International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 6, No. 3, 2024

The Application of Time Series Models in the Prediction of Hand-Foot-Mouth Disease Incidence

Nana Zhao [1], Wenming Lu [2], Zhongyuan Zhang [3, *]

[1] Shinan District Center for Disease Control and Prevention, Qingdao Shandong, 266071, China
[2] School of Computer Science and Technology, Qingdao University, Qingdao, Shandong, China
[3] Shibei District Health Development Center, Qingdao, Shandong, China

[*] **Corresponding author**: Zhongyuan Zhang (Email: 544731291@qq.com)

Abstract: Hand-Foot-Mouth Disease (HFMD) is a contagious illness predominantly affecting infants and children under five years old, caused by human enteroviruses. Over the past five decades, HFMD has rapidly spread across the Asia-Pacific region, gradually evolving into a significant public health challenge for many countries within this area. Currently, HFMD has emerged as an increasingly severe public health issue in our country. Therefore, analyzing the influencing factors of HFMD and predicting its incidence trends are of paramount importance for the prevention of the disease. With the rapid advancement of artificial intelligence technology, predictive models employing deep learning techniques have demonstrated superior performance among various infectious disease prediction models. This paper aims to construct a predictive model using deep learning methods to further enhance the accuracy of HFMD incidence predictions. We compared the effectiveness of Long Short-Term Memory (LSTM) networks, Transformer, and Informer models in HFMD prediction. The research findings indicate that the Informer model, by utilizing self-attention mechanisms and convolutional neural networks, can more effectively address long-term dependencies in time series data, thereby showing better performance in HFMD prediction compared to the LSTM and Transformer models. This has led to improvements in prediction accuracy and generalization capability.

Keywords: Hand-Foot-Mouth Disease; LSTM; Transformer; Informer.

1. Introduction

Hand-Foot-Mouth Disease (HFMD) is a common disease in children worldwide, caused by various enteroviruses, especially Coxsackievirus A16 and Enterovirus 71 (EV-71). While the disease is generally mild in most cases, it can lead to severe complications such as encephalitis and myocarditis, and can even be fatal. Understanding the epidemiological patterns of HFMD and establishing robust early warning mechanisms to scientifically predict its incidence trends are essential for effective prevention and control.

In recent years, with the rapid development of artificial intelligence and machine learning technologies, more and more studies have begun to explore using these advanced technologies for disease trend prediction. Traditionally, AutoRegressive Integrated Moving Average (ARIMA) or Seasonal AutoRegressive Integrated Moving Average (SARIMA) models are used to predict future incidence. However, ARIMA and SARIMA models typically assume that time series data are linear or stationary, which may not perform well with non-linear, non-stationary, or complex seasonal data.

Long Short-Term Memory (LSTM) networks, a special type of Recurrent Neural Network (RNN), can better capture long-term dependencies and non-linear characteristics in time series data, providing a new solution for disease prediction. In addition to these models, recent research has explored more efficient deep learning models such as Transformer and Informer, which have shown significant advantages in processing large-scale sequence data. Specifically, the Informer model, with its unique sparse transformer mechanism, has demonstrated high efficiency and accuracy in long-sequence time prediction tasks.

Therefore, this paper explores and compares the effectiveness of LSTM, Transformer, and Informer models in predicting HFMD incidence, providing a reference for the prevention and control.

2. Data

2.1. Data Source

The data used in this study is sourced from the HFMD incidence records in Shinan District, Qingdao, Shandong Province, covering the period from January 2014 to December 2023. Qingdao, a major coastal city in Shandong Province, is located in eastern China along the Yellow Sea and has a permanent population of approximately 10.342 million.

2.2. Statistical Analysis

According to statistics, from 2014 to 2023, a total of 4352 HFMD cases were reported in the Shinan District, with a male-to-female incidence ratio of 1.50. The incidence age mainly concentrated between 0 to 6 years (87.5%), with scattered children being the majority (44.92%).

