Water International

ISSN: 0250-8060 (Print) 1941-1707 (Online) Journal homepage: https://www.tandfonline.com/loi/rwin20

# Understanding water disputes in Chile with text and data mining tools

Mauricio Herrera, Cristian Candia, Diego Rivera, Douglas Aitken, Daniel Brieba, Camila Boettiger, Guillermo Donoso & Alex Godoy-Faúndez

**To cite this article:** Mauricio Herrera, Cristian Candia, Diego Rivera, Douglas Aitken, Daniel Brieba, Camila Boettiger, Guillermo Donoso & Alex Godoy-Faúndez (2019) Understanding water disputes in Chile with text and data mining tools, Water International, 44:3, 302-320, DOI: 10.1080/02508060.2019.1599774

**To link to this article:** https://doi.org/10.1080/02508060.2019.1599774

Published online: 30 Apr 2019.

WATER INTERNATIONAL
2019, VOL. 44, NO. 3, 302–320
https://doi.org/10.1080/02508060.2019.1599774

Routledge
Taylor & Francis Group

RESEARCH ARTICLE



# Understanding water disputes in Chile with text and data mining tools

Mauricio Herrera[a], Cristian Candia[a,b,c], Diego Rivera[d], Douglas Aitken[a], Daniel Brieba[e], Camila Boettiger[a,f], Guillermo Donoso[g] and Alex Godoy-Faúndez[a]

[a]Centro de Investigación en Sustentabilidad y Gestión Estratégica de Recursos, Facultad de Ingeniería, Universidad del Desarrollo, Santiago, Chile; [b]Collective Learning Group, MIT Media Lab, Massachusetts Institute of Technology, Network Science Institute & Northeastern University, Boston, MA, USA; [c]Centro de Investigación en Complejidad Social, Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile; [d]Department of Water Resources, Laboratory of Comparative Policy in Water Resources Management, University of Concepción, Chillán, Chile; [e]Escuela de Gobierno, Universidad Adolfo Ibañez, Santiago de Chile, Chile; [f]Centro de Justicia Constitucional, Facultad de Derecho, Universidad del Desarrollo, Santiago, Chile; [g]Centro de Derecho y Gestión de Aguas, Departamento de Economía Agraria, Facultad de Agronomía e Ingeniería Forestal, Pontificia Universidad Católica de Chile, Santiago

**ABSTRACT**

This article provides a multidimensional study based on data and text mining of prosecuted disputes on water rights in Chile, and an analysis of the state's capacity, particularly of the institutions related to water regulation. This study shows not only a substantial increase of legal disputes regarding water rights over the years (1981–2014), but also clear patterns in the geographic location of these conflicts, as well as in the types of legal actions, arguments and strategies used in their pursuit. Through a topic analysis, we find a growing diversification over time of the subjects contained in the legal claims, suggesting an increase in structure and complexity.

**ARTICLE HISTORY**
Received 8 January 2018
Accepted 23 March 2019

**KEYWORDS**
Text mining; network theory; geographic analysis; data mining; Chile

## Introduction

### *The Chilean water market and the legal framework around water rights: a brief contextualization*

The Chilean system of water management and allocation is an interesting case study due to its unusual tradable water rights system (Bauer, 2015). Few other places, such as some of the western states of the United States, South Africa, Australia, Iran and Spain's Canary Islands, also have water trading schemes (Saleth & Dinar, 2005). This system has attracted considerable analysis by proponents and critics because of its strongly free-market approach, reducing to a minimum the interference of the state as the main actor to drive water allocation (Bauer, 2005, 2015; Budds, 2004, 2009; Hearne & Donoso, 2014; Vergara, 2002, 2015).

Since 1981 the water market framework in Chile has been dictated by the Chilean Water Code. The Chilean Water Code regulates water allocation as a de facto private property right separated from land, freely traded and subject to minimal state

**CONTACT** Alex Godoy-Faúndez alexgodoy@ingenieros.udd.cl

regulation, even though water resources are defined as a national good for public use under the Chilean Constitution (Section 19, paragraph 23; see also Section 595 of the Civil Code and Section 5 of the Water Code). Under the Water Code, the Chilean state grants a petitioner's rights for the use of surface water and groundwater. Once all water rights have been allocated within a basin, the management of water rights at the watershed scale is overseen by water users' organizations (WUOs). These entities are supposed to manage the shared resources from the same river, channel, aquifer or reservoir and act as decentralized private organizations independent of the state (Retamal et al., 2012; Vergara, 2015).

A state agency, the General Water Directorate (Dirección General de Aguas, DGA), is mandated to protect the resource and maintain an updated nation-wide inventory of water resources, as well as a registry of granted water rights. Thus, the Chilean state is legislator, regulator and surveyor of water resources and should provide updated and reliable information regarding water uses and management. The state also has a role in settling water rights conflicts, both through the DGA and through the judiciary branch of the Republic of Chile, including the Supreme Court and the Courts of Appeal.

The overall legal framework thus seeks to foster water markets as a secondary allocation mechanism among users with no direct governmental intervention. The primary allocation depends on a nonmarket system under governmental management. If there is available water within a basin or watershed after governmental intervention, it can be reallocated through the exchange of water rights, i.e., the right to use certain amounts of water, either for a limited time (lease) or in perpetuity (sale).

