Chapter 14
Modeling Dynamic Processes with Memory by Higher Order Temporal Models

Anna Łupińska-Dubicka and Marek J. Druzdzel

Abstract Most practical uses of Dynamic Bayesian Networks (DBNs) involve temporal influences of the first order, i.e., influences between neighboring time steps. This choice is a convenient approximation influenced by the existence of efficient algorithms for first order models and limitations of available tools. In this paper, we focus on the question whether constructing higher time-order models is worth the effort when the underlying system's memory goes beyond the current state. We present the results of an experiment in which we successively introduce higher order DBN models monitoring woman's monthly cycle and measure the accuracy of these models in estimating the fertile period around the day of ovulation. We show that higher order models are more accurate than first order models. However, we have also observed over-fitting and a resulting decrease in accuracy when the time order chosen is too high.

14.1 Introduction

While all real world systems change over time, modeling their equilibrium states or ignoring change altogether, when it is sufficiently slow, is sufficient for solving a wide spectrum of practical problems. In some cases, however, it is necessary to follow the change that the system is undergoing and introduce time as one of the model variables.

We concentrate in this chapter on models that belong to the class of probabilistic graphical models, with their two prominent members: Bayesian networks (BNs) [7] and dynamic Bayesian networks (DBNs) [3]. BNs are widely used practical tools for knowledge representation and reasoning under uncertainty in equilibrium systems. DBNs extend them to time-dependent domains by introducing an explicit notion of time and influences that span over time. Most practical uses of DBNs involve temporal influences of the first order, i.e., influences between neighboring time steps. This choice is a convenient approximation influenced by existence of efficient algorithms for first order models and limitations of available tools. After all, introducing higher

order temporal influences may be costly in terms of the resulting computational complexity of inference, which is NP-hard even for static models. Limiting temporal influences to influences between neighboring time periods is equivalent to assuming that the only thing that matters in the future trajectory of the system is its current state. Many real world systems, however, have memory that spans beyond their current state.

The question that we pose in this chapter is whether introducing higher order influences, i.e., influences that span over multiple steps, is worth the effort in the sense of improving the accuracy of the model. The idea of increasing modeling accuracy by means of increasing the time order of a dynamic model was beautifully illustrated by Shannon. In his seminal paper [11], outlining the principles of theory of information, he shows sentences in the English language, generated by a series of Markov chain models of increasing time order, trained by means of the same corpus of text. The following sentence was generated by a first order model:

```
OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH
EEI ALHENHTTPA OOBTTVA NAH BRL.
```

Compare this with the following sentence generated by a sixth order model:

```
THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH
WRITER THAT THE CHARACTER OF THIS POINT IS
THEREFORE ANOTHER METHOD FOR THE LETTERS
THAT THE TIME OF WHO EVER TOLD THE PROBLEM
FOR AN UNEXPECTED.
```

The resemblance of the latter sentence to ordinary English text, an informal measure of the model's accuracy, has increased dramatically between the first and the sixth orders. A first order model was essentially impotent in its ability to learn and model the language.

While generation of English sentences may be too hard of a problem, the vehicle for our experiments with varying time order is the problem of monitoring the woman's monthly cycle, a problem central to human fertility. Every couple seeking help in a fertility clinic is asked to monitor the monthly cycle before any medical intervention is undertaken. An accurate monitoring model can be a great aid in natural family planning, indicating optimal days for sexual intercourse. There exist methods for fairly precise determining of the day of ovulation (e.g., blood hormone level tests or ultrasonographic analysis of the ovaries), but they either require laboratory visits or expensive testing kits. What is important from the perspective of the question posed in this chapter is that woman's monthly cycle is a system with memory going most certainly beyond one day and probably spanning over a period of roughly a month.

We report the results of an experiment in which we successively introduce higher order DBNs modeling the monthly cycle and measure the accuracy of these models in estimating the fertile period around the day of ovulation. We train our models on real time series data obtained from a longitudinal study of fecundability conducted in several European centers [2]. We show that increasing the time order of the model

greatly improves its accuracy but only up to a certain point. Too high order of a model decreases accuracy, probably though over-fitting the training data.

The remainder of the chapter is structured as follows. Section 14.3 reviews what we know about woman's monthly cycle. Section 14.5 describes the data that we used in training our models. Section 14.4 describes our DBN models, Sect. 14.6 describes our experiments, and Sect. 14.7 summarizes the results of our experiments with the models. Finally, Sect. 14.8 offers some advice to knowledge engineers building DBN models in practice.