To provide a clearer visualization of the changes in HFMD case numbers, the monthly incidence in the Shinan District was statistically analyzed. The figure shows that from 2014 to 2019, the number of HFMD cases in Shinan District, Qingdao City, generally tended to be stable, with no significant upward or downward trend, but there was a clear seasonal fluctuation cycle, with the peak incidence period occurring in June to August each year.

However, from April 2020 to October 2022, the data exhibited relatively minor fluctuations, with monthly peaks significantly lower than those of previous years. Additionally, the increase in cases during this period followed a more regular pattern, lacking the large variations seen in earlier years. This consistent trend is likely attributable to the

COVID-19 pandemic and may be related to vaccination, increased natural immunity, social distancing measures, or other public health interventions.

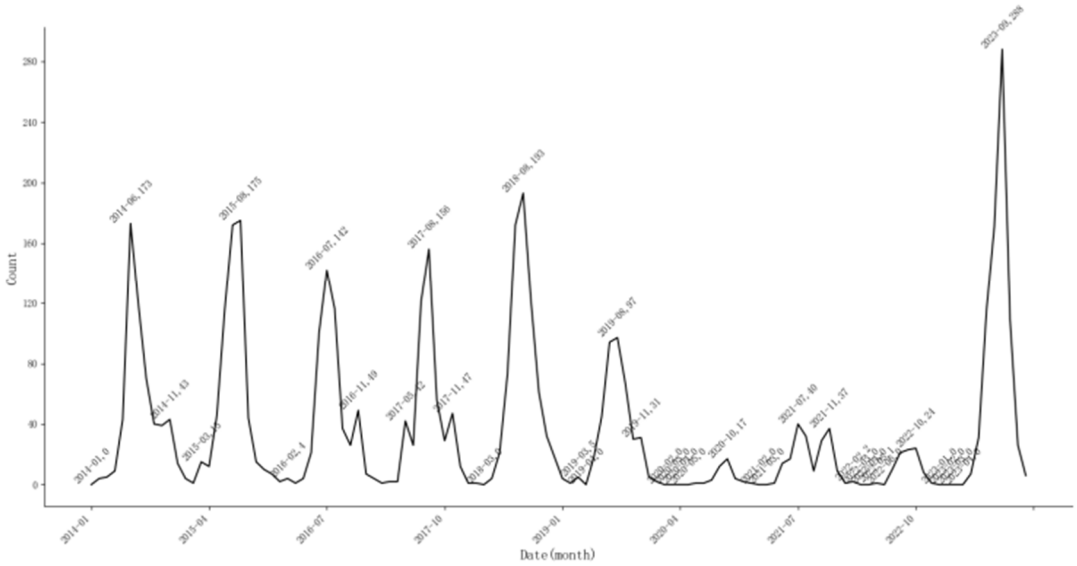

Figure 1. Monthly case counts in Qingdao City from 2014 to 2023

The models are established based on historical data and the assumption of normal epidemic trends. However, the abnormal conditions caused by COVID-19 may render these assumptions invalid. Consequently, during this period, models originally used to predict the number of HFMD cases might be disrupted by these anomalies, leading to a decrease in predictive accuracy.

3. Model and Evaluation

This chapter introduces how to convert the prediction problem into a supervised learning problem, the principles of LSTM, Transformer, and Informer models, and the multi-step prediction method of the models.

3.1. Problem Conversion

In the analysis of the time series data for HFMD, this paper adopts the sliding window technique, a commonly used method in time series prediction, particularly suitable for supervised learning frameworks. The weekly HFMD incidence forms a series of data points over time, denoted as $s = [s[0], s[1], \ldots, s[T]]$, where T represents the total length of the observation period. Under this framework, the sliding window method captures local patterns in the time series by continuously moving the window, each containing data from n_t consecutive time points.

