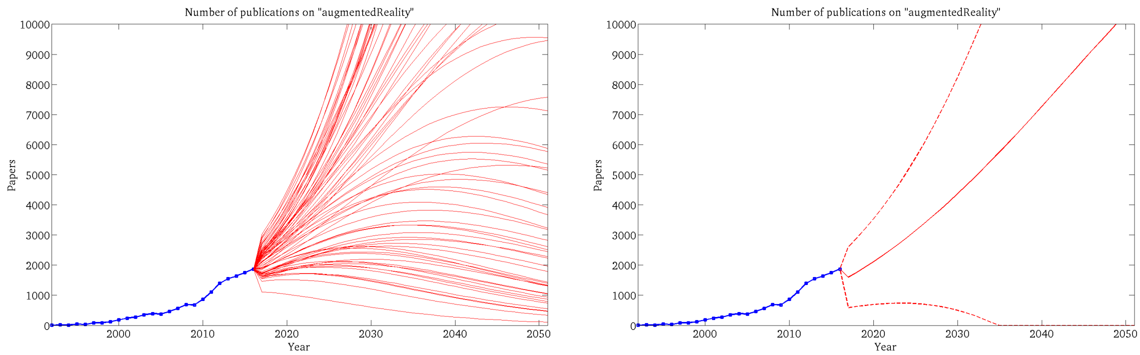

Figure 30. Prediction for "augmented reality".

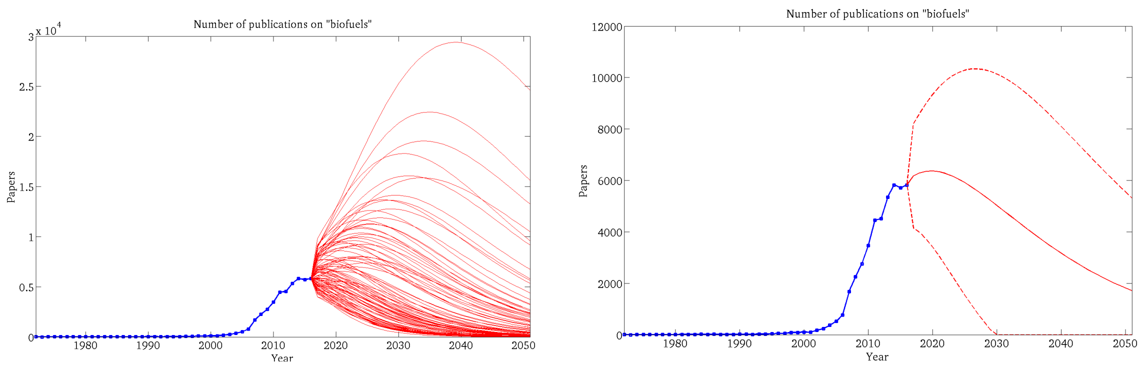

Figure 31. Prediction for "biofuels".

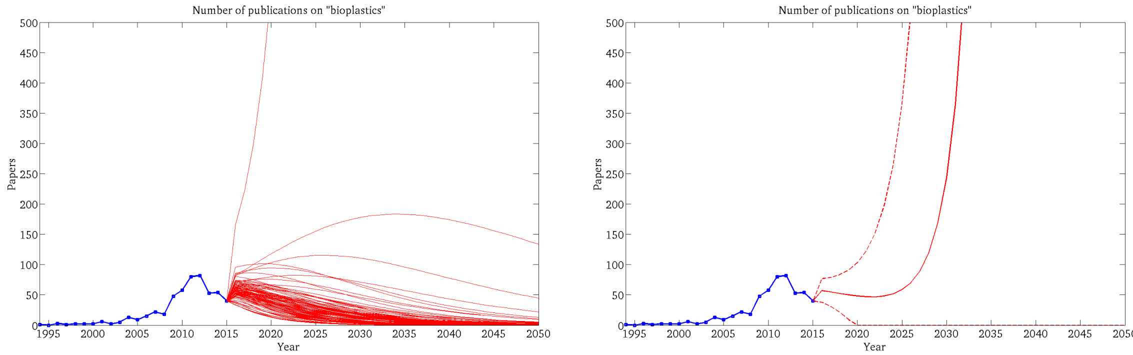

Figure 32. Prediction for "bioplastic" technologies.

