

A Model for Automated Rating of Case Law

Marc van Opijnen
Ministry of the Interior and Kingdom
Relations (KOOP)
P.O. Box 20011
2500 EA The Hague, The Netherlands
+31-6-18464907
marc.opijnen@koop.wmrijk.nl

ABSTRACT

The increasing volumes of case law available in publicly accessible databases cause serious problems if users are not able to filter on the authoritativeness of the decisions. In this paper we present a model by which large numbers of judgments can be rated, (nearly) without any human intervention. For most cases the legal importance of judgments can be derived from data on publication and citation, reflecting the opinion of the legal community. For newly rendered judgments endogenous variables are used, which turn out to be good predictors of authority.

Categories and Subject Descriptors

I.2.1 [Artificial Intelligence]: Application and Experts Systems
Language Constructs and Features – *law*.

General Terms

Algorithms, Measurement, Reliability, Experimentation, Standardization, Theory, Verification.

Keywords

Case law databases, Case law authority, Information overabundance.

1. INTRODUCTION

Judicial decisions are being published in rapidly growing numbers, especially within the public domain: not only because judiciaries are taking advantage of modern text processing and electronic publishing equipment, but also because they are considering a generous publication policy as a way to meet the growing demand for transparency, to reinforce their presence in the public debate and to improve their visibility in society as a whole.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICAL '13, June 10 - 14 2013, Rome, Italy.

Copyright 2013 ACM 978-1-4503-2080-1/13/06 ...\$15.00.

The drawback of these expanding public case law databases is the risk of information overabundance. Contrary to popular belief, search engines are, out-of-the-box, not able to sift the wheat from the chaff. Although modern search aids like faceted search, snippets, query expansion and legal thesauri meet many user needs, they do not solve one of the most persistent problems: the user of case law databases isn't told which decisions are the landmark cases and which are the trifles.

Even in databases for which a selection is made, considerable differences in legal importance exist. And even if a rating is made, such a rating is not consistent in time: what's legally authoritative today might fall into oblivion tomorrow.

In this paper we will discuss a Model for Automated Rating of Case law ('MARC') by which an unlimited number of judicial decisions can be classified into five consecutive classes of legal authority, at any point in time. In this paper we use 'Legal authority' as a synonym for 'legal importance' or 'legal relevance', which we define as the general opinion of the legal community on the significance of a case for legal theory and practice. The relevance of a judgment for a particular user query or a specific legal issue is not discussed here. Hence, our model could e.g. be added as an extra filtering option in a search engine, but it is not meant to replace other search functionalities.

We will start with evaluating research by others and our own earlier work (§ 2). In § 3 we will give an account of the composition and construction of our research database. Next, we will describe the big picture and the theoretical foundation of our model (§ 4). In this model we make a distinction between three periods in the life of a judicial decision. These periods and their distinctive variables and calculations will be discussed in § 5 to § 7. Since the outcomes of these calculations are not very human readable, they need to be transformed into a user friendly classification system (§ 8). In § 9 we will perform some additional tests, and in § 10 we will finish with some conclusions and suggestions for future work.

2. RELATION WITH EARLIER WORK

Research on legal authority of case law is, especially when compared with the benefits it might offer to legal professionals, quite scarce. Most research is done in common law systems, especially the United States, where the main focus has been on the analysis of citation networks, i.a. by Marx [1], Tapper [2], Post & Eisen [3], Smith [4] and Fowler e.a. [5]. A broader review of this research can be found in our earlier paper [6], which also discusses European research by Geist [7], Malmgren [8] and Winkels e.a. [9]. To summarize, most of these authors have focused on the analysis of citation networks, but with a limited number of algorithms (PageRank, HITS, in-degree) and limited

volumes of case law, mostly from the highest jurisdictions only. Other variables were hardly taken into account.

In our earlier paper [6], we used our extensive database (*infra*, § 3) to analyse the performance of various social network algorithms (*infra*, § 5.2.3), and for an exploratory survey on the possible relevance of various other variables. Although from this exploratory work we learned a lot about the usability of various variables, no integral model was developed yet. In the current paper we leave the exploratory phase, and are able to present a complete and working model to rate the legal importance of judicial decisions, in a way that can be understood by users of case law databases.

3. PREPARATORY WORK

Academic research is often performed on small datasets, and therefore runs the risk of being too small-scaled to extrapolate its results to real life environments. To minimize this risk, we built – given the means – the largest legal database possible.

In more detail this database is described in [6], here we highlight its main features. It contains 854.000¹ unique judicial decisions from the Netherlands, the Court of Justice of the EU and the European Court of Human Rights, gathered from the public website of the Dutch judiciary ('Rechtspraak.nl'), commercial publishers and an internal database of the judiciary. Legal literature was collected from 184 commercial titles (556.000 files) and an internal Wiki-site of the judiciary (13.000 entries). Also 11.000 press releases from the Rechtspraak.nl-website were used.

Because most of these texts lack any computer readable references, citations to case law were extracted using a parser and a canonicalization algorithm, described in [10]. To discover legislative references a search engine was used, together with a parser (with regard to the parsing of European references described in [11]). In total, 412.000 cross case law citations were detected, 673.000 case law citations in scholarly writings and 5.659.000 references from case law to legislation.

4. THEORETICAL FRAMEWORK

4.1 The Legal Crowd

Judicial decisions are not rendered to share intellectual insights or to make public statements on the desirable state of the law. Instead, they are solely meant to solve a conflict that is brought forward by the parties to the conflict. Still, the specificities of the conflict, the interpretation of existing law or the development of new rules, might make a judicial decision of interest not only to the parties involved, but also to the legal community or society as a whole.

To rate this legal relevance of a judicial decision is everything but easy; because of the subtleties and the complexity of the legal domain it requires legal expertise to assess the case-overriding importance of an individual judgment. Using lawyers, specialized in all fields of law, to explicitly assess and rate – on a continuous basis – the legal relevance of hundreds of thousands of judicial decisions making up modern case law databases, not only leads to disputable results [12, 13], it would also be a financial and organizational nightmare.