Some economists argue that water trading can promote more efficient water distribution because theoretically a market-based price acts as an incentive for users to reallocate resources from low-value activities to high-value activities. In practice, however, water markets have also created distortions, such as frequent stockpiling of water rights, where users hold water rights without using them, for speculative purposes (Scott, 1995). Difficulties in correctly assessing the social and environmental outcomes of water trading schemes, and the ethics of applying economic principles to a resource such as water, have raised concerns regarding the fairness of water markets in secondary allocation (Donoso, 2011).

### *The Water Code reform*

In 2005, the Water Code was significantly amended. The main concerns this reform sought to address were the insufficiency of state regulation to ensure effective functioning of water markets, and, relatedly, the need to discourage the stockpiling of water rights (Hearne & Donoso, 2005). The reform sought to reduce possible monopolistic distortion in the water and energy markets and to mitigate the effect of unused water rights on the economic development of the basins (Bauer, 2015). To achieve these goals, the reform established a fee applicable to unused surface and groundwater rights, to reduce hoarding. The reform also introduced tools for improving the management of aquifers and for expanding the powers of administrative evaluation concerning requests of water rights. For instance, it allowed the DGA to deny entirely or in part a water right petition if the volume requested was not justified by its proposed use (Article 147 *bis* of the Water Code), thus matching the authorized use of the resource to the amount

of water actually required. Finally, the reform allowed users to solve disputes in local courts, instead of having to litigate in a single court in the capital (Santiago), facilitating their access to the judicial system. In these ways, structural changes were made to water legislation in 2005 with the intent of enhancing sustainability, efficiency and equity in the use of this resource.

## Sources of conflicts and new approaches to understanding them

Chile's economy, particularly in the agricultural and mining sectors, is greatly dependent on reliable freshwater resources. Agriculture is the dominant consumer by a considerable margin; the demand for water by the agricultural sector is estimated at 82% of total use, according to the DGA Water Atlas 2016. In the years 2010 to 2016, Chile's central and central-northern (Norte Chico) zones were hit by a 'mega drought', which has been linked to a combination of climate change and natural variability (Boisier, Rondanelli, Garreaud, & Muñoz, 2016; Garreaud et al., 2017). This phenomenon affected over 60% of the agricultural industry's export products. The mining industry also suffered, particularly in the region of Antofagasta, where the industry is the largest water consumer (Aitken, Rivera, Godoy-Faúndez, & Holzapfel, 2016). Due to the demands of the mining and industrial sectors, many of Chile's regions are under considerable hydric stress due to low water availability, variability in supply and the increasing impact of climate change (OECD, 2015). Water scarcity increases the costs of copper production (which amounts to 13% of GDP), due to higher consumption of desalinated water (water–energy trade-off). It also lowers farm productivity because of higher crop losses and lower crop yields. These effects increase food prices, destroy jobs and increase operational costs for both mining and agriculture companies. This situation puts a substantial proportion of Chile's economy at considerable risk and must be addressed accordingly.

Aitken et al. (2016) showed that all of Chile's central-north and northern regions currently have water stress indices greater than 1 (high overexploitation), with a value as high as 51.6 for the region of Antofagasta (the main mining region), once environmental flow demands were considered. In most regions, agriculture accounted for around 70–90% of water demand. In Antofagasta, however, mining was calculated to account for about 64% of total water demand. This situation in which demand in specific sectors far exceeds water availability has a potential for severe conflict.

The unequal distribution of water rights, coupled with increasingly unbalanced water budgets, is generating tension among water users amidst more intense competition for this resource. Thus, rising water demand and climate/geographical variability create conflicts both between water right holders and between them and local communities, who often feel powerless due to their lack of financial resources to press their claims. These conflicts involving water rights holders need to be resolved within a WUO, whether or not the holders share the same water source (Hearne & Donoso, 2014; World Bank, 2013). Water users can now appeal decisions of the WUO to ordinary courts. Therefore, when legal disputes occur between parties, it is the ultimate responsibility of the courts to settle the conflict. According to the Water Code, there are

several institutions (e.g., WUOs, DGA, Courts of Appeal, Supreme Court) with diverse degrees of authority who are empowered to render decisions on water uses.

In terms of the country's overall governance framework for water issues, the World Bank (2011, 2013) finds that the complexity of the current institutional framework often leads to duplication in the execution of functions, gaps of omission, and problems of coordination between different agencies and organizations. There are also several institutional problems linked to low levels of funding, an inefficient system for generating relevant information, poor definition of policies and sectorial objectives, and the lack of institutional coordination. Thus, uncoordinated actions among these institutions, not based on reliable data, could produce inefficiencies and diminish the state's capacity to resolve conflicts.

However, perceptions about water conflict resolution have so far been mostly based on fragmentary evidence (Bauer, 2015) or on users' and academics' opinions (Vergara, 2015). For instance, water users' perceptions of institutional capacity to solve conflicts are mainly based on their experiences. Thus, users may conclude that institutions are making subjective judgements and that rules are being bent or broken depending on litigants' relative (and highly asymmetrical) economic power.

