Logist. Res. (2012) 5:133–139
DOI 10.1007/s12159-012-0087-8

ORIGINAL PAPER

Modeling production planning and transient clearing functions

Dieter Armbruster · Jasper Fonteijn · Matt Wienke

Received: 10 August 2012 / Accepted: 24 August 2012 / Published online: 7 September 2012

Abstract The production planning problem, that is, to determine the production rate of a factory in the future, requires an aggregate model for the production flow through a factory. The canonical model is the clearing function model based on the assumption that the local production rate instantaneously adjusts to the one given by the equilibrium relationship between production rate (flux) and work in progress (wip), for example, characterized by queueing theory. We will extend current theory and modeling for transient clearing functions by introducing a continuum description of the flow of product through the factory based on a partial differential equation model for the time evolution of the wip-density and the production velocity. It is shown that such a model improves the mismatch between models for transient production flows and discrete event simulations significantly compared to other clearing function approaches.

Keywords Production planning · Transient clearing functions · Continuum models

1 Introduction

The production planning problem is a well-studied problem in industrial engineering. Fundamentally, it involves finding the correct starts into a factory such that production meets demands. The problem is complicated by two different major issues: stochasticity and nonlinearity. Stochasticity manifests itself through the uncertainty of the demand and the variation of any demand realization. In addition, variations in the production speed and quality introduce other fundamental stochastic processes. While demand fluctuations are covered via suitably sized and placed inventories, stochasticity in the production process leads to variable lead times to refill these inventories.

Note that stochasticity is the more fundamental issue than nonlinearity, since the latter is generated by the former via queueing: Nonlinearity is generated by the fact that the variable lead times do not only depend on the stochastic processes that impact the production, it is mostly generated by waiting in queues. Such waiting depends crucially on the amount of material produced concurrently, that is, the wip. Specifically, the lead times increase dramatically together with the lengths of the queues, if the flux through the factory approaches the capacity limit of the factory. A typical scenario goes like this: Demand is projected to increase at a certain time in the future. Meeting demand requires increasing the start rate into the factory by a lead time earlier than the requested delivery time. However, increasing the start rate will increase the wip in the factory and as a result increases the cycle time—the time a product needs to completely go through the factory increases. The resulting nonlinear optimization is at the core of the production planning problem.

1.1 Clearing function

The baseline for all modeling in production systems is given by discrete event simulation (DES) models where every part, every machine and every production step are modeled with (different) probability distributions

D. Armbruster (✉) · M. Wienke
School of Mathematical and Statistical Sciences,
Arizona State University, Tempe, AZ 85287-1804, USA
e-mail: armbruster@asu.edu

J. Fonteijn
Department of Mechanical Engineering, Eindhoven University of Technology, 513, 5600MB Eindhoven, The Netherlands

characterizing the specific stochastic process responsible for uncertain steps. As the characterization of these stochastic processes is nontrivial and as such simulations are very expensive, aggregate deterministic models that represent *average* behavior have been developed.

The canonical aggregate model has become known as a clearing function, first introduced by Karmarkar [9], and as Betriebskennlinen in German by Wiendahl [12]. The clearing function can be defined for any size of production unit, that is, a group of machines, a production line, a full factory or even a supply chain. It is a state equation that defines the outflux F of the production unit as a function of the wip W in steady state in that unit, that is,

$$F = \Phi(W). \tag{1}$$

The functional form of the clearing function Φ has been determined in many different ways: Measured in real factories, modeled via an $M/M/1$ queue (i.e., a queue with exponentially distributed arrivals and exponentially distributed machine processing times), modeled after the fundamental diagram of a traffic model [10], etc. (see e.g., [5, 6]) ,

$$\begin{array}{ll} \Phi = \frac{\mu_0 W}{1+W} & M/M/1, \\ \Phi = \mu_0 W - W^2 & \text{fundamental diagram of traffic.} \end{array} \tag{2}$$

Aouam et al. [2] notice that the clearing function can be approximated by piecewise linear functions, making the production planning problem an Integer-LP optimization problem.

