Examination of the Relationships Between Original, Real and Apparent Extracts, and Alcohol in Pilot Plant and Commercially Produced Beers

Anthony J. Cutaia[1], Anna-Jean Reid[2] and R. Alex Speers[2,3]

ABSTRACT

J. Inst. Brew. 115(4), 318–327, 2009

The historical development of equations used to relate alcohol and real extract to apparent extract and original gravity, as well as ratios between the corrected Real (RDF) and Apparent Degrees of Fermentation (ADF), were examined in light of modern polynomial and non-linear regression techniques. Comparisons were performed using an extensive data set of 532 brews obtained from commercial and pilot fermentations with statistical error analysis of these empirical relationships. New predictions of apparent extract were calculated as a function of alcohol and real extract analogous to the Improved Tabarie's formula. In addition, attempts at improving Balling's original equation model estimating original gravity from alcohol and real extract are detailed and discussed. The statistical analyses of relationships between $A_{w/w}$ (alcohol by weight) and functions of OE (Original Extract), AE (Apparent Extract) and RE (Real Extract) as well as ratios between the corrected Real and Apparent Degrees of Fermentation (RDF/ADF) are reported. It is expected that this paper will be useful for brewers to more accurately estimate $A_{w/w}$ and real extract values.

Key words: ADF, alcohol, apparent extract, brewing calculations, original extract, RDF, real extract.

INTRODUCTION

As early as 1784[25], almost a century before the time of Balling[2], brewers were interested in and reported on the density relationships between wort and fermented beer[12,13,22,27]. Most of these researchers were concerned with the relationships between wort and beer density and alcohol content and published their work in German prior to 1933. One notable exception was a chapter in the first edition of the Handbook of Brewing by Weissler[28] that reported the approximate relations between original extract (OE, °P), apparent extract (AE, °P), real extract (RE, °P) and alcohol, on a weight/weight basis $(A_{w/w})$[22].

[1] Science Source Consulting, Ballwin, MO.
[2] Food Science and Technology, Dalhousie University, 1360 Barrington Street, D401, Halifax, NS B3J 2X4 Canada.
[3] Corresponding author. E-mail: Alex.Speers@Dal.Ca.

As was the practice of the period, little discussion, justification or assumptions were presented in these reports. Nor was any analysis of possible errors inherent in these formulas undertaken. Viewed from our current perspective, these omissions may be considered serious criticisms. One should remember, however, that statistical routines and calculators were essentially non-existent. For instance, regression techniques were most certainly not well known until some time after Pearson's publication of 1896[23]. In any case, without access to a personal computer, any relationships would have been derived via tedious graphical techniques. The tables produced by authors such as Balling[2], Wahl and Henius[27], Holzner[13], and Pawloski and Doemens[22] relating OE, AE, RE and $A_{w/w}$ involved an almost unbelievable amount of calculation. In the course of this study we examined Table III of Holzner[13] which tabulates 6252 predictions of real extract and alcohol content for given original and apparent extracts. This massive effort was obviously done by hand!

The most well known expression relating OE, RE and $A_{w/w}$ is the Balling equation[2] which relates the original extract to the real extract and alcohol (% w/w):

$$OE = 100 * (2.0665 * A_{w/w} + RE) / (100 + 1.0665 * A_{w/w}) \qquad (1a)$$

This relationship is an approximate one as noted by Cutaia[4]:

> Two major assumptions are used in the derivation of Balling's formula. The first assumption is that 0.11 g of carbohydrate is converted to yeast mass for each gram of ethanol produced in fermentation. The second assumption is that all fermentable dissolved wort solids are monosaccharides. This assumption is implicit in the composition of the Balling constant…

While relied upon by the brewing industry, it should be emphasized that the Balling equation is an approximation violating various theoretical assumptions and is a "working estimate" only. For instance, it is a common fallacy that Balling's formula can be used to back-calculate a diluted beer's original (undiluted) extract. In fact, a back-calculation of the original extract knowing dilution rates, $A_{w/w}$ and RE always gives a larger estimate of OE than the true undiluted value.

Neilson et al.[21] provided a critique of Balling's formula and listed six effective but minor objections to the formula. Surveys by the Carlsberg brewery in 1942 and 1971 showed the formula overestimated OE by 0.2[21]. Similarly, with a limited range of data, Navarro et al.[18] reported the Balling formula fit the data with a coefficient of determination of 0.995 and a residual mean square of 0.032. They also reported on alternate formulas, which improved the prediction of original extract with their limited 36 fermentation dataset. A recent survey by Neilson et al.[21] of 50 fermentations (with a OE specification of 14.3°P) indicated the Balling formula underestimated the true OE when applied to fermenting worts with developing $A_{w/w}$ levels of up to ~3.66 and overestimated the true OE by 0.1°P at the cessation of fermentation.

There are also other calculations in the literature that include[22]:

$$OE = [(RE - AE) / q] + RE \qquad (2a)$$

$$A_{w/w} = a * (OE - AE) \qquad (3a)$$

$$A_{w/w} = b * (OE - RE) \qquad (4a)$$

$$A_{w/w} = c * (RE - AE) \qquad (5a)$$

These equations were apparently taken from Balling's original work[2] as they use the German phrase "nach Balling" that can be translated as "according to Balling". Confusingly, however (at least to the English reader), the coefficients q, a, b and c of equations 2a–5a (hereafter referred to as "brewing values") in Balling's 1865 paper (Table IX p. 243) are different than those reported by later authors[6,12,22]. Interestingly, both the brewing values listed by Balling[2] and Pawloski and Doemens[22] were reported to vary linearly with OE (Table X, p. 251). Table I lists the Pawloski-Doemens values which agree with Weissler[28]. Curiously, the values for q published by Pawloski and Doemens in 1932[22] were first presented by Holzner in 1877[12]. It is worth noting that that small brewers and home brewers can seldom measure real extract or alcohol content easily and thus formulas containing only OE and AE values are more than curiosities.

Furthermore, an expression analogous to equation 3, but using specific gravity rather than extract values, is legislated for use by small brewers in Great Britain to calculate alcohol levels on a volume/volume basis ($A_{v/v}$) for excise purposes[11]:

$$A_{v/v} = f * (OG - PG) \qquad (6)$$

where OG and PG are the original and present (apparent) gravities multiplied by 1000 and f is a factor 'mandated' to vary from 0.125 to 0.135 as (OG – PG) varies from 6.9 to 100.7. Other estimations using specific gravity units have been reported by early British researchers[10,26]. One simple estimation of $A_{v/v}$ is reported by multiplying by the drop in specific gravity by 102[14].