In an attempt to go beyond perceptions and understand regional variations in water conflicts through data, Rivera et al. (2016) used a new approach based on geotagging of the public legal records of judiciary judgments, as a proxy for the geographical distribution of conflicts. They found that the number of water conflicts varies regionally, which yielded spatial patterns correlated with economic activities and the availability of water resources. For example, in northern regions, conflicts over water rights usually occur between agricultural and mining operations. In Chile's more densely populated central zone, conflicts are mostly related to the ownership of water rights, both between farmers and the water authority (DGA), and between farmers seeking water for irrigation use and companies in charge of the drinking water supply.

These results, however, concern the general pattern of water-related conflicts and do not consider the details, evolution or complexity of the conflicts, the relationship between conflicts and increasing demand of water among different sectors, or the impacts of climate change and of successive regulatory reforms. Therefore, further analysis is needed to address these unanswered questions. Concretely, how have legal demands – which are viewed as a proxy for conflict – evolved due to the increasing tension between industrial, agricultural and drinking water supply sectors arising from climate change and the unequal distribution of water rights? What was the real impact of regulatory reforms in the water rights market? How effective is the state in solving water-related conflicts?

## *Legal records as a tool for the spatial-temporal characterization of conflicts related to water*

The 1981 Water Code is based on a strong role of the judiciary as arbitrator in resolving disputes and enforcing management decisions. Therefore, one of the best ways to bridge science and policy, and to support the decision-making process in water-related issues, could be to analyze all available legal texts and data, using data mining tools (Sunil, 2013, 2014).

We propose a multidimensional study based on data and text mining of disputes (Waltl, Matthes, Waltl, & Grass, 2016) involving water rights conflicts brought before the courts, together with an analysis of the capacity of the state's institutions to manage these conflicts. These cases can be used as a proxy for water-related conflicts and therefore to analyze the dynamics and content of these conflicts for a better understanding of typical actions, resources and fundamentals of the claims in water rights lawsuits. This approach can also provide results based on data analysis to confirm or reject perceptions among water users regarding these issues.

## Material, methods and exploratory data analysis

### *Data sources for this study: water stakeholders and legal actions*

This work is based on historical case study research dating back to 1981, for which we analyzed almost 4000 legal demands (public legal records) made by stakeholders and water users. All these records are judgments and decisions issued by the Supreme Court (in Santiago, the capital city) and by Courts of Appeal, which are regional courts located in the major cities of the country. These data are available at the judiciary system website (http://www.pjud.cl) and were compiled by an external entity (Microjuris, 2017). For clarity, in this study, we considered the regions (the largest political-administrative divisions) of Chile as territorial units for the analysis. We also used records of registered water rights (DGA, 2017a) and records from the Public Registry of User Organizations reported by the General Water Directorate on their website (DGA, 2017b), as well as DGA personnel data (from http://transparencia.dga.cl/).

### *Computer-assisted analysis of jurisprudence*

Generally, legal action is the result of an unresolved conflict that has escalated due to failure of the parties to reach an agreed resolution in a first instance. Therefore, lawsuits can be taken as a proxy for an underlying conflict. Thus, we used legal actions as proxies to understand the dynamics, complexity, diversity and content of these conflicts (Wyner, Mochales-Palau, Moens, & Milward, 2010). This work differs from Rivera et al. (2016) in that it uses text mining techniques to analyze discourses instead of merely registering the occurrence of conflicts.

Using text mining tools, we extracted information from water lawsuit documents such as the Water Code's sections, regulations, and codes used by lawyers and judges in legal proceedings. We call this procedure computer-assisted analysis of jurisprudence. The *feature extraction algorithm* identifies articles related to the water code, the mining code, the civil code and civil procedure that appeared in the text of pleadings filed in the lawsuit procedures, as well as the final verdict, which is used as a label to classify the result as failed/rejected or successful/not rejected. With this information we build *vectors of features*, which are used to search for patterns by data mining techniques.

The number and diversity of legal concepts in the lawsuit procedures registered in the vectors of features were used to discover the different legal strategies, the types of actions and the fundamental arguments of the claims in each ruled case. With the computer-assisted analysis of jurisprudence we found that most water right legal

procedures can be put under several major themes related to the protection of property and the environment, property regularization, and legal formalism, in agreement with other studies (Hearne & Donoso, 2014; Rivera et al., 2016).

## Topic modelling

To better understand the nature of these conflicts, we applied topic modelling techniques from text mining to legal texts (Firdhous, 2010). Topic modelling is a type of statistical modelling used to discover the latent topics in a corpus or collection of documents. We used it as a procedure for deriving high-quality information from the analyzed texts to understand in greater detail the issues covered by water-related legal demands over time (Merkl & Schweighofer, 1997). This kind of analysis allows us to delve into the context and local reality of each region, including inferring the causes that ultimately lead to legal action.

Using the texts of almost 4000 legal rulings provided by the Microjuris database (Microjuris.com, 2017), we constructed a corpus, and from this corpus we extracted 'hidden variables' or subjects. With the help of some probabilistic topic models (Steyvers & Griths, 2007), such as the latent Dirichlet allocation (LDA) model, we uncovered the thematic structure of the corpus built from the documents of the judicial proceedings.

The LDA model (Blei & Lafferty, 2009) assumes that each document in a corpus contains a mix of topics. However, the topic structure is hidden (or 'latent') because we can only observe the documents and words, not the topics themselves. This method infers the topic structure from the known words and documents.