An interesting alternative is to make use of the already existing 'wisdom of the crowd' [14]. Legal professionals, making up a 'legal crowd', are constantly reading and assessing case law, and thereby creating a wealth of implicit ratings. We can distinguish between five of these 'crowd variables':

1. Publication in jurisprudence magazines ('PubMagazines'). Editors of these case law magazines are focused on selecting judgments having relevance for their target audience, sometimes consisting of legal scholars and professionals in general, sometimes of highly specialized jurists.
2. Publication of annotations ('PubAnnotations'). Annotations (i.e. scholarly notes) are always published in conjunction with the judgments in the abovementioned magazines, generally on the most important or legally disputed cases.
3. Incoming citations from case law ('InCitationsCaseLaw').
4. Incoming citations from one-off literature ('InCitationsOneLit'). We make a distinction between two types of legal literature. 'One-off literature' are all publications that are published only once, and are not edited or updated later on. Examples are PhD theses and articles in law reviews.
5. Incoming citations from 'permanent literature' ('InCitationsPermLit'). The second type of legal literature is formed by the vast publications that are continually updated to reflect the state of play on a specific field of law. Traditionally these were handbooks or loose-leaf paper editions; nowadays they are often replaced by websites, sometimes as a wiki-system.

4.2 Phases in the Public Life of a Judgment

These five crowd variables surface in a specific order, as displayed in Figure 1. One could say the public life of judgment is depicted here.

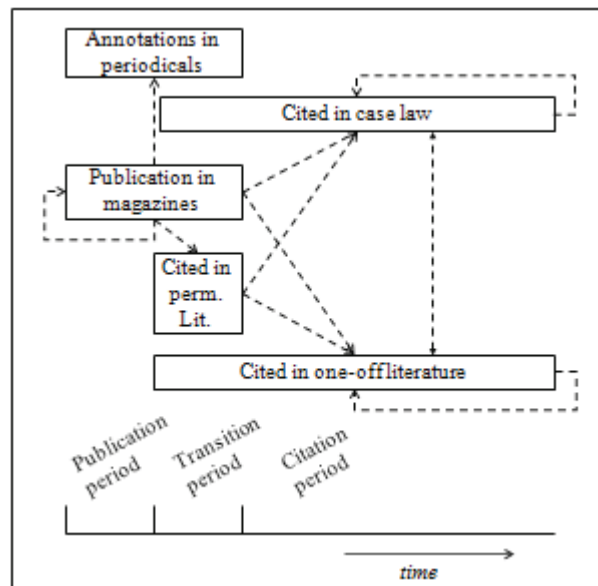


Figure 1. Theoretical model of the crowd variables. The dotted lines reflect the main influences between the distinctive variables.

¹ All numbers in this paragraph are rounded to the nearest 1000. Data were gathered until June 2010.

This public life starts with publication in one or more periodicals (publicly accessible databases included). Annotations are dependent on these publications, but take place at the same time. Only after a decision is published for the first time, it starts being studied by scholars and judges, who might cite it in permanent literature, one-off literature or judicial decisions. Meanwhile the decision keeps being published and annotated in other magazines. After a certain period though, all periodicals considering the decision relevant for their collection have published and annotated it, and all permanent literature has been updated with references to the new case. After that, the judgment is only cited in one-off literature and case law – citations that can continue infinitely.

The timeline at the bottom of Figure 1 shows the three periods in which we can divide the life of a judgment. The last period, in which a judgment is only re-cited, is the ‘citation period’. The publication period starts at the moment a decision can technically be published, which nowadays coincides with the day of rendering. The boundary between publication period and transition period is determined by the time authors and judges need to start citing and discussing the judgment.

For all three periods we need a different model to calculate legal authority. In the citation period we can fully use the opinion of the crowd, as expressed by the five crowd variables (to be discussed in § 5). In the publication period (in which a case might, of course, also *not* be published) the legal crowd hasn’t been able to give its opinion yet, so we have to use other indicators to establish legal authority (§ 6). The transition period is in between, and needs its own algorithm (§ 7).

To establish the duration of the distinctive periods we reasoned as follows. The publication period is assumed to start the day the judgment is pronounced. On many occasions the document is published on Rechtspraak.nl that same day. Practice shows that it takes about a week before (other) magazines start publishing the judgment, and people start discussing and referring to it. So, after a week the transition period starts. Furthermore, statistics on the jurisprudence periodicals reveal that the average publication term for all magazines is 127 days.² This term varies over the years; in 2010 it was 103 days. The median for that year was 81 days. We assume that a comparable term is needed by authors of permanent literature to update their works.³

Therefore, a duration of 100 days for the transition period seems reasonable. As will be explained in § 7, the influence of the exact length of the transition period is mitigated by the way the MARC-indicator for this period is calculated.

4.3 A Note on the Statistical Model

The calculations in this paper are conducted – unless stated otherwise – with the ‘generalized (non)linear model’. This is a statistical algorithm related to regression analysis, but with the advantage of being able to combine variables of different measurement levels and to work on collections of different distributions and scales.

Basically, like the regression analysis, the generalized (non)linear model calculates the probability that something (the ‘dependent

variable’ or ‘regressor’) will have a certain value, given specific values of (one or more) ‘independent variables’ or ‘predictors’.

Unfortunately, due to age, origin, or poor drafting of the documents used for our database, and in spite of several intelligent scripts and parsers used to improve the data, we were confronted with quite some missing variables (e.g. on judges). In the generalized (non)linear model all records having missing data were discarded.

Next, we will discuss which predictors and regressors were used for all three periods.

5. CITATION PERIOD

5.1 Introduction

We start with the last period, the citation period, because it is the most natural state. For this period our initial hypothesis is that the legal authority of a judgment can be fully established by assessing the opinion of the crowd, supplemented with three other exogenous variables (i.e. variables not in the text of the judgment itself). Facts that are within the judgment (‘endogenous variables’) are considered to be evaluated by the crowd, and therefore do not have to be taken into account separately. Only at a later stage (*infra*, § 9.3) we will test this initial hypothesis.

