Forecasting the View of Mt. Fuji Using Earth Observation Data

Mitsuru KAKIMOTO\(^{(a)}\), Hisaaki HATANO\(^{(b)}\), Members, and Yosoko NISHIZAWA\(^{(c)}\), Nonmember

SUMMARY In this paper, we present a forecasting method for the view of Mt. Fuji as an application of Earth observation data (EOD) obtained by satellites. We defined the Mt. Fuji viewing index (FVI) that characterises how well the mountain looks on a given day, based on photo data produced by a fixed-point observation. A long-term predictor of FVI, ranging from 0 to 30 days, was constructed through support vector machine regression on climate and earth observation data. It was found that the aerosol mass concentration (AMC) improves prediction performance, and such performance is particularly significant in the long-term range.

key words: earth observation data, support vector machine regression, singular value decomposition

1. Introduction

The Earth Observing System (EOS), which includes a series of Earth-observing satellites, produces a huge amount of remote sensory data every day. These data have unique characteristics in that they have wide coverage of the Earth’s surface and are collected continuously through a long period of time. Therefore, the data have the potential to benefit many aspects of human activities. These satellites and the sensors loaded on them usually have specific predetermined missions. There is no reason, however, to limit the range of applications of the data obtained. It is desirable to extend their application field in order to rationalize the cost of maintaining the system.

In this paper, we present a forecasting method for the view of Mt. Fuji as an application of Earth observation data (EOD). Mt. Fuji is a major sightseeing destination in Japan. Knowing in advance how well it will look on a given day is of critical importance, for example, in planning a travel itinerary.

It is obvious that the view of Mt. Fuji is strongly influenced by climate conditions, such as temperature and the amount of cloud. Therefore, in attempting to construct a forecasting method, it is natural to first consider the use of climate data (CD). It is unclear, however, how the climate will influence the view in the long-term range. Physical quantities obtained by EOS may influence the view in a manner different from that of CD. For example, it is known that the light scattering property of aerosol in the air strongly influences visibility [1]. If aerosol remains in the region of interest over the long term, we may take advantage of that fact. Thus, we expect the forecast will make good use of EOD.

We built a framework for forecasting the view of Mt. Fuji based on a regression using predictors both from CD and EOD. Generally, when building a regression model, the choice of predictors is of critical importance. Although use of multiple predictors makes it possible to take advantage of their synergy, too many of them may harm the prediction reliability owing to overfitting. Predictors were carefully chosen from among those quantities available in CD and EOD. A support vector machine with radial basis function (RBF) was employed as a prediction model [2]. Non-linearity of the model was expected to enhance prediction performance. Long-range forecasts, up to 30 days, were performed based on this model. Points of interest addressed are as follows:

- How does the performance depend on a choice of predictors?
- How does the forecast behave in the long range?
- How high is the forecast performance?
- Is a non-linear regression model suitable for the forecast?

The remainder of this paper is organised as follows. In Sect. 2, the forecasting framework is described. First, how we measure the view of Mt. Fuji is explained. Then the organization of the CD and EOD used is described. Finally, we show the method of forecasting and how performance is evaluated. Section 3 first gives performances of forecasts using each of the quantities available from CD and EOD. How we handled each multidimensional quantity in EOD is also described. Then the building process of the forecasting model with multiple predictors is detailed. Section 4 evaluates the performance of the forecasts and discusses the significance of predictor selection and prediction models. Section 5 gives our conclusions.

2. Method

2.1 Characterization of the View of Mt. Fuji

The first step toward forecasting the view of Mt. Fuji is to numerically characterize how well it looks on a given day. We defined the Mt. Fuji viewing index (FVI) based on photo
data produced by a fixed-point observation. The fixed-point observation we used has following properties.

- The camera is placed at Fuji city hall, which is located at 20 km south of Mt. Fuji.
- A photo of Mt. Fuji is taken every hour, every day from 9AM to 5PM.
- Photos have been taken since 1999.

The photos are publicly available on the Web [3]. We used those photos taken from July 2002 to Dec. 2004.

For each day during the period, we selected the photo that provides the best view for that day. Then these best photos are again compared with one another, and ranked into the following four levels.

- 3: Mt. Fuji is clearly visible.
- 2: Mt. Fuji is visible.
- 1: Mt. Fuji is barely visible.
- 0: Mt. Fuji is invisible.