For LDA, it is necessary to specify in advance the number of topics in the underlying topic structure. We used perplexity (Berthard, Ghosh, Martin, & Summer, 2009) to determine the optimal number of themes. Perplexity is a statistical measure of how well a probability model predicts a sample. As applied to LDA, for a given value of the number of topics, the parameters for the LDA model are estimated. Then, given the theoretical word distributions represented by the topics, the result is compared to the actual theme mixtures, or the distribution of words in the documents. This statistic is somewhat meaningless on its own. The benefit comes from comparing perplexity across different models with varying numbers of topics. The model with the lowest perplexity is generally considered the best.

Once we defined the number of topics, we analyzed the complete corpus built from all the available documents, from 1981 to 2014, to infer the structure of topics covered by all legal claims. After that, for a more detailed analysis – to understand the context of legal demands, and the contents and causes that lead to the most frequent type of legal actions – we split the corpus into three parts to balance them according to the number of conflicts. The 1980s and 1990s provided a significantly smaller number of documents per year. Therefore, to have a statistically significant sample for topic modelling, we included all documents up to the year 2000 in our first period of analysis. Specifically, we separated the corpus into three time intervals: $t_1$ = 1981–2000, $t_2$ = 2001–2008 and $t_3$ = 2009–2014.

For each set of documents in each period, we chose an optimal number $k$ of topics according to the lowest perplexity criterion and used LDA to identify the topic representation in each document and the words associated with each topic. For this, we used Gibbs sampling (Casella & George, 1992). A basic summary of this algorithm is as follows:

(1) The algorithm runs through each document in the corpus, and randomly assigns each word in the document to one of $k$ topics. This random assignment already provides both topic representations of all the documents and word distributions of all the topics. But because it's random, this is not yet a very accurate structure.
(2) To improve it, for each document $d$, the algorithm goes through each word $w$ in $d$ and for each topic $t$ computes
   - the proportion of words in document $d$ that are currently assigned to topic $t$, i.e., the conditional probability $p(t|d)$; and
   - the proportion of assignments to topic $t$ over all documents that come from this word $w$, $p(w|t)$.
   - It then reassigns $w$ to a new topic, where the topic $t$ is chosen with probability $p(t|d) \times p(w|t)$. This is the probability that topic $t$ generated word $w$. In other words, in this step, we are assuming that all topic assignments except for the current word in question are correct, so the assignment of the topic by adding the newest word is updated. After repeating the previous step many times, the algorithm eventually reaches a roughly steady state, where the assignments could be considered good.
(3) Using these assignments, it is possible to estimate
   - the topic mixture of each document (by counting the proportion of words assigned to each topic within that document); and
   - the words associated with each topic (by counting the proportion of words assigned to each topic overall).

In what follows, the analysis proceeds in two steps. First, through an exploratory data analysis, we show some major patterns and trends in the granting of water rights and in judicial rulings over time. We connect these trends with the state's ability to mediate conflict through administrative and judicial channels, as well as with the apparent effects of the 2005 reform to the Water Code. In the second step, we proceed to the topic modelling.

## Results

### *First screening of data on the structure of the Chilean system of water management and allocation*

A first screening of data showed that Chile has more than 3000 WUOs across the country, 50 Surveillance River Boards, and nearly 200 Associations for Water Channels (canal users). These stakeholder organizations related to water distribution and consumption create a complex system of interrelationships.
In addition to these organizations, there are state agencies responsible for granting water use rights, regulating the market and developing infrastructures for water use.

Figure 1 shows the number of water-use rights granted by the DGA throughout the country. Figures 1(a,b) are choropleth maps of Chile for regions and communes, respectively, with darkening shades representing a greater number of water rights granted by the DGA, the state agency in charge of this process. Chile has a high diversity of ecosystems, leading to four climatic regions. From Figure 1(a), it is apparent