Aside these relationships, there are the relatively well-known factors of Apparent Degree of Fermentation (ADF):

$$ADF = 100 * (OE - AE) / OE \qquad (7)$$

and Real Degree of Fermentation (RDF)[1]:

$$RDF = [100 * (OE - RE) / OE] * [1 / (1 - .005161 * RE)] \qquad (8)$$

The correction term in equation 8 '[1 / (1 – .005161 * RE)]' was introduced in 1979 and corrects for mass lost by CO_2 and yeast uptake during fermentation[5]. The ratio of RDF to ADF is also mentioned in many texts:

$$RDF = ADF * Q^{\dagger} \qquad (9a)$$

Note the factor $Q^{\dagger}$ should reflect the corrected RDF value. Of course the original definitions of RDF (and thus equation 9a) did not account for the '[1 / (1 – .005161 * RE)]' correction term. It was noted that Balling first proposed calculating the ratio of RDF/ADF[20] and it is termed 'Balling's shorter formula'. The uncorrected form of factor $Q^{\dagger}$ (i.e., [(OE – RE) / OE] / [(OE – AE) / OE] or RDF uncorrected/ADF) has been often cited in the past as 0.819[6] or 0.81[28]. Kunze[16] and Neilson and Erdal[20] both reported a value of 0.81 and attributed this value to a publication by Balling. It is noteworthy that the uncorrected value of $Q^{\dagger}$ (i.e., RDF *uncorrected*/ADF) is equal to "1 / (q + 1)", q being a constant introduced in equation 2a that was reported to be dependent on original extract as shown in Table I[22].

Given the development in brewing sciences and advances in our chemical and statistical analyses, one might argue it is well overdue to re-examine the brewing values and relationships reported in the 1800s. This paper will report on the relationships between these brewing values and various brewing parameters and comment on the inherent error when using these brewing formulas.

Table I. Empirical values published by Pawloski and Doemens[22] and Weissler[28].

OE (°P)	q (%/°P)	a (%/°P)	b (%/°P)	c (%/°P)
1	0.221	0.3983	0.4864	2.2010
2	0.222	0.4001	0.4889	2.2024
3	0.223	0.4018	0.4915	2.2041
4	0.224	0.4036	0.4941	2.2058
5	0.225	0.4054	0.4967	2.2076
6	0.226	0.4073	0.4993	2.2096
7	0.227	0.4091	0.5020	2.2116
8	0.228	0.4110	0.5047	2.2137
9	0.229	0.4129	0.5074	2.2160
10	0.230	0.4148	0.5102	2.2184
11	0.231	0.4167	0.5130	2.2209
12	0.232	0.4187	0.5158	2.2234
13	0.233	0.4206	0.5187	2.2262
14	0.234	0.4226	0.5215	2.2290
15	0.235	0.4246	0.5245	2.2319
16	0.236	0.4267	0.5274	2.2350
17	0.237	0.4288	0.5304	2.2381
18	0.238	0.4309	0.5334	2.2414
19	0.239	0.4330	0.5365	2.2448
20	0.240	0.4351	0.5396	2.2483
21	0.241	0.4373	0.5427	2.2519
22	0.242	0.4395	0.5458	2.2557
23	0.243	0.4417	0.5490	2.2595
24	0.244	0.4439	0.5523	2.2636
25	0.245	0.4462	0.5555	2.2677
26	0.246	0.4485	0.5589	2.2719
27	0.247	0.4508	0.5622	2.2763
28	0.248	0.4532	0.5656	2.2808
29	0.249	0.4556	0.5690	2.2854
30	0.250	0.4580	0.5725	2.2902

Table IIa. Polynomial fit of specific gravity function (SG-1) and apparent extract (AE) values of ASBC Table 1[a].

Statistical value	Eq. 10 prediction of SG	Eq. 11 prediction of AE	Eq. 12 prediction of AE
Dependent variable	SG-1	AE	AE
Coefficients[1]			
b_0	1.3084×10^{-5}	-4.63374×10^2	-
b_1	3.86777×10^{-3}	6.68723×10^2	2.56899×10^2
b_2	1.2745×10^{-5}	-2.05349×10^2	6.7126
b_4	6.3×10^{-8}	-	-1.44816×10^4
b_5	-	-	5.17578×10^5
b_6	-	-	-1.07464×10^7
b_7	-	-	1.3011×10^8
b_8	-	-	-8.50786×10^8
b_9	-	-	2.3231×10^9
Adjusted multiple R^2	0.999	0.999	0.999
Standard error of estimate	2.36×10^{-6}	1.59640×10^{-3}	2.0098×10^{-4}

[a] $b_{0,1...9}$ are the intercept, linear, squared, and cubed terms, and so on. n = 1661.

DATA COLLECTION AND ANALYSIS

Data collection

In the course of this investigation, five brewing datasets were obtained. The first two data collections were from a pilot brewery (sets RPB-I and RPB-II) of 92 and 112 fermentations respectively. A third dataset was a 'proprietary' set of 71 commercial lager fermentations. The first three data collections contained directly measured values of original extract, apparent extract and alcohol (% w/w). A fourth 'CMBTC' dataset consisted of 221 directly measured values of original extract, (apparent) specific gravity (SG), and alcohol (% v/v) from pilot scale fermentations. These were obtained from the Canadian Malting and Brewing Technical Centre, Winnipeg, MB. This laboratory routinely produces pilot beers in the course of evaluating malting barley. While measurement methods were proprietary, RPB-I and RB-II dataset values were derived with a SCABA™ instrument and an Anton-Parr™ apparatus was used to collect the Proprietary and CMBTC datasets. A fifth dataset was formed from the published values of Navarro et al.[18] (original, apparent and real extracts and alcohol levels % w/w). While Navarro et al. reported 142 data points, they were in fact replicates of 36 fermentations measured using European Brewing Convention procedures. Thus, the replicate values were averaged for each of the 36 fermentations resulting in 36 values for the original, apparent and real extracts, as well as the alcohol level (assumed to be % w/w). All the values collected were measured prior to any post-fermentation dilution.

It is noteworthy that during the course of data collection, it became apparent that the direct measurement of the real extract is now *very rarely* undertaken. Only one dataset[18] contained *measured* real extract values[8]. The correct calculation of the real extract values from directly measured alcohol and apparent extract levels is discussed later in this paper.