14.2 Bayesian Networks

Bayesian networks (BNs) are probabilistic graphical models that offer a compact representation of the joint probability distribution over a set of random variables $X = x_1, \ldots, x_n$. Formally, a Bayesian network is a pair (G, Θ), where G is a acyclic directed graph (ADG) in which nodes represent random variables $x_1, i \ldots, x_n$ and edges represent direct dependencies between pairs of variables. The second component of a Bayesian network, Θ, represents the set of parameters that describes a conditional distribution for each node x_i in G, given its parents in G, i.e., $P(x_i | Pa(x_i))$. Very often, the structure of the graph is given a causal interpretation, convenient from the point of view of knowledge engineering and user interfaces. Bayesian networks allow for computing probability distributions over subsets of their variables conditional on other subsets of observed variables. This can be given the interpretation of computing the probability of a hypothesis in light of evidence. BNs are widely applied in decision support systems, where they typically form the central inferential engine.

Consider the simple Bayesian network shown in Fig. 14.1. This is a simplified example, illustrating various causes of allergy in children. The tendency to develop allergies is often hereditary. Allergic parents are more likely to have allergic chil-
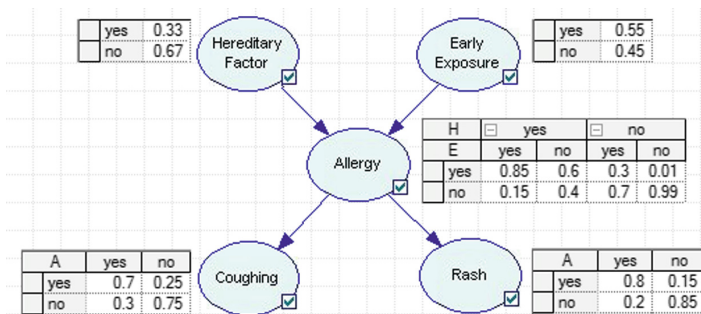

Fig. 14.1 A simple Bayesian network illustrating selected causes and effects of allergy in children.

dren, and their allergies are likely to be more severe than those from non-allergic parents. Exposure to allergens, especially in early life, is also an important risk factor for allergy. When an allergen enters the body of an allergic child, the child can cough or develop a rash. Figure 14.1 shows the dependency structure among the variables and the conditional probability distributions for each of the variables. All variables in this example are Boolean. At the roots, we have the prior probabilities (e.g., that one or both of the parents suffer from allergies or a child had a contact with allergen in early life). The conditional probabilities for the non-root nodes give the probability distributions over the nodes conditional on various outcomes of the direct predecessors in the graph (e.g., probability distribution over the variable coughing given that a child has allergy).

Dynamic Bayesian networks (DBNs) are an extension of Bayesian networks for modeling dynamic systems. In a DBN, the state of a system at time t is represented by a set of random variables $X_t = (X_{1,t}, \ldots, X_{n,t})$. The state at time t generally dependents on the states at previous time steps. Typically, we assume that each state only depends on the immediately preceding state (i.e., the system is first-order Markov), and thus we represent the transition distribution $P(X_t|X_{t-1})$. This can be done using a two-slice Bayesian network fragment (2TBN) B_t, which contains variables from X_t whose parents are variables from X_{t-1} and/or X_t, and variables from X_{t-1} without their parents. The term *dynamic* means that we model the state of a system over time, not that the model structure and its parameters change over time (even though the latter is theoretically possible). A DBN is typically defined as a pair of Bayesian networks $(B_0, B_\rightarrow)$, where B_0 represents the initial distribution $P(X_0$, and $B_\rightarrow$ is a two time slice Bayesian netwok, which defines the transition distribution $P(X_t|Xt-1)$ as follows [3]:

$$P(X_t|X_{t-1}) = \prod_{i=1}^{N} P(X_{i,t}|Pa(X_{i,t}))$$

Consider a two years old child whose parents suffer from allergy and who has been exposed to some allergens. We know that this child has not developed any symptoms of allergy in the previous year. Suppose that we want to know the probability that allergy appears in the third year. If we use the BN pictured in Fig. 14.1, we omit all historical information except for the previous year. Figure 14.2 (a) shows a DBN of first temporal order, which means that we take into consideration not only present observations but also these from the previous year.