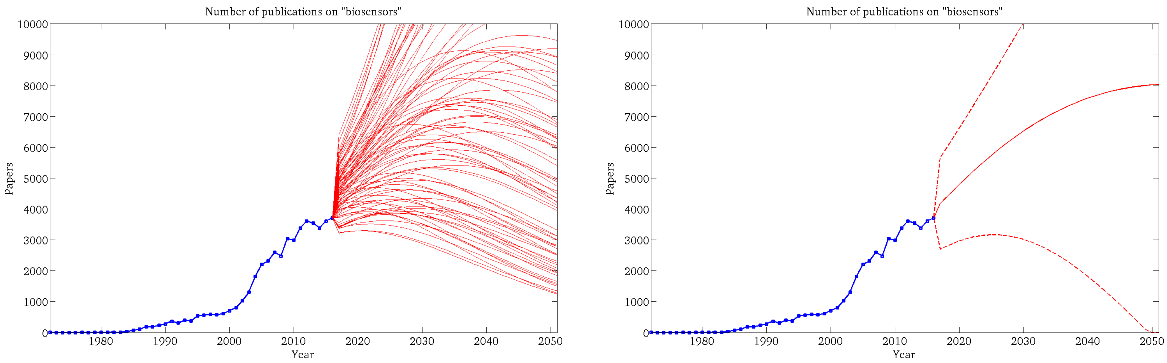

Figure 33. Prediction for "biosensors".

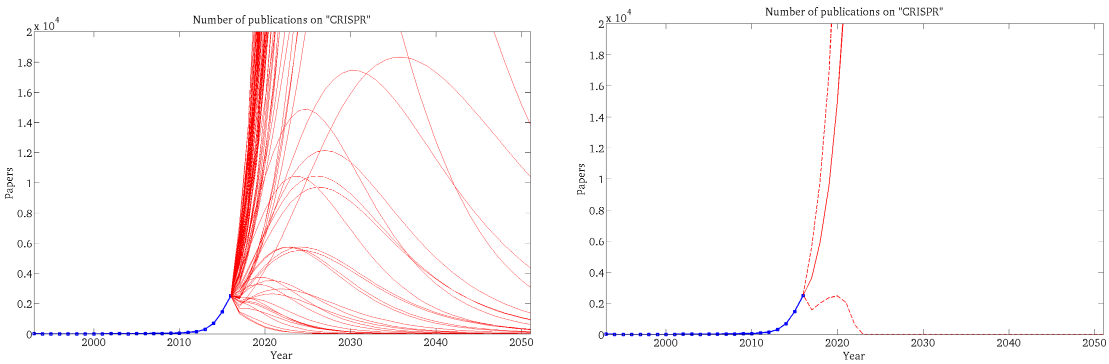

Figure 34. Prediction for "CRISPR" technologies.

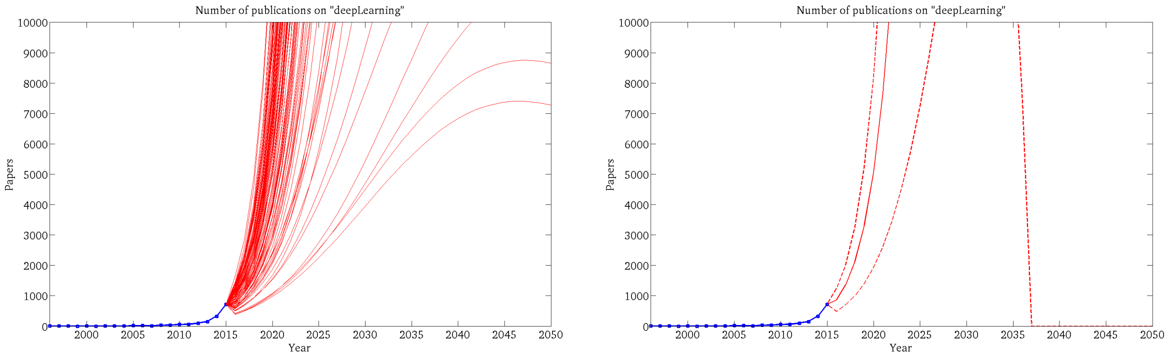

Figure 35. Prediction for "deep learning" technologies.