We refer to these values as FVI. Table 1 lists the numbers of days that fall into each class. Note that the distribution is well balanced, which is preferable in a regression model construction step.

### Table 1: Numbers of days at each level of FVI.

| FVI | # of days |
|-----|-----------|
| 0   | 266       |
| 1   | 156       |
| 2   | 264       |
| 3   | 259       |
| total | 945       |

2.2 Climate Data (CD)

As described in the following sections, our forecasting method is based on a regression scheme. The first set of predictors for the regression scheme is chosen from CD quantities. Diverse quantities are available daily from observation performed at the meteorological observation stations in Shizuoka city[3]. These include average pressure, average/maximum/minimum temperature, average water vapor pressure, average/minimum humidity, maximum/average wind speed, wind direction, average cloudiness, duration of sunshine, and amount of precipitation. We wanted to select those quantities likely to contribute to the forecast. This was done by a preliminary selection algorithm described below.

- Step 1
  $x_i \leftarrow$ each one of the observed quantities listed above.
  $y \leftarrow$ FVI.
- Step 2
  Correlation $\gamma_i$ between $x_i$ and $y$ is calculated.
  \[
  \gamma_i = \frac{\sum_{t=1}^{N} (x_i(t) - \bar{x})(y(t) - \bar{y})}{\sqrt{\sum_{t=1}^{N} (x_i(t) - \bar{x})^2 \sum_{t=1}^{N} (y(t) - \bar{y})^2}}
  \]
  $\bar{x} = \frac{1}{N} \sum_{t=1}^{N} x_i(t)$
  $\bar{y} = \frac{1}{N} \sum_{t=1}^{N} y(t)$

Some of the physical quantities obtained as a result of the remote monitoring of the atmosphere influence visibility. For instance, it is known that visibility degradation results from light scattering and absorption by atmospheric particles and gases [1]. Although a number of EOD quantities are available, incorporating all of them into the forecasting scheme would have required a considerable amount of work, including downloading a number of large data files, analyzing their formats and programming to extract the necessary information. Since it would have been prohibitive to try many of them, we selected quantities likely to contribute to the forecasting, based on our knowledge of the physical process of visibility.

### Footnote

[3] Shizuoka city is 50 km southwest of Mt. Fuji. Its meteorological observation station is the closest to Mt. Fuji.
Aerosol Mass Concentration (AMC)
AMC describes the mass concentration of aerosol in the atmosphere. The aerosol is particles in the atmosphere that have nearly the same size as the wavelength of visible light. The aerosol therefore strongly influences the visibility [1].

Water Vapor Near Infrared (WVNI)
WVNI describes the amount of water vapor in the atmosphere measured using the near-infrared band.

Water Vapor Infrared (WVI)
WVI denotes the amount of water vapor in the atmosphere measured using the infrared band.

Aerosol Optical Depth (AOD)
AOD characterizes the entire attenuation of solar radiation in a vertical air column due to the presence of aerosol particles [5].

Rain Rate (RR)
RR is a satellite-based estimate of rain rate relying on cloud temperatures and information about vertical profiles [6].

The sensors, satellites, product names and providers of these quantities are summarized in Table 2. We obtained these data from NASA's archive and distribution system. It should be noted that MODIS consists of two separate instruments mounted on two satellites, AQUA and TERRA. Although data were provided in distinct products, we combined them and treated them as a single quantity.

Data files provided by NASA cover Japan and a wide area surrounding it. We trimmed the files to fit into the area including Mt. Fuji and its vicinity designated in Fig. 1. Then the area was divided into nine blocks as is shown in the figure. Each block has boundaries exactly coinciding with constant longitude (latitude) lines with integral values. For each quantity in EOD, the values inside each block are averaged to make the quantity nine-dimensional. This number of blocks is chosen so as to maintain EOD's spatial nature without going into too much detail. Although a finer grid could be used, it would increase the computational cost.

2.4 Prediction Method

Our forecast is based on a regression method formalized as

\[
FVI(t + \tau) = f(p(t)) = f(p_1(t), p_2(t), \ldots, p_n(t)). \tag{1}
\]

In this scheme, the target variable is FVI at a given day in the future. \(t\) refers to the day when the forecast is made. \(\tau\) specifies the time between the day the forecast is made and the day for which FVI is forecasted. This \(\tau\) is referred to as TLP (Time Lag of Prediction) and varies from 0 to 30 days. It should be noted that TLP = 0 means predicting FVI on a given day based on the physical quantities of the same day, which is not a forecast in a strict sense.

\[ p = (p_1, p_2, \ldots, p_n) \]

is a set of predictors in the regression scheme. These quantities are chosen from among CD and EOD as mentioned above. \(n\) refers to the number of predictors. The main task in construction of the regression scheme is selection of predictors.