As we mentioned above, one often assumes in practice that each state depends only on the immediately preceding state. In most cases, taking into consideration only the first-order dependence is probably sufficient. However, in general, we can specify layers from $t - n$ to n. There is a possibility that some phenomena could be modeled with higher efficiency if they also take account of the influence of states earlier than immediately preceding the current state of the model. To our knowledge, the question whether such simplification of dynamic models leads to incomplete and even erroneous results has never been studied systematically.
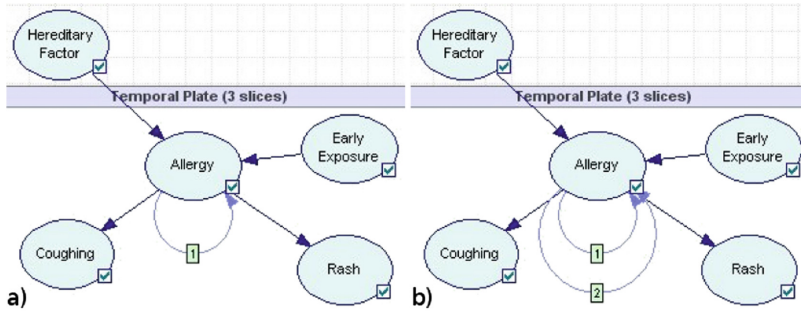
Fig. 14.2 Dynamic Bayesian networks modeling causes and effects an allergy in children: (a) first order DBN, (b) second order DBN. Number of slices is the number of steps for which we perform the inference. In this example, one step means one year. Temporal plate is the part of dynamic network that contains the temporal nodes. Hereditary Factor is time independent; the values of remaining the nodes can change over time.

Figure 14.2(b) shows a second order dynamic network, i.e., in which there are two temporal arcs from node *Allergy*, the first order takes the information from one step before, the second from two steps before. Typically, the older the child the lower the probability of allergy appearing. And, generally, the child that has not developed allergy two years in a row, has a lower chance of developing allergy in the third year.

14.3 Woman's Monthly Cycle

Woman's monthly cycle is driven by a highly complex interaction among hormones produced by three organs of the body: the hypothalamus, the pituitary gland, and the ovaries. There are five main hormones involved in the menstrual cycle process: estrogen, progesterone, gonadotropin releasing hormone (GnRH), follicle stimulating hormone (FSH), and luteinizing hormone (LH).

Estrogen refers to a group of hormones that stimulate growth and strengthen tissues. It is needed to build up the lining of the uterus so that it may nourish and sustain a fertilized egg. Progesterone is produced by the follicle from which the mature egg has been released (the follicle that has released an egg is called *corpus luteum*). Progesterone helps make the endometrial lining ready for implantation if an egg is fertilized during the cycle. It also prevents the egg follicles from developing any further. GnRH, produced by the hypothalamus in the brain, is responsible for the production and levels of estrogen in the body. FSH is secreted by the pituitary gland, which is stimulated by the hypothalamus' production of GnRH. Increased levels of FSH help to stimulate egg follicles. LH, produced by the pituitary gland, is needed to trigger the ovulation.

The woman's monthly cycle consists of four phases (Fig. 14.3 shows these four phases along with the associated hormone levels): (1) menstruation, (2) the follicular phase, (3) ovulation, and (4) the luteal phase. Counting from the first day of the
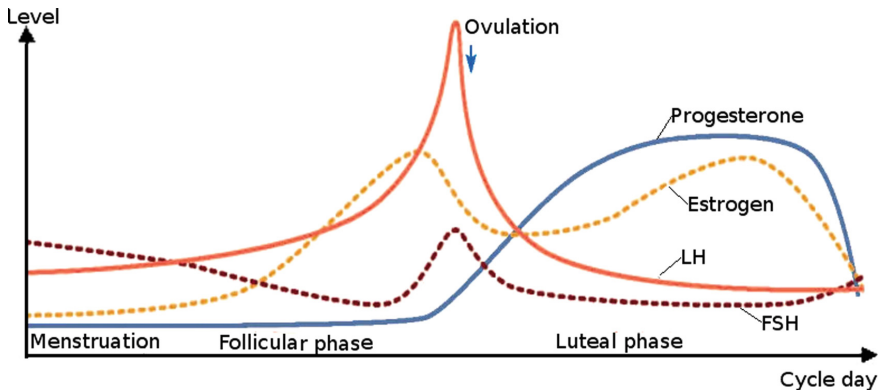
Fig. 14.3 Levels of hormones during the phases of the woman's monthly cycle [13]

menstrual flow, the length of each phase may vary from woman to woman and from cycle to cycle, although the entire cycle takes typically between 24 and 32 days.