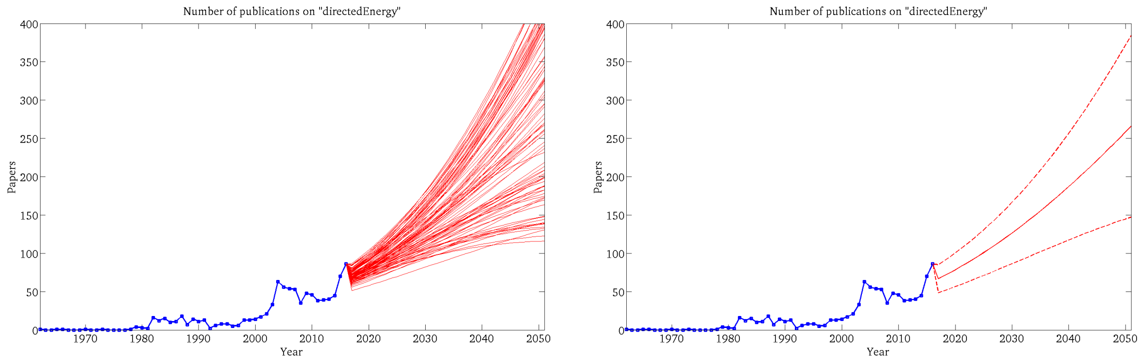

Figure 36. Prediction for "directed energy" technologies.

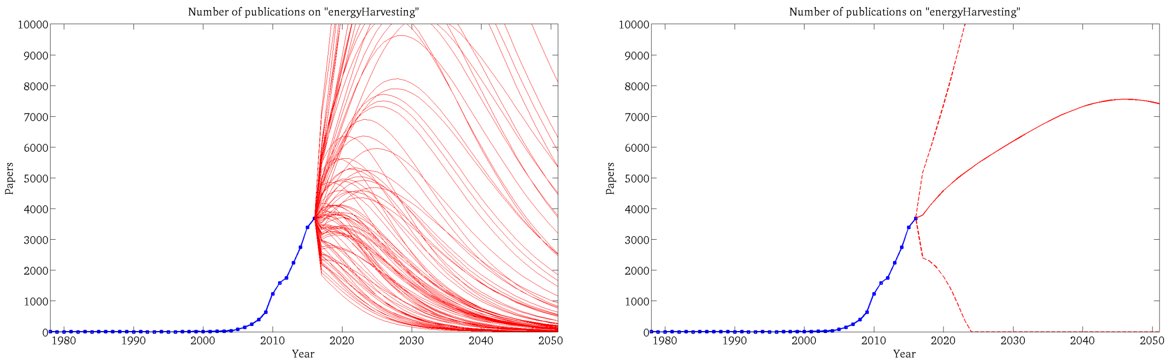

Figure 37. Prediction for "energy harvesting" technologies.

