Grind Size Is Not a Guess: Method and Evidence Behind the mikron. Grind Recommendations

Max J. Hassenstein, PhD

November 2026

mikron. Coffee Tools | mikron.coffee

Contact: support@mikron.coffee

Summary

mikron. recommends a grind setting from a model whose predictions can be checked, and it learns from the corrections a brewer logs after tasting. This paper describes how the recommendation is produced and states, for each part, what evidence stands behind it: one part is fit to brew data, one part learns from the brewer and checks itself on the brewer's phone, and the remainder is reasoned from the literature and not yet tested. All brew data are my own, from one taster and one grinder. On 74 drip brews made after the model's base equation was fixed, the model alone missed the grind I settled on by 1.0 clicks on average, and by 0.6 clicks once the app had learned from my earlier corrections.

---

1. The problem

Among the variables a home brewer controls, the grind setting is one of the hardest to transfer. A setting is specific to one grinder. It shifts with the amount of coffee brewed, and it depends on conditions most brewers do not track, such as the age of the beans and the hardness of the water. Many brewing apps and grinder charts answer this with a table: one brew method, one range of settings. Such tables are useful, and community knowledge can be more current than any single dataset. However, a table cannot tell a brewer what to change between a 500 ml and a 1,000 ml batch. Furthermore, a table cannot be tested in the same way, as it makes no prediction for a specific brew.

I built mikron. on the opposite premise: a recommendation should be a prediction, and a prediction should be checked against what happened in the cup. In this paper, I describe how

the recommendation is produced, report how well it has predicted so far, and name what remains untested. The base equation is given in full, so that a reader can verify at least one claim by hand.

---

2. The idea: predict, then check

The conceptual target of every recommendation is a balanced extraction. The usual shorthand reads a sour cup as too little extraction and a bitter cup as too much, which is a starting point rather than a full description [3]. For filter coffee, the classic reference is the ideal region of the Coffee Brewing Control Chart, which goes back to Lockhart [1]: an extraction yield of 18–22% and 1.15–1.35% total dissolved solids. Extraction yield is the share of the dry coffee that dissolves into the cup, and total dissolved solids measure the strength of the brew. Sensory work has shown how sourness, bitterness, and sweetness move across that space [3]. I use this region as a conceptual frame only: no refractometer was used at any point, and I do not report extraction yields.

What the app records instead is the brewer's verdict after tasting, such as "spot on", "one click finer", or "two clicks coarser". A click is one step of a grinder's adjustment, and it is the unit in which a brewer experiences an error. The brewed setting plus the verdict gives the grind the brewer settled on. For example, a brew ground at 40 clicks and judged "one click finer" has a settled grind of 39 clicks. For any verdict other than "spot on", this is the setting the verdict implies, not a second cup that was tasted. This settled grind is the yardstick of this paper: every part of the system that can be scored is scored by how far its recommendation was from it, and Section 7 names the parts that cannot be scored yet.

The recommendation is produced and corrected in four steps:

1. Predict: a model converts the brew (method, amount, bean, water) into a grind size on a micron scale (µm, thousandths of a millimeter; Section 3).
2. Translate: the micron target is converted into a setting on the brewer's own grinder.
3. Learn: logged verdicts adjust the next recommendation for that brewer.
4. Diagnose: when a cup misses, described symptoms are mapped to the changes most likely to help, the most likely first.

Three rules govern the system. First, the model's prediction remains the anchor: what is learned is recomputed from the logbook each time and never overwrites the model. Second, the app names on screen the factors that shaped the number, shows the size of any personal adjustment in microns, and lets the brewer switch personal learning off. Third, a constant of the drip model changes only if a rule fixed before the analysis is met; Section 5 reports one case in which this

rule did not permit a change I was inclined to make.

The parts of the system differ in how much evidence they have (Table 1). The sections that follow are ordered by that evidence, strongest first.

Table 1. The parts of the system and the evidence behind each as of November 2026.