Data analysis

In addition to the analysis of the datasets discussed above, the relationships of °Plato (AE) versus specific gravity (sugar solutions) and specific gravity versus alcohol (w/w) in ASBC Tables 1 and 2[1], real extract and finally alcohol estimations were determined using the Systat 11 package (Systat Software Inc., San Jose, CA). As the purpose of these fits was to develop predictive equa-

Table IIb. Polynomial fit of specific gravity (SG_{OH}) and alcohol values of ASBC Table 2[a].

Statistical value	Eq. 13 prediction of SG_{OH}
Dependent variable	SG_{OH}
Coefficients[1]	
b_0	1.000011541
b_1	-1.92647×10^{-3}
b_2	3.0266×10^{-5}
Adjusted multiple R^2	0.999
Standard error of estimate	1.0165×10^{-5}

[a] b_0, b_1 and b_2 are the intercept, linear, and squared terms. n = 501.

tions, colinearity and variance inflation factors were not considered[29].

The empirical relationship between original gravity, alcohol (w/w) and real extract was determined for equations in Balling's original format and related variations for the combined datasets and each subset individually. The non-linear curve-fitting statistical package used was STATISTICA (StatSoft, Inc., Tulsa, OK). Comparisons of statistically significant differences between data subset parameters were performed using a Monte Carlo bootstrapping technique[7] implemented on Microsoft's EXCEL spreadsheet package and the @RISK package (Palisades Corp., Ithaca, NY). This method was employed to eliminate spurious results due to non-normal distributions of the estimated subset parameters. In this instance, distributions of 10,000 differences between all possible subset pairs of parameter estimates were generated by random sampling, with replacement of all subset dataset observations. By inspection of the resulting distributions, it was determined whether zero (indicating no difference in parameters) was contained between the 2.5% and 97.5% region of the distribution of differences. If indeed it was not, then a statistical difference at the 95% confidence level was determined for the indicated parameter estimate comparison.

Relations between specific gravity and extract (as °Plato, ASBC Table 1)

In order to compare and recalculate constants in the brewing equations, accurate relationships between °Plato (extract) and specific gravity were required. Relationships between these two values have been reported in the past by Seibert[24], Lincoln[17] and Hackbarth[9]. Seibert[24] reported a third degree polynomial that related specific gravity to Plato values using selected data from ASBC Table 1[1].

Lincoln reported a calculation for Plato values as a function of specific gravity equivalent to a second-degree polynomial[17]. Hackbarth[9] reported the use of an eighth degree polynomial to also calculate for Plato values as a function of specific gravity.

To update the calculations of Seibert[24] and Lincoln[17], third and second degree polynomials were fit to all the values from ASBC Table 1, using a forward stepwise procedure ($P > 0.15$ to include). A form of Hackbarth's formula, an eight-degree polynomial relating the apparent extract to the specific gravity less 1.0 (i.e., [SG-1], [SG-1]^2 ... [SG-1]^8) was also examined using a polynomial regression model. The best-fit results are shown in Table IIa and can be represented by the following equations:

$$SG-1 = 1.308 \times 10^{-5} + 3.868 \times 10^{-3} * extract$$
$$+ 1.275 \times 10^{-5} * extract^2$$
$$+ 6.300 \times 10^{-8} * extract^3 \qquad (10)$$

$$Extract = -4.6337 \times 10^2 + 6.6872 \times 10^2 * SG$$
$$- 2.0535 \times 10^2 * SG^2 \qquad (11)$$

$$Extract = 2.569 \times 10^2 * (SG-1) + 6.7126 \times 10^0 *$$
$$(SG-1)^2 - 1.4482 \times 10^4 * (SG-1)^3$$
$$+ 5.1758 \times 10^5 * (SG-1)^4 - 1.0746 \times 10^7 *$$
$$(SG-1)^5 + 1.3011 \times 10^8 * (SG-1)^6$$
$$- 8.5079 \times 10^8 * (SG-1)^7 + 2.3231 \times 10^9 *$$
$$(SG-1)^8 \qquad (12)$$

The coefficients for the specific gravity prediction differ slightly from those of Seibert[24] presumably due to machine error of the number of significant figures that could be carried by the machine calculators of the time. Our regression of apparent extract as a function of specific gravity confirms previous equations. The standard error of the estimate of these equations equal 2.36×10^{-6} for equation 10, 1.59×10^{-3} for equation 11 and 2.01×10^{-4} for equation 12. Considering the adjusted multiple R^2, the standard error of the estimate and the residual ranges, the most exact formula to estimate extract from specific gravity is obtained with equation 12 as noted in Table IIa.

These relations allowed calculation of apparent extract values from the specific gravity values for the CMBTC dataset using equations 12 and 10 respectively.

Another quadratic expression relating the specific gravity of ethanol solutions (SG$_{OH}$) to a given alcohol level (A$_{w/w}$) from Table 2[1] was developed:

$$SG_{OH} = 1.1541 \times 10^{-5} - 1.9265 \times 10^{-3} * A_{w/w}$$
$$+ 3.0266 \times 10^{-5} * A_{w/w}^2 \qquad (13)$$

The residual error arising from the use of this equation ranges from –0.00003 to 0.00003 for prediction of SG$_{OH}$ (equation 13). The corresponding regression analysis is presented in Table IIb.

Calculation of alcohol (w/w basis)

In order to convert alcohol to a percent weight/weight basis from alcohol in a percent volume/volume basis for future calculations, we used (beer) specific gravity and alcohol level values (A$_{v/v}$) and the formula of Weisler[28] and the exact conversion factor of the International Bureau of Legal Metrology[15]:

$$A_{w/w} = A_{v/v} * 0.7907 / SG \qquad (14)$$

Calculation of real extract

While the measurement of real extract is simple, it is tedious and now rarely undertaken in brewing laboratories. Rather, it is calculated by formulas derived by either Balling or Tabarie[3]. Recently, a new calculation was reported of an 'Improved Tabarie's formula' which relates the specific gravity of beer to the specific gravity of the real extract and the alcohol[19]:

$$SG = SG_{RE} + [SG_{OH} - 1 + 2.96 * (1 - SG_{OH})^2 - 381 *$$
$$(1 - SG_{OH})^3] \qquad (15a)$$

or

$$SG_{RE} = SG - [SG_{OH} - 1 + 2.96 * (1 - SG_{OH})^2 - 381 *$$
$$(1 - SG_{OH})^3] \qquad (15b)$$

The SG$_{Real}$ value can first be converted into g$_{extract}$ per 100 g (of water) via equation 12 and then to RE via ASBC method 5a:

$$RE = [(g_{extract} / 100\ g_{water}) * SG_{Real}] / SG \qquad (16)$$

It is worth noting that the Tabarie or 'Improved' Tabarie formula involves conversion of A$_{v/v}$ (not A$_{w/w}$!) to SG$_{OH}$ and that SG$_{RE}$ calculated from that value is by volume.