Notice that the clearing function is used with a wip level that is a function of time and hence models the outflux as a function of time. The fundamental assumption here is known as the adiabatic or quasi-steady assumption: The wip level changes slowly relative to the damping time of the underlying stochastic process. Hence, the outflux is never transient and instantaneously relaxes to its steady-state behavior.

Missbauer [11] extends the clearing function concept to capture transient phenomena in a three-parameter clearing function. He shows that the outflux of a system depends on the initial wip of the system, the expected number of arriving lots and the probability distribution for sampling the initial wip. He studies an $M/M/1$ single-server queue with infinite buffer, a mean arrival rate λ and a mean machine process rate $\mu = 1$. The number of lots in the system (queue plus machine) at the beginning of period t is denoted by $W(t)$ the wip in the workstation. Missbauer studies a version of the clearing function that characterizes the expected outflux at time t over a time interval $[t, t + T]$ denoted by X_t. The expected outflux is a function of the expected load $E[L_t]$ in the system given by the initial wip at time t and the new arrivals over the time interval

$$E[L_t] = W_t + A_t, \tag{3}$$

$$A_t = \int_t^{t+T} \lambda(s) ds. \tag{4}$$

For constant arrival rate $\lambda = \lambda_c$ we get $\lambda_c = \frac{A_t}{T}$. For time varying influx a DES was used. For further reference, we define an decreasing and a increasing influx

$$\lambda_D(t) = \frac{\lambda_c}{(t/T + 1/2)ln(3)}, \tag{5}$$

$$\lambda_I(t) = \frac{\lambda_c}{(t/T - 3/2)ln(3)}, \tag{6}$$

corresponding to a linear interpolation of the inter-arrival times between the two steady states related to the initial queue and the queue associated with $E[L_t]$. Missbauer's experiments were done for a time interval of $T = 5$. Figure 1 shows the expected output generated as averages of DES simulations for constant λ with five different initial wips. The dependence on the initial wip is obvious.

2 Transient clearing functions

While Missbauer restricted himself to a constant arrival rate, we have studied arrival rates that vary over the simulated time period but that lead to the same expected total load. We investigated five cases and generated clearing functions as a function of the initial wip like Fig. 1 for five different influx protocols: (1) constant influx, (2) instantaneous influx at the beginning of the time period, (3) instantaneous influx at the end of the time period, (4) a monotonic decreasing influx rate (Eq. 5) and (5) a monotonic increasing influx rate (Eq. 6). Figures 2a, b shows the

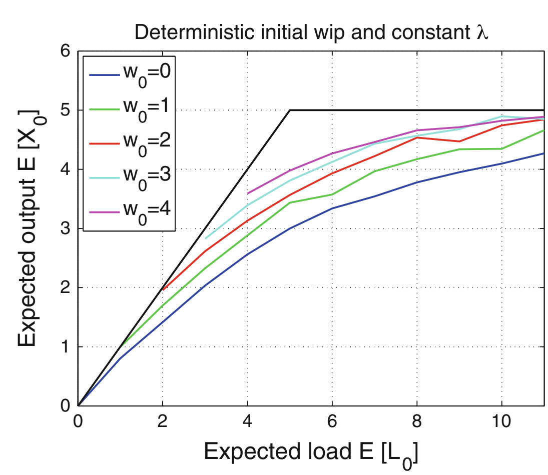

Fig. 1 Expected total ouflux for an $M/M/1$ queue for a time interval of five mean cycle times as a function of the expected total load

