ABSTRACT

Electric load modeling during the design of ships has historically relied upon a minimal amount of information on each load: connected load and load factor for various operating conditions. While this approach has worked in the past, modern power-electronics based power systems call for more advanced modeling to properly account for cycling behavior, load aggregation at lower levels, common mode currents, grounding impacts, and control system interactions.

INTRODUCTION

Electric load modeling is a critical activity in the design of the electrical system of a ship. Electric load modeling is part of the process for conducting an Electric Power Load Analysis (EPLA) as described in T9070-A3-DPC-010/310-1 (also referred to as DPC 310-1). The EPLA is used to determine the required power or current rating of elements of the electrical system as well as predicting the amount of fuel that will be consumed to generate the electrical power over a period of time. DPC 310-1 defines the following methods for conducting an EPLA:

- Demand Factor Analysis
- Load Factor Analysis
- Zonal Load factor Analysis
- Stochastic Load Analysis
- Modeling and Simulation Load Analysis
- Quality of Service Load Analysis
- 24 Hour Average Load Parametric Equation

This paper describes different ways of modeling electric loads and aggregating these models to produce estimates of the load a given power system component will experience for a specific operating condition, ambient condition, and time scale. In forming aggregations of loads, it is important to understand the power system components for which the aggregation applies, and the specific loads that make up the aggregation.

The estimated load is typically an average over a given time scale. For fuel consumption calculations, the time scale is typically 24 hours. For determining the rating of power system components, the time scale depends on the ability or tolerance of the component to safely operate within a tolerance band around the estimated load. For many power system components, the time scale is based on thermal limitations. Power electronic components have much lower thermal capacities than steel and copper-based components; thus the time scale for power electronic components should be much less than for steel and copper based components.

Additional considerations including common mode behavior, grounding systems, and controls will also require new load models in the future.

CONNECTED LOAD MODEL

A very straightforward model of an electric load is to simply use its connected load. The connected load is the identification plate or rated power of the load. The total connected load at an aggregation point is the sum of the connected load of each individual load within the aggregation. Demand Factor Analysis multiplies this total connected load by a demand factor obtained from a curve that plots the demand factor vs the total connected load.

Demand factor analysis has traditionally been used to determine the current rating of cables, circuit breakers, and switchgear. All loads physically connected to the power system component are part of the aggregation. The basis for the curve used to determine the demand factor is not understood, but its use has generally been
satisfactory if there are many loads in the aggregation and no single load is a large fraction of the total connected load.

**LOAD FACTOR MODEL**

A load factor model estimates the load’s operating load by multiplying the load’s connected load with a load factor. For each load, load factors are assigned for combinations of ambient conditions and ship operating conditions. The total operating load for an aggregation of loads (for a given ambient condition and ship operating condition) is calculated by summing the applicable operating loads of all the loads within the aggregation. Margins and Service Life Allowances (SLAs) are typically applied to the total operating load to establish minimum power ratings for power system components. Margins account for uncertainty in the estimates and SLAs account for load growth while the ship is in service.

Determining the value to use for the load factor of a particular load under a particular operating condition and ambient condition is not always obvious and depends on the intended use of the calculated total operating load. Additionally, the choice of loads to include in an aggregation also depends on the intended use of the calculated total operating load. The load factor model facilitates the calculation of the total operating load for an aggregation by simplifying the model of a load to a factor applied to the connected load. Unfortunately, the complexity of many loads complicates the establishment of a reasonable load factor, particularly in early stages of design.

The current version of DPC 310-1 does not directly address the importance of specifying the time scale. The importance of doing so can be illustrated by considering a cycling load such as the one depicted in Figure 1. Presuming the connected load for this particular load is 12 kW, then if the time scale is 24 hours which is much greater than the cycling load period, the load factor can be set equal to the duty cycle of the load (1 minute on / 6 minute period) or 0.17. MIL-E-7016F suggests that in addition to 24 hours (or continuous), time scales of 5 seconds and 5 minutes should also be considered. Five seconds may be appropriate for loads where the source is power electronics based, while 5 minutes may be appropriate for when sources are generators and transformers. If loads are buffered by energy storage of sufficient power and energy capability, then the time scale is derived from the energy storage and not the actual source; 5 minutes may be appropriate for loads with power electronic sources.