\(f\) in Eq. (1) denotes functional form of the prediction method. It is common practice to use a linear equation such as

\[
f(p) = a \cdot p + b = a_1 p_1 + \cdots + a_n p_n + b. \tag{2}
\]

Although this type of model is easy to handle, its predictive ability is sometimes limited. We therefore used a function with the radial basis function (RBF) kernel instead.

\[
f(p) = \sum_v a_v \exp \left( -\gamma ||p - p_v||^2 \right) + b. \tag{3}
\]

We chose the RBF kernel since it has some attractive properties, such as clarity of meaning and parsimony of parameters. Although some other types of kernel could be

Table 2: Predictors

| Quantity | Sensor | Satellite | Product Name | Provider |
|----------|--------|-----------|--------------|----------|
| AMC      | MODIS  | AQUA      | MYD04_L2     | GSFC     |
| WVNI     | MODIS  | AQUA      | MYD04_L2     | GSFC     |
| WVI      | MODIS  | AQUA      | MYD05_L2     | GSFC     |
| AOD      | MODIS  | AQUA      | MYD04_L2     | GSFC     |
| RR       | AMSR-E | AQUA      | AE_Rain      | NSIDC    |

MODIS: Moderate Resolution Imaging Spectroradiometer
AMSR-E: Advanced Microwave Scanning Radiometer for EOS
GSFC: Goddard Space Flight Center (NASA)
NSIDC: National Snow and Ice Data Center (NASA)
used, they would require much more care in parameter tuning. The regression equation contains some parameters that should be adjusted to fit the observed data. The RBF kernel is non-linear in nature, which makes it difficult to optimize. The support vector machine, with its margin maximization principle, allows optimization of the regression equation without overfitting [2]. We will discuss the benefit of non-linearity compared with linearity in Sect. 4. The support vector machine with the RBF kernel has two extra parameters, namely, violation cost parameter and size of kernel (γ in Eq. (3)). These were optimized using the cross-validation methods. The algorithm was implemented using a library, LIBSVM [7].

2.5 Prediction Performance Measurement

A prerequisite for building a good prediction model is the ability to measure the accuracy of the prediction. We measure the prediction performance in terms of covariance between observed \( FVI \) and predicted \( FVI \), which is referred to as CFP (covariance between observed \( FVI \) and its predicted value). We estimate CFP using the \( K \)-fold cross-validation method, which proceeds as follows. We have \( N \) days of \((\text{predictors, } FVI)\) pairs from our data set, which is denoted as

\[
((p(t), v(t + \tau)) \mid \tau = 1, \cdots, N].
\]

These are separated into \( K \) partitions \( P_i \) \((i = 1, \cdots, K)\), each containing \( N/K \) pairs.

\[
P_i = \{(p(t), v(t + \tau)) \mid \tau \in T_i\}.
\]

\( T_i \) here denotes a set of days on which observations were made corresponding to \( P_i \). We further divide these partitions into two sets. One is referred to as a learning set, which consists of \( K-1 \) partitions \( P_1, \cdots, P_{K-1} \).

\[
\bigcup_{i=1}^{K-1} P_i = \{(p(t'), v(t' + \tau)) \mid t' \in \bigcup_{i=1}^{K-1} T_i\}.
\]

We first build a prediction model using this learning set. The other consists of a single partition \( P_K \), and is referred to as a validation set.

\[
((p(t), v(t + \tau)) \mid \tau \in T_K).
\]

We apply the prediction model obtained above to the predictors in the validation set to obtain \( N/K \) predicted values of \( FVI \),

\[
\hat{v}(t) \mid t \in T_K.
\]

Then the role of partitions in the scheme above is cyclically changed. We have \( K \) ways of separating the original data into a learning set and a validation set. By iterating the above process for each of the ways of separation, we get \( N = N/K \times K \) predicted values,

\[
\hat{v}(t + \tau) \mid \tau = 1, \cdots, N.
\]

CFP is estimated as

\[
\text{CFP}(\tau) = \sum_{\tau=1}^{N} \{v(t + \tau) - \bar{v}\} \{\hat{v}(t + \tau) - \bar{v}\}.
\]

It should be clear that CFP captures the intuitive notion of goodness of prediction. Setting \( K = N \) makes the \( K \)-fold cross-validation the usual leave-one-out jack-knife method. We chose, however, \( K = 128 \) to save computational cost.