Menstruation begins with the first day of bleeding. Contraction of the muscle layer occurs, expelling blood and endometrial cells through the vagina. During the follicular phase (or the proliferative phase), the follicles in the ovary mature. The main hormone controlling this stage is estrogen. Just before the ovulation, the level of estrogen is high enough to cause an increased release of luteinizing hormone and, as a result, the egg is released from the ovary. The luteal (or the secretory) phase is the latter phase of the menstrual cycle. The main hormone associated with this stage is progesterone, which occurs at significantly higher levels during the luteal phase than during the other phases of the cycle.

In addition to measurable blood hormone levels, there are several readily accessible indicators of the phase of the cycle, two of which we will use in our models. The basal body temperature (BBT) is defined as the body temperature measured immediately after awakening and before any physical activity has been undertaken. It should be measured every day at the same time. Before ovulation, BBT is relatively low. Following the ovulation, as a result of an increased level of progesterone in the body, women typically experience an increase in the basal body temperature (BBT) of at least 0.2 °C. This shift indicates that ovulation has occurred. The BBT charting may provide valuable information about woman's monthly cycle, such as duration of the cycle, length of the follicular and luteal phases, and the pattern of the timing of ovulation. Sometimes BBT can rise due to causes other than ovulation. This atypical rise is treated as disturbance and can be caused by a change in conditions around the measurement, such as later measurement time, lack of sleep, different thermometer, high stress, travel, or illness. As the cycle progresses, due to hormonal fluctuations, the cervical mucus increases in volume and changes texture. When there is no mucus or the mucus discharge is small, the day is considered infertile. There can be also a feeling of dryness around the vulva. Around the ovulation, mucus is the thinnest, clearest, and most abundant, resembling egg white. In the luteal phase, it returns to

the sticky stage. During the monthly cycle, the cervix changes its position, firmness, and openness, in response to the same hormones that cause cervical mucus to be produced and to dry up. At the beginning of the cycle, cervix is located low in the vaginal canal and the os (the orifice of the uterus) is relatively small or closed. As ovulation approaches, cervix moves up the vaginal canal and becomes softer, with the os opening up. After ovulation cervix moves down and closes.

The menstrual cycle is a fairly noisy temporal process with memory spanning over the entire cycle. This means that the current state is not only influenced by the previous state but also by prior days, going back even to the beginning of the phase.

14.4   The Model

Accurate prediction of the fertile phase of the menstrual cycle is crucial for couples who want to conceive or couples who want to avoid pregnancy using natural methods. The fertile phase of the menstrual cycle is defined as the time when an intercourse has a non–zero probability of resulting in conception. Because the fertile period starts roughly five days before ovulation (this is essentially due to the fact that sperm can live up to five and fertilize the egg when ovulation happens, prediction has to be made in advance and, hence, asks for models that include an explicit notion of time.

Our model (Fig. 14.4), combines information retrieved from BBT charting with observations of the cervical mucus secretions. It contains a variable *Phase* with four states: *menstruation*, *follicular*, *ovulation*, and *luteal*. We included three observation variables: *Basal Body Temperature* (*BBT*), *Bleeding* and *Mucus observation*. All variables are discrete. *BBT* has two possible values: lower range and higher range, representing temperature before and after the BBT shift respectively. *Bleeding* describes whether on a particular day the woman had menses or not. *Mucus observation* can be in one of four states (*s1* through *s4*), described in detail in [4]. We modeled time explicitly as n time steps, where n is the number of days of the longest monthly cycle of the modeled woman. The model is of k-order, i.e., it contains temporal influences between 1 and k. Figure 14.4 shows an example DBN of 3rd order. Furthermore, while any DBN model should contain at least one first order influence, a model of order k does not need to include influences of all orders between 1 and $k - 1$.