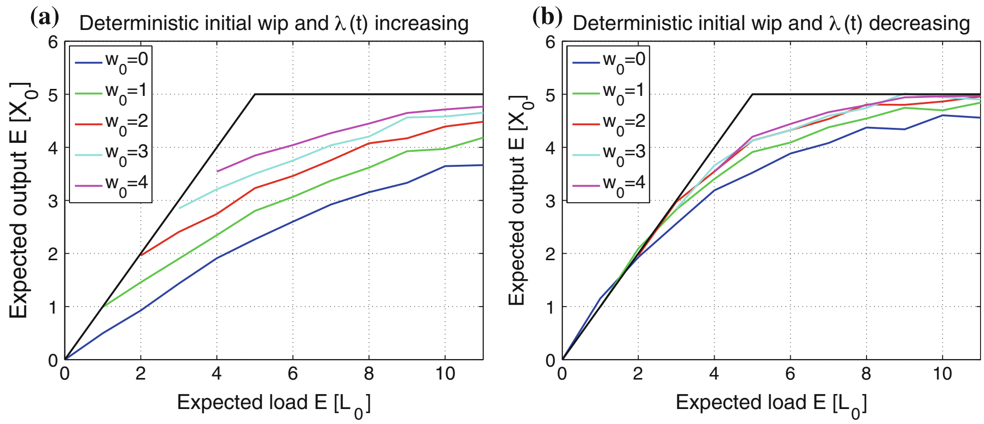

Fig. 2 Clearing functions for an *M*/*M*/1 queue with different initial wips and (**a**) increasing, (**b**) decreasing influx protocols

expected outflux for an increasing and a decreasing influx. Since a decreasing influx introduces more of the load into the system early in the time period, it is not surprising that the variance in the outflux due to the initial wip becomes smaller than for increasing influx. The extremes of these cases are instantaneous influx at the very beginning of the time interval leading to high outflux almost independent of initial wip, and instantaneous influx at the end of the time interval leading to outflux based entirely on the initial wip. The conclusion of these DESs is that instead of extending the clearing function concept to three parameters as suggested in [11], the influx–outflux relationship in a transient setting is much more complicated: In addition to the total load, the functional form of the influx over the time interval of interest is highly important, and therefore, the clearing function cannot be just a parametric relationship between input and output.

The applicability of Missbauer's result [11] that the probability distribution of the initial wip has a big influence on the clearing function needs to be clarified further: We can imagine two fundamentally different scenarios for the experiments described above:

1. Production has been halted, and the state of the system can be examined. Hence, the wip in the system is known exactly. When production is resumed, the initial wip is known deterministically.
2. Alternatively, one might want to plan a transition of the state of the factory from a steady state to another steady state, and initial wip may only be known in the mean but no specific sample will be taken to determine the actual initial wip at the beginning of the planning period. In that case, the initial wip follows the geometric probability distribution associated with the probability of finding an *M*/*M*/1 queue at a particular level for a given arrival process and a given exit process.

Figure 3 shows the clearing functions for different mean initial wips for constant influxes. Figures 1 and 3 report on the same experiment—the difference is that in the former the initial wip is known, whereas in the latter the clearing function is an average over many samples taken from the steady-state distribution associated with the mean initial wip. We confirm that increasing variance of the initial wip leads to a lower outflux. However, the striking result of Fig. 3 is the fact that the dependence of the clearing function on the mean initial wip is almost completely gone.

3 Continuum models

3.1 Transport equations

A clearing function model gives an instantaneous relationship between outflux and wip in steady state. Since it is used to model an influx that changes in time, it will not be able to model the delay associated with the production time and waiting in the factory [5]. This is the fundamental reason why the clearing function cannot be parameterized by a finite number of parameters but depends on the complete history of the influx function. Attempts for transient clearing functions will therefore always be restricted to the special experimental setups.

In an attempt to design a complete time-dependent theory of production flows, we have therefore in recent years developed an aggregate theory of production flows based on standard transport equations studied in physics, especially in fluid mechanics and in some traffic models [3, 4]. Transport equations are partial differential equations that describe the time and space evolution of a density under an influx. In our case the spatial variable is given by the degree of completion of the part or the stage of the production. We scale the stage or completion variable $x \in [0, 1]$ and define