First, in § 5.2, we will describe the nature and calculation of the predictors used for the citation period. Subsequently, in § 5.3, we will discuss the choice and calculation of the regressor. In § 5.4 we will review the outcome of the statistical calculations.

5.2 Predictor Variables

For the citation period we use eight predictor variables. The first five are the crowd variables as mentioned in § 4.1. Also, we use the hierarchical position of the court, the field of law and the age of the judgment.

5.2.1 Publication in Jurisprudence Magazines (PubMagazines)

Judgments that are considered to be legally relevant, are selected by the editors of case law periodicals. Some magazines have a written list of selection criteria, but mostly it comes down to ‘legal gut feeling’ [15]. The more magazines decide to publish a decision, the more agreement can be assumed to be in the legal crowd on the importance of the case. Apart from the number of magazines a decision is published in, it is also relevant to assess in which magazines exactly the case is published. While ca. 20.000 judgments are published yearly on Rechtspraak.nl, just around 600 are published in ‘NJ’, presumably the most influential case law magazine in the Netherlands. This quite general magazine publishes the landmark decisions, mostly from the Supreme Court and European courts. Tens of other magazines republish the same judgments or pay interest to judgments on more substantial issues, in various fields of law. Since not all magazines and databases have the same standing, a rating system for case law magazines had to be developed.

To achieve this, we tried to assess the ability of the editors to select exactly those cases that later turn out to be the most cited. Being published in a magazine with a high score on this scale implies that the decision is potentially important.

The formula to calculate this ‘Jurisprudence Periodical Weight Factor’ (JPWF) reads:

² To neutralize the effect of outliers, those rare decisions published after more than one year were ignored.

³ We were unable to verify this assumption, for reasons explained in footnote 7.

$$JPWF = \frac{\sum \log_2(C_{in})}{U}$$

In which ‘C_{in}’ is the total number of incoming citations – multiplicity⁴ taken into account – to a judgment published in this magazine. The logarithmic scale is used to mitigate the influence of outliers. ‘U’ stands for the total number of judgments published in the magazine. The JPWF values of the 59 magazines of which data were available range between 0,31 and 4,71. Rechtspraak.nl scores 0,85. The ‘NJ’ scores 3,16.

Our next step is to calculate the value of the ‘PubMagazines’ variable for each individual decision. Also here we use a logarithmic scale; as not to overrate cases published in many (important) magazines. The formula for PubMagazines reads:

$$PubMagazines = \log_2\left(1 + \sum JPWF\right)$$

5.2.2 Annotations (PubAnnotations)

Annotations are scholarly notes, commenting on a particular judgment and published in conjunction with that judgment in a jurisprudence periodical. Not all periodicals publish annotations, and not all judgments published in periodicals that do publish annotations, are annotated. Having an annotation is thus quite rare: 7,5% of the judgments in our database has one or more annotations. Being selected to be annotated could therefore be considered as an indication of authority. Although one could try to establish the relevance of an annotation by its length, incoming citations or the standing of its author, we simply used the formula of PubMagazines to calculate PubAnnotations. A judgment with an annotation in a periodical with a high JPWF might therefore score higher on PubAnnotations than a judgment with annotations in two magazines with a low JPWF.

On Rechtspraak.nl press releases are often published about cases that attracted public attention or are considered of special relevance for the legal community. These press releases are counted as annotations, scoring the same JPWF of 0,85 as judgments published on the website.

5.2.3 Incoming Citations from Case Law (InCitationsCaseLaw)

Judgments regularly cite other judgments, both in common law and civil law systems. The reasons for citing an earlier judgment differ: the cited decision can e.g. be overthrown, confirmed, used for an analogy or deemed not relevant for the case at hand. In this research we haven’t taken the specific reason for citation into consideration, but whatever this reason, a citation always implies some authority being attributed to the cited case. In [6] various social network analysis algorithms to assess the role of a judgment in a citation network have been evaluated. Most of these network algorithms, tested against four different benchmarks, turned out to be poor performers. To meet the specific characteristics of a case law citation network a special algorithm, ‘MARC-in-degree’ (MID), was developed. This algorithm takes multiplicity into

account, mitigates outliers and is based on direct citations only. The formula for MID reads:

$$MID = \log_2(1 + C_{in})$$

Where C_{in} is the total number of non-distinct incoming citations. We will use MID as our starting point for the calculation of InCitationsCaseLaw.

As a next step we have to take the element ‘time’ into consideration. Judgments cited many years ago, and never since, clearly lost relevance. First, we decided to leave out all citations older than ten years. But even then: if judgments A and B both have been cited twenty times in the last ten years, there is a difference in relevance of these citations today, if A was cited twenty times in 2003 and never since, and B was cited twice a year, for ten years in a row. To express that recent citations are more relevant than older citations, we introduce a weighted moving average of (at maximum) ten years.

The formula for the variable InCitationsCaseLaw thus reads:

$$InCitationsCaseLaw = \frac{\sum_{0 \leq n < 10}^L ((10 - n) * MID_{T-n})}{\sum_{0 \leq n < 10}^L (10 - n)}$$

Where L is the age of the judgment on moment T, and T the year from which we look. When e.g. calculating InCitationsCaseLaw at the end of 2009 for a judgment of 2006, with MID-values for 2009, 2008, 2007 and 2006 of 2, 3,32, 0 and 2,81 respectively, we calculate:

$$\frac{(10 * 2) + (9 * 3,32) + (8 * 0) + (7 * 2,81)}{(10 + 9 + 8 + 7)} = \frac{69,55}{34} = 2,05$$

5.2.4 Incoming Citations from One-off Literature (InCitationsOneLit)

Like judicial decisions, one-off literature is published once and never updated afterwards. Both also have the tendency to reflect on the relevance of previous decisions for the current state of the law. The reason to treat InCitationsCaseLaw and InCitationsOneLit as different variables lies in the fact that scholars and judges have different considerations for citation. While judges are bound by the case at hand, scholars have the freedom to discuss whatever interests them. The ‘wrongful life’-case of the Dutch Supreme Court might serve as an example here.⁵ This case was much debated – also abroad – because it touches upon fundamental issues of human life, tort, and the ethics of medical profession. This is reflected by the high number of 108 citations in one-off literature. But while important for scholarly legal debate, the case hardly has any relevance for day-to-day judicial life, and therefore received only sixteen citations in case law. On the other hand many judgments can be found that are routinely cited in case law, but don’t attract much attention in legal literature.⁶

For the calculation of InCitationsOneLit we can use the same formula as was used for InCitationsCaseLaw (§ 5.2.3): citations

⁴ A citation from A by B can be counted distinct and non distinct. In the former situation multiple citations are discarded, in the latter situation (‘multiplicity taken into account’) all citations are added.