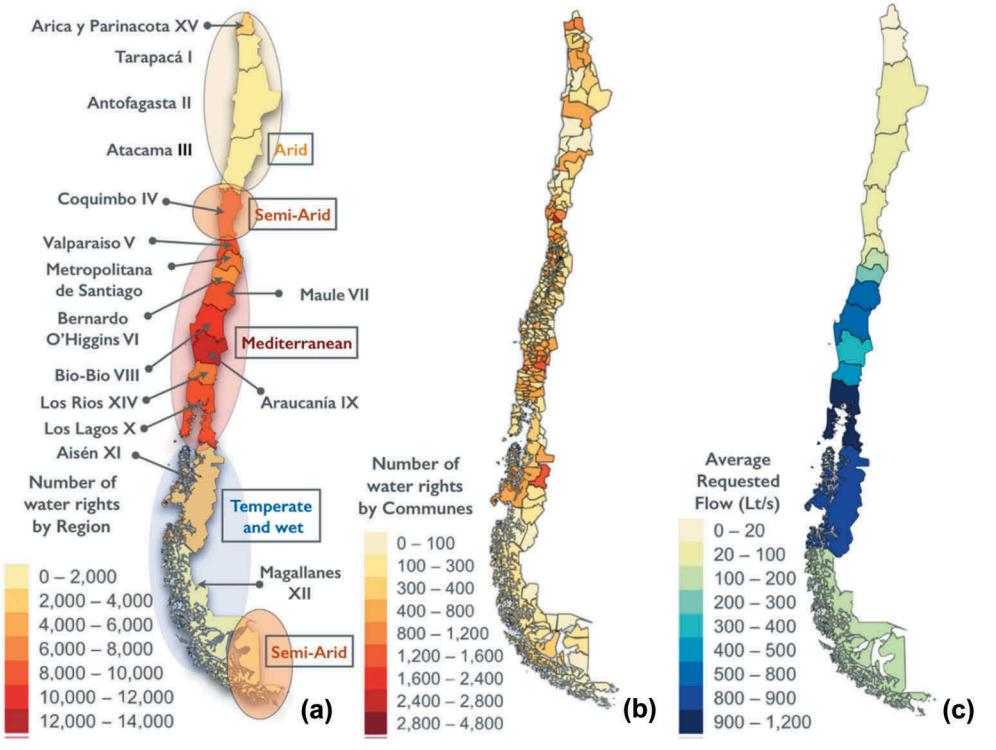


**Figure 1.** (a) Chile's administrative regions and climatic zones (labelled in boxes). Darker shades indicate a greater number of water use rights that have been granted by the General Water Directorate. (b) Water rights distribution by commune within each region. The requests for water rights are accompanied by demands for flows. (c) Average flow requested in each region, with darker shades indicating greater requested flows.

that the number of granted water rights correlates with water availability in different climatic regions. This is not a coincidence. There is little availability of water in the north, so the number of water rights is also low, while the central zone has a more humid climate, with more water, and therefore more water rights for trading.

## *More water rights granted, despite the Water Code reform*

The 2005 Water Code reform introduced tools to improve the management of aquifers and to expand the powers of administrative evaluation concerning requests for water rights, matching the use of the resource with the amount of water required. Because the reform strengthened regulatory authority over future grants of water use rights by establishing additional requirements for the petition of new rights, fewer grants of water rights might have been expected. Moreover, in 2005 Chile began to be affected by a severe drought, which caused a significantly less availability of water resources. In this sense, our results (Figure 2) are counterintuitive.

The main chart in that figure shows the evolution in the number of judicial rulings, while the inset shows the number of water rights granted over time. After the 2005 amendment to the Water Code there was a significant *increase* in the number of newly

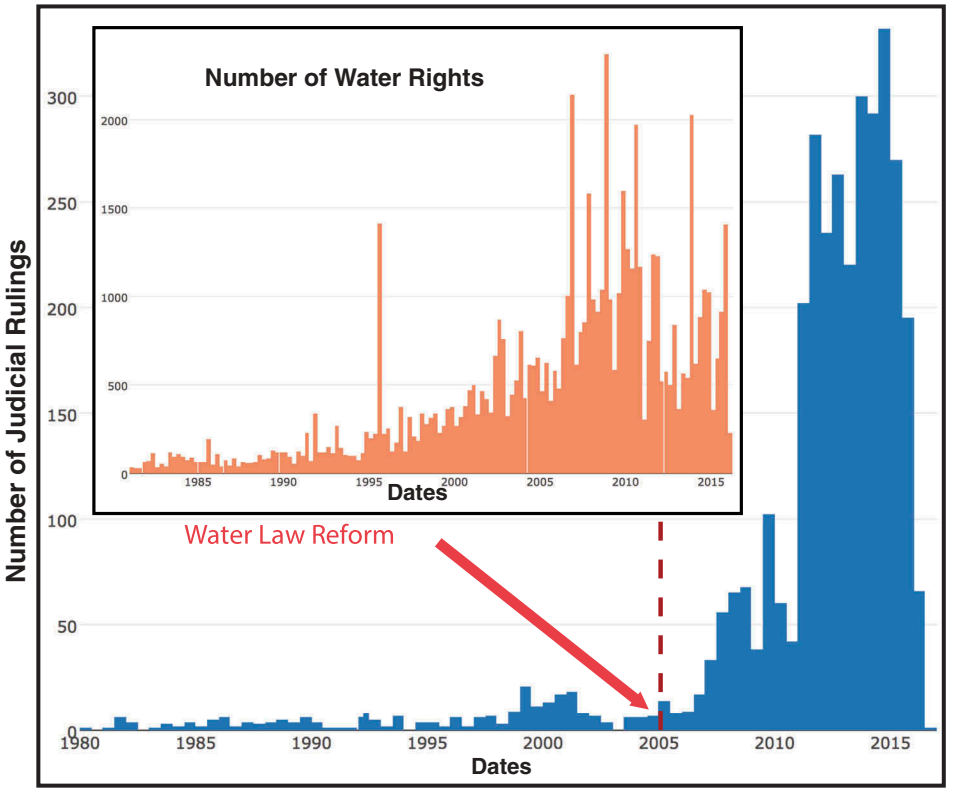


**Figure 2.** Larger panel: evolution of legal demands over time (DGA, 2017a). Inset panel: evolution in number of water rights over time.

granted water rights in practically all regions of the country. Two additional patterns are also apparent: the 'explosion' of legal claims after the modification of the Water Code in 2005; and the correlation (with some delay) between the granting of water rights and the explosion in legal disputes.

