Abstract—There are many models of time series description. The most popular are trend + seasonal + cyclic component decomposition and the Box-Jenkins methodology. The big advantage of decomposition models is very intuitive mapping of real world into mathematical model. These models are easily understandable even to common users without deep mathematical background. On the other hand, current decomposition models are not so much suitable for modeling of complex time series with more periodic components and asymmetric behavior during the period. The mode proposed in this paper aims to better fit more deep structured time series.

Keywords—time series, weekly cycle, calendar dependency.

I. INTRODUCTION

We are dealing with time series describing additive calendar dependent processes like volume of data transported over internet line, number of telephone connections per given time period and similar. Processes of this type have periodic behavior with a priori known periods. In such processes we can observe periods given by time of the day, day of the week and week of the year. There is often a month period as well but the week one is “stronger” in terms of lower residual noise. The periodicity of the process is given by its nature. E.g. it is not expectable that phone calls will be evenly distributed during day. There will be probably more or less no calls during night, many calls during peak hours, much less during lunch time and very few during evening. The overall number and the distribution of phone calls during the day also strongly depend on the day of the week. There will be another situation during weekdays and weekend and also the beginning and end of the week influence the peoples need for phone calls.

It is also not expectable that the distribution of the number of call per day of the week will be very similar during the whole year. There will be different situation during Christmas and New year, common working months, state or school holidays, a couple of weeks before and after the summer holidays etc.

Nowadays, the time series analysis is used mostly for economical applications like sales or manufacture volume analysis and prediction. The most meaningful time periods in economy are month and year. In technical applications, however, we don’t rely so strong on the month period which is a little bit artificial and can more emphasize the week period which has not so big sense in economy but more precisely correspond to the real behavior of the human oriented processes.

II. MOTIVATION

Masaryk University in Brno is operating a regional PACS archive serving to mostly all hospitals in Brno metropolis and a lot of remote medicine institutions. The communication between modalities (i.e. CT, ultrasound, X-ray, etc.), PACS archive and viewing stations is based on DICOM standard.

The PACS - Picture Archiving and Communications System - is a currently used procedure and methodology for processing medical multimedia data obtained from picture acquisition machines like computer tomography, ultrasound, x-ray etc. Multimedia medicine data obtained from these machines in PACS terminology called modalities - are stored in central PACS server. The PACS server then provides these multimedia data to viewing stations. Viewing stations serve to radiologists for analyzing the multimedia data. This approach offers much more capabilities than former film medium. Viewing stations allow image transformation, combination of images from more modalities etc.

The Shared Regional PACS project MediMed [1]-[5] started as a collaborative effort among Brno hospitals to process medical multimedia data. The system serves for transmitting, archiving, and sharing medical image data originating from various medical modalities (computer tomography, magnetic resonance, ultrasound, mammography, etc.) from hospitals. The central PACS serves as a metropolitan communications node as well as a long term archive of patients' image studies.

Outsourcing of the hospitals' archiving and communications technology permits cooperation among hospitals and the use of existing patient multimedia data. The Shared Regional PACS is more than just a computer network. Gradually, it changes the thinking of medical specialists and gets them to cooperate and share data about patients in electronic form. It builds a network of medical specialists. The impact of this work is not only in patient care but also in the education of medical specialists. The data stored in the shared archive can be anonymized (i.e. personal data of patients is replaced by fictitious data) and used for educational purposes. The realization of the project facilitates fast communication among individual hospitals, allows decision consultations, and brings
various other advantages due to direct connections via optic networks. The MeDiMed project is clearly designed to support society-wide healthcare programs in the Czech Republic as well as programmed implemented by other countries. The system is also supposed to serve as a learning tool for medical students of the Masaryk University as well as physicians in hospitals.

III. THE DWY MODEL

A time series decomposition model oriented to weekly period instead of a monthly one more precisely describes processes depending on human behavior. People are mostly keeping the weekly periodicity (five working days and two days off). Also the yearly period is meaningful for people.

The main drawback of currently used decomposition methodology is low respect to variability of “higher frequency” periodic components. The member $X(t)$ of a time serie can be expressed as

$$X(t) = T(t) + S_i(t) + S_j(t) + \ldots + S_n(t) + E(t) \quad (1)$$

Where $T(t)$ is the trend, $S_i(t)$ are periodic components and $E(t)$ is the white noise representing the stochastic behavior of the random variable. Let assume that the period of $S_i$ is longer than that of $S_j$ in case $i < j$. In the incumbent time series decomposition methodology we deal with only one $S_i$ for given frequency. Our experience says that this is not the case for many human based processes.