⁵ Dutch Supreme Court, 18-3-2005, LJN AR5213.

⁶ E.g. Dutch Council of State, 27-1-2003, LJN AF5566, with 1924 citations in case law and just 23 in one-off literature.

are counted non-distinct on a binary logarithmic scale, and then a weighted moving average is calculated.

5.2.5 Incoming Citations from Permanent Literature (*InCitationsPermLit*)

In § 4.1 permanent literature was defined as legal literature that is continuously updated to reflect the current state of play on a specific field of law.

Because permanent literature is updated continuously, it offers us a good insight in what the legal crowd considers to be the most relevant jurisprudence. New case law can be expected to be incorporated rather quickly, and if a case loses its importance, e.g. because of new legislation or overriding case law, it will be deleted from permanent literature.

For the very reason that permanent literature is assumed to reflect continuously the relevance of a judgment, it is not necessary to calculate a (weighted) moving average. We can simply use the MARC-in-degree to calculate the variable *InCitationsPermLit*.⁷

5.2.6 Hierarchical Position of the Court (*Hierarchy*)

Apart from the five crowd-variables discussed above, three other variables are used as predictors in the citation period. First, we discuss the hierarchical position of the rendering court.

Although in continental law systems the system of *stare decisis* is unknown, decisions of higher courts are generally considered to be more important than those of district courts, if only because not following the highest court will lead to successful appeals.

Leaving aside specific rules of competence for deviant proceedings, we based our typology of judicial hierarchy on the standard classification. We created five categories. In the first, lowest, category we find tribunals which are not part of the Dutch judiciary, but have a formal role in legal proceedings, like disciplinary committees, administrative appeal tribunals and foreign courts. The second category contains all district courts and other (mostly former) courts of first instance, while the third category contains the courts of appeal. The four highest courts make up the fourth category: Supreme Court (civil, criminal and tax law), Central Appeals Tribunal (social security law), Administrative High Court for Trade and Industry (social-economic administrative law and competition law) and the Council of State (all other categories of administrative law). The fifth category contains supranational jurisdictions: Benelux court, Court of Justice of the European Union (CoJ EU) and European Court for Human Rights (ECHR).

Statistically, the variable ‘Hierarchy’ is ordinal in nature, but we transformed it into an interval variable by using a (binary) logarithmic scale. As a result the distance between e.g. a district court and a court of appeal is bigger than the distance between the Supreme Court and the CoJ EU.

5.2.7 Field of law

The behaviour of other variables might depend on the field of law: tax law judgments could fall into oblivion faster than judgments on criminal law, or the percentage of annotated judgments could

⁷ To calculate *InCitationsPermLit* for any point in time though, we need historical versions of the permanent literature. In our database this information is not available (it is not offered by legal publishers as a standard service), and therefore we used the 2010-version of all permanent literature publications.

be higher within civil law than within administrative law. To take such differences into account, the field of law is introduced as a variable of its own. A typology on the field of law can be quite detailed, but due to differing, lacking or minimal metadata, we brought it down to three values: criminal law, civil law and administrative law.

The variable ‘FieldOfLaw’ is a nominal or categorical variable: there is no hierarchy between the values.

5.2.8 Age

The last predictor variable for the citation period is the age of the judgment. Although temporal aspects were already taken into account when calculating *InCitationsCaseLaw* and *InCitationsOneLit*, it is also considered to be relevant on its own. And because the difference between one and two years is considered to be bigger than between 27 and 28 years, also this variable is measured on a logarithmic scale.

5.3 Regressor Variable

Choosing a regressor variable is one of the trickiest parts of the model. How to quantify the legal importance of any judicial decision, at any point in time during the citation period? Looking at Figure 1 again, we notice that in this period a decision will not be published or annotated anymore, and its citation rate in permanent literature can assumed to change only because of changes in the legal authority and not because of authors lagging in their editorial activities. The only variables that are changing in this period are the citations in one-off literature and case law, and therefore these could constitute our regressor. But of course we are not using *InCitationsCaseLaw* and *InCitationsOneLit*, since these (predictor) variables state something about the past, and not about the future. And this future is important, because we could measure the legal authority of the judgment as the interest it is likely to receive tomorrow. In other words, we ask ourselves: what are the odds of a judgment being cited in future one-off literature and case law, given the predictor variables? Because of the fluctuations in citation frequencies, already noted in § 5.2.3, we define ‘the future’ as a period of three years.

To calculate this regressor we could simply add up all citations in literature and case law, but then we would be too dependent on the accidental composition of the database: because of the, already noted, substantial differences in citation frequencies between case law and scholarly writings, the value of the regressor could change substantially if we added more judgments or more literature to the database.

To determine the relative weight of both variables is not an easy choice; actually we are asking which source is reflecting legal importance of judicial decisions best. Lacking a clear indicator and avoiding disputes, we weighted them both for 50%. Therefore, the formula to calculate this new variable, the regressor for the citation period (‘RegressorCP’) reads for judgment ‘x’:

$$\text{RegressorCP}_x = \left(\frac{\text{COL}_x(Y1 + Y2 + Y3)}{\sum \text{COL}(Y1 + Y2 + Y3)} \right) + \left(\frac{\text{CCL}_x(Y1 + Y2 + Y3)}{\sum \text{CCL}(Y1 + Y2 + Y3)} \right)$$

In which COL is the number of (non-distinct) incoming citations from one-off literature, CCL is the number of (non-distinct) incoming citations from case law and Y is the year.