## *Water Code reform and real water demand*

What might have caused this almost simultaneous explosion in the allocation of water rights and in legal disputes? The granting of new water rights has been accompanied by an increase in market demand for more water. Figure 3 shows the average flow requested by users in each region. In this figure, if we compare the requested flow associated with the water rights market with the demand for water calculated from information from specialized government agencies, we can see that in all regions the water rights market demand for flow (Market) is significantly higher than the demand for flow reported by government agencies (Demand). In other words, allocated water rights systematically exceed physical water flows actually consumed by users. Therefore, we can formulate at least two, mutually non-exclusive hypotheses: water supply calculations made by government agencies are not correct; and users ask for, and get, more

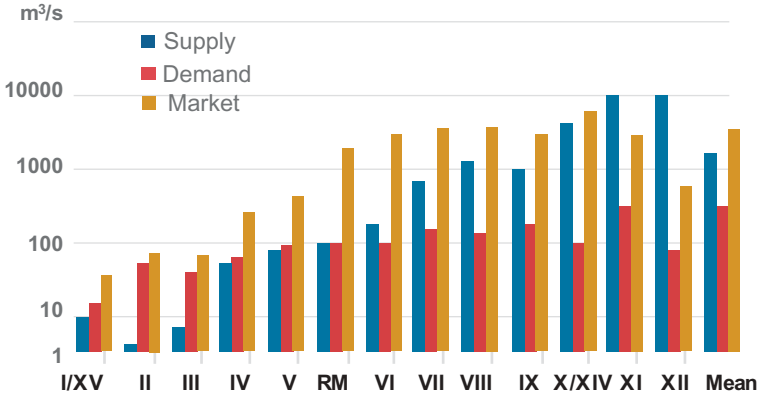


**Figure 3.** Water supply and demand in 2010, by region. Adapted from World Bank (2011).

water rights than they need, and therefore the market for water rights contains a high degree of speculation.

If the first hypothesis is true, one possible explanation is that governmental agencies miscalculated the water supply, because the official data were provided by the last water balance from 1987, itself based on records taken between 1951 and 1980. Those data, for example, did not include climate change and variability impacts, such as the high decadal variability of rainfall amounts (Montecinos & Aceituno, 2003; Quintana & Aceituno, 2012; Valdés-Pineda et al., 2014). Thus, they over-allocated water rights because they overestimated the available hydric resources.

If we assume the second hypothesis to be dominant, we conclude that the state agency in charge of regulating the market has failed to match the granting of new water rights to actual demand, thus encouraging market speculation in water rights. This is a powerful stimulus for conflict under conditions of water scarcity.

### *Water Code reform and non-local resolution of conflicts*

One important achievement of the reform was introducing the possibility for stakeholders to resolve conflicts through the local civil courts, giving more users access to justice. Indeed, as we saw above, our results confirm an increase in the number of lawsuits in local civil courts, which suggests the reform has worked in the intended direction.

But the inefficiency of the system is also evident from the analysis of the data. Our exploratory data analysis shows that more than 25% of legal claims end up in the Appeals and Supreme courts, which are both extremely centralized and outside the local framework supposedly favoured by the code. For instance, as shown by Rivera et al. (2016), most appeals are settled in a single court located in Santiago, where the DGA is, regardless of the location of the conflict. This is due to a double judicial review system, in which the decisions of the Courts of Appeal can finally be decided by the Supreme Court if the dissatisfied party appeals (appeal of cassation). Thus, the losing party can always appeal the decision of the regional court to the Supreme Court. This could overload the judicial system in the long run. Indeed, and inasmuch as cases 'travel upwards' to regional courts and then to Santiago, the final resolution of conflicts

through local courts – as intended by the Water Code reform – has not been achieved. This non-local resolution of conflicts suggests excessive centralization of the current institutional framework, which is paradoxical for water markets considered as a tool for decentralized conflict resolution.

According to the Water Code, many legal claims must be directly filed against the DGA. In our study, nearly 50% of registered litigations were conducted in this manner. Our exploratory data analysis finds that the DGA has lost up to 73% of these cases. This is highly suggestive of the low capacity of state organizations (Fukuyama, 2013), in the sense of having low technical and/or financial resources to execute its tasks adequately. Indeed, the DGA may lose so many cases either because its initial grant (or denial) of water rights to consumers was technically or legally deficient – forcing users to appeal to the courts for rectification – or because, having made reasonable allocations in the first place, its legal capabilities are so weak that the DGA can easily be defeated by users (particularly those with considerable legal support) in the courts.

Indeed, and notwithstanding the plausibility of the first possibility just discussed, the water authority's legal resources seem highly unequal to their task. A review of the DGA's personnel – publicly accessible thanks to transparency laws – reveals that in a typical month (June 2015), 39 lawyers were working at the DGA (data from http://transparencia.dga.cl/). Only 27 of the 39 lawyers were hired on a permanent or semi-permanent basis (as opposed to fee-for-services schemes), and only they can represent the DGA in court. Moreover, only 10 of the permanent or semi-permanent lawyers had functions which plausibly included representing the DGA in court, and only two of them had job descriptions which specified 'litigation' as their main occupation. Considering the several hundreds of outstanding legal cases at any one time, these figures seem low. Moreover, of these 10 possible litigation lawyers, 9 were employed in Santiago, which suggests that the water authority's capacity to litigate effectively at the local and regional levels may be particularly weak. Most lawyers are, in any case, assigned to the legal tasks associated with the allocation of water rights. Finally, if the DGA must resort to the ordinary courts of justice to request compliance with the sanctions they imposed and file complaints with the courts for infractions such as illegal extractions, the increased number of lawsuits could well overwhelm the DGA's capacity to resolve these cases.