Part	How it works	Evidence to date
Grind model for drip brewers (Sections 4 and 5)	One equation plus four corrections	Fit on 114 brews and tested on 74 later drip brews, one taster
Personal learning (Section 6)	A standing personal correction and a recall of a recent verdict	The same logbook, re-run in order; checks itself on each phone
Grinder translation (Section 7)	Step size per grinder; a fitted curve where none is published	Sourced figures, none measured by me
Other brew methods (Section 7)	Grind ranges per method with the same corrections	Published ranges and brewing literature
Diagnosis and guided dial-in (Section 7)	Symptom scoring; one change per brew in the guided dial-in	Brewing literature and fixed test cases

3. The logbook

All brew data in this paper come from my own logbook: one taster, one grinder (Comandante C60 Baracuda), and mostly one machine, the Moccamaster KBG Select. Grind sizes are stated on a reference scale defined by this grinder, namely a setting in clicks times 21 µm, the step size its maker gives; no particle size was measured. The Moccamaster is an automatic drip brewer with a wedge-shaped filter basket of the Melitta type and a single off-center outlet, and it holds the brew temperature itself. I used its half-jug setting for batches below 800 ml and the full-jug setting from 800 ml, so that this setting followed the batch size. The water was untreated municipal tap water in Hanover, Germany (total hardness 214 mg/L and alkalinity 114 mg/L, both as $CaCO_3$, according to the supplier's annual report). This is hard by the app's scale and by US conventions, and medium on the German scale. I brewed drip coffee with soft water only during one period of less than two weeks, and once with water of medium hardness.

The record analyzed here ends on November 3, 2026, and holds 247 brews, 200 of them on

the Moccamaster. They come from 28 bags; one further brew has no bag linked. I fixed the model's base equation (Section 4) in July 2026 on the 114 Moccamaster brews logged on this grinder by then, which came from 17 bags roasted medium-light to medium-dark. A brew logged afterwards is one the base equation has never seen, and such brews form the test set of this paper. Of the 107 later brews with a verdict, 95 can be replayed; the other 12, none of them on the Moccamaster, carry no saved recommendation. These 95 span all brew methods, and 74 of them are drip brews from 15 bags, 11 of which were first brewed after the fit. Soft-water brews enter the test set only if they were made after the soft-water correction (Section 4) was refit.

---

4. The model

The machine holds the brew temperature, and over the 87–93 °C range tested, temperature had little sensory effect once strength and extraction were matched [2]. The main variable I model is therefore the amount of coffee in the basket. A deeper bed tends to slow the water and to extract more at the same grind, and grind size acts on extraction through surface area, diffusion distance, and bed resistance [4,5]. The grind therefore coarsens as the batch grows. I model this as

$$g = a + b \cdot \ln\!\left(\frac{V \cdot c}{60}\right), \qquad a = -49.8, \quad b = 146$$

where g is the base grind in µm, V the brew water in ml, c the brew ratio in g of coffee per liter, and 60 g/L the reference ratio; $V \cdot c/60$ is thus the water in ml that the same amount of coffee would take at 60 g/L. Since $V \cdot c$ is proportional to the coffee mass, the equation states that the grind depends on the logarithm of that mass: doubling the batch at the same ratio coarsens the grind by $146 \cdot \ln 2 \approx 101$ µm, about five clicks on my grinder. For example, 750 ml at 60 g/L on water of medium hardness gives a base of $-49.8 + 146 \cdot \ln(750) \approx 917$ µm, which is 44 clicks on the Comandante C60 and 31 clicks on a Comandante C40 MK4 at 30 µm per click. The app rounds the drip result to the 21 µm steps of my grinder and therefore shows 924 µm, which gives the same clicks. I chose the logarithm over a straight line as it fit the data clearly better (Note 1), which may reflect the diminishing effect of each additional gram on a bed that is already deep.

Four additive corrections adjust this base: bean freshness, processing method, a blend term, and water hardness. The drip model has no term for roast level (Section 5). Fresh beans hold CO_2 that they release over weeks [6], and in my data very fresh bags needed a coarser grind and old bags a finer one; I report this as an empirical pattern and make no claim about its mechanism. Dissolved minerals change what water extracts [7]. The model grinds coarser for hard water and

finer for soft water, with water of medium hardness as the reference that carries no correction; the direction agrees with [7], but the size of each step comes from my data. The water term uses hardness only and does not model alkalinity. The freshness steps and the soft-water correction for drip were refit on logged brews, whereas the processing terms rest on one or two bags each, and the honey term is set from brewing knowledge.

Since October 2026, a constant of the drip model changes only under a rule that I fixed before running the analysis of Section 5. The change must improve the prediction for bags the fit has not seen by at least a quarter of a click. In addition, the 95% interval of that improvement must exclude zero (Notes 2 and 3).

---

5. How well the model predicted

I report errors as the mean absolute error, that is, the average size of the miss regardless of its direction, and each range in parentheses is a 95% interval (Note 3). On the 114 brews it was fit to, the model reproduced the grind I settled on with a mean absolute error of 18.1 µm at the time of the fit, which is less than one click. That figure flatters the model, since the equation was tuned to exactly those brews. The more informative figure comes from the test set: on the 74 later drip brews, the mean absolute error was 1.03 clicks (0.73–1.35).