However, importantly and most recently, it was noted[9] that neither of these relationships account for ethanol-sucrose interactions and a routine was developed that accounted for these interactions. Using this routine[9], the dependency of real extract on A$_{w/w}$ and AE was determined by varying these values at 0.5 intervals from 0–7% A$_{w/w}$ and 0–10°P respectively. A response surface-type function was then developed to predict the real extract when the alcohol and apparent extract values are known:

$$RE = 0.496815689 * A_{w/w} + 1.001534136 * AE$$
$$- 0.000591051 * A_{w/w} * AE - 0.000294307 * AE^2$$
$$- 0.0084747 * A_{w/w}^2 + 0.000183564 * A_{w/w}^3$$
$$+ 0.000011151 * AE^3 + 0.000002452 * A_{w/w}^2 *$$
$$AE^2 \qquad (17)$$

The function estimated the prediction of the real extract[9] over the range of A$_{w/w}$ 0–7% and AE 0–10°P. The residual error of the function was below ±0.0015°P for all but 6 (of 278) cases and was never greater than ±0.0035°P. Since one of the objectives of this study was to examine Balling's formulas (i.e., equations 1–5 and 9) equation 17 was applied to calculate real extract values in appropriate datasets.

Aside from the measurement of real extract leading to Eq 17 by Hackbarth[9], the authors were unable to find laboratories that currently measure real extract by distillation. In our collection of data only one report[18] presented measured OE, A$_{w/w}$, AE and RE values. A comparison of these 36 measured and predicted real extract values[18] (via equation 17) gave an error range from –0.02 to 0.05°P and a standard deviation of these residuals of 0.019°P.

Final datasets

To summarize, the use of the above calculations resulted in four datasets of 92, 112 and 71 fermentations and a CMBTC dataset containing the results of 221 fermentations with measured original extract, apparent ex-

tract and alcohol (w/w) values and calculated real extract amounts. A fifth dataset of Navarro of 36 fermentations contained the means of four measured original, apparent and real extract and alcohol ($A_{w/w}$) levels. Table III shows the minima, maxima and average values of original and apparent extracts, alcohol ($A_{w/w}$) and real extract values for the five datasets. Figure 1 shows the distribution of the datasets as a function of their real extract and alcohol levels.

RESULTS AND DISCUSSION

Functional modelling of original wort extract estimation

Using the first five datasets, it was possible to determine the accuracy of the Balling equation constants. Balling's original equation (equation 1a) was based on an implied functional relationship between alcohol ($A_{w/w}$) and unfermented solids (RE or real extract) remaining in

Table III. Original extract, alcohol (% w/w), calculated real extract and apparent extract values used in this study.

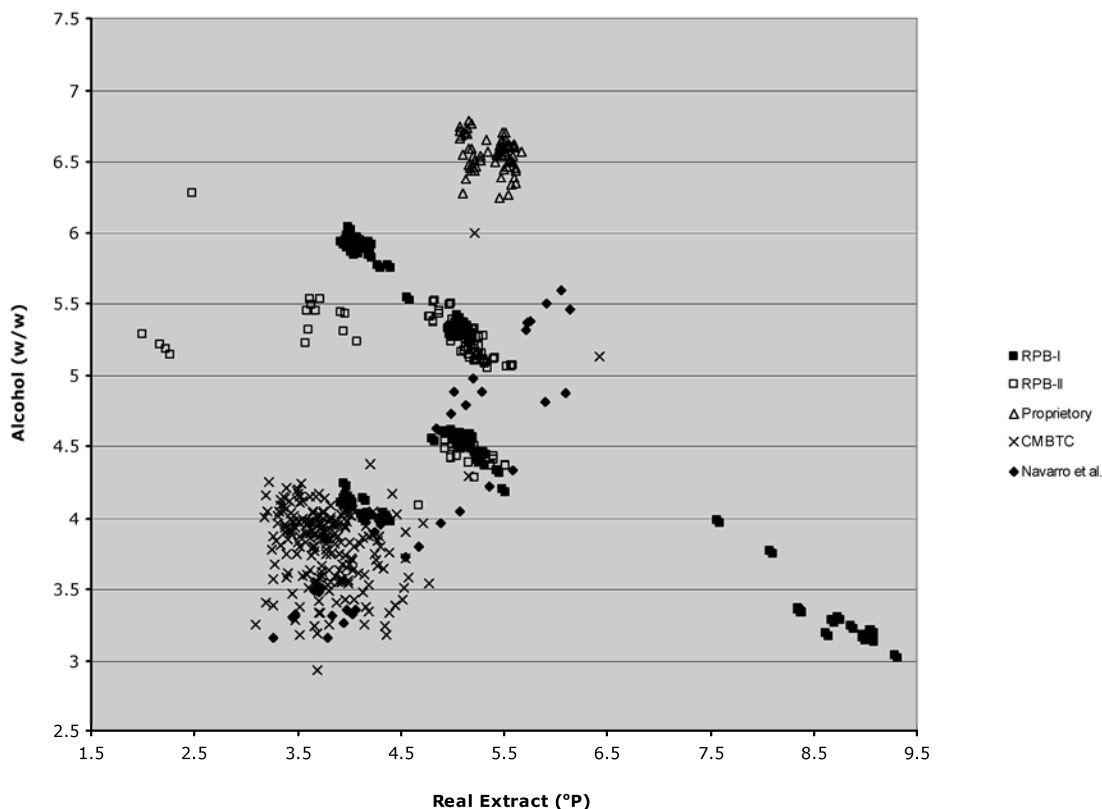| Dataset	Original extract (°P)	Alcohol (% w/w)	Real extract[a] (°P)	Apparent extract (°P)
RPB-I, n = 92				
Minimum	11.62	3.04	3.92	1.24
Maximum	15.11	6.04	9.28	7.86
Mean	14.08	4.72	5.28	3.11
RPB-II, n = 112				
Minimum	12.14	4.09	1.99	–0.42
Maximum	15.16	6.27	5.59	3.49
Mean	14.54	5.14	4.88	2.53
Proprietary, n = 71				
Minimum	17.09	6.24	5.07	2.06
Maximum	18.03	6.79	5.67	2.77
Mean	17.67	6.56	5.38	2.45
CMBTC, n = 221				
Minimum	9.44	2.94	3.10	1.26
Maximum	17.12	5.99	6.42	4.08
Mean	11.34	3.82	3.80	2.01
Navarro, n = 36				
Minimum	9.51	3.16	3.27	1.79
Maximum	16.49	5.59	6.14	3.85
Mean	12.75	4.21	4.74	2.79

[a] Calculated by equation 17.