The load factor should be chosen to reflect the highest operating load over the time scale. For the 5 second time scale, the load factor would be 1.0 since the load could be on for the full duration of the 5 seconds. For the 5 minute time scale, the maximum time on is 1 minute (even though the load could be off for the full 5 minutes) which translates into a 1 min / 5 min = 0.2 load factor. If the load period (on time + off time) is shorter than the time scale, then using the load duty cycle is acceptable, even though strictly the highest operating load divided by the time scale could be somewhat larger.

![Cycling Load Profile](image)

Figure 1: Cycling Load with 6 minute period

Hence for the example depicted in Figure 1, the load factor as a function of time-scale is depicted in Figure 2. Note that if the power used while on is less than the connected load, the load factor must be scaled accordingly.
Figure 2: Load factor for cycling load (60 seconds on 300 seconds off) vs time-scale

Since the load factor method removes the time variable by averaging over a time interval, the resulting total operating load is not sensitive to the phase relationship of loads; the method assumes the worst case phasing of loads. Consequently, the total operating load is a worst case estimate for the average power of the aggregation of loads over the time scale.

DPC 310-1 provides a description of the loads that should be incorporated into aggregation of loads for both total ship calculations and for zonal calculations. For loads that are not constant (such as cycling loads) zonal load factors are modified from the load factors to account for greater variability of load due to having fewer non-constant loads in the aggregation. The greater variance in average load due to having fewer loads can be derived from the Bienaymé formula. For the special case where all the loads have the same variance, the variance of the mean value of the sum of n uncorrelated loads is equal to the variance of each load divided by n. Hence, as the number of loads increases, the variance of the sum of the loads decreases.

From the above discussion, when capturing data to form a load factor model of a cycling load, one should capture the cycling profile. For each operating condition and ambient temperature, the amount of time the equipment is online and offline should be recorded. Insight from the amount of time a load is online and offline should help in deciding whether to include a load into a specific aggregation or not. In addition to the connected load, for the time when the equipment is online, the “on” time, “off” time, load while “on” and load while “off” should all be recorded.

Let:

**t\_ave\_on** When online, the average time (sec) the equipment is in the “on” mode

**t\_ave\_off** When online, the average time (sec) the equipment is in the “off” mode

**P\_ave\_on** When online, the average power (kW) for the equipment in the “on” mode

**P\_ave\_off** When online, the average power (kW) for the equipment in the “off” mode

**P\_connect** The connected load for the load.

The load factor \( L_f \) for a given time scale \( t_s \) is given by:

If \( t\_ave\_on \geq t_s \)

\[
L_f = \frac{P\_ave\_on}{P\_connect}
\]

If \( t\_ave\_on + t\_ave\_off \geq t_s \)

\[
L_f = \frac{P\_ave\_off}{P\_connect} + \left( \frac{P\_ave\_on - P\_ave\_off}{P\_connect} \right) \left( \frac{t\_ave\_on}{t_s} \right)
\]

If \( t\_ave\_on + t\_ave\_off < t_s \)

\[
L_f = \frac{P\_ave\_off}{P\_connect} + \left( \frac{P\_ave\_on - P\_ave\_off}{P\_connect} \right) \left( \frac{t\_ave\_on}{t\_ave\_on + t\_ave\_off} \right)
\]

For loads that have multiple “on” modes, the time and load for each mode should be recorded, as well as the transition logic from one mode to another. For these loads with multiple modes, it may first be advantageous to produce a stochastic model, then derive a load factor from the stochastic model. In any case, making changes to the cycling profile to reflect differences in the specific loads or the load application from the ship the data was captured from is likely more traceable than attempting to adjust the load factor directly.
If the “on time” and “off time” have a significant variance with respect to their averages, then an alternate, more conservative approach to establishing the load factor may be warranted. In this alternate method, a time sequence of load power samples (for when the load is online) is divided into a sequence of $m$ time windows of length equal to the time scale $t_s$. Ideally, $m$ should be large, on the order of 100 or more. For each time window, the average power $P_{\text{ave}_w}$ is calculated. The load factor is set equal to the largest of the $m$ calculated $P_{\text{ave}_w}$ values divided by the connected load $P_{\text{connect}}$.

If $m$ is not sufficiently large, but still greater than about 10, the maximum of the moving average of the load divided by the connected load may be used.

When analyzing a time sequence of load measurements, it may be beneficial to calculate the load factor for several time scales; the results can be plotted similar to Figure 2.