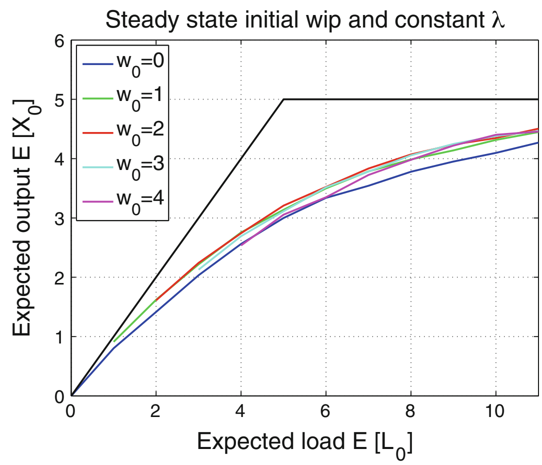

Fig. 3 Expected ouflux for an *M*/*M*/1 queue as in Fig. 1. Here, the outflux represents an ensemble average over a steady-state probability distribution for the initial wips with a mean as indicated

density of parts at stage x at time t by $\rho(x, t)$. If the fluid moves with a velocity field $v(x, t)$, then the flux is described as $F(x, t) = v(x, t)\, \rho(x, t)$. Mass conservation then is given by the partial differential equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial F}{\partial x} = 0. \tag{7}$$

Since $v(x, t) \geq 0$, the fluid moves from left to right, allowing a boundary condition to be imposed at $x = 0$. Typically, the boundary condition is $F(0, t) = \lambda(t)$, that is, the local flux at zero is the arrival rate of the parts into the factory. Together with an initial wip profile $\rho(x,0) = \rho_0(x)$, this sets up a well-defined hyperbolic problem. Notice that we are describing a flow that is continuous in its parts and continuous in its spatial direction. This should be distinguished from the so-called fluid equation models of queueing theory [8] which are continuous in its parts but describes a flow through a finite and distinct number of queues, leading to a set of ordinary differential equations (ODEs).

In [4] we extended the fluid analogy even further and derived macroscopic transport equations from kinetic models leading to Boltzmann equations, which is akin to deriving the Euler equations of fluid dynamics from first principles based on Newton's law. Defining a particle density $f(x, v, t)$ describing the number of parts at state x at time t moving with a velocity v in completion space, we derive equations for the first moments of this density of the form:

$$\frac{\partial \rho(x,t)}{\partial t} + \frac{\partial v(x,t)\rho(x,t)}{\partial x} = 0, \tag{8}$$

$$\frac{\partial v(x,t)}{\partial t} + v(x,t)\frac{\partial v(x,t)}{\partial x} = 0. \tag{9}$$

An initial value problem appropriate for the DES experiments described in Sect. 1.1 can be defined by setting $\rho(x,0) = w_0$ and $v(x,0) = v_0$ with w_0 and v_0 constants.

3.2 Boundary conditions

Determining the right boundary conditions to describe the DES experiments is the major modeling issue here. Equations (8, 9) are a set of hyperbolic partial differential equations whose solutions travel from left to right as long as $v(x,t) > 0$. Hence, boundary conditions have to be imposed at the boundary $x = 0$, and the outflux at the other boundary $x = 1$ is a result of the transport. Clearly, the flux has to be given by the production start rate, hence

$$\rho(0,t)v(0,t) = \lambda(t). \tag{10}$$

The other boundary condition is based on the relationship between queuing theory and Eq. (9): Equation (9) is Burgers equation and can be solved via characteristics. Hence, ignoring the initial conditions, after a while the solution $v(x, t)$ is determined by the value of the velocity at the boundary. As a result, a mass $\rho\, dx$ arrives at the boundary and travels downstream with the velocity it acquires at the moment of arrival at the boundary. Translating this into the *M*/*M*/1 setting and defining the velocity at the boundary as $v(0,t) = \frac{1}{\text{cycle time}}$ we see that the velocity $v(0,t)$ should depend on the queue length $w(t) = \int_0^1 \rho(x,t)\mathrm{d}x$ at the moment a part arrives at the end of the queue.

The problem therefore reduces to finding the expected cycle time, conditioned on the length of the queue. For a steady-state queue, the cycle time is determined by the PASTA (Poisson Arrivals See Time Averages) property of *M*/*M*/1 queues: In steady state a part arriving at the end of queue will find an average queue length w_0, and the resulting cycle time for this part will be $\tau = \frac{1}{\mu}(1 + w_0)$. Hence,

$$v_{ss}(t) = \frac{\mu}{1 + w(t)} \tag{11}$$

is the velocity related to the well-known *M*/*M*/1 clearing function (cf Eq. 2).