5.4 Calculations

With the eight predictors and the regressor in place we can start our calculations, using the generalized (non)linear model. The database has data till June 2010, and three full years are needed to calculate the regressor. Therefore we choose 31-12-2006 as our 'sichttag'.

As a first step, we calculated all possible subsets (i.e. combinations) of variables to discover which predictors are most suited to explain the variability in the regressor. Although we cannot display all 255 possible subsets here, we can summarize some highlights. When just one variable is used, InCitationsCaseLaw scores best, followed by InCitationsPermLit and InCitationsOneLit. Hierarchy, FieldOfLaw and Age are the worst performers. But when asked which variable can most easily be left out if all other are included, a different picture emerges: PubAnnotations and InCitationsOneLit are most easily left out, obviously because they have a high correlation with PubMagazines, respectively InCitationsCaseLaw. Also, while Hierarchy on its own can explain hardly any variability (subset 252), it can not be easily left out: all 26 best scoring subsets have Hierarchy included. By this measure, InCitationsCaseLaw seems to be the most important variable: the first 116 subsets all have it included, and together with just PubMagazines it performs better than all seven other variables together.

Still, all variables turn out to be statistically significant, and therefore we decided to use them all.⁸ This resulted in an R^2 (i.e. the portion of variability of the regressor explained by the predictors) of 0,47, which is quite a good performance. Given the values of the predictor variables, and the formula calculated by the statistical model, we can calculate the predicted value for each judgment, by multiplying the value of each variable by the 'estimate' displayed in Table 1.

Table 1. Wald statistic and estimate for all variables of the citation period.

Variable	Wald statistic	Estimate
Intercept ⁹	121381	-10,3484
InCitationsCaseLaw	454806	0,6209
InCitationsPermLit	27191	0,2014
Hierarchy	20738	1,9475
Age	19944	-0,2915
PubMagazines	6067	0,2901
FieldOfLaw-Criminal		-0,1996
FieldOfLaw-Civil	4701	0,2179
FieldOfLaw-Administrative		-0,0183
InCitationsOneLit	2615	-0,1057
PubAnnotations	1513	0,1119

⁸ The generalized (non)linear model allows for various tuning options, i.a. on the effects between individual variables. Various options were tried, but no significant improvements were found.

⁹ The intercept is the point where the axis is crossed. It is not relevant in the comparison of the variables, but displayed for reasons of completeness.

Because the model uses a logarithmic link function, we have to calculate the exponent of this summation. Finally, because of further calculations in § 8 we have to standardize this predicted value, using the general statistical formula:

$$Z = \frac{(X - \mu)}{\sigma}$$

Where Z is the standardized value, X is the calculated value, μ is the mean, and σ is the standard deviation. We label the result (Z) as the MARC-indicator for the citation period ('MARC-CP').

Also in Table 1 the Wald statistic is displayed, which measures the importance of each variable within the model. InCitationsCaseLaw scores best by far, followed by InCitationsPermLit, Hierarchy and Age. Like with the comparison of the subsets, also here PubAnnotations and InCitationsOneLit have a low score, due to their overlap with PubMagazines, respectively InCitationsCaseLaw.

6. PUBLICATION PERIOD

6.1 Introduction

During the publication period a judgment is not yet known to the legal crowd, and hence no crowd variables can be calculated. Because this period is very short, also Age is useless. And since we already noted that Hierarchy and FieldOfLaw are the poorest performers as sole predictors, we have to look for other variables for the publication period. To a large extend these variables have to be found in the judicial decision itself. We will discuss these endogenous variables in § 6.2. For the publication period we will also use another regressor, to be discussed in § 6.3. Calculations will be reviewed in § 6.4.

6.2 Predictor Variables

6.2.1 Outgoing Case Law Citations (OutCitationsCaseLaw)

While judgments receiving many incoming citations have legal authority, judgments having many outgoing case law citations can be called 'well-founded in law' [5]. Such decisions apparently needed a lot of jurisprudential research to come to the verdict, and therefore are presumably also important themselves. The variable 'OutCitationsCaseLaw' is calculated with multiplicity taken into account, but – contrary to the InCitationsCaseLaw variable – a linear scale turns out to perform better than a logarithmic scale, probably due to the fact that a judge cannot control the number of citations he will receive, but he is in full control of his outgoing citations.

6.2.2 References to Legislation (OutCitationsLegislat)

Case law citations are references to specific sources of law, and so are citations of legislation. Caution is required though, since many paragraphs of law are cited for procedural or habitual reasons, and are no indication of legal complexity.¹⁰ To filter out these procedural citations, we looked for the most frequently cited paragraphs of law. To establish a threshold we used a Pearson

¹⁰ Some case law is also just cited for formal reasons (e.g. a Court of Appeal citing the decision in first instance), but these citations were filtered out in the citation parser ([6]).

correlation test; it turned out that best results are obtained by leaving out all those articles being cited more than 2.000 times in the whole database: roughly the top 250 of 57.000 different paragraphs of law cited. Also all laws and regulations cited without a referral to a specific paragraph were to be ignored.

To calculate the variable ‘OutCitationsLegislat’, the remaining references were counted non-distinct, on a linear scale.

6.2.3 Number of Judges (Njudges)

In many proceedings the law prescribes imperatively whether a case has to be dealt with by a single judge or in full court. In other cases it is up to the court to decide on this. In both cases though proceedings before a single judge can be assumed to be of a more routinely and less complex nature, while full-court decisions are more likely to have complicated substantial or legal issues involved.

Because the magnitude of chambers differs between various types of courts, the number of judges was rescaled, so that e.g. three judges on the Supreme Court equal a single judge on a district court. This rescaled (ordinal) variable ‘Njudges’ has two values.