When calculating the zonal load factor with this alternate method, the peak load during the time window of the maximum average power should be used for the peak operating load.

This alternate method is suitable for use by any load but will likely result in a conservative estimate for the total operating load of an aggregation because it assumes all the loads experience their highest average power over a time window at the same time. While this is possible, it is highly unlikely if there are many cycling loads in an aggregation.

**VOLTAGE IMPACT OF OVERLOAD**

Before discussing stochastic models, it is important to gain an understanding of what happens to the system voltage due to an overload. How much of an overload can be tolerated for a short time period without causing violations of power quality requirements?

If all the loads in an aggregation are resistive, then the loads can be combined into a single equivalent resistance $R$. For d.c. systems, at the nominal system voltage $V_n$, the power $P$ is given by:

$$P = \frac{V_n^2}{R}$$

If however, the source is current limiting to $i_{\text{limit}}$, then the power consumed by the loads is simply

$$P = i_{\text{limit}}^2 R$$

When the source is current limiting but at the nominal system voltage, the source is providing its rated power $P_{\text{rated}}$:

$$P_{\text{rated}} = i_{\text{limit}} V_n$$

$$i_{\text{limit}} = \frac{P_{\text{rated}}}{V_n}$$

The voltage is simply:

$$V = i_{\text{limit}} R = R \frac{P_{\text{rated}}}{V_n}$$

Now the power demanded by the load is the power of the load at the nominal systems voltage

$$P_{\text{demand}} = \frac{V_n^2}{R}$$

$$R = \frac{V_n^2}{P_{\text{demand}}}$$

Or combining ...

$$\frac{V}{V_n} = \frac{P_{\text{rated}}}{P_{\text{demand}}}$$

If the demand is 10% greater than the rating of the source, the voltage will drop about 9%. This only applies for the case of resistive loads and current limiting sources. For this case, if a 9% voltage drop is at the limit of being tolerable, one would want the possibility of any overload to be small (say 5%) and the possibility of a 10% overload to be very very small (say 0.1%).

Constant power loads will lead to a greater voltage drop and a general voltage collapse if the power demand from the constant power loads is greater than the rated power of the source. If the constant power loads are only a fraction of the
total load and less than the rated power of the source, the voltage can be calculated as:

$$V = \frac{P_{\text{rated}} - P_{\text{CPL}}}{P_{\text{demand}} - P_{\text{CPL}}}$$

Note that the demand power includes the power required by the constant power loads. Based on this relationship, one would want the probability of the constant power loads to exceed the rated power of the source to be nearly zero. The probability of power quality violating requirements should be very, very small. In general, the mean time between service interruptions (i.e. the voltage not meeting the interface standard for a specified period of time) should meet Quality of Service (QOS) requirements. See IEEE 45.3 and T9300-AF-PRO-020 for more details on QOS. Since many of the load aggregation methods do not allow for a direct approximation of the mean time between service interruptions, keeping the probability of such interruptions extremely low enhances the chances of meeting QOS requirements.

Some loads will have a behavior that can be modeled as a resistive load and a constant power load in parallel.

**STOCHASTIC MODELING**

In a stochastic model of a load, random variables, expressed as probability density functions (PDFs) and cumulative distribution functions (CDFs) are used to describe the behavior of a load. While there are many ways in which this can be done, this paper will focus on three.

a. **Pure Stochastic.** A pure stochastic model represents the load as a single random variable. This model is created using a time sequence of load samples by:

i) Creating a sequential set of “bins” to divide the total range of power from 0 to the connected load into $k$ sequential bins of equal size (kW). Each bin therefore has a size equal to the connected load divided by $k$. $k$ should be chosen so that the resulting bin size is several times the measurement error for each load sample. One would expect $k$ to fall between roughly 20 and 100. Each bin has an assigned minimum and maximum power level such that any power between 0 to the connected load can be assigned to one and only one bin.

ii) For each bin, the number of samples in the time sequence that fall between the minimum and maximum bin power levels is counted and assigned to the bin.

iii) Once all samples from the time sequence have been assigned to a bin, the probability of a bin is calculated by dividing the number of samples assigned to the bin divided by the sum of all the samples in all the bins. The PDF of the load is a discrete PDF where the load value for a bin is typically assigned the average of the minimum and maximum bin power level with probability equal to that assigned to the bin. This distribution may also be called a Probability Mass Function.