Furthermore, the current legal and administrative framework provides important limitations on the DGA that prevent it from carrying out its work effectively. For example, the DGA must resort to the ordinary courts of justice to request compliance with the imposed sanctions; it must appeal to the Public Ministry, like any other individual; and it must eventually request the legal support of the State Defence Council (Consejo de Defensa del Estado) to register complaints with the courts on illegal extractions. Also, to add to the litany of weaknesses, the DGA does not have guaranteed access to the groundwater wells to carry out its enforcement function (World Bank, 2013). Of course, this is how the Water Code was initially constructed, relying on a strong role of the judiciary, with administrative agencies controlled at the end by the judges as arbitrators for resolving disputes.

## *Understanding legal complexities using topic modelling*

Using 3827 legal text documents (legal rulings) provided by the Microjuris database (Microjuris, 2017), we build one complete corpus for our topic analysis. We estimated

the optimal number of topics in the corpus by using the perplexity criterion and obtained more than 40 topics for the entire corpus.

Figure 4 shows the results of the topic analysis using the LDA model (Jurafsky & Martin, 2000). In Figure 4(a) the density graph shows an increase over time in the number and diversity of topics treated in the analyzed legal texts, which in turn suggests an increase in the number, diversity and complexity of underlying conflicts. Figure 4(b) shows the thematic structure of the corpus. Topics are numbered for easy viewing and are represented by nodes in a network, with sizes proportional to the frequency of each topic's occurrence over time. The links represent the interrelation between topics, and the 'clouds' around the set of nodes, with different colours,[1] separate the communities, that is, clusters with common global legal themes.

Topic analysis provides rich information that allows us to understand different sources of conflicts and at the same time to look for plausible explanations of the nature of these conflicts. In turn, this could help suggest proposals for future legal reforms. For example, one of the most frequent topics is Topic 1, represented by a large node in the upper central part of Figure 4(b). This topic is *groundwater use*. As mentioned, the DGA does not have guaranteed access to groundwater wells for their control (World Bank, 2013). On the other hand, with an extreme drought underway, the search for new sources of water is critical, so many conflicts can be expected given this combination of factors. This fact could be considered in future reforms of DGA responsibilities.

Topic 17 is another frequent topic, *cassation resources*. A cassation resource is an extraordinary resource with the purpose of nullifying a judicial sentence that contains an incorrect interpretation or application. The decisions of the regional courts may eventually be modified by the Supreme Court if the dissatisfied party interposes a cassation resource. The frequent occurrence of this specific topic is a confirmation that many claims end up, first, in the regional Appeals courts, and then in the Supreme Court. Both instances are centralized and outside the local framework of conflict

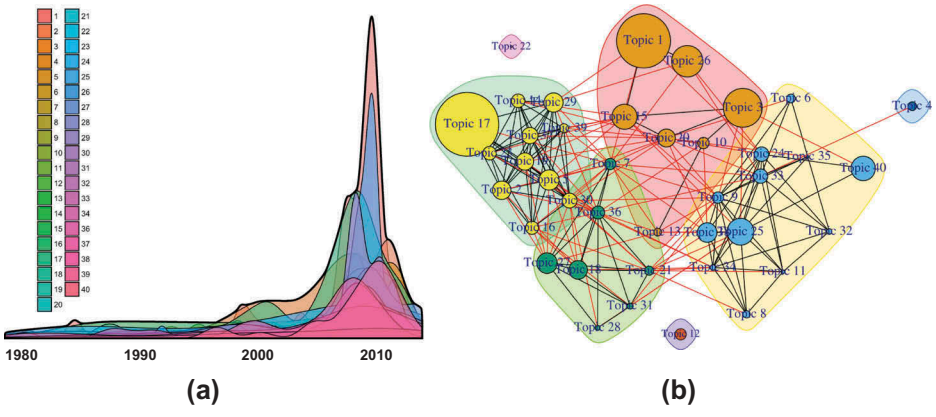


**Figure 4.** Topic modelling using latent Dirichlet allocation of a corpus formed from 3827 documents of legal rulings provided by the Microjuris database (Microjuris.com, 2017). (a) Density graph for 40 topics treated in legal text through time. Topics are numbered and represented with different colours. (b) Community structures for the network of topics. The size of the network nodes is related to the frequency with which the topic appears. Seven communities are shown.

resolution intended by the reform of the Water Code. In turn, *Topic 3* addresses issues of property protection and regularization. If we continued making new analyses in this way, we could uncover the detailed structure of underlying conflicts related to water use in the entire corpus.

We also performed a more detailed analysis of the data through a time segmentation of the corpus. The main goal was to reduce the number of topics in the analyzed corpus, to understand, as in the previous analysis, the context, contents and causes of frequent legal actions, while at the same time capturing some variations (dynamics) in the structure of conflict over time. For this, we split the corpus into three time intervals: $t_1$ = 1981–2000, $t_2$ = 2001–2008 and $t_3$ = 2009–2014. With this segmentation, we have balanced samples of the corpus for each time interval. Of course, other segmentations are possible. For example, it may be useful to split the data (the corpus) according to periods with relevant Chilean historical, social, administrative or climatic events to study their effects on conflicts. However, for the purpose of demonstrating variances over time in the structure of the corpus and hence capturing some dynamic aspects of the underlying structures of conflict, we believe it is sufficient to consider the proposed data partition.