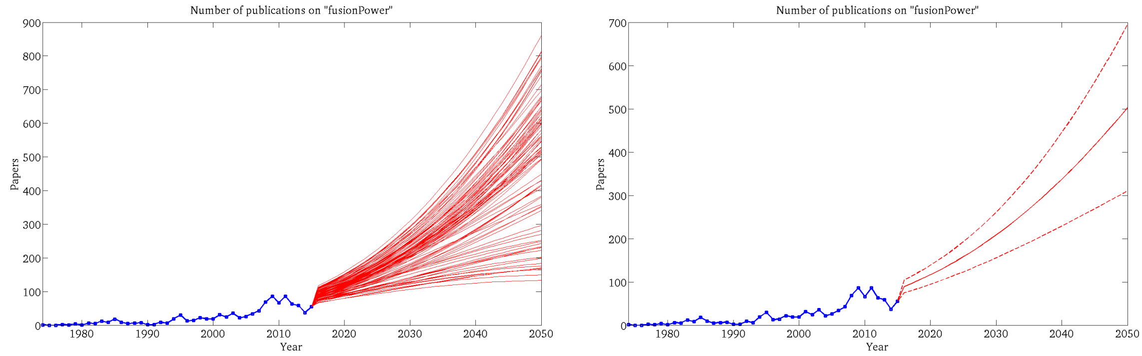

Figure 38. Prediction for "fusion power".

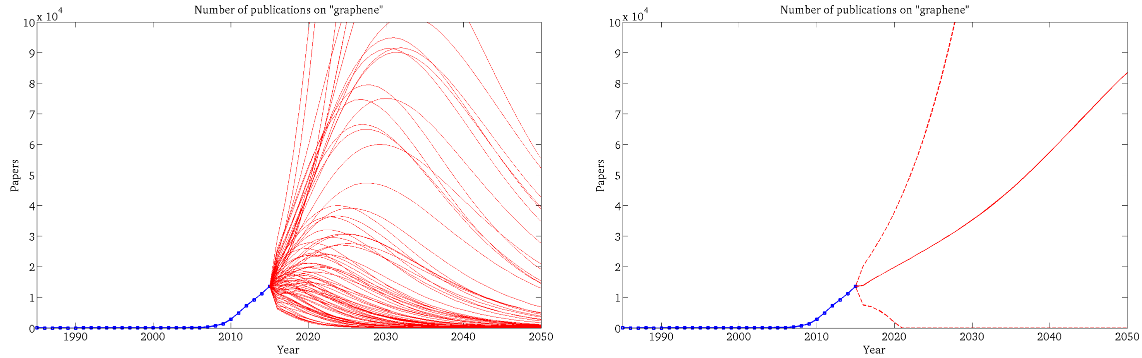

Figure 39. Prediction for "graphene" technologies.

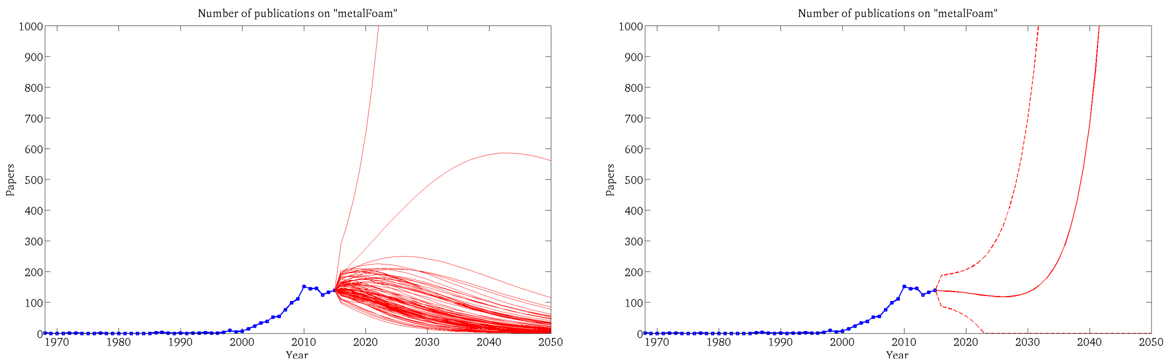

Figure 40. Prediction for "metal foam" technologies.