However, the later brews also show a bias. On the hard-water brews among them (n = 70), the model's number was 16.4 µm too coarse on average (6.9–26.1 µm), which is about three-quarters of a click, and the shift arose from September 2026 onward. I was inclined to move the model, and I tested the move under the rule of Section 4 on the hard-water drip brews with a verdict. Refitting the constant a with each bag held out in turn improved the prediction for that bag by 0.06 clicks (−0.09 to +0.21). When I refit a on the brews before each month and predicted that month, the gain over the later half of these brews was 0.18 clicks (0.03 to 0.30), which excludes zero but remains below the threshold. Hence, the model stayed as it was.

One observation, made after the fact, is consistent with that decision. On hard water, brews ground two clicks finer than the model's number were judged "spot on" in 11 of 11 cases (10 of 10 among those made after the fit). This is consistent with a taster who tolerates a finer grind, although it does not exclude a model that runs too coarse (Section 8). If my verdicts are lenient on the fine side, a verdict of "finer" persists within a bag without a counterweight: once a brew has moved finer, a tolerant palate answers "spot on" rather than "coarser", and the grind I settle on can drift finer regardless of what the beans require. Moving the base would then have shifted every new user's first recommendation in order to match one taster's tolerance.

I also set a term to zero in August 2026, before the rule of Section 4 existed. The model once

coarsened substantially for robusta, on the assumption that a pure robusta should need a coarser grind than a robusta blend. However, the first pure-robusta bag I brewed needed a grind nearly three clicks finer than the term implied. The remaining blend term is confounded with roast level and with bag identity, and my data cannot separate them.

---

6. Learning from the brewer

A general model cannot know one brewer's palate, technique, or grinder. Two mechanisms therefore learn from the brewer's own verdicts.

The first is a personal offset, a standing correction to the model's number. Each logged brew contributes the grind the brewer settled on. An explicit verdict is weighted more strongly than a high star rating without a verdict, and recent brews more strongly than old ones. The estimate is then pulled toward the model: a single brew moves the recommendation only part of the way, and the offset approaches the brewer's own mean as consistent evidence accumulates (Note 4). Once the bag in use has enough brews, only its brews count; before that, the offset draws on other bags of the same roast level or, failing that, on all bags from the same grinder, and it is pulled more strongly toward the model. The offset is bounded. Each brew enters as its difference from the model under the conditions it was made in, so that a verdict given on a small batch of fresh beans carries over to a large batch of older ones under the additive form of the model (Section 8).

The second mechanism recalls a single verdict and infers nothing. If a recent brew of the same bag carried a verdict that the recommendation does not yet reflect, the app recalls it. At the same batch size and strength, it offers the setting that verdict implies; at a clearly different batch size or strength, it only reminds the brewer of the verdict and offers no number. A "spot on" is recalled as well if the recommendation has since moved away from it. The recall is withdrawn once the personal offset has absorbed the verdict, so that no correction is counted twice.

To test whether learning helps, I replayed my logbook in order. Each brew was predicted only from the brews logged before it, and the prediction was scored against the settled grind before that brew was used for learning (Note 5). On the 74 later drip brews, the two mechanisms together reduced the mean absolute error from 1.03 clicks to 0.64, a paired difference of −0.39 clicks (−0.65 to −0.15; Table 2). On the drip brews of bags first brewed after the fit, from the fourth brew of a bag onward, the difference was −0.63 clicks (−0.91 to −0.32). Across all brew methods (n = 95), misses of two clicks or more fell from 27% (15–41%) with the model alone to 12% (5–19%) with both mechanisms, a paired difference of −16 percentage points (−33 to −3). A simpler correction, which adds the model's mean error over the last 15 drip brews to

its number, came close: 0.73 clicks on the same 74 brews, or 0.09 clicks (−0.13 to +0.30) worse than the two mechanisms. The two mechanisms are therefore a design choice that this logbook does not show to be necessary.

Table 2. Mean absolute error in clicks on drip brews made after the base equation was fixed, with 95% intervals. The replay scored each brew before learning from it.