The same PASTA property gives us the initial condition: At the beginning of the experiment, we have an initial wip w_0 that we assume is a known deterministic quantity. This initial wip is the length of the queue. For a Markov process, the history of arrivals—whether they arrived in packets or spaced out—is not important. Hence, we can use the average cycle time formula for a *M*/*M*/1 queue or the heuristic extensions discussed below to determine the initial condition for the velocity in the factory.

To improve on the steady-state result requires significantly more queuing theory machinery: Since we are describing transient phenomena, the system is not ergodic any more and hence ensemble averages and time averages are not the same. We therefore need to be more specific about the "expected" cycle time. The natural setup following the experiments in Sect. 2 is to determine the probability distribution of the cycle times given that we restart a factory with a given initial wip w_0 over many instances of this scenario—that is, we are interested in the ensemble average, conditioned on the initial wip.

Our current models for the expected value of the cycle time are preliminary and based on fitting heuristic boundary condition models to the DES. We distinguish two regimes:

1. If the production start rate $\lambda(t)$ is less than the mean production rate μ, then we expect that any initial wip distribution exponentially fast decays to the wip distribution associated with the steady state related to the arrival rate. Hence, the boundary condition is determined by the solution to an ordinary differential equation

$$\frac{\mathrm{d}v(0,t)}{\mathrm{d}t} = -\sigma(v(0,t) - v_{ss}(t)) = -\sigma\left(v(0,t) - \frac{\mu}{1+\int_0^1 \rho(x,t)\mathrm{d}t}\right) \quad (12)$$

where the decay constant σ will be determined experimentally.

2. If the production start rate $\lambda(t)$ is bigger than the mean production rate, there is no associated steady state since the queue length will become unbounded. In this case, the cycle time at arrival of a part at a queue length of $w(t)$ will become just $\tau = \frac{1}{\mu}w(t)$ which would lead to a velocity equation of

$$v_{hw}(t) = \frac{\mu}{w(t)}. \quad (13)$$

It turns out that for small wip and for $\lambda - \mu << 1$, this model creates a velocity that is too high and hence the production in the PDE simulations is overestimated relative to the DES. This is due to the fact that basing the ensemble average only on the stochastic properties of the exit process it not a good model in these cases since for small wips, machines do occasionally idle as a result of missing arrivals. We settled for a model that averages between the steady state Eq. (11) and the high wip model (13) of the form

$$v(t) = \frac{\mu}{0.5 + w(t)}. \quad (14)$$

Hence, the full boundary conditions for Eq. (9) become

$$\begin{aligned} v(0,t) &= \frac{\mu}{0.5+\int_0^1 \rho(x,t)\mathrm{d}x} && \text{for } \lambda \geq \mu, \\ \frac{\mathrm{d}v(0,t)}{\mathrm{d}t} &= -\sigma\left(v(0,t) - \frac{\mu}{1+\int_0^1 \rho(x,t)\mathrm{d}t}\right) && \text{for } \lambda < \mu, \\ v(0,0) &= \frac{\mu}{0.5+\int_0^1 \rho(x,t)\mathrm{d}x}. \end{aligned} \quad (15)$$

The last equation describes the initial condition for the ordinary differential equation. It is based on the assumption of a deterministic initial condition, that is, the initial wip is exactly known, and hence, the ensemble average will be mostly affected by the stochasticity of the machine process and little affected by the stochasticity of the arrival process.