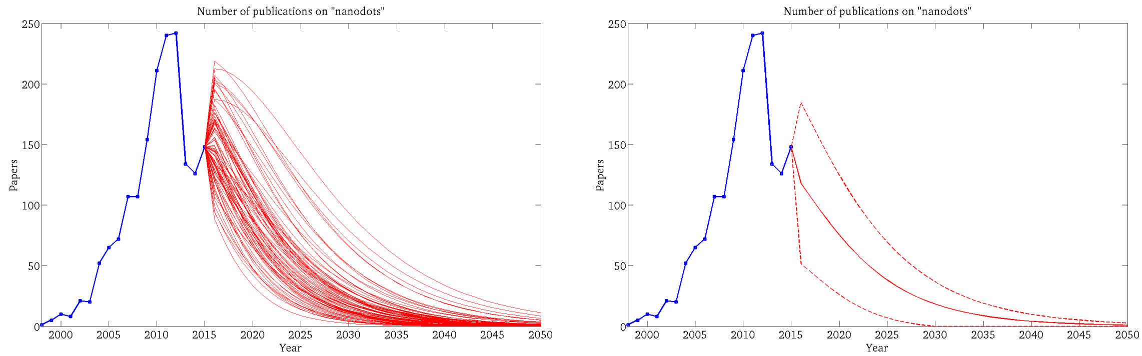

Figure 41. Prediction for "nanodots" technologies.

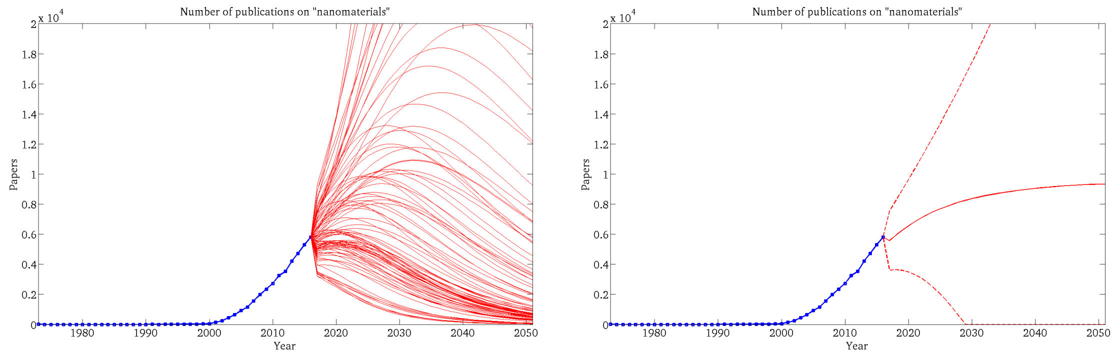

Figure 42. Prediction for "nanomaterial" technologies.

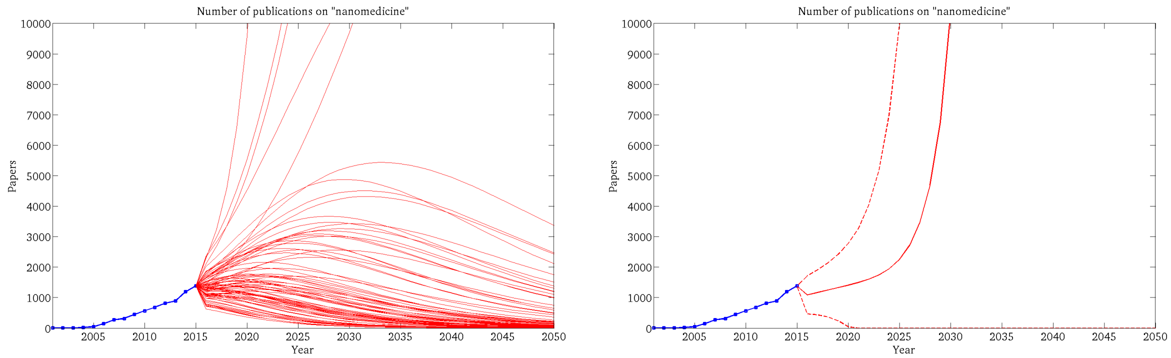

Figure 43. Prediction for "nanomedicine".