Fig. 1. Distribution of alcohol and real extract values for the five datasets used in this study.

the product after fermentation, the mass lost to carbon dioxide and yeast, and the solids content of the original wort (OE):

$$OE = 100 * (2.0665 * A_{w/w} + RE) / \\ (100 + 1.0665 * A_{w/w}) \quad (1a)$$

His empirically determined conversion coefficient of 2.0665 used in the numerator of his equation indicated the grams of fermentable material necessary to produce one gram of ethanol. The 1.0665 coefficient used in the denominator of equation 1a represented the mass in grams of fermentable material lost to carbon dioxide and yeast mass increase (2.0665 grams of fermentable material in original wort minus 1 gram of ethanol remaining in fermented product).

Two other variations on the Balling equation as discussed below were also examined. Table IV shows the fit of equations predicting original extract values (OE) using the original Balling equation and the two variations to the five datasets. The models shown in Table IV were derived from the following equations:

$$OE = 100 * (k_1 * A_{ww} + RE) / \\ (100 + (k_1 - 1) * A_{ww}) \quad (1b)$$

$$OE = 100 * (k_1 * A_{ww} + RE + k_2) / \\ (100 + .9565 * A_{ww} + k_2) \quad (1c)$$

where the variables are as defined above and k_1 and k_2 are coefficients estimated by non-linear regression.

Equation 1a is the original Balling equation with his familiar coefficients inserted. Figure 2 shows the residual fit of equation 1a. Equation 1b uses Balling's assumptions as stated previously, but with its conversion coefficient derived from non-linear curve fitting to observed data. Equation 1c is similar to equation 1b but was augmented by an additional assumption that solids converted to yeast mass are constant and are not proportional to the amount of mass converted to ethanol and CO_2. Thus, a second coefficient directly representing yeast mass uptake was incorporated. This coefficient is, in itself, an approximation since it strictly should be corrected to a wort basis. The fixed factor of 0.9565 in the equation denominator is the molar ratio of CO_2 produced (and lost) for each gram of ethanol generated. This molar factor was used since the fermentable mass converted to alcohol is strictly used for this purpose and fermentable mass not producing alcohol is accounted for by the second coefficient.

In equations 1a–1c, it was implicitly assumed that ethanol and other losses such as moisture were non-existent, that the utilization of mass to produce other volatile metabolites is insignificant, and that the product is measured undiluted.

In these fitted models, the data were initially subjected to non-linear curve fitting procedures with all sets pooled as a single population. Calculation of root mean squared error, mean deviation of model estimate from observed original extract and explained variation (r^2) was performed for the pooled data and for the individual datasets where appropriate. In Eq 1c, however, the non-linear fit analysis for the individual datasets determined individual dataset yeast mass uptake coefficients, (k_2), using a dummy variable technique while simultaneously employing all datasets for conversion coefficient estimation, (k_1). This approach was based on the assumption that the conversion factor was relatively stable across disparate fermentations, while the apportionment of solids to yeast mass, being highly dependent on process factors such as aeration and wort nutrients, would be expected to be less stable across various processes.

Referencing Table IV, equation 1a (i.e., the original Balling model), is seen to possess a root mean square error (RMSE) of 0.237°P and an average mean deviation of –0.043°P for the pooled datasets. This negative mean deviation would indicate a consistent, though slight, overestimation of measured original extract by Balling's formula. The elevated RMSE appears to be strongly influenced by the large RMSE associated with the largest dataset. If this dataset were eliminated, the precision

Table IV. Fit of equations 1a-c to estimate original extract values.

Population	n	k_1	k_2	Statistical grouping	RMSE (°P)	Statistical grouping	Obs-Est deviation (°P)	Explained variance (°P)
Equation 1a: OE = 100 * (2.0665 * A_{ww} + RE) / (100 + (2.0665 – 1) * A_{ww})								
Pooled	532	2.0665	-	-	0.237	-	–0.043	-
RPB I	92	2.0665	-	-	0.102	b	–0.230	-
RPB II	112	2.0665	-	-	0.077	a	–0.153	-
Proprietary	71	2.0665	-	-	0.164	c	–0.031	-
CMBTC	221	2.0665	-	-	0.276	d	0.098	-
Navarro	36	2.0665	-	-	0.108	b	–0.109	-
Equation 1b: OE = 100 * (k_1 * A_{ww} + RE) / (100 + (k_1 – 1) * A_{ww})								
Pooled	532	2.0518	-	-	0.234	-	0.000	0.9901
RPB I	92	2.0052	-	a	0.075	a	0.000	0.9961
RPB II	112	2.0294	-	b	0.072	a	0.000	0.9898
Proprietary	71	2.0602	-	c	0.164	c	0.000	0.3285
CMBTC	221	2.0951	-	d	0.278	d	0.000	0.8570
Navarro	36	2.0329	-	b	0.094	b	0.000	0.9982
Equation 1c: OE = 100 * (k_1 * A_{ww} + RE + k_2) / (100 + .9565 * A_{ww} + k_2)								
Pooled	532	1.9629	0.4219	-	0.221	-	0.000	0.9911
RPB I	92	1.9298	0.3847	a	0.064	a	0.000	0.9917
RPB II	112	1.9298	0.5374	c	0.068	a	0.000	0.9914
Proprietary	71	1.9298	0.8963	e	0.156	b	0.000	0.3917
CMBTC	221	1.9298	0.6497	d	0.272	c	0.000	0.8627
Navarro	36	1.9298	0.4605	b	0.065	a	0.000	0.9991

would be significantly enhanced though it would strongly accentuate the overestimation of measured original extract. Examination of individual dataset results shows that the RMSE ranges between 0.077 and 0.276°P while the mean deviation ranges between –0.230 and 0.098°P. Measured original extract was overestimated in four of the five datasets while two of the datasets showed significantly greater RMSE. Statistical grouping of RMSE values using F-ratio testing is presented and indicates that they are not equivalent. This further accentuates the elevated values for the Proprietary and CMBTC dataset.