Brews	Model alone	With personal offset	With offset and recall
Later drip brews (n = 74, 15 bags)	1.03 (0.73–1.35)	0.82 (0.63–1.02)	0.64 (0.51–0.77)
New bags, fourth brew onward (n = 41, 11 bags)	1.17 (0.76–1.58)	0.76 (0.53–0.98)	0.54 (0.39–0.71)

Across the full logbook, after a correction of at least one click, the next brew of the same bag at the same volume applied it in 26 of 33 cases, and 22 of those 26 were then judged "spot on". However, learning did not help on the first brew of a bag, where nothing specific to that bag is known yet. On the first brews of the 11 bags first brewed after the fit, the mean absolute error was higher with learning (1.18 clicks) than with the model alone (1.00 clicks). Both are descriptive counts on few brews.

This replay covers one brewer. mikron. contains no analytics, and I receive no brew data automatically; a brewer can choose to share an export. I therefore could not notice whether the personal offset made someone else's recommendations worse. The app runs the same kind of replay on the device, on the brewer's own logbook, each time it assembles a recommendation. If the personal offset has predicted the brewer's settled grind worse than the model alone over the recent brews, by more than a fixed margin, the app pauses the offset, says so, and resumes it when the data support it again. This check covers the built-in brew methods and user-defined methods with a fixed base; user-defined volume-aware methods are not yet graded.

---

7. Reasoned, not yet tested

The remaining parts rest on published work and on reasoning from brewing practice. Automated tests check that they behave as specified, which does not establish benefit.

The model's output in microns is converted to a grinder setting through a database of 55 grinders, each with a step size and a reachable range. None of these figures was measured by me. They are manufacturer specifications, values computed from published ranges, community measurements,

or fits to the makers' own pour-over charts, and 48 of the 55 are flagged as approximate in version 1.0.5 of the app. The chart fits replaced the specifications of 22 hand grinders in November 2026: their makers state how far the burr moves per click, which is not the change in grind size the model works with, and the two differed by a factor of 1.3 to 3.9. For a grinder with no published scale, the app reconstructs a personal curve from dial setting to microns, based on the settings the owner already uses for known brew methods (Note 6). The curve is constrained to rise steadily, since a coarser dial setting cannot produce a finer grind; this is the one property I can require without measurement.

For pour-over, immersion, and espresso, recommendations come from grind ranges per method with the same family of corrections. However, these ranges follow established practice for each method and were not fitted to my own data, most of which are drip brews. I therefore treat these recommendations as reasoned starting points, which the brewer's verdicts then correct as described in Section 6.

When a cup misses, three diagnostic engines (espresso, percolation such as pour-over and drip, and immersion) map reported symptoms to recommended changes by weighted scoring and rank them. The espresso engine distinguishes 13 root causes and gives precedence to distinctive combinations. For instance, the engine reads sourness together with bitterness as a sign of uneven flow through the puck, the bed of coffee in the portafilter, and physical models show how uneven flow can over-extract one part of a bed while under-extracting another [8,9]. The engine therefore addresses puck preparation before grind size. A guided dial-in then changes a single variable per brew and narrows the step when the brewer reports that the cup got worse. About 1,600 fixed test cases check that the engines behave as specified; each pairs a combination of symptoms with the diagnosis it must produce. They do not measure how often a followed diagnosis improves the next cup.

---

8. Limitations and open questions

This work has limitations, and they follow from its data: one taster, one grinder, mostly one machine, and mainly one water. First, the yardstick is my own verdict, which may be lenient on the fine side (Section 5); the personal offset learns from the same verdicts and would inherit that leniency. Second, the intervals rest on the 11 to 15 bags of the two test sets and are probably too narrow. Third, the replay of Section 6 uses brews the base equation has never seen, but the settings of the personal offset were tuned on the same logbook. Fourth, the four soft-water brews in the test set come from three bags that were also used to refit the soft-water correction, so they are not an independent test of it. Fifth, the additive form of the corrections assumes that they do not interact, and my data are too few to test that assumption.

I see five open questions, and for each I state what would settle it.

1. The fitted base may not hold for other palates and grinders. The later shift could be a model error or my own tolerance. A deliberate series on one bag, ground progressively finer until "coarser" is first given, would locate my acceptable band. Regardless of its result, changing the base for other brewers would require at least a second taster or a second grinder.
2. The brew-ratio term is not tested. The model responds to the amount of coffee, so at the same amount of coffee a different ratio leaves its prediction unchanged. About four in five of my Moccamaster brews were made at exactly 60 g/L, so this behavior is carried by the model's mathematical form; the data neither confirm nor contradict it. Pairs of brews from one bag with the same amount of coffee at clearly different ratios would test it.
3. The water terms are barely tested outside hard water, as all soft-water drip brews stem from one period of less than two weeks. For each family of brew methods, one bag brewed with one recipe on two waters would test the term.
4. Personal learning may not help brewers other than me. This cannot be answered from my logbook, and the app collects no data automatically. It requires logbooks that users choose to share.
5. The benefit of a diagnosis is unknown. The measure is whether the brew after a followed diagnosis was rated better, which the logbook can record.