To train a complex model we need a large number of observations. Learning models from data is based on strong theoretical foundations. Having sufficient amount of data, we can reliably learn numerical parameters of the model. In practice, however, the number of data records is often limited and generally making it challenging to learn reliable estimates of the parameters. Collecting data in case of a woman's monthly cycle problem will never result in sufficiently large data sets. Assuming that a woman is fertile during 40 years of her life, with roughly 13 cycles each year, she can collect at most 520 records. When these 520 records have been accrued, they are useless, as the woman is no longer fertile. In practice a woman will have not more than a couple of years worth of reliable data, i.e., roughly twenty-something records.
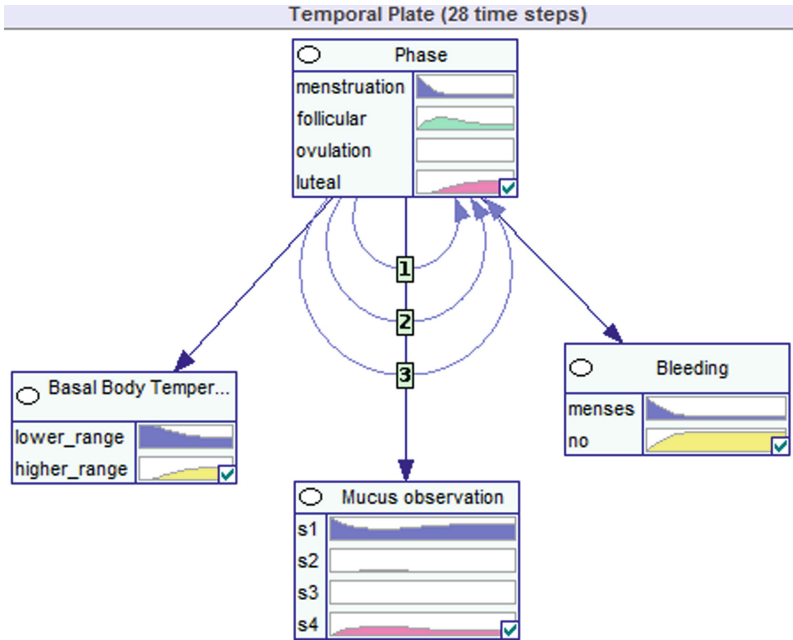
Fig. 14.4 A 3rd order DBN model of woman's monthly cycle. The plots inside the rectangles show the marginal probability distributions over the variables that they represent.

Typically, a model that aids in conception or in avoiding pregnancy, needs to rest on a handful of records.

Learning conditional probability distribution tables amounts essentially to counting data records for different conditions encoded in the network. The number of parameters required to specify a CPT for a node grows exponentially in the number of its parents, and thus the higher the order the more complex its structure and the more data are needed to learn parameters. In case of a fifth order DBN network of woman's monthly cycle for the node *Phase*, we need to estimate 1, 024 parameters. Even if we take into consideration that due to the specifics of the domain many columns of the CPTs represent unlikely cases, we are still dealing with a problem of insufficient amount of data. Please note, that most practical fertility awareness methods advise to consider charting at least six cycles to become familiar with a method. This means two problems: (1) Constant struggle against over-fitting the model to the data, and (2) Necessity to use prior knowledge, as a handful of records will never be enough to learn a complex probabilistic model.

When we learn the network parameters from such a small amount of data, some of the CPT entries might be learned from an insufficient number of records or there might even be no data records to learn distributions for some combination of the outcomes of the parents in a node. In order to provide more meaningful results and to compensate for the small amount data, we have based the initial structure of a

model and its parameters on the domain knowledge. This procedure can be described as follows. We randomly divided all women into five equal subsets. For each woman the training data set was the sum of four subsets, excluding this which the woman belonged to. We learned the initial model parameters based on the population of women. Then we applied these population–based model as the *a priori* parameters in all woman–specific models. And as our intention was to simulate usage of a model by woman who wants to become pregnant or wants to avoid pregnancy, we adjusted the initial model to each woman using data for her first six cycles.

14.5 The Training Data

Our training data are drawn from an Italian study of daily fecundability [2], which enrolled women from seven European centers (Milan, Verona, Lugano, Düsseldorf, Paris, London and Brussels) and from Auckland, New Zealand. To our knowledge, this is one of the most comprehensive data sets describing woman's monthly cycle. Between the years 1992 and 1996, 881 women recorded a total of over seven thousand monthly cycles. Women participating in the study satisfied the following five entry criteria: (1) experienced in use of a Natural Family Planning method, (2) married or in a stable relationship, (3) between 18th and 40th birthday at admission, (4) had at least one menses after cessation of breastfeeding or after delivery, (5) not taking hormonal medication or drugs affecting fertility. In addition, neither partner could be permanently infertile and both had to be free from any illness that could affect fertility.