Non-linear fitting to obtain the conversion coefficient using equation 1b for the pooled population resulted in an insignificant improvement in the RMSE at 0.234 though, as expected, the mean deviation is zero. Notably, the fitted conversion coefficient of 2.0518 was remarkably close to the 2.0665 derived by Balling over 150 years ago. Estimation of the conversion coefficient for the individual datasets yielded the values presented in Table IV. Although these fitted coefficients appeared to be close in value, ranging from 2.0052 to 2.0951, statistical grouping using Monte Carlo bootstrap analysis showed that the five values formed four statistically separate groups with only the RPB II and Navarro datasets being statistically equivalent. Additionally, the RMSE values of the individual datasets for this model also grouped into four statistically unique sets, though in a different pattern. The "explained variance" or r^2 values were for the most part, quite reasonable. The exception was the value for the Proprietary dataset. An r^2 of 0.3285 implies that the selected model would be a very poor predictor. If one examines Table III, however, the reason for this may be attributed to the very narrow range of both dependent and independent variables. This low r^2 value for the Proprietary dataset carried over into the equation 1c analysis in Table IV.

Monte Carlo bootstrap analysis was used to determine grouping significance in order to circumvent multiple comparison problems associated with asymptotic variances, widely disparate variance estimates, and the large differences in dataset sample sizes.

Equation 1c represents a departure from previous concepts of original extract estimation, incorporating not only a mass conversion coefficient, but also a yeast mass uptake coefficient, thus divorcing yeast proliferation proportionality from ethanol formation. The conversion coefficient resulting from the fitting of the total population was 1.9629 while the yeast uptake coefficient was 0.4219. Unfortunately, the RMSE was only slightly less than that of either Balling's original equation equation 1a or the fitted equation 1b. Examination of the individual datasets, however, is more encouraging.

The original assumption that a pooled model for the conversion coefficient was warranted was justified in that the asymptotic variance-based error was only 0.0154 for an estimate of 1.9298. The individual dataset estimates for the yeast uptake coefficients ranged from 0.3847 to 0.8963 and all were shown to be statistically distinct from each other by Monte Carlo bootstrap analysis. The RMSE values from the RPB I, RPB II and Navarro datasets were statistically identical and were the lowest of the three models. The Proprietary and CMBTC datasets, however, continued to exhibit the largest RMSE values. Elimination of these datasets and refitting the remaining data did not materially improve the already low RMSE values for the
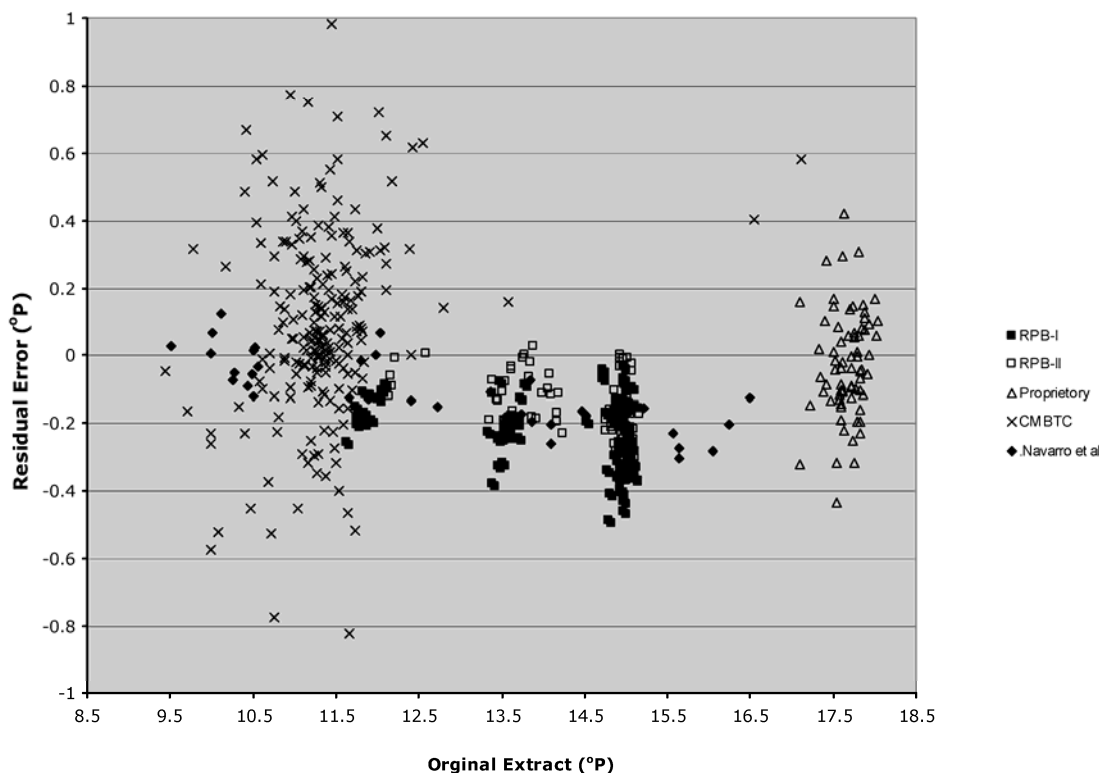

Fig. 2. Distribution of error in calculation of original extract from the Balling equation.

RPB I, RPB II and Navarro datasets. Additionally, the estimates for the yeast uptake coefficients also changed little and remained statistically distinct. Again, r^2 values were quite good, although the Proprietary dataset still had a poor value at 0.3917.

The conclusion one can obtain from these results is that a general equation, which is solely a function of ethanol and real extract in a fermented product, seems to be unable to reliably predict the measured original wort extract with sufficient accuracy. Examination of the dataset coefficients for equation 1c in Table IV offers some insight as to why this may be the case. The amounts of solids incorporated into yeast were significantly different for each dataset, possibly reflecting process differences that were not apparent from measuring only ethanol and real extract. If that were the case, equations incorporating only those measures or any measures derived from them, such as apparent extract, would be incomplete and would not be reliable predictors of original wort gravity.

Nonetheless, it is tribute to the analytical skills of Balling that he could determine the Balling constants of 2.0665 and 1.0665 in the mid-1800s. The original Balling values are embedded in the brewing literature and their use, as shown in equation 1a, is reported to slightly over-estimate[21] the original extract. Application of Balling's original equation to the datasets in this paper similarly over-estimates original wort extract in all datasets but one. However, OE, RE and OH_{ww} measurement errors may overshadow differences in the Balling constants. For these reasons and since the Balling equation is embedded in the brewing literature (and culture!) we do not expect (nor recommend) adaptation of a new Balling equation.

Modelling of traditional brewing values

Notwithstanding the finding of differences in various datasets reported above, the authors were both (1) curious

as the applicability of equations 3a–5a and 9a to the total range of brewing measurements collected and (2) cognisant of the need by industrial, craft and home brewers for a method to estimate alcohol levels given OE, AE and RE values. Table V shows a calculation of the brewing values a, b, c, and $Q^\dagger$ from equations 3a–5a and 9a using the combined datasets. They did not vary substantially with coefficients of variation less than 4%.