For any given time point t, data is extracted from point $t - n_t + 1$ to point t, constructing an input vector $x_t = [s[t - n_t + 1], s[t - n_t + 2], \ldots, s[t]]$. This vector reflects the data from the past n_t time points and is used to predict the incidence of HFMD for the next n_0 time points. The predicted output vector is defined as $y_t = [s[t + 1], s[t + 2], \ldots, s[t + n_0]]$, which includes the incidence from time point $t + 1$ to $t + n_0$.

Through the sliding window technique, the time series forecasting issue is adeptly converted into a supervised learning framework. Within this framework, the sequence of historical data points is harnessed as the feature vector, with the sequence of prospective data points designated as the target vector. This methodology empowers the model to discern both short-term and long-term dependencies within the time series, thereby enhancing the precision of predictive outcomes.

3.2. LSTM Model

LSTM is a special type of RNN, proposed by Hochreiter et al. It uses memory cells to replace the neural units in the RNN hidden layer, effectively solving the gradient vanishing problem in RNN. Each cell contains three gate states: input gate, output gate, and forget gate. In each cell, the input gate selectively records new information into the current cell state, while the forget gate decides which information to discard. Then, the output gate partially outputs the processed information to other cells. The specific structure of LSTM is shown in the figure.

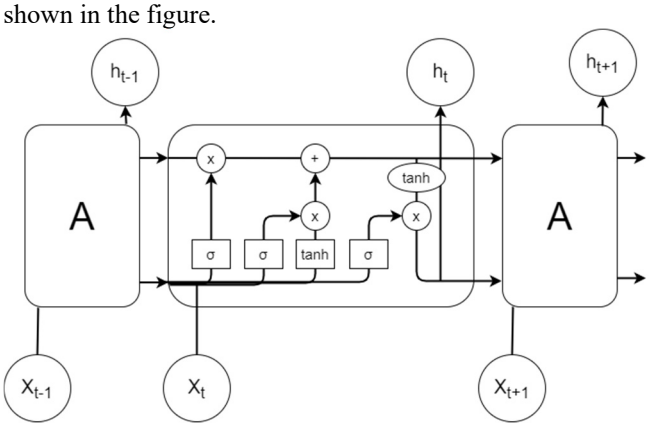

Figure 2. Architecture of LSTM.

LSTM network has three types of inputs: the state of the neural unit at the previous time step, the input content of the neural unit at the previous time step, and the input content of the neural unit at the current time step. The operational process of the LSTM network is as follows:

The first step involves using the forget gate to calculate the historical information that needs to be forgotten, which is crucial for adapting to the seasonal variations in HFMD data, ensuring that the model can ignore past information that is no longer relevant. The specific calculation is shown in Equation (1).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

The second step calculates the information that needs to be preserved within the cell state and updates this information, primarily through control of the input gate. This allows the model to selectively update its internal memory, which is sensitive to changes in HFMD incidence trends. The specific calculations are shown in Equations (2) and (3). Here, i_t represents the content that needs to be preserved in the cell state, while $\widetilde{C}_t$ represents the forgetting weight of the information.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\widetilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

The third step updates the cell state to include information for predicting future incidence rates. The computation is shown in Equation (4).

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \quad (4)$$

Finally, the model uses the updated cell state, combined with control from the output gate, to generate predictions for future HFMD incidence rates. The output results are compressed using the tanh function and multiplied with the output from the sigmoid gate, as outlined in Equations (5) and (6):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

where $W_f, W_i, W_c, and\ W_o$ represent the weight matrices for the forget gate, input gate, memory cell, and output gate, respectively; $b_f, b_i, b_c, and\ b_o$ denote the biases for the forget gate, input gate, memory cell, and output gate, respectively.

3.3. Transformer Model

In forecasting HFMD incidence, conventional time series analytical techniques might prove inadequate for capturing the intricate dynamics inherent in the progression of the disease. To address this challenge, the advent of Transformer models introduces an innovative methodological paradigm, significantly augmenting capacity to adeptly manage and forecast HFMD time series data.