In each menstrual cycle, the woman was asked to record the days of her period, her basal body temperature, and any disturbances such as illness, disruption of sleep, or travel. She was also asked to observe and chart her cervical mucus symptoms daily during the cycle and to record every episode of coitus, with specification whether the couple used contraceptives or not.

A menstrual cycle was defined as the interval in days between the first day of menstrual bleeding in two neighboring cycles, where day 1 was the first day of fresh red bleeding, excluding any preceding days with spotting. The day of ovulation was identified in each cycle from records of basal body temperature and mucus symptoms. The daily mucus observations were classified into four classes; ranging from a score of 1 (no discharge and dry) to 4 (transparent, stretchy, slippery) [4]. The cervical mucus peak day was defined as the last day with best quality mucus, in a specific cycle of the woman. If there were different mucus observations on one day, the most fertile characteristic of the mucus observed determined the classification. To determine the BBT shift, the "three over six" rule (popular among fertility awareness methods or FAMs) was used: The first time in the menstrual cycle when three consecutive temperatures were registered, all of which were above the average temperature of the last six proceeding days.

In our analysis, we included only 3, 432 (of 7, 017) cycles from 236 (of 881) women. We excluded all women who collected fewer than seven cycles, because a

woman needs at least six cycles to become familiar with a chosen fertility awareness method. We also excluded cycles with no uniquely identified mucus peak or the BBT shift days, because our model uses these values to determine the beginning of the post–ovulatory infertility. We also excluded women with very long cycles (longer than 40 days).

14.6 Experiments

For each woman, we created seven DBNs of temporal orders ranging from 1 to 7. Additionally, for each woman we created a model, with a structure that can change after each cycle. We changed the structure of that model by adding or removing temporal arcs, bearing in mind that first order arc is necessary and cannot be removed. For the last 12 cycles, we calculated the minimal and most frequent day of the ovulation. Dividing these values by two we received the order of temporal arcs that should appear in the model. Typically these orders were between six and nine. We determined the initial parameters of all models based on domain knowledge. We personalized each model using data for the first six cycles. After each cycle we re-evaluated the model's parameters based on previous cycles of the woman. Because a woman's body can also change over time and with it the characteristics of the cycle, we updated the structure and parameters using not more than the last 12 monthly cycles.

In case of monitoring a woman's monthly cycle, the main goal is to predict the day of ovulation and based on it to determine the fertile window. The number of fertile days during a menstrual cycle is difficult to specify, as it depends on the life span of the ovum and sperm, which varies from person to person and from cycle to cycle. Most menstrual cycles start with infertile days (pre–ovulatory infertility), a period of fertility and then several infertile days until the next menstruation (post–ovulatory infertility). It is generally believed that an ovum can be fertilized only within the first 24 h after ovulation [10]. Many authors agree that the start of the fertile interval is strictly connected with changes in vaginal discharge and, in particular, estrogenic–type cervical mucus secretions. However, they differ in their estimates of the length of the fertile window. Potter [8] calculated that there are only two days during the menstrual cycle when a woman can become pregnant. Wilcox *et al.* [14] found that the maximum sperm life span equals approximately five days (in presence of sufficient level of estrogenic–type mucus), which comes down to a fertile period of six days, including the day of the ovulation. The results of a multi–center study conducted by the World Health Organization [6] estimate the fertile period to be as many as 10 days before ovulation. Some of the fertility awareness methods assume this interval to be as long as 13 days or even longer [1, 5, 9, 12].

Our intention was to simulate the usage of DBN model by women who want to become pregnant or want to avoid pregnancy. At every time step (i.e., every day of the cycle), our model computed the most probable day of ovulation. If a time interval between the current day and the day with the highest probability of the ovulation was shorter than seven days, we marked the current day as fertile. To find the beginning
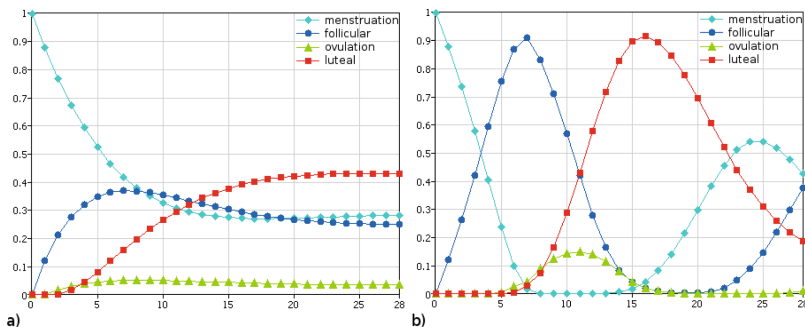
Fig. 14.5 Probabilities of each phase during the monthly cycle: (a) order 1, and (b) order 7 DBNs

of the post–ovulatory phase, our model used the BBT shift: We considered the third day after the BBT shift as infertile.