We get CFP(\( \tau \)) for each value \( \tau \) of TLP ranging from 0 to 30. As a measure of overall performance of a given prediction model, we use cumulative CFP (cCFP) defined as

\[
c\text{CFP} = \sum_{\tau=0}^{30} \text{CFP}(\tau).
\]

It should be noted that CFP and cCFP depend on the set of quantities selected as predictors. Let us denote a set of predictors as follows,

\[
(\text{quantity1, quantity2}, \cdots)
\]

For example, \((\text{Tmp, WVNI})\) denote prediction based on multiple regression with two predictors, i.e., Tmp and WVNI. To specify the set of predictors selected in calculating CFP (or cCFP), we denote it as

\[
\text{CFP}(p_1, p_2, \cdots).
\]

For example, cCFP (Tmp, AMC) refers to cCFP calculated using Tmp and AMC as predictors.

3. Result

First, we look at the performance of the prediction using each physical quantity from CD and EOD.

3.1 CD

Figure 2 shows CFP for TLP between 0 and 30 for each physical quantities from CD. Notice that Tmp is by far the best predictor through the entire range of TLP. Cld and Ssn have a short-term effect.
3.2 EOD

As described in the previous section, each EOD quantity is multidimensional. This implies that a considerable amount of noise may be contained in EOD. It is therefore necessary to extract from this multidimensional data a few components that are likely to contribute to the forecast. We used principle component regression (PCR), which we now describe. Each quantity in EOD can be represented by a matrix $X$ with row and column representing the spatial and the temporal dimension, respectively.

$$X = (x_{ij}) \in \mathbb{R}^{m \times n}. \tag{8}$$

Here $m$ denotes the number of spatial components (9 in our case) and $n$ denotes the number of temporal components. We applied singular value decomposition (SVD) to $X$, which describes $X$ as a product of three matrices.

$$X = USV^T \tag{9}$$

$$U \in \mathbb{R}^{m \times m}, S \in \mathbb{R}^{m \times n}, V \in \mathbb{R}^{n \times n}.$$ 

SVD represents a decomposition of the quantity into $\min(m, n)$ components orthogonal in space and time. We assume here that $m < n$, which holds for our case. The $i$-th column vector $u_i$ of $U$ represents a spatial form of the $i$-th component. The $i$-th column vector $v_i$ of $V$ represents a temporal variation of the $i$-th component. The vectors $u_i$ and $v_i$ are called left-singular and right-singular vectors, respectively. These vectors are orthogonal to each other.

$$u_i^T u_j = \delta_{i,j}, \quad v_i^T v_j = \delta_{i,j}. \tag{10}$$

$\delta_{i,j}$ denotes the Kronecker delta. The matrix $S$ has the following diagonal form.

$$S = \begin{pmatrix} s_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & s_m \end{pmatrix}. \tag{11}$$

The value $s_i$ denotes a weight of $i$-th component and is referred to as a singular value. So SVD expresses $X$ as a superposition of $m$ components as

$$X = u_1 s_1 v_1^T + \cdots + u_m s_m v_m^T. \tag{12}$$

These components are referred to as SVD components. It is possible to think, without loss of generality, that $s_i$s are all non-negative and ordered by decreasing values.

$$s_1 \geq s_2 \geq \cdots \geq s_{m-1} \geq s_m.$$ 

We denote $i$-th component as $P_i$. $P_1$ represents the largest variation in $X$, $P_2$ represents the second largest variation, and so on. It is therefore expected that characteristics of $X$ are represented with much fewer components than its actual dimension by selecting some components with the largest singular values.

SVD can be used in our prediction scheme as follows.

Since each right-singular vector $v_i$ represents a temporal variation of $P_i$, it can be used as a predictor in our regression scheme. For each quantity of EOD, we performed three types of regression based on the SVD components extracted as described above.

1. Type I: Single predictor (using only $P_1$).
2. Type II: Two predictors (using $P_1 P_2$).
3. Type III: Three predictors (using $P_1 P_2 P_3$).