In the prediction of HFMD incidence rates using Transformer models, appropriate preprocessing of the disease incidence data is essential before application. This involves converting raw incidence rate data into a format that the model can accept, typically in the form of time series. The Transformer model is particularly suitable for handling multivariate time series data, which means besides case numbers, it can incorporate external factors such as temperature and humidity that may influence disease transmission. By integrating these multidimensional inputs, the model can provide more comprehensive and accurate forecasts.

The Transformer model effectively captures long-term dependencies through its self-attention mechanism, which is crucial for understanding seasonal and interannual variations in HFMD incidence trends. Unlike the sequential processing approach of traditional RNNs, the self-attention mechanism allows the model to directly establish connections between any two points, thereby comprehensively capturing correlations in the time series.

Positional encoding is another important feature of the Transformer model, addressing the lack of sequential awareness when processing sequence data. By directly embedding positional information into the input data, the model can recognize the temporal relationships between different time points in the time series, which is crucial for predicting the incidence of diseases like HFMD that exhibit pronounced seasonal characteristics.

Utilizing an encoder-decoder structure, the Transformer model first processes input time series data through the encoder to capture dependencies between different time points. Subsequently, the decoder uses the encoder's output and previous predictions to gradually generate forecasts of future case numbers. This method of learning from past and present data to predict the future provides a powerful tool for analyzing HFMD incidence trends.

In predicting HFMD incidence rates, the self-attention mechanism allows the model to consider not only the absolute positions of each data point in the time series (e.g., months in a year) but also assesses the relative importance of these data points relative to each other. This relative importance is achieved through attention scores computed between queries (Q), keys (K), and values (V).

With data processed using the self-attention mechanism, the Transformer model can more accurately predict future HFMD incidence rates. This predictive method, weighted by the importance derived from historical data, demonstrates higher efficiency and accuracy in handling complex time series data compared to traditional time series analysis techniques. This capability is particularly crucial for diseases with distinct seasonal patterns and influenced by multiple environmental factors, highlighting the importance of the self-attention mechanism.

Multi-head self-attention mechanisms are particularly advantageous for capturing complex relationships and patterns from multidimensional time series data. By parallelly processing multiple self-attention processes, this mechanism allows the model to simultaneously learn multiple features and dependencies of the data in different representation subspaces.

Leveraging multi-head self-attention mechanisms enables the Transformer model to comprehensively understand and utilize the intricate patterns embedded in historical data when predicting HFMD incidence rates. Such capability is crucial for forecasting disease development trends, especially when disease transmission is influenced by multiple factors. For example, the model may identify increases in incidence rates under specific climatic conditions or changes in incidence rates following public health interventions such as vaccination campaigns.

3.4. Informer Model

The Informer model consists of two primary components: an encoder and a decoder. Each component includes position information embedding layers, multi-head probabilistic sparse self-attention layers, and convolutional distillation layers. The operational workflow is structured as follows: initially, the model encodes the input sequence by incorporating temporal information. It then utilizes a sophisticated neural network comprising multi-head probabilistic sparse self-attention layers and convolutional layers to extract deep features. Lastly, these features are aggregated through fully connected layers to generate predictions of case counts.

The Informer model is an enhanced variant based on the Transformer, using a probabilistically sparse self-attention mechanism that significantly boosts processing efficiency. This mechanism not only optimizes computational time complexity but also facilitates a deeper focus on key temporal

points, thereby enhancing the model's temporal sensitivity. Furthermore, Informer's generative decoding strategy further accelerates the prediction speed. In terms of its core architecture, Informer maintains a framework similar to Transformer's, comprising both an encoder and a decoder. Within this framework, the encoder transforms the input sequence $X = (x_1, x_2, \ldots, x_n)$ into an intermediate continuous sequence $Z = (z_1, z_2, \ldots, z_n)$. Subsequently, the decoder leverages this sequence to generate the final predicted outcome $Y = (y_1, y_2, \ldots, y_n)$ in a single step.