Just to give an idea of the capability of such models to reproduce the monthly cycle, we present the probabilities of the four phases of the monthly cycle as a function of time in Fig. 14.5. These probabilities were generated by DBNs models of (a) first and (b) seventh order DBNs, trained on monthly charts of one of the women in the data set. We entered no observation into the models, except for anchoring the first time step to the first day of menses, i.e., first day of the monthly cycle. Please note the increased similarity of the shape of the curves to that of the hormone levels in Fig. 14.3, which are direct indications of phases of the monthly cycle.

To compare the accuracy of different models, we used two measures: (1) the percentage length of the infertile period (the union of the pre–ovulatory and the post–ovulatory phase), and (2) the percentage length of the fertile window. We determined the number of fertile and infertile days in all cycles and divided this number by the total length of the cycle for each woman and for each cycle. Effectively, we obtained the percentage of all days that were classified as infertile and percentage of all days that were classified as fertile. In our opinion, these two numbers (they add up to 100 %) are a good indication of the precision of each model.

From the practical perspective, for a model of a monthly cycle to be useful, it has to predict the day of ovulation and, ultimately, to determine the fertile window. Days inside the fertile window that were classified as infertile are false negatives. Please note that because of a possible application of a model like this in natural family planning, false negatives may be much more serious than false positives, so the model should minimize its false negative rate to zero. This is essentially the case with all fertility awareness methods. Days that were marked as fertile and were outside the fertile window are false positives. The smaller the false positive rate, the closer the predicted day of ovulation is to the real day of ovulation, which can be helpful for couples seeking pregnancy. In our experiment, as the gold standard, we followed Wilcox *et al.* [14], who define the fertile window as the period between day of ovulation minus five days and day of ovulation plus one day.

14.7 Results

Table 14.1 and Fig. 14.6 show the average percentage of fertile and infertile days during a woman's monthly cycle sorted in the descending order (i.e., the longest to the shortest infertile period). The number of days in which a woman should abstain from intercourse to prevent unplanned pregnancy is larger for lower order models. The smaller the false positive rate, the closer the predicted day of ovulation is to the real day of ovulation, which can be helpful for couples seeking pregnancy. The higher the order of the model, the lower the percentage of the false positives. The 7–th order DBN model was most precise and indicated the longest infertile periods and the shortest fertile periods.

False negatives (Table 14.1 and Fig. 14.7) are an important measure of accuracy of a FAM, because on one hand they may lead to unplanned pregnancy and on the other hand to less likely conception in case of couples seeking pregnancy.

Table 14.1 Average percentage of fertile and infertile days and false negatives/false positives during the monthly cycle for each of the compared DBN models.

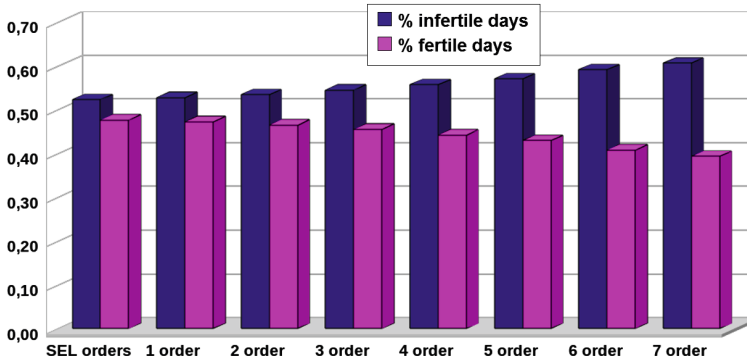| Method	% infertile days	% fertile days	% false negatives	% false positives
SEL orders	0.52 %	0.48 %	0.0000 %	22.84 %
1st order	0.53 %	0.47 %	0.0000 %	22.39 %
2nd order	0.53 %	0.47 %	0.0000 %	21.71 %
3rd order	0.54 %	0.46 %	0.0000 %	20.68 %
4th order	0.56 %	0.44 %	0.0008 %	19.42 %
5th order	0.57 %	0.43 %	0.0008 %	18.13 %
6th order	0.59 %	0.41 %	0.0008 %	16.00 %
7th order	0.61 %	0.39 %	0.0008 %	14.54 %

Fig. 14.6 Average percentage of fertile and infertile days during the monthly cycle for each of the compared DBN models.