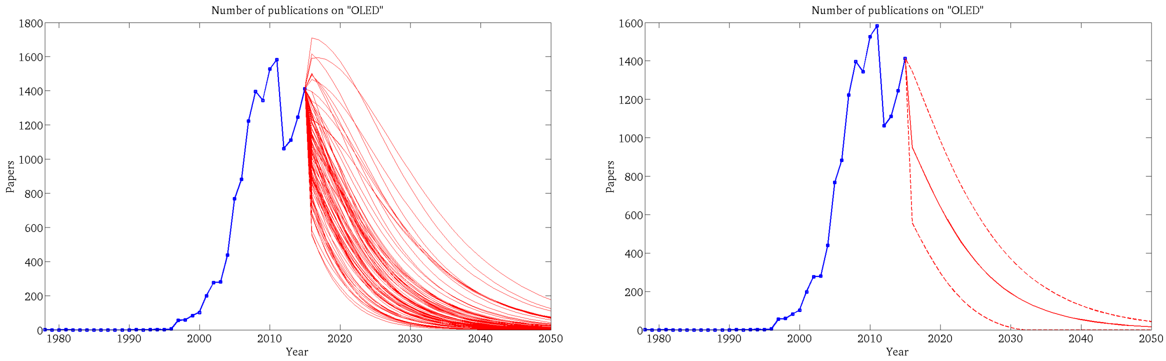

Figure 44. Prediction for "OLED" technologies.

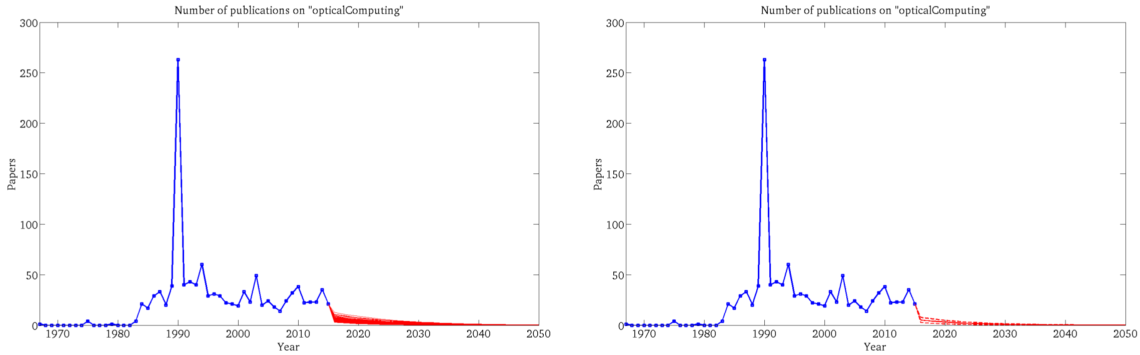

Figure 45. Prediction for "optical computing".

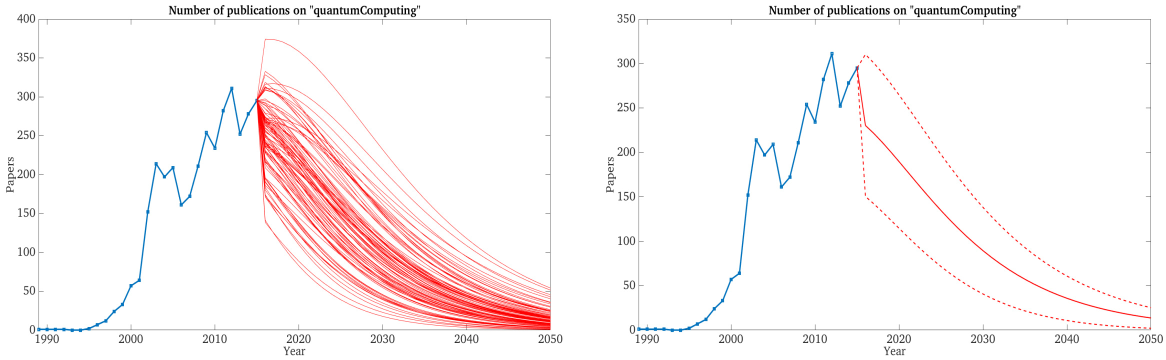

Figure 46. Prediction for "quantum computing".