Given previous reports of these values correlations with original extract, further relationships were examined with the Systat statistical analysis software. A difficulty in examining these datasets was their distribution with respect to the original gravity. While the datasets were randomly selected, the distribution of data with respect to the original extract was not. For this reason we considered weighting the data to account for uneven sampling (noted in Fig. 1) However, weighting datasets is a debatable and non-trivial procedure. As a first estimate, ordinary least squares regression was employed to estimate the brewing values a, b, c, and $Q^\dagger$. Specifically, the dependence of brewing values on original extract was examined by linear regression using variations of equations 3a–5a and 9a where the terms "$b_o + b_1 * OE$" were substituted for a, b, c or $Q^\dagger$ as appropriate:

$$A_{w/w} / (OE - AE) = b_{ao} + b_{a1} * OE \qquad (3b)$$

$$A_{w/w} / (OE - RE) = b_{bo} + b_{b1} * OE \qquad (4b)$$

$$A_{w/w} / (RE - AE) = b_{co} + b_{c1} * OE \qquad (5b)$$

$$RDF / ADF = b_{Q\dagger o} + b_{Q\dagger 1} * OE \qquad (9b)$$

Table VI also lists the linear dependence of the same brewing parameters cited by Pawloski and Doemens[22] in Table I. The results of the linear regressions of equations 3b–5b and 9b are also shown in Table VI. While the original extract significantly ($p < .0001$) affects the dependent variable of equations 3b–5b ($A_{w/w} / (OE - AE)$, etc.) the left-hand terms are only weakly dependent on OE.

Our findings are in general agreement with previous reported brewing values. The initial literature on this topic[2,12] is over 100 years old, in German, and apparently does not report on how the relationships were developed. We note that equation 4 is a transformation of the Balling equation. If one assumes the Balling equation is correct, then the b value is equivalent to $1 / (2.0665 - 0.010665 * OE)$ and agrees with values in Table I.

Table V. Average brewing values (a, b, c and $Q^\dagger$) of equations 3a–5a and 9a[a].

	a	b	c	$Q^\dagger$
Minimum	0.371	0.449	2.05	0.805
Maximum	0.455	0.579	2.24	0.841
Mean	0.421	0.522	2.17	0.825
Standard deviation	0.013	0.019	0.034	0.005

[a] n = 532.

Table VI. Estimation of brewing values (a, b, c and $Q^\dagger$) as a function of original extract.

Statistical term[a]	Brewing values			
	a	b	c	$Q^\dagger$
Coefficients[a]				
b_o	3.72×10^{-1}	4.59×10^{-1}	1.99×10^{-1}	8.14×10^{-1}
b_1	3.57×10^{-3}	4.69×10^{-3}	1.34×10^{-2}	8.62×10^{-4}
Standard error of the estimate	1.00×10^{-2}	1.52×10^{-2}	1.26×10^{-2}	4.84×10^{-3}
Adjusted r^2	0.410	0.343	0.860	0.148
Pawloski and Doemens coefficients[b]				
b_o	3.95×10^{-1}	4.81×10^{-1}	2.19	8.20×10^{-1c}
b_1	2.1×10^{-3}	3.0×10^{-3}	3.1×10^{-3}	-6.6×10^{-4c}

[a] Coefficients b_o and b_1 are estimates of intercept and slopes for a, b, and c and $Q^\dagger$ as shown in equations 3b–5b and 9b. All regressions are significant $p < 0.0001$, n = 532.
[b] Linear dependence of previously published brewing values from the 1932 Table I of Pawloski and Doemens[22].
[c] An uncorrected estimate of $Q^\dagger$ calculated as $1 / (1 + q)$.

Prediction of alcohol levels from original extract and apparent extract levels

Small brewers and home brewers almost always know their original and apparent extract values, but find measurement of alcohol levels difficult. While equation 3a or 3b can be used with original and apparent extract values to predict $A_{w/w}$, an attempt was made to find a more accurate and precise expression to predict $A_{w/w}$. Using all five datasets (n = 532, for ranges of $A_{w/w}$, OE and AE values see Table III), a number of transforms of OE and AE terms were regressed on the alcohol level $A_{w/w}$ to attempt to find a prediction equation. The best prediction as evidenced by high correlation coefficients and no trend in residual errors was found using a forward stepwise regression technique (Systat software; p = 0.15 to enter terms). The $A_{w/w}$ value was used as the dependent variable and OE − AE and $(OE - AE)^2$ as independent terms. The regression analysis is presented in Table VII.

$$A_{w/w} = 0.38726 * (OE - AE) + 0.00307 * (OE - AE)^2 \qquad (18)$$

Using this expression the standard error of the estimate was 0.106. Figure 3 shows the ability of both equations 3b and 18 to predict final alcohol levels. Examination of these figures as well as statistics from tables indicates that equation 3b fits the data slightly better. Thus, it is hoped that function 3b will be utilized by small and craft brewers who do not, or are not able to determine the alcohol or real extract content of every brew.

Table VII. Prediction of $A_{w/w}$ as a function of original and apparent extract levels by equation 18[a].

Independent variable	$A_{w/w}$
Coefficients	
OE − AE	3.8726×10^{-1}
$(OE - AE)^2$	3.07×10^{-3}
Adjusted r^2	0.999
Standard error of estimate	1.06×10^{-1}