Comparison of percentage of false negarites and false positives
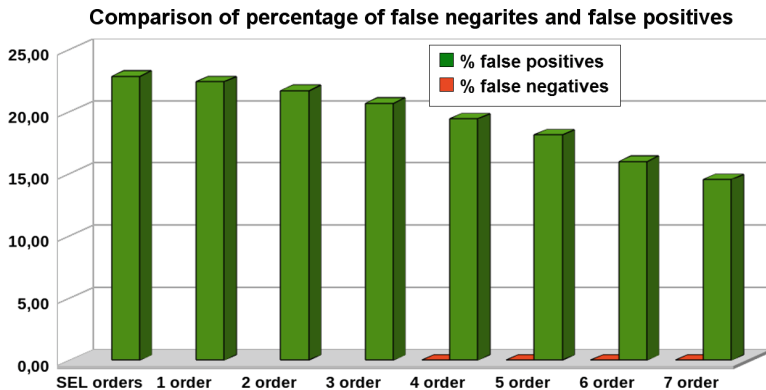

Fig. 14.7 False negatives and false positives during monthly cycle for each of the compared models.

Our results show that higher order models (4th through 7th) show non-zero false negative rate. We investigated this further and found that in each case there was an anomalous cycle, not recognized by the model. It seems that higher order models have the tendency to over-fit the data and be unable to deal with monthly cycles that deviate from typical cycles.

14.8 Conclusion

We have presented the results of an experiment with a series of DBN models monitoring woman's monthly cycle. We have shown that higher order models are more accurate than first order models, as summarized in Fig. 14.6. The lengths of the fertile period for higher order models were shorter, which indicates a better ability of the model to predict ovulation. The percentage of *false negatives* for all models was zero or very close to zero (0.0008 %). Higher order models tend to over-fit the data and have difficulty with anomalous cycles. While we advise to use higher order temporal models for systems with memory, we caution against too high order models when the system exhibit significant noise, as such models may over-fit the data and perform poorly when the course of events departs from typical.

References

1. Barron, M.L., Fehring, R.J.: Basal body temperature assessment: is it useful to couples seeking pregnancy? Am. J. Matern. Child Nurs. **30**(5), 290–296 (2005)
2. Colombo, B., Masarotto, G.: Daily fecundability: first results from a new data base. Demographic Res. **3**(5), (2000)
3. Dean, T., Kanazawa, K.: A model for reasoning about persistence and causation. Comput. Intell. **5**(2), 142–150 (1989)
4. Dunson, D.B., Sinai, I., Colombo, B.: The relationship between cervical secretions and the daily probabilities of pregnancy effectiveness of the TwoDay algorithm. Hum. Reprod. **16**(11), 2278–2282 (2001)
5. Kippley, J., Kippley, S.: The Art of Natural Family Planning, 4th edn. The Couple to Couple League, Cincinnati (1996)
6. World Health Organization: A prospective multicentre trial of the ovulation method of natural family planning. III. Characteristics of the menstrual cycle and of the fertile phase. Fertil. Steril. **40**(6), 773–778 (1983)
7. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Mateo (1988)
8. Potter Jr., R.G.: Length of the fertile period. Milbank Q. **39**, 132–162 (1961)
9. Rötzer, J.: Supplemented basal body temperature and regulation of conception. Archiv für Gynäkologie **206**(2), 195–214 (1968)
10. Royston, J.P.: Basal body temperature, ovulation and the risk of conception, with special reference to the lifetimes of sperm and egg. Biometrics **38**(2), 397–406 (1982)
11. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(1), 379–423 (1948)
12. Szymański, Z.: Płodność i Planowanie Rodziny. Wydawnictwo Pomorskiej Akademii Medycznej, Szczecin (2004)
13. Weschler, T.: Taking Charge of Your Fertility: The Definitive Guide to Natural Birth Control, Pregnancy Achievement, and Reproductive Health. Collins (2006)
14. Wilcox, A.J., Weinberg, C.R., Baird, D.D.: Timing of sexual intercourse in relation to ovulation effects on the probability of conception, survival of the pregnancy, and sex of the baby. N. Engl. J. Med. **333**(23), 1517–1521 (1995)