4 Numerical results

We have been reproducing the DES of Sect. 2. Since there are only a small number of lots involved in these simulations, the discretization error between the DES and the partial differential equation becomes an issue. In the discrete case, wip measures whole lots whereas the continuous model registers infinitesimally small lots. This is not a problem for large wips, but for these experiments, partial lots in the PDE are counted earlier than they really appear in the DES, and hence, they lead to lower velocities than in the DES. We compensate for this by calculating wip with a floor function, that is, $w(t) = \frac{1}{2}\lfloor 2\int_0^1 \rho(x,t)\mathrm{d}x\rfloor$. In that way the PDE system observes partial lots only after half of the lot has already appeared. Figure 4a and b compares the outflux of the PDE simulations for different constant influxes and initial wip of $w_0 = 0$ and $w_0 = 3$ with the corresponding DES. The clearing functions of the DES for these wips as well as others for different initial wips are very well reproduced by the PDE simulations. The decay constant σ has been adjusted to give the best fit of the two curves over all the data points. Notice that the best fit depends on the initial wip: For $w_0 = 0$ the best fit is $\sigma = 2.3$, indicating fast relaxation to the steady state, and for $w_0 = 3$ the best fit is $\sigma = 0.3$, indicating a very slow relaxation that has not yet equilibrated after the five time intervals used in the experiment.

The advantage of a PDE simulation becomes apparent in the Figs. 5a and b which shows the clearing functions for an influx that corresponds to a linearly increasing inter-arrival time and a linearly decreasing inter-arrival time for an initial wip of $w_0 = 1$. Although the decreasing case Figs. 5a shows a slight overproduction of the PDE compared to the DES, the overall trend of the PDE simulations captures the DES simulations very well.

5 Conclusion

We have developed a PDE model for a transient *M*/*M*/1 queuing experiment representing the most simplified case of a production model. We used a coupled system of evolution equations for the part density and the velocity to describe the production system as transport equation and showed that the crucial modeling aspect of the problem is the boundary condition of the velocity equation.

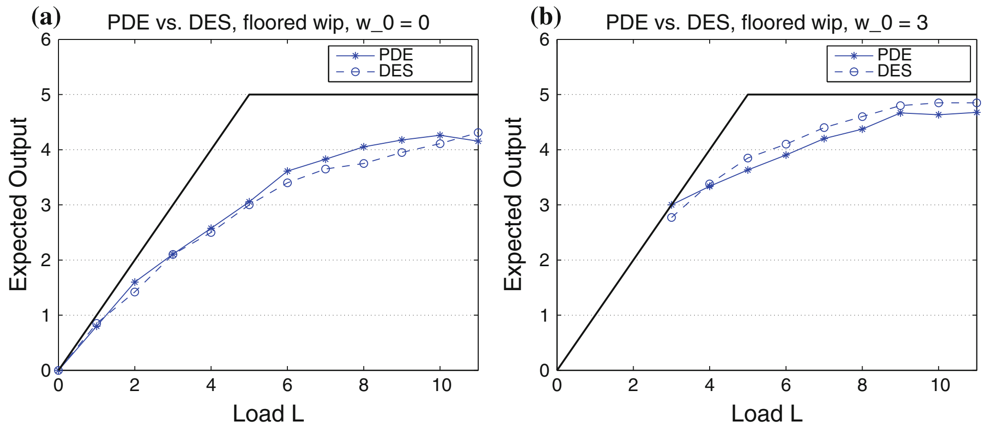

Fig. 4 Outflux over five time intervals as a function of the total expected load for the PDE model (8, 9) with boundary conditions (10) and (15) and constant influx. **a** shows the DES simulation for an initial wip of $w_0 = 0$ and the corresponding PDE simulation, **b** Initial wip $w_0 = 3$