Figure 3 shows CFPs for these three types of prediction using (AMC) as a predictor set. The figure shows that increasing the number of components does not necessarily improve prediction performance. Similar results were observed for other quantities as is summarized in Table 3. Since we wanted to avoid the trouble of handling too many predictors, we used the first SVD component only for each quantity in EOD in the latter part of the paper. The impact of multiple components will be discussed in Sect. 4.

Figure 4 shows CFPs for TLP between 0 and 30 for each physical quantities of our EOD. The best performance is obtained when WV1 is used as a predictor. Its performance degrades, however, with increase of TLP. It should be noted that prediction performances of aerosol related quantities (AMC and AOD) do not seem to degrade.

### 3.3 Multiple Predictors

It is often possible to improve prediction performance by using multiple quantities in a regression scheme. This is done by selecting a set of predictors. Increasing the number of predictors expands availability of features that benefit the prediction performance. On the other hand, too many
Fig. 4 CFP based on EOD.

Fig. 5 Predictor selection process.

predictors risk ending up in overfitting. Therefore, predictors should be carefully chosen. We used a stepwise forward addition procedure to construct the regression model with multiple predictors. First, we selected a single predictor and measured performance of a regression model based on that predictor. The one which resulted in the best performance was chosen and added to the set of predictors of the regression model. Then we picked up another predictor variable, constructed a regression model with it and all the predictors chosen previously, and evaluated the performance of the model. This process was iterated until addition of any new variable did not improve prediction performance significantly.

Figure 5 summarizes steps taken in the predictor selection process. Figure 5 (a) shows cCFPs for predictions with single predictor. It is clear from the figure that Tmp gives the best prediction performance. We therefore fixed Tmp in the predictor set.

Figure 5 (b) shows cCFPs obtained by adding a second predictor. It should be noted that the vertical axis is magnified to clarify difference among them. The solid circles indicate cCFP obtained by adding another predictor beside Tmp. The open circle shows cCFP for the best prediction performance with single predictor, i.e., (Tmp) in this case, for comparison. The best prediction performance was obtained when we added AMC as the second predictor. It is clear that a significant gain is obtained over (Tmp).

Figure 5 (c) shows cCFPs obtained by adding the third predictor to (Tmp, AMC). It is not clear, however, if (Tmp, AMC, Prs) has a significant gain over (Tmp, AMC). In order to evaluate the hypothesis that (Tmp, AMC, Prs) is superior to (Tmp, AMC), we have performed a test based on a bootstrap technique,
whose procedure is illustrated in Fig. 6. The original sample is described as a triplet set \((p_A; p_B; o)_i\), where 
\(p_A\) : predicted value of FVI using a predictor set A,
\(p_B\) : predicted value of FVI using a predictor set B,
o : observed value of FVI.
In the present case, A and B stand for \((\text{Tmp, AMC, Prs})\) and \((\text{Tmp, AMC})\), respectively. \(i\) denotes the day when the prediction is made. Out of this original sample, we built 1000 bootstrap samples, using resampling with replacement. From each bootstrap sample, the values of cCFP were replicated. In Fig. 6, cCFP(A) and cCFP(B) stand for cCFP for predictor set A and B, respectively. Then an achieved significance level
\[
\text{ASL} = \text{Prob}(\text{cCFP}(\text{Tmp, AMC, Prs}) > \text{cCFP}(\text{Tmp, AMC})_i),
\]
is evaluated by counting how often \((\text{Tmp, AMC, Prs})\) outperforms \((\text{Tmp, AMC})\) in the bootstrap sampling. ASL in this case was 0.648 which is low enough to reject the hypothesis with a 5% risk. The predictor selection process for multiple regression stopped here with the final choice of \((\text{Tmp, AMC})\) as the optimal predictor set.

4. Discussion

4.1 Optimal Predictor Selection

Figure 5 (a) shows that prediction performance of single predictor cases is ordered as follows:
\[
\text{Tmp} > \text{WVI} > \text{AOD} > \text{AMC} > \text{Cld} > \text{Ssn} > \text{Prs} > \text{Wd} > \text{WVNI} > \text{RR}.
\]
The best performance in the single predictor cases is obtained when Tmp is chosen as a predictor. Figure 2 and Fig. 4 show that (Tmp) has the best prediction performance in the entire TLP. This is natural since seasons strongly influence the atmospheric visibility: i.e., usually Mt. Fuji is clearly seen in winter, and it is hazy in summer. In addition, it deteriorates slightly as TLP increases, which is also natural considering temporal change of seasonal climate condition.