[a] n = 532.
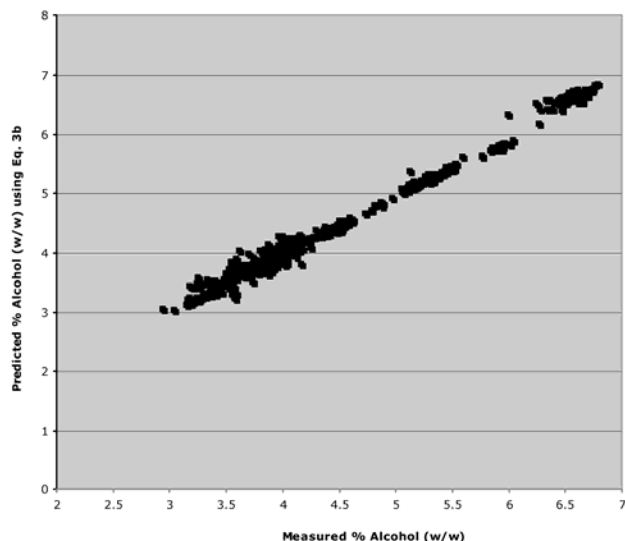
REFERENCES

1. American Society of Brewing Chemists. Tables related to determinations on wort, beer, and brewing sugars and syrups. Beer-5A, Beer-6A, Extract of Wort: Tables I and II. The Society: St. Paul, MN, 2004.
2. Balling, C. J. N., Die Bierbrauerei. Verlag von Friedrich Temski: Prague, CHZ, 1865.
3. Beckett, M. H. and Pavelsek, V. L., Automatic alcohol and real extract determination. *J. Am. Soc. Brew. Chem.*, 1986, **44**, 78-80.
4. Cutaia, A. J., Estimation of yeast mass increase in production brewing fermentations by calculations of carbohydrate utilization. *J. Am Soc Brew Chem.*, 2007, **65**, 166-171.
5. Cutaia, A. J. and Munroe, J. H., A method for the consistent estimation of real degree of fermentation. *J. Am Soc Brew Chem.*, 1979, **37**, 188-189.
6. DeClerck, J., A Textbook of Brewing. Vol. 2. Translated by K. Barton-Wright. Chapman and Hall Ltd: London, 1958.
7. Efron, B. and Tibshirani, R., An Introduction to the Bootstrap. CRC Press: Boca Raton, FL, 1993.
8. European Brewery Convention, Original Extract of Beers - Real Extract. Analytica - EBC. 3rd Ed. Method 7.1. P. E56. Schweizer Brauerei-Rundschau: Zurich, 1975.
9. Hackbarth, J. J., The effect of ethanol-sucrose interactions on specific gravity. *J. Am. Soc. Brew. Chem.*, 2009, **67**, 146-151.
10. Her Majesty's Stationary Office. The Alcoholic Liquor Regulations. 1979. Statutory Instrument, No. 1146. HMSO: London, 1979, pp. 2908-2913.
11. Her Majesty's Revenue and Customs. Notice 226, Beer Duty. Plymouth: UK, 2005, pp. 85-86.
12. Holzner, G., Über den Zusammenhang Zwischen dem Scheinbargen und Wirklichen Vergahrungsgrande. *Zeitschrift für das Gesammte Brauwesen*, 1877, **12**, 104-107.
13. Holzner, G., Tabellen zur Berechnung der Ausbeute aus Malze und zur Saccharometririchen Bieranalyze. Druck und Verlag von R: Oldenburg, Munich, 1904.
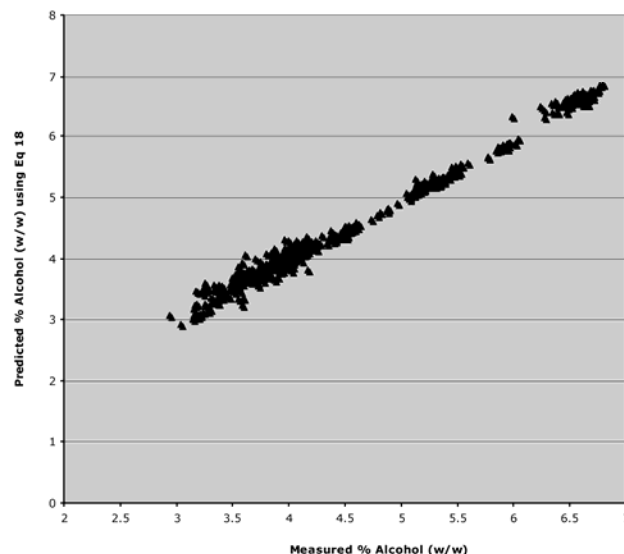
Fig. 3. Prediction of alcohol (w/w) levels using original and apparent extract values with equations 3b and 18.

14. Hough, J. S., Briggs, D. E., Stevens, R. and Young, T. W., Malting and Brewing Science. Vol. II, Chapman and Hall: London, 1971.

15. International Bureau of Legal Metrology. International Alcoholic Tables. The Society: Paris, 1973.

16. Kunze, W., Technology of Malting and Brewing. 7th Ed. Translated by T. Wainwright. VLB: Berlin, 1996, pp. 353-354.

17. Lincoln, R. H., Computer compatible parametric equations for basic brewing computation. *Tech. Q. Master Brew. Assoc. Am.*, 1987, **24**, 129-132.

18. Navarro, A., Carbonell, J. V., Pinaga, F. and Lopez, F., Study for the accuracy of Balling's formula and proposal for alternative equations to calculate the original extract of beer. *Cerveza y Malta*, 1987, **24**, 44-48.

19. Neilson, H. and Aastrup, S., Improving Tabarie's formula. *Scandinavian Brewers Review,* 2004, **61(4)**, 30-34.

20. Neilson, H. and Erdal, K., The degree of fermentation. *Scandinavian Brewers Review,* 2005*,* **62(3)**, 34-35, 37-39.

21. Neilson, H., Kristiansen, A. G., Krieger Larsen, K. M. and Erikstrom, C., Balling's formula - scrutiny of a brewing dogma. *Brauwelt Int*., 2007, **25**, 90-93.

22. Pawloski, R. and Doemens, A., Die Brautechnischen Untersuchungs-Methoden. Verlag von R: Oldenburg, Munch, 1932.

23. Pearson, K., Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philosophical Trans. Royal Soc. London*, 1896. **187**, 253-318.

24. Siebert, K. J., Routine use of a programmable calculator for computing alcohol, real extract, original gravity and calories in beer. *J. Am. Soc. Brew. Chem.,* 1980, **38**, 27-33.

25. Teich, M., A chapter on the history of transfer of information on attenuation. *The Journal of the Brewing History Society Online.* 2005, **121**, 40-46. www.breweryhistory.com/journal/archive/121/bh-121-040.htm (last accessed November 2009).

26. Thorpe, E. and Brown, H. T., Reports on the determination of the original gravity of beers by the distillation process. Part I. *J. Instit. Brew.,* 1914, **20**, 569-593.

27. Wahl, R. and Henius, M., American Handy Book of the Brewing, Malting and Auxiliary Trades. 3rd Ed. Vol. II. Wahl-Henius Institute: Chicago, IL, 1908.

28. Weissler, H. E., Brewing calculations. In: Handbook of Brewing. 1st Edition. W. A. Hardwick, Ed. Marcel Decker Inc: New York, NY, 1995, pp. 643-706.

29. Yu, C. H., An overview of remedial tools for collinearity in SAS. Proceedings of 2000 Western Users of SAS Software Conference. 2000, pp. 196-201.

30. www.creative-wisdom.com/pub/collin.pdf (last accessed November 2009).

(Manuscript accepted for publication November 2009)