Let imagine e.g. the number of patients going to the hospital. The distribution of the quantity of patients arriving during the given hour depends on the day of the week. Some people just don't intend to use the weekend for medical examination and treatment and simply wait for the weekday. In Monday some people prefer to stay at cottage longer and arrive to the medicine treatment at later time. The opposite situation comes on Friday when most of people intend to start the weekend as soon as possible and prefer not to visit a doctor later than noon. Also in Saturdays most people prefer to solve stuff other than relaxing during morning while on Sunday the preferred time is the afternoon/evening.

The result of our observation is: for proper description of the time serie behavior it is better to use not only one periodic component describing distribution of the e.g. number of patients coming to the hospital at a given time of the day but a set of such components: One for Mondays, the second one for Tuesdays, Wednesdays and Thursdays, third one for Fridays and yet two more for Saturdays and Sundays. This leads to the following proposal:

Our proposal follows from commonly used additive decomposition model. The expected value of a random variable $X$ at the time $t=h+d+w$, where $h$ is the hour of the day, $d$ is the days of the week and $w$ is the week of the year can be expressed as (1).

$$X(t) = T(t) + Y(w) + W(d) + D_i(h) + E(t) \quad (2)$$

Where $T(t)$ is the trend, $Y(w)$ is the yearly period, $W(d)$ is the set of weekly periods, $D_i(h)$ is the set of daily periods and $E(t)$ is the white noise representing the stochastic behavior of the random variable. The $i \in \{1,\ldots,52\}$ represents the week in the year, and the $j \in \{1,\ldots,7\}$ represents the day in the week. This way we can use different weekly period for distinguished weeks in year (with respect to the main school holidays, Christmas, etc.). The same holds for the daily period which is dependent on the day of the week in common.

The principle of this methodology can be easily explained on the example provided in the following section.

IV. EXAMPLE

We can demonstrate advantages of the proposed model on the time serie describing the volume of data stored to the shared regional PACS archive by an average district hospital. We have collected the volume of data stored per day for about three years. The time serie itself is depicted on the figure 1. The negative value of the stored data volume is caused by cleaning procedure and removal of unusable material like unfocussed or fuzzy pictures.

![Fig. 1 The volume of PACS data stored by average district hospital per day](image)

When using the incumbent decomposition methodology we can use the weekly average (figure 2) or weekly summary (figure 3) to cover the yearly periodicity and to focus on the weekly one. If we try to construct a model of the weekly period by the incumbent decomposition model, we just make an average of data stored at all Mondays, all Tuesdays, etc. as is easy to see from figure 4. This approach doesn’t take into account discrepancies of the overall week data volume spread across the year (lower volume during summer holidays or Christmas). This is solved by the weekly summary or average as depicted on figures 2 and 3.
Our approach follows from the longest period observed (in this case the yearly period as depicted on figure 5) and constructs a set of shorter period (weekly period) approximations. Of course it is not necessary to construct 52 weekly period approximations because the behavior of most of the weekly periods is the same. Mainly we have regular work weeks, holidays and in some cases transitional ones i.e. weeks just before or after longer holidays. On the figure 5 is easy to see lower data volume during summer holidays, Eastern and begin/end of the year. So we construct who weekly approximations, one for on-peak weeks and the second one for the off-peak ones.

Our methodology can slightly better model the behavior of large time series and provides more precise results comparing to incumbent decomposition model. The main source of this improvement is different approach or viewpoint to the time serie. Incumbent decomposition methodology is oriented towards economical applications which are strongly based on monthly and annual periodicity. On monthly and annual basis we pay taxes, get salaries and count expenses of manufacturing.

This method, however, does not fit so exactly to the real behavior of time series depending on human activities like visiting hospital, shopping etc. which are more oriented to the weekly period. For this reason we have bring a modification of the decomposition methodology which respects this weekly oriented behavior. This modified methodology fits a slightly better for our main intended applications in the field of prediction of used space on data storage and detecting anomalies in data storage which may indicate either technical or security issues.

V. Future Work

The DWY model can be generalized to accommodate just any set of periods not necessarily dependent on the calendar or a priori known periodicity of the time series under study. Our intention is to scale up this methodology to more general case and to produce something like a pseudo-orthogonal time series decomposition model. The basic idea is to start the time series decomposition from the trend and the longest period and go down to the shortest period by keeping a set of short period models based on the phase of the longer period.

VI. Conclusion

The presented DWY time series decomposition methodology seems to be a viable tool for modeling time series depending on human behavior based on the calendar. For technical applications like prediction of the stored data
volume or detection of suspicious network behavior it seems to be better than existing incumbent statistical methodologies which are more oriented to econometry.

ACKNOWLEDGMENT

This work is supported by Czech Technology Agency fund project number TA01010268 - "Maintenance-free PACS system for small and mid-sized healthcare institutions".

REFERENCES