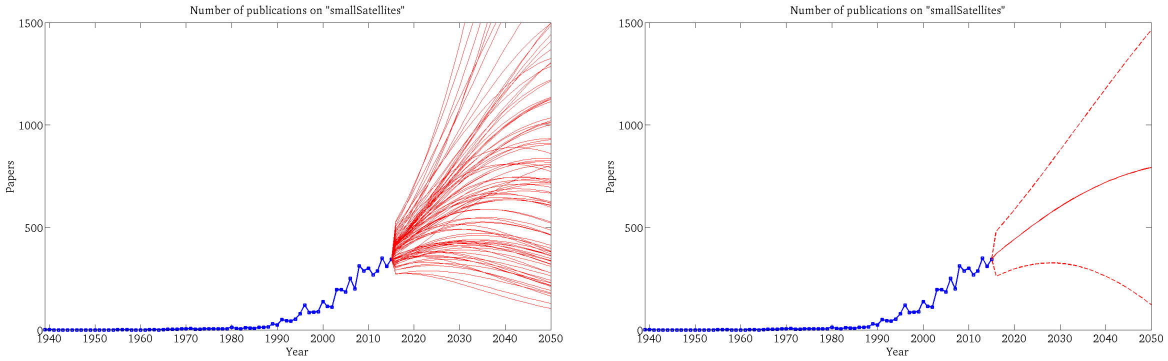

Figure 47. Prediction for "small satellites".

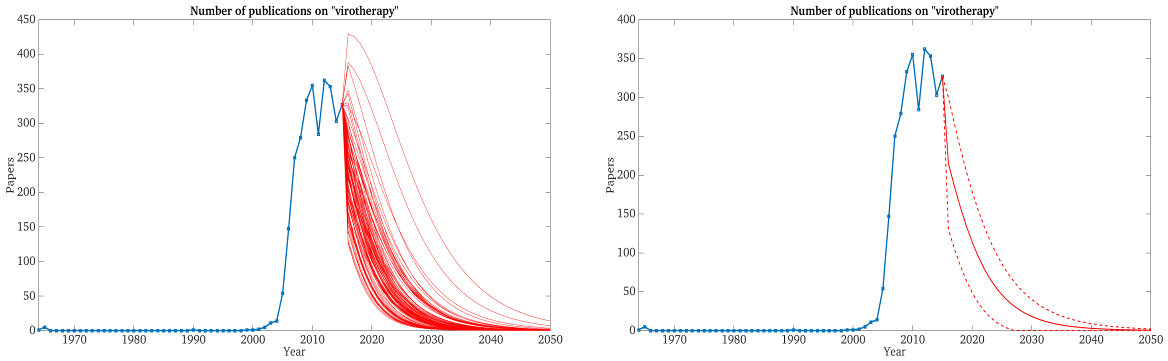

Figure 48. Prediction for "virotherapy" technologies.

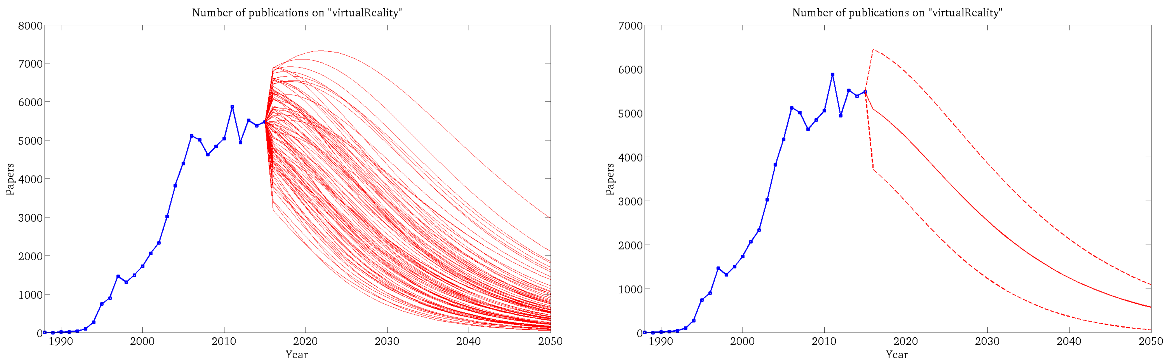

Figure 49. Prediction for "virtual reality".