The Informer model introduces an improved probabilistic sparse self-attention module. The attention given by the $i-th$ query to all keys is defined as the probability $p(k_j|q_i)$, with the output being a combination of this probability and the corresponding value v. The primary dot-product pairs cause the attention probability distribution for a matching query to deviate from a uniform distribution. If $p(k_j|q_i)$ equals the uniform distribution:

$$q(k_j \mid q_i) = \frac{1}{L_k} \tag{7}$$

The self-attention then reduces to a trivial sum over the values v, which is superfluous for existing inputs. At this point, the similarity between distributions p and q can be utilized to distinguish 'important' queries. Ignoring constants, the $i-th$ sparsity measure is defined as follows:

$$M(q_i, K) = \ln \sum_{j=1}^{L_k} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_k} \sum_{j=1}^{L_k} \frac{q_i k_j^T}{\sqrt{d}} \tag{8}$$

where the first term is the Logarithmic Sum Exponent (LSE) of q_i across all keys, while the second term denotes their arithmetic mean. Consequently, the probabilistically sparse self-attention mechanism is represented as:

$$A(Q, K, V) = Soft\max(\frac{\overline{Q}K^T}{\sqrt{d}})V \tag{9}$$

Finally, by sampling the dot-product pairs according to $L_k \ln L_k$, the computational complexity of attention is reduced to $O(L_k \ln L_k)$ per layer.

To address the potential redundancy issue in feature mappings of the encoder introduced by the multi-head probabilistic sparse self-attention mechanism, the Informer model incorporates a distillation mechanism. This mechanism, by assigning higher weights to critical data features, not only effectively optimizes the quality of information transfer but also facilitates the extraction of more salient features in subsequent layers. Specifically, the distillation process employs one-dimensional convolutions and max-pooling techniques to downsample along the time dimension of the input data, thereby reducing both the temporal dimension and the lengths of the Q and K sequences. This approach not only reduces the model's computational complexity but also enhances its capability to capture key information. The computation formula from layer j to $j+1$ is as follows:

$$X_{j+1}^t = MaxPool(ELU(Convld([X_j^t]_{AB}))) \tag{10}$$

where $[X_j^t]_{AB}$ encompasses fundamental operations within the multi-head sparse self-attention and attention blocks. Conv1d represents a one-dimensional convolution operation applied along the time series, followed by activation through the ELU function. Subsequently, a max-pooling layer is employed, which halves the length of the input sequence at each decoder layer, thereby drastically reducing memory overhead and computation time for the encoder.

In the Informer model, the input sequence to the decoder consists of two parts: one is the sequence to be predicted, and the other is a segment of known sequence preceding the time point to be predicted. In this context, where the prediction task involves forecasting the incidence of hand-foot-and-mouth disease for the next 7 days, the cases from the 7 days prior to the prediction time point constitute the sequence to be predicted. Additionally, the number of cases from the previous week also forms part of the input to the decoder. The decoder input is thus represented by the following vector:

$$X_{feed\_de}^t = Concat(X_{token}^t, X_0^t) \in R^{(L_{token}+L_y)\times d_{model}} \tag{11}$$

where X_{token} is the start vector, X_0^t serves as a placeholder, L_{token} represents the length of the historical known sequence, and L_y denotes the length of the sequence to be predicted. After passing through the decoder, each position to be predicted yields a vector, which is then fed into a fully connected layer to obtain the predicted case count results.

3.5. Evaluation Criteria

Two metrics were employed in this paper to assess prediction accuracy: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Smaller values of MAE and RMSE indicate better predictive performance of the model. The formulas for MAE and RMSE calculations are provided below:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |(y_i - \hat{y}_i)| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{13}$$

where y_i represents the actual values, $\hat{y}_i$ denotes the predicted values, and n is the number of samples.

4. Experimental Results and Analysis

During the experimental process, LSTM, Transformer, and Informer models were implemented using the Python and PyTorch frameworks, leveraging GPU acceleration. Table 1 presents the comparative 2023 prediction results of these different models.