The probability of an aggregation is the sum of the random variables of the loads in the aggregation. The Monte Carlo method described in DPC 310-1 is likely the most straightforward way of calculating the PDF of the total load.

The question now becomes how to use the PDF / CDF of the total load to determine the rating of the equipment associated with the aggregation of loads. As a consequence of the previous section, it seems to make sense to use the CDF in the following way:

iv) For the highest bucket load value with a non-zero probability, determine what rated power below this value will not result in the voltage falling out of the tolerance range of the interface standard.

v) Using the CDF determine the power level associated with being able to serve the total load for a large fraction of the time (typically between 95% and 99%).
vi) Use the higher of the values from (iv) and (v) above as the total load. To this number apply a reduced margin (to account for missing loads) and a service life allowance to calculate the minimum power rating for the equipment associated with the aggregation of load.

The advantage of this method is that it is independent of time scale. For a given load, the PDF and CDF can be calculated once and used for any application without consideration for the properties of the source.

The disadvantage of this method is that it is independent of time scale and does not take advantage of the short-term overload capability of a source. The overload capability is derived from the interaction of the loads to the overload behavior of the sources.

b. **Time Scale Stochastic.** This model attempts to correct the disadvantage of the Pure Stochastic model. Instead of using each sample in the time sequence directly, the time sequence is divided into a series of windows of duration equal to the time scale, and then use the average values of the time sequence within each window to create the PDF: In this way, very high power levels for very short durations are averaged out. This model may result in a lower minimum power rating for the equipment associated with the aggregation of load.

c. **Time Domain Stochastic:** The pure stochastic and time scale stochastic models work well for converting a time sequence of load measurements for a specific load. However, the resulting load model, while accurate for the ship from which the data was taken, may not apply directly to the ship design under consideration. It may be desirable to adjust the load profile to reflect different operational concepts, system configurations, and environments. Making these adjustments in the time domain, while still capturing the stochastic nature of the load may prove beneficial.

We revisit the simple cycling load represented in Figure 1 and define the following random variables:

- $x_{on}$ When online, the PDF of the time duration (sec) when the equipment is in the “on” mode
- $x_{off}$ When online, the PDF of the time duration (sec) when the equipment is in the “off” mode
- $x_{Pon}$ When online, the PDF of the load power (kW) when the equipment in the “on” mode
- $x_{Poff}$ When online, the PDF of the load power (kW) when the equipment in the “off” mode
- $P_{connect}$ The connected load for the load.

Each of the PDFs can readily be determined from the time sequence of load measurements. If needed, the PDFs can be adjusted as needed to reflect the design under study. The above model can be sampled to create a synthetic time sequence of load measurements from which another model can be created. Alternately the model can be used directly in a Monte Carlo simulation.

If the load has more than one “on” state, a Markov Chain model may be better suited. Markov Chains are described in Appendix A (Doerry and Koenig 2017). To apply Markov Chains to a time sequence of load measurements, the range of power levels from 0 to the connected load is divided into $j$ different load power ranges (kW) where each range corresponds to a unique load power state. The size of the load power ranges (kw) need not be equal. The time sequence of load measurements is partitioned in a series of time window of time duration equal to the time scale. The average value over the time scale is computed for each time window then assigned the appropriate load power state.

Once every time window is assigned a load power state, the following is accomplished:
For each load power state, the PDF and CDF are calculated for the average power level for the applicable time windows.
ii) The probability of being in each load power state is calculated across all the time windows
iii) The transition matrix which indicates the probability of transitioning from one state to another at the window boundaries.

Once again, this model can be modified to reflect changes in the application for the current study.

From this model one can calculate a load factor (sum the products of the probability of being in each state and the mean load of the state and divide the results by the connected load)

One can also create a synthetic time sequence of load measurements from which another model can be created. Alternately the model can be directly used to calculate the PDF of an aggregation of loads.

**MODELING AND SIMULATION**

DPC 310-1 provides the following recommendation for modeling and simulation load analysis:

“Situations where specific loads are large compared to the generation or power system component capacity, have unusual electrical characteristics, require large amounts of rolling reserve not normally reflected in load averages, or when the correlation of many loads is complex and cannot be adequately modeled using one of the other methods described here, may benefit from the use of modeling and simulation load analysis.”