6.2.4 Length of the decision

In general, complex decisions can be expected to require more words than the more routine judgments, which are often produced with the help of standardized text modules. The length of the decision might therefore be a relevant variable. To obtain a workable measurement, a grouping of the values was necessary. Five groups for the variable ‘Length’ were created: 100-400 words, 400-1000 words, 1000-2000 words, 2000-5000 words and >5000 words. Decisions having less than 100 words were considered to be without text (and thus discarded as having missing data, *supra* § 4.3)

6.2.5 Other Predictors

Apart from the four endogenous variables, discussed in the previous paragraphs, we used four other predictors, which are similar or closely related to predictors for the citation period.

FieldOfLaw and Hierarchy will be used for the publication period. They are the same as used for the citation period.

In § 5.2.1, the case law database of Rechtspraak.nl was considered to be one of the jurisprudential periodicals, relevant for the calculation of the PubMagazines variable. This is reasonable, since all these publications occur before the start of the citation period. But in § 4.2 we considered an immediate publication on Rechtspraak.nl to be in the publication period, while publications in other magazines are assumed to take place in the transition period. Therefore, the insertion of a judgment in the case law database of Rechtspraak.nl can be regarded as a predictor in the publication period. We abbreviate this (ordinal) variable as ‘PubWebsite’; the possible values are 0 and 1.

The press release sometimes accompanying the judgment on Rechtspraak.nl (*supra*, § 5.2.2), can similarly be regarded as a predictor variable for the publication period. This variable ‘PubPress’ can have three values: ‘0’ (no press release), ‘1’ (press release on the sub-website of the rendering court) or ‘2’ (press release on the national homepage of Rechtspraak.nl).

6.3 Regressor Variable

Immediately following the publication period a judgment can start being published, annotated and cited. Although the publication on Rechtspraak.nl is part of the publication period and from that

point in time the judgment is freely accessible for all, it turns out that most cases that are cited in case law or literature are not only published on the website, but also in one or more commercial periodicals.¹¹ So, either jurists still rely heavily on these magazines to have a good overview of relevant new decisions, or the editors of these magazines are making the right pick indeed. Either way, we can create the fiction that selection for periodicals is the starting point for the assessment work by the rest of the legal crowd. And thus the PubMagazines variable can be used as the regressor variable for the publication period: based on our eight predictors we calculate the odds that a judgment will score a certain value for PubMagazines.

One adjustment has to be made though. Since the publication on Rechtspraak.nl is now a predictor (*supra*, § 6.2.5), we cannot use a regressor that has this publication as one of its constituents (*supra*, § 5.2.1). As a result, we create ‘PubMagazines_1’, which excludes the publication on Rechtspraak.nl.

6.4 Calculations

As we did for the citation period, we calculated all 255 possible subsets for the publication period. If only one variable is used, Hierarchy is explaining most variability, followed by PubWebsite and Njudges. Worst scores are for PubPress and OutCitationsLegislat. PubPress is also the predictor that can be ignored most easily. PubPress doesn’t even turn out to be statistically relevant at all, and is therefore left out of the model. As with the subsets for the citation period, Hierarchy cannot be easily left out: it is present in all 110 best subsets.

When all variables, except PubPress, are used, R^2 is 0,28. Although less than the R^2 of the citation period, it is still an acceptable score. As with the predicted value in the citation period (*supra*, § 5.4), we standardized the predicted value for the publication period, which we label as ‘MARC-PP’, the MARC-indicator for the publication period.

Table 2. Wald statistic and estimate for all variables of the publication period.

Variable	Wald	Estimate
Intercept	17102	-6,0644
Hierarchy	12130	2,5877
FieldOfLaw-Criminal	5447	-0,3607
FieldOfLaw-Civil		0,4361
FieldOfLaw-Administrative		-0,0754
Njudges	5367	0,7013
OutCitationsCaseLaw	4654	0,0142
PubWebsite	1278	0,2866
OutCitationsLegislat	1095	0,0053
Length	358	0,0730

¹¹ According to [16], 21,3% of the cases cited in judicial decisions and 17,3% of cases cited in scholarly writings are published on Rechtspraak.nl only, and not in commercial periodicals. These judgments are receiving significantly less citations though than those judgments which are (also) published in commercial magazines.

In Table 2 the Wald statistic for the various variables of the publication period are displayed. Compared to the citation period (*supra*, Table 1), it is noticeable that the values are more evenly distributed: instead of one overriding variable (InCitationsCaseLaw for the citation period), all variables contribute to the model more equally. Hierarchy scores best, followed by Field of Law, Njudges and OutCitationsCaseLaw.

7. TRANSITION PERIOD

The transition period is between the publication period and the citation period. It is the period in which a judgment – if considered to be legally important – will be published, annotated, cited and discussed. Without a transition period, the citation period would start on one specific day, and the change from MARC-PP to MARC-CP could be a wild jump. The transition period is meant to smooth such sudden changes. In § 4.2 we defined it as a 100-day period. During this period the MARC-indicator (MARC-TP) will be calculated as a weighted average of MARC-CP and MARC-PP, depending on the day within the transition period. In formula:

$$MARC_{tp} = \frac{(100 - TD) \cdot MARC_{pp} + TD \cdot MARC_{cp}}{100}$$

Where TD is the number of the day in the transition period.

8. TOWARDS A USER-FRIENDLY MODEL

The MARC-indicators calculated thus far can have thousands of different values, ranging from something like -0,4894170847 to 32,663963198. Displaying such a number to the end user wouldn't make sense though. A classification with five categories, on the internet also used for rating movies, hotels, recipes, videos and books, would be more user friendly. We label these categories MARC-1 to MARC-5, the latter being the highest.

Setting the boundaries between these five categories will always be a little arbitrary, but this arbitrariness can be reduced by a thorough analysis of the contents of the database for which the classification is used. If MARC is used for a collection just containing case law published in commercial periodicals, a distribution with 20% in each category might render justifiable results. In our database though 62,3% of the judgments has never been published at all. Although some of these judgments probably should have been published, a distribution with 20% in each category will not meet the expectations and perceptions of the end users.