Table 1. Comparative 2023 prediction results of different models

Model	MAE	RMSE
LSTM	1.13	2.07
Transformer	1.15	2.25
Informer	1.08	1.99

From the table data, it is evident that the Informer model demonstrates superior performance in predicting hand, foot, and mouth disease incidence rates, with errors lower than those of the LSTM and Transformer models.

HFMD, characterized by distinct seasonal patterns, exhibits significant variations in prevalence over different time periods. Therefore, comparisons were made between actual values during peak incidence periods of the disease and predictions from each model. Figure 3 provides a visual comparison between the predicted values of each model and the actual values, while Figure 4 details the distribution of errors between model predictions and actual values.

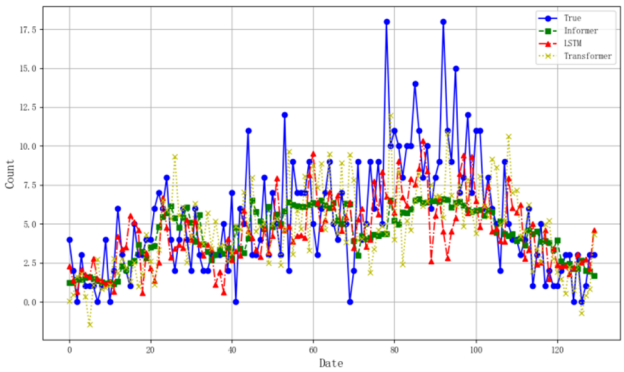

Figure 3. Comparison of model predictions.

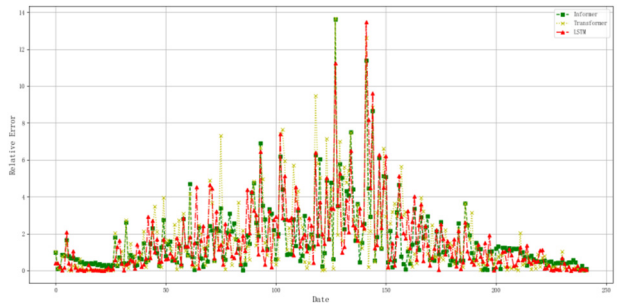

Figure 4. Comparison of relative errors

From the figures, it is evident that in the early stages, the LSTM model exhibits strong tracking ability with the actual data, closely aligning its predicted curve with the actual data curve. However, during peak disease periods, the LSTM model's predictions significantly underestimate the actual incidence rates, suggesting potential limitations in capturing dynamic disease developments.

The Transformer model demonstrates good tracking of data variations in certain periods but exhibits noticeable prediction deviations in others. This uneven performance implies potential weaknesses in stability and generalization, which may affect its long-term predictive accuracy.

In contrast to the LSTM and Transformer models, the Informer model's predicted curve shows a smoother trend, particularly during peak periods of hand, foot, and mouth disease. It avoids the low prediction values seen in the other two models. This stable trend suggests better stability and robustness of the Informer model. While this conservative prediction strategy might indicate a tendency to be overly cautious during peak periods, it also mitigates large fluctuations in predicted values, potentially making it more suitable for public health strategies requiring robustness.

5. Conclusion

This paper delves into the prediction of hand, foot, and mouth disease (HFMD) case numbers, comprehensively discussing the performance of three deep learning models in forecasting HFMD incidence rates. Through thorough experimentation and comparative analysis, the study reveals that while the LSTM model exhibits the smallest error in predicting the 2023 incidence rates, all models have certain limitations. HFMD incidence rates are influenced by various factors such as climatic conditions, public health policies, and population mobility, which were not fully considered in the dataset.

Future work could enhance model generalization by integrating more relevant factors to improve prediction accuracy. Additionally, models need to adapt to the impacts of sudden public health events like the COVID-19 pandemic on HFMD incidence rates to achieve more precise forecasts. Overall, LSTM, Transformer, and Informer models demonstrate unique strengths and potential in predicting disease incidence rates, warranting further exploration and optimization in future research.