This type of analysis requires time domain quasi-static models that include quasi-static control system behavior. The addition of the control system behavior differentiates this type of modeling from the other methods described in this document. Essentially, the other analyses presumed the loads within an aggregation all acted predominately independent of each other.

In modeling and simulation, the cross-dependency of loads is modeled.

The modeling and simulation environment can contain elements of the other types of load models described above. Only those loads that are correlated with other loads, or are dependent on control system behavior, need to have the requisite additional modeling detail.

The modeling and simulation model of the power system should be exercised many times with the initial conditions randomly varied. Using the results from the many runs of the model, the PDF of the total operating load can be calculated.

**LOAD MODEL CHALLENGES**

This paper is focused on modeling loads. Historically, only the differential mode models of electric loads within a ship have been considered. The differential mode reflects how the electric load is intended to operate. Hence, the differential mode model evolves from the design of the load and is well understood.

Consider the different approaches to modeling loads that are discussed above. The differential mode models of electric loads are still necessary; however, they are no longer sufficient to ensure that electric power system capacities are adequate.

a. **Challenges Brought by Power Electronic Converters:** Building upon the relevant published works, Doerry and Amy (2018) discuss the nature of common mode behavior that arises in systems containing power electronic converters. Figures 3 – 5 shows this for the simple example of a three phase uncontrolled rectifier; the common mode voltage is the voltage offset between the neutrals of the a.c. and d.c. sides of the rectifier. Common mode behavior is a consequence of power electronic converters - even as simple as a diode rectifier, parasitic
coupling of equipment and circuits and ground, and circuit imbalances. 

![Diagram of three phase rectifier](image)

Figure 3: Three phase rectifier

![Waveforms and Neutral Offset of d.c. System with Respect to Neutral of a.c. System](image)

Figure 4: D.c. Waveforms and Neutral Offset of d.c. System with Respect to Neutral of a.c. System (Doerry and Amy 2018)

![Waveforms and Neutral Offset of a.c. System with Respect to Neutral of d.c. System](image)

Figure: 5 D.c. Waveforms and Neutral Offset of a.c. System with Respect to Neutral of d.c. System (Doerry and Amy 2018)

i) Common Mode Sources: The simple three phase uncontrolled rectifier example referred to above is a source of common mode currents with frequencies which are lower order harmonics of the system fundamental frequency (60 Hz usually). Advanced Pulse Width Modulated, and other types of switched power electronic converters are common mode sources that inject currents with frequencies that reflect the converters’ switching frequency and its harmonics, spanning perhaps from 10s of kHz to MHz. Electric loads that contain rectifiers, switched power supplies, inverters et cetera are all common mode current sources with potentially a wide, rich frequency spectrum.

ii) Capacitance, and inductance, at inputs / outputs: Many electric loads on ships are designed with ‘low-pass’ filters on their inputs to avoid ‘electromagnetic interference’ entering from the power system which contain capacitance and inductance. These filters are designed to be effective for a specified range of frequencies. Many electric loads on ships are designed with features, sometimes filters, intended to reduce the injection of current harmonics from that load. Many sources in power systems have output filters which are intended to ‘smooth’ an output waveform, reducing the frequency spectrum of the output waveform. In addition to these discreet capacitances and inductances, the parasitic capacitances, and leakage inductances, of electric loads and the cables that supply them must be considered as well when one is concerned with understanding the common mode current paths.

iii) The switching frequency, \( f_{\text{sw}} \) determines the \( \Delta t \) in \( \frac{\Delta v}{\Delta t} \) and \( \frac{\Delta i}{\Delta t} \): Consideration of a PWM waveform demonstrates how switching frequency develops high-frequency injection into a power system. For a simplistic two-level, 50% duty cycle, and 10% rise / fall time, the frequency content of such a PWM waveform as deduced from its Fourier coefficients provides a relatively benign spectrum of injected frequencies (figure 6). If this waveform has a switching frequency of 10kHz, the period is 100\( \mu \)s. Even-numbered harmonics within a Fourier series representing the waveform do not contribute. Odd-numbered harmonics have an amplitudes greater than on the order of 1\%
of the fundamental through the 17th harmonic, which for the example is 170kHz. At 170kHz and 1% of the fundamental current/voltage amplitude, the di/dt or dv/dt of the 17th harmonic component would be on the order of 17,000A/s or V/s. As expected, the ratio of the derivative of a time scaled waveform is proportional to the waveform frequency (See Figure 7). Less benign duty cycles and steeper rise / fall times coupled with higher switching frequencies increase the number of harmonics which contribute energy to the system.