Although WVI is the second best single predictor, as Fig. 5 (b) shows, its addition to the predictor set does not improve the prediction performance. The climate in the Mt. Fuji area is wet in summer and dry in winter. WVI has therefore multicollinearity with Tmp, and overlapping predictive power. The best performance in the case of two predictors is obtained when AMC is chosen as a second predictor. Figure 7 compares CFP for (Tmp) and (Tmp, AMC). Although Tmp has the dominant effect over the entire TLP, AMC substantially improves the prediction performance. It should be noted that the improvement is significant in the long-term prediction range (TLP \( \geq 20 \)). This is in accord with the prediction characteristics of (AMC) shown in Fig. 4.

4.2 Contribution from Extra SVD Components

In Sect. 2, we saw that utilizing more SVD components enhances single quantity prediction performance for some EOD. We check here if this is the case when Tmp coexists in the predictor set. Table 4 compares prediction performances of single vs. multiple SVD components for CD used along with Tmp. This clearly shows that extra components are of no benefit, which justifies our decision to utilize the first SVD component only.

4.3 Prediction Performance

So far we have discussed measurement of the predictive performance in terms of CFP (or cCFP). It is hard to imagine, however, how this measure is related to a performance that would be readily apparent to a human observer. Here, we evaluate the performance in terms of prediction hit/miss, which is easier to understand. The view of Mt. Fuji is classified into two classes, “good” and “bad”, depending on whether \(\text{FVI} > 1.5\) or not. The prediction hits if the predicted class coincides with that observed, otherwise it misses. We define two hit rates as follows. \(p_{\text{good} | \text{good}}\) denotes a probability of predicting that the view is “good” on the condition that the actually observed view was “good”. In the same manner, \(p_{\text{bad} | \text{bad}}\) denotes a probability of “bad” on the condition that the actually observed view was “bad”. Figure 8 (a) displays the prediction hit rate for our optimal prediction model, i.e., \((\text{Tmp, AMC})\). The horizontal axis and the vertical axis represent \(p_{\text{bad} | \text{bad}}\) and \(p_{\text{good} | \text{good}}\), respectively. Each plot shows the hit rates for given TLP between 0 and 30. The dashed diagonal line represents performances of random predictions. Any sensible prediction must be positioned above this line. Substantial gain obtained by our prediction can be clearly observed.

Figure 8 (b) displays the hit rates in a magnified scale. Following the first rapid decay in TLP range of a few days, hit rates tend to stabilize and do not seem to collapse to the
level of random predictions. This suggests the significance of the long-range prediction.

4.4 Advantage of EOD

CD quantities have several advantages over EOD quantities. CD quantities are observed densely, i.e., daily, hourly or even continuously. They are collected over a long period of time and are easy to obtain. As shown in Sect. 3, the dominant contribution to the prediction performance is obtained from CD quantities, i.e., Tmp.

Since we are interested in the utility of EOD, we want to evaluate how much improvement is obtained by using EOD. Figure 9 displays the improvement in hit rates obtained by the optimal model, i.e. (Tmp, AMC), over the sub-optimal model that does not include any quantity of EOD in its predictor set, i.e., (Tmp, Cld). Note that the scale is magnified so as to clarify fine differences. Most of the improvements in hit rates lie in the range of 0.0–0.05. This corresponds to an increase of several hits in the course of a year.

4.5 Merit of Non-linear Model

All the results given so far were obtained by a non-linear regression method (SVMRBF). Here, we discuss the impact of such a non-linear method and justify its usage. Figure 10 compares CFP (Tmp, AMC) obtained by SVMRBF and that obtained by linear SVM. The figure clearly shows the consistent advantage of SVMRBF.

Figure 11 shows plots of observed data points for TLP = 29 on Tmp-AMC plane in an arbitrary chosen scale. Open circles and closed circles represent observed data points with FVI > 1.5 (good) and FVI ≤ 1.5 (bad), respectively. Solid curves and a dashed line represent contours at (predicted FVI) = 1.5 for SVMRBF and linear SVM, re-
respectively. In the ideal case, all the open circles should reside on one side of the curves, leaving the closed circles on the other sides. The intricate mixture of open and closed circles reveals difficulty of the regression problem. “Good” points tend to locate at the low Tmp and high AMC area. This tendency can be captured even by the linear model. However, a small cluster of “good” points can be found at the high Tmp and relatively low AMC region, which cannot be captured by the linear model. The non-linear model can take care of this small cluster and enhances its prediction performance.