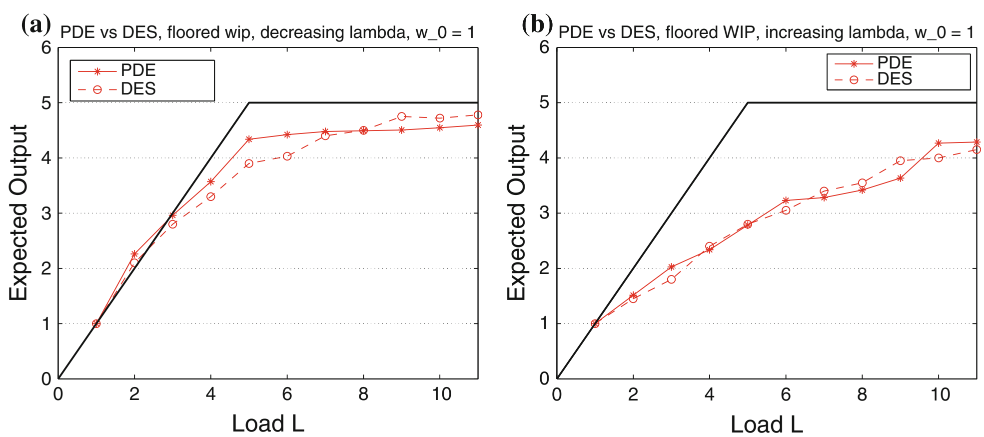

Fig. 5 As in Fig. 4 with initial wip $w_0 = 1$. **a** shows a decreasing influx rate, **b** an increasing influx rate

We showed that a heuristic model based on exponential relaxation of initial queue lengths to their steady-state values given by $M/M/1$ queuing theory in addition to a high wip limiting model of queueing behavior leads to very good agreement between the PDE simulations and the DES. Specifically, we compare the two approaches visually via plots presenting the expected throughput over five time units for an ensemble average of repeated experiments as a function of the average total load in the factory similar to clearing function models. The mean relative error for each of the clearing functions experiment is of the order of 6 %, which is far better than any other modeling approach for these experiments. Overall, the heuristics requires data fitting of a single decay parameter (σ) and a global choice of the functional dependence of the ensemble average of the velocity for the case when there is no steady state.

The current state of the project to model the ensemble average of transient behavior of production systems is clearly unsatisfactory. While the heuristic model presented here is a clear improvement for any practical considerations of the production planning problem that can easily be implemented for a practical code, the theoretical state of the model is very unsatisfactory. Research based on exact and approximate solutions of transient queueing theory [1, 7] is currently underway to bridge the gap between theoretical and heuristic models.

Acknowledgments This project was started when D.A. was a part time professor at the Department of Mechanical Engineering, Eindhoven University of Technology. Helpful discussions with Erjen Lefeber and Ivo Adan on transient queueing models are gratefully acknowledged. Discussions with Hubert Missbauer and Reha Uzsoy helped to clarify the experiments on transient queuing and the relationship between clearing function models and the production planning problem. This research was supported by a grant from the Volkswagen Foundation under the program on Complex Networks.

References

1. Abate J, Whitt W (1991) Transient behavior of the M/G/1 workload process. Oper Res 42(4):750–764

2. Aouam T, Uzsoy R (2012) An exploratory analysis of production planning in the face of stochastic demand and workload-dependent lead times. In: Armbruster D, Kempf K (eds) Decision policies for production systems. Springer, London, pp 173–209
3. Armbruster D, Marthaler D, Ringhofer C, Kempf K, Jo T-C (2006) A continuum model for a re-entrant factory. Oper Res 54(5):933–950
4. Armbruster D, Marthaler D, Ringhofer C (2004) Kinetic and fluid model hierarchies for supply chains. SIAM Multiscale Model. Simul 2(1):43–61
5. Armbruster D (2012) The production planning problem: clearing functions, variable leads times, delay equations and partial differential equations. In: Armbruster D, Kempf K (eds) Decision policies for production systems. Springer, London, pp 289–302
6. Asmundsson JM, Rardin RL et al (2009) Production planning models with resources subject to congestion. Naval Res Logist 56:142–157
7. Berstimas D, Mourtzinou G (1997) Transient laws of non-stationary queueing systems and their applications. Queueing Syst 25:115–155
8. Bramson M (2008) Stability of queueing networks, lecture notes in mathematics, 1950. Springer, Berlin
9. Karmarkar US (1989) Capacity loading and release planning with work-in- progress (wip) and lead-times. J Manuf Oper Manag 2:105–123
10. Lighthill MJ, Whitham GB (1955) On kinematic waves. II. A theory of traffic flow on long crowded roads. Proc R Soc Lond A 229:317–345
11. Missbauer H (2010) Order release planning with clearing functions: a queueing-theoretical analysis of the clearing function concept. Int J Prod Econ. doi:10.1016/j.ijpe.2009.09.003
12. Wiendahl H-P (1995) Load oriented manufacturing control. Springer, Berlin