b. Challenges Brought by ‘The Unintended Circuit’: As mentioned earlier, the EPLA is used to determine the required power or current rating of elements of the electric system. In the case of ground and common mode circuits, the questions arise, what should the current carrying capacity be for ground circuits, and, what is an appropriate current level to endeavor to limit common mode current? These two questions apply just as well to load equipment as to power system equipment. Ground circuits are typically explicitly designed; hence, for example, high-resistance ground resistors and leads can be designed to handle an arbitrary current level. Common mode current, on the other hand, frequently flows in parasitic paths and shields. These paths have not hitherto been explicitly designed to carry current. How then to ensure that these paths can carry the current and not cause deleterious effects in the load and power system equipment? What do loads ‘look’ like at frequencies well above fundamental and even fsw? How to model this? What does the power system ‘look’ like at those same frequencies? Are there frequencies where the power system and some of the loads interact through ‘the Unintended Circuit’?

c. Challenges Brought by Controls: When considering models of electric loads, their dynamic behavior over a millisecond time scale is relevant to control system actions. Robinett and Wilson (2011) and Wilson et al. (2014) assign three time scales, “Agent/Informatics Energy Management & Reasoning Mode” (update steady state set-points / ethernet network) – o(seconds), “Energy Storage Hamiltonian Controls” (network communications and hardwired feedback) – o(msecs), and “Servo Layer” (converters, accepts set-points, hardwired) – o(μsecs). Doerry and Amy (2016) also assign three time scales, “Supervisory Control Time Scale” (100msec – 10seconds), “Outer Loop Control Time Scale” (1msec – 1 second), and

Different waveforms driven by different types of modulation will have different injected spectra. At some point, modeling the spectral content of electric loads will be necessary. This will be complicated further if fast converter controls switch between different modulation schemes.
“Inner Loop Control Time Scale” (100nsec – 1msec).

For an example system with a switching frequency of 20kHz, the switching period is 50µs. If an “Inner Loop Control Time Scale” occurs at 10x the switching period, then its time scale is on the order of 0.5ms (500µs). If an “Outer Loop Control Time Scale” occurs at 100x the switching period, then its time scale is on the order of 5ms. What this implies is that as shipboard power systems’ networks and supervisory controls become faster with their time scales moving closer to tens of milliseconds, electric loads with “Outer Loop Control Time Scales” that are closer to tens of milliseconds create the situation where control systems may begin to interact in unplanned ways.

d. Implications for Electric Load Models in the Future
To address the challenges in designing future power systems, electric load models in the future must incorporate a much better characterization than the simple load factor models of the past:

i) Common Mode plus Differential Mode models will be required. Swept Frequency Response Analysis and its digital model analog, or some similar tool, are needed to characterize loads, and the power system.

ii) Impedance characterizations across a wide frequency spectrum are necessary (but not sufficient) to assess system performance and system level stability.

iii) The models must reflect an understanding of controls time scales and their interactions

OBSERVATIONS AND CONCLUSIONS
For any given aggregation, the modeling method requiring the least effort while still providing satisfactory results should be chosen. Easier said than done. Experience will provide guidance as to which method to use under which conditions. If a less complex but conservative method forces a step increase in source cost or ship impact, modeling using one of the alternate methods may result in a lower cost system and should be pursued. On the other hand, if the extra fidelity gained in increased modeling effort is not likely to result in a favorable change to the design, then one should probably forgo the extra fidelity modeling.

REFERENCES


Dr. Norbert Doerry is the Technical Director of the Naval Sea Systems Command Technology Office. In addition to leading special projects, he facilitates the transition of technology from industry and academia into naval warships. He retired from active duty in the U.S. Navy in 2009 as a Captain with 26 years of service operating, designing, constructing, and repairing naval ships and naval power systems.

Dr. John Amy Jr. P.E. is the Senior Technologist for Naval Power Architectures Technologies at the Naval Surface Warfare Center Philadelphia, PA, USA.