---

9. Conclusion

In this work, I have described a grind recommendation that is a prediction and can therefore be checked, and I have stated for each of its parts what evidence exists. The empirical evidence rests on one taster's data, which is the principal limitation and the reason this paper reports methods and evidence side by side.

From my point of view, the position a tool of this kind can defend is neither "the model knows" nor "it is all taste". A model can supply a reasoned first setting, the brewer's verdict can correct it, and each claim made along the way can be stated together with the evidence behind it. mikron. is available for iPhone on the Apple App Store; further information is at mikron.coffee.

Statistical notes

1. Model form. The logarithmic and the linear form were compared by the Akaike information criterion (AIC), a measure of model fit that penalizes complexity [10]. On the same 83

brews from 17 bags, drawn from a 102-brew set and fitted with a separate intercept for each bag, the AIC of the logarithmic form was lower by 24–31, depending on the linear variant it was compared with.

2. Change rule. The improvement is measured leave-one-bag-out: the candidate is refit with each bag held out in turn and scored on the held-out bag. The 0.18-click figure in Section 5 is a walk-forward comparison: the constant a is refit on all brews before each month and scored on that month, and the gain is reported over the later half of these brews (n = 45).
3. Intervals. Where I report an interval, it is a 95% bootstrap interval from 2,000 resamples. Whole bags are resampled, not single brews, since the brews of one bag are not independent. A wider interval means a less certain estimate.
4. Personal offset. The estimate is a shrinkage estimator in the spirit of empirical Bayes [11], with fixed shrinkage constants.
5. Replay. The evaluation is prequential: each prediction uses only earlier outcomes and is scored before its own outcome is used [12].
6. Grinder curve. The curve is a monotone piecewise cubic interpolant [13].

References

1. Lockhart EE. The soluble solids in beverage coffee as an index to cup quality. Publication No. 27. New York: Coffee Brewing Institute; 1957.
2. Batali ME, Ristenpart WD, Guinard J-X. Brew temperature, at fixed brew strength and extraction, has little impact on the sensory profile of drip brew coffee. *Scientific Reports.* 2020;10:16450.
3. Frost SC, Ristenpart WD, Guinard J-X. Effects of brew strength, brew yield, and roast on the sensory quality of drip brewed coffee. *Journal of Food Science.* 2020;85(8):2530–2543.
4. Cordoba N, Fernandez-Alduenda M, Moreno FL, Ruiz Y. Coffee extraction: a review of parameters and their influence on the physicochemical characteristics and flavour of coffee brews. *Trends in Food Science & Technology.* 2020;96:45–60.
5. Moroney KM, Lee WT, O'Brien SBG, Suijver F, Marra J. Modelling of coffee extraction during brewing using multiscale methods: an experimentally validated model. *Chemical Engineering Science.* 2015;137:216–234.
6. Wang X, Lim L-T. Effect of roasting conditions on carbon dioxide degassing behavior in coffee. *Food Research International.* 2014;61:144–151.
7. Hendon CH, Colonna-Dashwood L, Colonna-Dashwood M. The role of dissolved cations in coffee extraction. *Journal of Agricultural and Food Chemistry.* 2014;62(21):4947–4950.
8. Cameron MI, Morisco D, Hofstetter D, Uman E, Wilkinson J, Kennedy ZC, et al. System-

atically improving espresso: insights from mathematical modeling and experiment. *Matter.* 2020;2(3):631–648.

9. Lee WT, Smith A, Arshad A. Uneven extraction in coffee brewing. *Physics of Fluids.* 2023;35(5):054110.
10. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control.* 1974;19(6):716–723.
11. Efron B, Morris C. Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association.* 1975;70(350):311–319.
12. Dawid AP. Present position and potential developments: some personal views. Statistical theory: the prequential approach. *Journal of the Royal Statistical Society, Series A.* 1984;147(2):278–292.
13. Fritsch FN, Carlson RE. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis.* 1980;17(2):238–246.

---

mikron. is an independent project with no affiliation with Moccamaster, Comandante, or any other manufacturer. Calibration data were collected by the author on personally owned equipment.