For our database we start with the assumption that (net) all published decisions, and ca. 5% of the unpublished decisions should be in MARC-2 to MARC-5, which places (rounded) 60% of the judgments in MARC-1. The size of each next category we set at about 40% of the previous one, while MARC-5 shouldn't be more than 0,5% (ca. 4.000 judgments). The suggested magnitudes are thus: 60–25–11–3,5–0,5.¹²

¹² This distribution is comparable to the one used in HUDOC, the case law database of the ECHR, in which all decisions of this Court are published. It has four categories (with manual classification), with the distribution (as of 7-1-2013): 62-19-13-6. The last category covers all judgments published in the Case reports.

9. ADDITIONAL TESTS

9.1 General Remarks

Because the regressor variable for the citation period is composed of the citation data of three consecutive years, and data are available till June 2010,¹³ we could use all decisions up to and including 2006. The number of usable records (not having missing data) from this period is 274.967 decisions. The distribution of unpublished versus published cases for this set is comparable to the whole database (64,3% versus 35,7%).

Although the predicted values for both periods are standardized, we have to be aware of the fact that the distribution in the both periods differ substantially. The boundaries between the five categories are therefore different between publication period and citation period.¹⁴

9.2 Comparing MARC-PP to MARC-CP

Apart from Hierarchy and FieldOfLaw, the publication period and the citation period share no predictor variables. It is therefore interesting to see whether judgments end up in the same MARC-category for both periods. In other words: is the MARC-indicator for the publication period a reliable indicator for legal authority in the citation period?

For a fair comparison we have to take into account that the value for MARC-PP is steady, while MARC-CP is changing over time (due to both Age and the time component in InCitationsCaseLaw and InCitationsOneLit). Because we calculated MARC-CP over 2006, we have to limit ourselves to judgments of the year 2006 for this comparison between MARC-PP and MARC-CP.

The classification of these 81.710 testable decisions from 2006 is displayed in a multiple response table (Table 3). Figures in bold are classified in the same MARC category both in publication and citation period; in total 87,5%. Of the remaining judgments 11,9% shifts only one category up or down and just 0,6% moves two categories.

Table 3. Distribution of judgments of 2006 over the five MARC categories in citation and publication period. N=81.710.

Citation period	Publication period					Total
	1	2	3	4	5	
1	71,1	0,1	0,0	0,0	0,0	71,2
2	3,9	11,1	0,9	0,0	0,0	15,8
3	0,0	4,8	4,8	1,2	0,0	10,9
4	0,0	0,5	0,7	0,4	0,2	1,7
5	0,0	0,0	0,1	0,1	0,1	0,3
Total	75,0	16,5	6,5	1,7	0,3	100,0

¹³ Because exact publication dates of scholarly writings were missing, the date component in the citation variables is based on full years only.

¹⁴ As a consequence, when classifying records in the transition period, not only the predicted values have to be weighted, but also the boundaries between the categories in publication and citation period have to be recalculated.

From this table we can conclude that the predictors used for the publication period render quite good results, and therefore MARC-PP is a quite reliable indicator of legal importance for judgments that haven't been assessed by the legal crowd yet.

9.3 Testing the Crowd Hypothesis

In § 5.1 we posed the hypothesis that during the citation period all variability in the regressor variable can be explained by the five crowd variables, Age, Hierarchy and FieldOfLaw. Holding on to this hypothesis enabled us to compare, in the previous section, the results of the citation period with the publication period. If we would have used also the endogenous variables in the citation period, the comparison would have been less fair because of the same variables being used for both calculations.

Still though, it is interesting to see if the endogenous variables are able to improve the results of the analysis in the citation period. Therefore we added OutCitationsCaseLaw, OutCitationsLegislat, Njudges and Length as predictors. PubWebsite and PubPress aren't added since they already are an integral part of, respectively, PubMagazines and PubAnnotations.

The resulting R^2 is 0,55, which is substantially better than the 0,47 scored originally (*supra*, § 5.4). We have to take care though, since in the generalized (non)linear model (nearly) every extra variable contributes to the outcome, while too many variables aren't contributing to the stability of the model. A balance therefore has to be sought. To assess the relative importance of all variables, Table 4 displays the Wald statistic for this extended model of the citation period. According to the Wald statistic the endogenous variables Length, Njudges, OutCitationsCaseLaw and OutCitationsLegislat are the worst performers, together with PubAnnotations and InCitationsOneLit, which we already determined as having a lot of overlap with PubMagazines and InCitationsCaseLaw. The low performance of

Table 4. Estimates and Wald Statistic for variables in extended model for citation period.

Variable	Wald statistic
Intercept	51688
InCitationsCaseLaw	339724
Age	22083
InCitationsPermLit	17615
Hierarchy	8458
PubMagazines	6160
FieldOfLaw-Criminal	3337
FieldOfLaw-Civil	
FieldOfLaw-Administrative	
Length	2220
InCitationsOneLit	1607
Njudges	1083
OutCitationsLegislat	1083
PubAnnotations	148
OutCitationsCaseLaw	5

OutCitationsCaseLaw is striking, since this variable plays quite an important role in the model for MARC-PP (*supra*, Table 2).

Given the very low Wald statistic of these endogenous variables, we could draw the conclusion that the hypothesis holds: the crowd variables suffice to establish legal importance.

10. CONCLUSIONS AND FUTURE WORK

In this paper we summarized the results of our research on the possibilities to develop a computer model (MARC) that can rate judicial decisions on their legal authority. MARC does not try to establish the uniqueness and specificity of single decisions by trying to understand the legal reasoning, but instead uses the wisdom and opinion of the legal crowd to establish whether a decision will play a role in future legal practice and debate. To this end, two basic models were developed. A first one for the citation period, in which the rating of a decision is determined by the extent to which it has been published, annotated and cited. The second model is for the – short – period in which scholars and judges haven't had time yet to review, comment and cite the judgment. For this period we use almost solely endogenous variables to predict its future authority. Also, an intermediary model was developed to smooth the transition from the publication period to the citation period. Finally, a model with five categories was devised to make the model comprehensible for end users of databases or integrated search engines. The statistical exercises helped us to gain insight in the relevance of the various variables for establishing legal importance.