APPENDIX A: INTRODUCTION TO MARKOV CHAINS (From Doerry & Koenig 2017)

A Markov process, named after Andrey Markov, is a stochastic process where the system under study has multiple states, and the transition from the current state to the next state in a time increment is stochastically dependent on the current state, but not upon any previous (or future) states. For example, Figure 8 depicts a three-state process where the states are represented by the letters A, B, and C. The arrows represent the possible state transitions and the associated number is the conditional probability that given that the system is in the state at the base of the arrow, the transition will occur to the state at the end of the arrow during the following time increment. The sum of the probabilities of the arrows leaving a state adds up to 1.0; there is a 100 percent probability of transitioning, including transitions to the same state. For Figure 8, if the current state of the process is state A, then there is a 70 percent chance that the process will remain in state A, a 20 percent chance that the process will transition to state B, and a 10 percent chance that the process will transition to state C.

![Figure 8: Example Markov Process](image)

In a number of references, the transpose of this matrix is called the transition matrix. In this format, $A$, then there is a 70 percent chance that the process will remain in state A, a 20 percent chance that the process will transition to state B, and a 10 percent chance that the process will transition to state C.

$x_n$ is a row vector rather than a column vector depicted above: $x_{n+1} = x_n P$


\[ P = \begin{bmatrix}
0.7 & 0.5 & 0.3 \\
0.2 & 0.3 & 0.3 \\
0.1 & 0.2 & 0.4
\end{bmatrix} \]

As a result of the constraint that the probability of transitioning from a state is 100 percent, the sum of the elements in each column of \( P \) is equal to 1.

If \( x_n \) is a stochastic vector with elements equal to the probability of system being in each of the three states at time \( n \), then the probability of the system being in each of the three states \( x_{n+1} \) at time \( n+1 \) is given by:

\[ x_{n+1} = Px_n \]

For example, if the system is currently in state A, then the current stochastic vector is

\[ x_n = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \]

In the next time increment the probability of being in each state is given by:

\[ x_{n+1} = \begin{bmatrix} 0.7 & 0.5 & 0.3 \\ 0.2 & 0.3 & 0.3 \\ 0.1 & 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0.2 \\ 0.1 \end{bmatrix} \]

Similarly, the probability of being in each state at time increment \( n+2 \) is given by:

\[ x_{n+2} = \begin{bmatrix} 0.7 & 0.5 & 0.3 \\ 0.2 & 0.3 & 0.3 \\ 0.1 & 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} 0.7 \\ 0.2 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 0.62 \\ 0.23 \\ 0.15 \end{bmatrix} \]

\( x_{n+2} \) can also be calculated by multiplying \( P \) by itself before multiplying it to \( x_1 \).

\[ x_{n+2} = \begin{bmatrix} 0.7 & 0.5 & 0.3 \\ 0.2 & 0.3 & 0.3 \\ 0.1 & 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} 0.7 & 0.5 & 0.3 \\ 0.2 & 0.3 & 0.3 \\ 0.1 & 0.2 & 0.4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.62 & 0.56 & 0.48 \\ 0.23 & 0.25 & 0.27 \\ 0.15 & 0.19 & 0.25 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.62 \\ 0.23 \\ 0.15 \end{bmatrix} \]

Note that the stochastic vector at \( n+m \) can be determined by applying the \( P \) transition matrix \( m \) times...

\[ x_{n+m} = P^m x_n \]

Hence if the current value of the state is known \( (x_n) \) this equation enables one to stochastically calculate the value at a desired starting year in the future \( (x_{n+m}) \) and thus provides a method for determining the initial value of a Markov chain. Another method is to observe the system over some time period, calculate the probability of being in each state, and apply the resulting probabilities to determine the initial value.

If \( m \) becomes very large, \( P^m \) converges to the following matrix

\[ \lim_{m \to \infty} P^m = \begin{bmatrix} 0.58 & 0.58 & 0.58 \\ 0.24 & 0.24 & 0.24 \\ 0.18 & 0.18 & 0.18 \end{bmatrix} \]

because the columns of this matrix are identical:

\[ \lim_{m \to \infty} x_{n+m} = \begin{bmatrix} 0.58 \\ 0.24 \\ 0.18 \end{bmatrix} \]

Hence the long-term steady state probability of being in each state is independent of the original state and a function only of the transition matrix. This can be shown through an eigendecomposition of \( P \) to be generally true for the types of transition matrices normally encountered (see for example Grinstead and Snell 1997). One can use this long-term steady state probability as an alternate way to determine the initial value for a Markov chain.

**REFERENCE**