5. Conclusion

In this paper, a method for forecasting the view of Mt. Fuji was presented. The forecast was based on a regression with predictors selected from quantities available in climate data and Earth observation data. Non-linear support vector regression with the RBF kernel was employed. Predictors were selected in a usual stepwise forward addition procedure. Among the combinations of predictors we have tried, Tmp and AMC proved to be optimal for the forecast. Tmp has the best performance in single predictor cases and plays a dominant role in the multiple predictors case. Its performance is attributable to the seasonal change of clearness of the view of Mt. Fuji. Therefore, the performance degrades as TLP increases. On the other hand, it was observed that AMC contributes to the long-range forecast. It has the effect of compensating Tmp. Hit rates of the forecast using AMC are a few % higher than those without using any EOD. Thus, we have demonstrated the advantage of using EOD for the forecast problem.

For our forecasting problem, SVMRBF was found to have an advantage over linear SVM. Highly skewed distribution of observed data points in the predictors’ value space makes the problem difficult to handle by linear models. Such characteristics are well taken care of by the non-linear model.

There still are some directions to follow in future work. Although the merit of AMC was confirmed statistically, it is still unclear what accounts for the merit. The contribution of AMC to the forecasts for TLP of more than 20 days is especially mysterious considering that weather conditions usually fluctuate in a much shorter period. There may be a geophysical explanation.

We have shown how to extract components relevant for the forecast from multidimensional EOD. We still feel, however, that our treatment of EOD requires further work to exploit its potential power. It was pointed out that SVD components with minor singular values are not necessarily irrelevant [9]. Exhaustive search of predictors in the space with minor singular value components may improve the forecast performance. Moreover, in our experiment, we used only values of those areas limited to the neighborhood of Mt. Fuji. For example, considering that weather flows from west to east in this area, it might be a good idea to use data from much further west. When we treat EOD, it is possible to use data from the entire surface of the Earth. Data at some locations might be related to the data which we are interested in. Techniques to find hidden relations in a huge data set might help.

Acknowledgement

This work is supported by the Earth Observation Research Center of the Japan Aerospace Exploration Agency grants 04RSTK-016207 and 05RSTK-017877.

References

[1] W. Malm, “Introduction to visibility,” Technical Report, Air Resource Division National Park Service and Cooperative Institute for Research in the Atmosphere (CIRA) Colorado State University, Fort Collins, CO., 1999.

[2] V.N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed., Springer-Verlag New York, 1995.

[3] http://www.fuji-net.co.jp

[4] M. King, Y. Kauffman, W. Menzel, and D. Tanré, “Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (modis),” IEEE Trans. Geosci. Remote Sens., vol.30, no.1, pp.2–27, 1992.

[5] E. Mészáros, Fundamentals of Atmospheric Aerosol Chemistry, AKADÉMIAI KIADÓ, Budapest, 1999.

[6] R. Adler, T. Wilheit, Jr., C. Kummerow, and R. Ferraro, “Amsr-e/aqua 12b global swath rain rate/type gsf profiling algorithm v001, march to june 2004,” Digital media, updated daily. Boulder, CO, USA: National Snow and Ice Data Center, 2004.

[7] C. Chang and C. Lin, “LIBSVM: A library for support vector machines,” Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.

[8] B. Eflon and R.J. Tibshirani, An Introduction to the Bootstrap, Chapman&Hall/CRC, 1993.

[9] I.T. Jolliffe, “A note on the use of principal components in regression,” J. Royal Statistical Society, Series C (Applied Statistics), vol.31, no.3, pp.300–303, 1982.
Mitsuru Kakimoto received the B.S., M.S. degrees and the Ph.D. in Physics from University of Tokyo, in 1982, 1984 and 1987, respectively. Currently he is a research scientist at the System Engineering Laboratory, R&D Center, Toshiba Corp. His research interests are in the development and application of intelligent image processing system, performance evaluation of computer systems.

Hisaki Hatano received the B.S. degree in control engineering in 1985, the M.S. degree in systems science in 1987, from Tokyo Institute of Technology, Tokyo, Japan. His interest is data analysis using statistics and datamining.

Yosoko Nishizawa received the B.S. and M.S. degrees in Physics from University of Kyushu, in 1984 and 1986, respectively. Currently she estimates the value of design for product at Design Center, Toshiba Corp. She is interested in consumer's choice about products.