Given that HFMD is susceptible to factors like temperature, humidity, and social conditions, there are still many areas for improvement in this study. Changes in prevention and control policies (such as lockdown due to the COVID-19 pandemic in 2020 and full reopening announced by the end of 2022) altered the disease trends in 2023, contributing to prediction errors. A deeper analysis of these factors and their incorporation into the models may enhance predictive effectiveness.

References

[1] G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[3] H. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1985, ch. 4.

[4] B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.

[5] E. H. Miller, “A note on reflector arrays (Periodical style-Accepted for publication),” IEEE Trans. Antennas Propagat., to be published.

[6] J. Wang, “Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),” IEEE J. Quantum Electron., submitted for publication.

[7] C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.

[8] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces (Translation Journals style),” IEEE Transl. J. Magn. Jpn., vol. 2, Aug. 1987, pp. 740–741 [Dig. 9th Annu. Conf. Magnetics Japan, 1982, p. 301].

[9] M. Young, The Techincal Writers Handbook. Mill Valley, CA: University Science, 1989.

[10] J. U. Duncombe, “Infrared navigation—Part I: An assessment of feasibility (Periodical style),” IEEE Trans. Electron Devices, vol. ED-11, pp. 34–39, Jan. 1959.

[11] S. Chen, B. Mulgrew, and P. M. Grant, “A clustering technique for digital communications channel equalization using radial basis function networks,” IEEE Trans. Neural Networks, vol. 4, pp. 570–578, Jul. 1993.

[12] R. W. Lucky, “Automatic equalization for digital communication,” Bell Syst. Tech. J., vol. 44, no. 4, pp. 547–588, Apr. 1965.

[13] S. P. Bingulac, “On the compatibility of adaptive controllers (Published Conference Proceedings style),” in Proc. 4th Annu. Allerton Conf. Circuits and Systems Theory, New York, 1994, pp. 8–16.

[14] G. R. Faulhaber, “Design of service systems with priority reservation,” in Conf. Rec. 1995 IEEE Int. Conf. Communications, pp. 3–8.

[15] W. D. Doyle, “Magnetization reversal in films with biaxial anisotropy,” in 1987 Proc. INTERMAG Conf., pp. 2.2-1–2.2-6.

[16] G. W. Juette and L. E. Zeffanella, “Radio noise currents n short sections on bundle conductors (Presented Conference Paper style),” presented at the IEEE Summer power Meeting, Dallas, TX, Jun. 22–27, 1990, Paper 90 SM 690-0 PWRS.

[17] J. G. Kreifeldt, “An analysis of surface-detected EMG as an amplitude-modulated noise,” presented at the 1989 Int. Conf. Medicine and Biological Engineering, Chicago, IL.

[18] J. Williams, “Narrow-band analyzer (Thesis or Dissertation style),” Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, 1993.

[19] N. Kawasaki, “Parametric study of thermal and chemical nonequilibrium nozzle flow,” M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993.

[20] J. P. Wilkinson, “Nonlinear resonant circuit devices (Patent style),” U.S. Patent 3 624 12, July 16, 1990.

[21] IEEE Criteria for Class IE Electric Systems (Standards style), IEEE Standard 308, 1969.

[22] Letter Symbols for Quantities, ANSI Standard Y10.5-1968.

[23] R. E. Haskell and C. T. Case, “Transient signal propagation in lossless isotropic plasmas (Report style),” USAF Cambridge Res. Lab., Cambridge, MA Rep. ARCRL-66-234 (II), 1994, vol. 2.

[24] E. E. Reber, R. L. Michell, and C. J. Carter, “Oxygen absorption in the Earth’s atmosphere,” Aerospace Corp., Los Angeles, CA, Tech. Rep. TR-0200 (420-46)-3, Nov. 1988.

[25] R. J. Vidmar. (1992, August). On the use of atmospheric plasmas as electromagnetic reflectors. IEEE Trans. Plasma Sci. [Online]. 21(3). pp. 876–880. Available: http://www. halcyon. com/pub/journals/21ps03-vidmar.