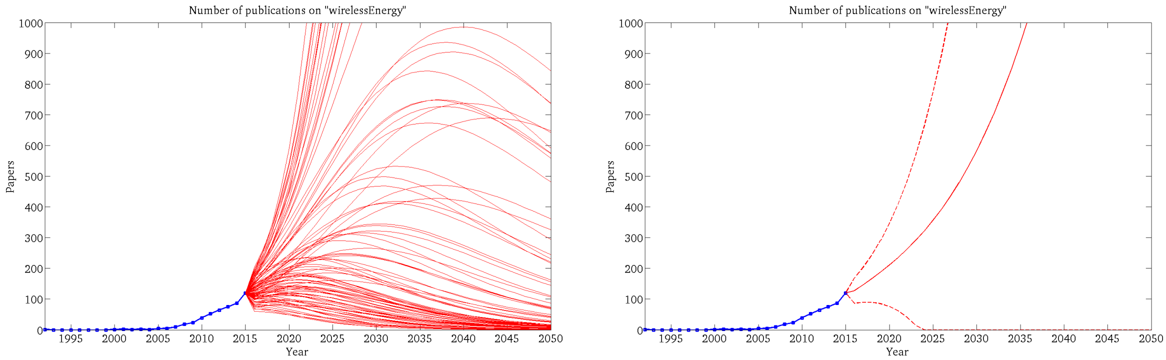

Figure 50. Prediction for "wireless energy" technologies.

<table>
<tr><td colspan="4">DEFENCE SCIENCE AND TECHNOLOGY GROUP
DOCUMENT CONTROL DATA</td><td colspan="2">1. DLM/CAVEAT (OF DOCUMENT)</td></tr>
<tr><td colspan="3">2. TITLE

Predicting Trends in Peer-Reviewed Publications</td><td colspan="3">3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED LIMITED RELEASE USE (U/L) NEXT TO DOCUMENT CLASSIFICATION)

Document (U/L)
Title (U)
Abstract (U)</td></tr>
<tr><td colspan="3">4. AUTHOR(S)

Dmitri Kamenetsky</td><td colspan="3">5. CORPORATE AUTHOR

Defence Science and Technology Group
DST Edinburgh
PO Box 1500
Edinburgh SA 5111</td></tr>
<tr><td>6a. DST GROUP NUMBER

DST-Group-TN-1881</td><td colspan="2">6b. AR NUMBER</td><td colspan="2">6c. TYPE OF REPORT

Technical Note</td><td>7. DOCUMENT DATE

June 2019</td></tr>
<tr><td colspan="2">8.TASK NUMBER

17/563</td><td colspan="2">9.TASK SPONSOR

N/A</td><td colspan="2">10. RESEARCH DIVISION

Joint and Operations Analysis Division</td></tr>
<tr><td colspan="3">11. MSTC

Strategy and Joint Force</td><td colspan="3">12. STC

Concepts and Futures</td></tr>
<tr><td colspan="6">15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111</td></tr>
<tr><td colspan="6">16. DELIBERATE ANNOUNCEMENT

Australian Department of Defence and Defence Force Personnel.</td></tr>
<tr><td colspan="6">17. CITATION IN OTHER DOCUMENTS

Yes</td></tr>
<tr><td colspan="6">18. RESEARCH LIBRARY THESAURUS

Technology; Futures Analysis; Data Analysis; Literature Reviews</td></tr>
<tr><td colspan="6">19. ABSTRACT

This work investigated a method for predicting future trends in peer-reviewed publication data related to science and technology. The developed method use non-linear regression to fit a particular type of s-shaped curve. The method’s prediction accuracy is measured on historical publication data of mature technologies. It was found that the prediction accuracy is acceptable for short-term (5-10 years) predictions, but declines for long-term (10+ years) predictions. The method was also used to predict the future publication trends of novel technologies. It was found that the publication rate of various upcoming technologies is expected to grow. In particular, CRISPR and deep learning are expected to grow rapidly. Some technologies however, are already at the peak of their trend and are expected to decline in the future.</td></tr>
</table>