The extent to which the classification of decisions in the publication period overlaps with that of the citation period surpassed expectations. Although the predictors used for the publication period seem to be of little added value in a model that uses all available variables for the citation period, they do have enough predictive value on their own.

When used in live environments, there are some issues for consideration. The first one was already mentioned: the boundaries of the categories have to be chosen with care, and closely monitored when more (collections of) decisions are added to the database. Second, a policy on missing data has to be developed. Records with missing data are automatically left out, leaving them without classification. If this is unwanted, missing fields have to be populated with e.g. mean, median or mode. Also, in more general terms, data quality is an important issue, and – as long as properly linked data are not supplied – the parsers used for extracting references to case law and legislation have to be monitored constantly, since identifiers and citation habits change.¹⁵ Refinements of the model are conceivable. Other predictors could be introduced, and integration with e.g. text comparison tools could be examined.

Also manual review and manipulation of the (outcomes of the) model are imaginable. The public (possibly limited to jurists) could be asked to rate decisions as to their perceived legal authority. Here, a word of caution seems appropriate though: opening such a rating system up to the public at large may lead to improper classifications, when people are mistaking the atrocity of a crime or the celebrity status of plaintiff or defendant for legal importance. Limiting the group of evaluators to the legal

¹⁵ E.g. the introduction of new identifiers like the European Legislation Identifier [17] and the European Case Law Identifier [18].

community mitigates this risk, but also lawyers can easily mix up 'rate' with 'like'. Using explicit user ratings makes the system manipulable and dependent on the commitment of a probably limited group of users. The same risk is attached to the use of usage data of case law databases.¹⁶

Since the results of our research look very promising, we are in favour of improving the system as described in this paper, instead of investing in the development of dedicated systems to activate the crowd for rating judicial systems. On further development, focus should be on collecting more data from which the (implicit) opinion of the legal crowd can be read, improving data quality and closely monitoring the performance of the model in a live environment.

11. REFERENCES

- [1] S. M. Marx, *Citation Networks in the Law*, *Jurimetrics Journal* 10 (1970): 121-137.
- [2] C. Tapper, *The use of citation vectors for legal information retrieval*, *Journal of Law & Information Science* 1 (1982): 131-161.
- [3] D. G. Post and M. B. Eisen, How Long Is the Coastline of the Law? Thoughts on the Fractal Nature of Legal Systems, *The Journal of Legal Studies* 29 (2000): 545-584.
- [4] T. A. Smith, *The Web of Law*, *San Diego Law Review* 44 (2007): 309-354.
- [5] J. H. Fowler, T. R. Johnson, J. F. Spriggs II, S. Jeon and P. J. Wahlbeck, *Network Analysis and the Law: Measuring the Legal Importance of Precedents at U.S. Supreme Court*, *Political Analysis* 15 (2007): 324-346.
- [6] M. van Opijnen, Citation Analysis and Beyond: in Search of Indicators Measuring Case Law Importance, in B. Schäfer, ed., *Legal Knowledge and Information Systems - JURIX 2012: The Twenty-Fifth Annual Conference*, IOS Press, Amsterdam, 2012, pp. 95-104.
- [7] A. Geist, *Using Citation Analysis Techniques For Computer-Assisted Legal Research In Continental Jurisdictions*, Graduate thesis, University of Edinburgh, 2009.
- [8] S. Malmgren, *Towards a theory of jurisprudential relevance ranking. Using link analysis on EU case law*, Graduate thesis, Stockholm University, 2011.
- [9] R. Winkels, J. de Ruyter and H. Kroese, Determining Authority of Dutch Case Law, in K. M. Atkinson, ed., *Legal Knowledge and Information Systems. JURIX 2011: The Twenty-Fourth International Conference.*, IOS Press, Amsterdam, 2011, pp. 103-112.
- [10] M. van Opijnen, Canonicalizing Complex Case Law Citations, in R. Winkels, ed., *Legal Knowledge and Information Systems - JURIX 2010: The Twenty-Third Annual Conference*, IOS Press, Amsterdam, 2010, pp. 97-106.
- [11] M. van Opijnen, Searching for References to Secondary EU Legislation, in S. Tojo, ed., *Fourth International Workshop on Juris-informatics (JURISIN 2010)*, 2010.
- [12] D. C. Blair and M. E. Maron, An evaluation of retrieval effectiveness for a full-text document-retrieval system, *Communications of the ACM* 28 (1985): 289-299.
- [13] W. B. Eldridge, An appraisal of a case law retrieval project, in D. Johnston, ed., *Proceedings of the Computers and the Law Conference 1968*, 1968, pp. 36-61.
- [14] J. Surowiecki, *The Wisdom of Crowds*, Anchor Books, 2005.
- [15] M. Malsch, G. Ijpelaar and H. Nijboer, Publicatie van strafrechtelijke uitspraken - evenwichtige rechtsontwikkeling of ijdelheid van de rechter?, in M. Malsch and N. van Manen, eds., *De begrijpelijkheid van de rechtspraak*, Boom, Den Haag, 2007, pp. 95-114.
- [16] M. van Opijnen, *Jurisprudentiestatistiek - over de publicatie, annotatie en citatie van rechterlijke uitspraken*, NJB 2011/1658.
- [17] M. van Opijnen, *The European Legal Semantic Web: Completed Building Blocks and Future Work*, 2012 <ssrn.com/abstract=2181901>.
- [18] M. van Opijnen, European Case Law Identifier: indispensable asset for legal information retrieval, in M. A. Biasiotti and S. Faro, eds., *From Information to Knowledge. Online Access to Legal Information: Methodologies, Trends and Perspectives*, IOS Press, 2011, pp. 91-104.

¹⁶ The use of usage data (of the public case law database) has been considered, but a first analysis revealed an excessive interest in gruesome rape and murder cases and other decisions that attracted media attention. We therefore decided to leave them aside.