For each set of documents in each period, we chose an optimal number $k$ of topics according to the lowest perplexity criterion and used an LDA model for topic modelling. Since the lowest calculated perplexity score was obtained for 15 topics in all periods ($t_1$, $t_2$ and $t_3$), the LDA model (Grüen & Hornik, 2011) was set to produce that number of topics. After running the code, the outputs showed diversification over time of the topics in legal claims (Figures 5–7), suggesting a certain increase in the complexity of conflicts over time. Although the optimal *number* of topics (calculated using the perplexity criterion) stays about the same (15), the subjects are mostly different in each period, and they are interconnected in different

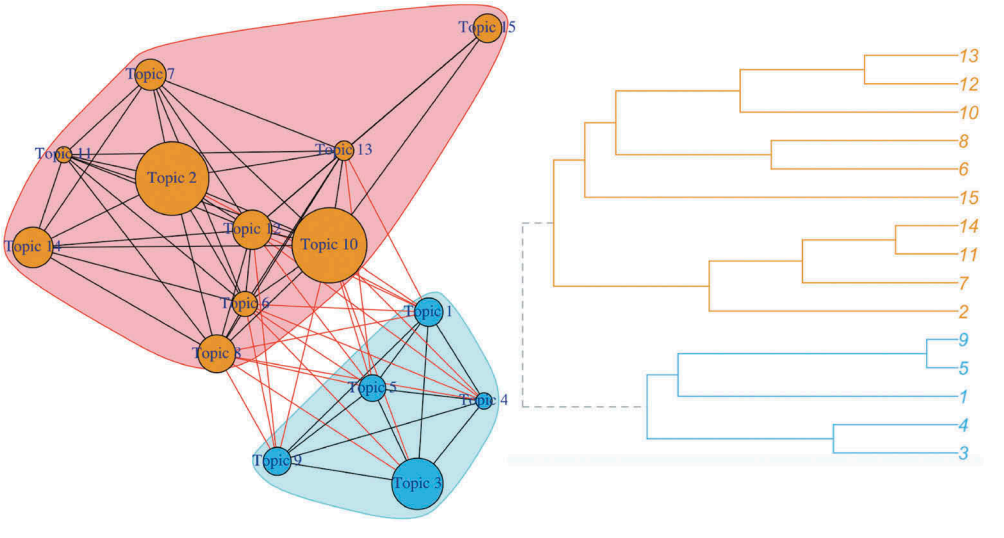


**Figure 5.** Community structure, 1980–2000. The communities and their topics are Protection of Property and the Environment (2, 6, 7, 8, 10, 11, 12, 13, 14, 15) and Claims and Property Regularization (1, 3, 4, 5, 9).

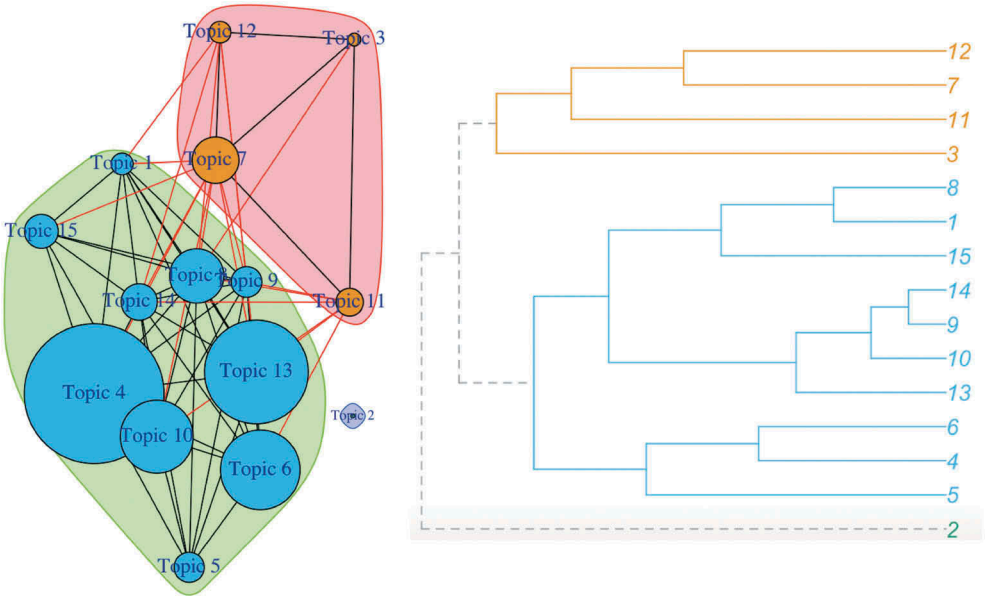


**Figure 6.** Community structure, 2001–2008. The communities and their topics are Protection of Property and the Environment (3, 7, 11, 12); Claims and Property Regularization (1, 4, 5, 6, 8, 9, 10, 13, 14, 15); and Bidding, Companies, Competition and Market (2).

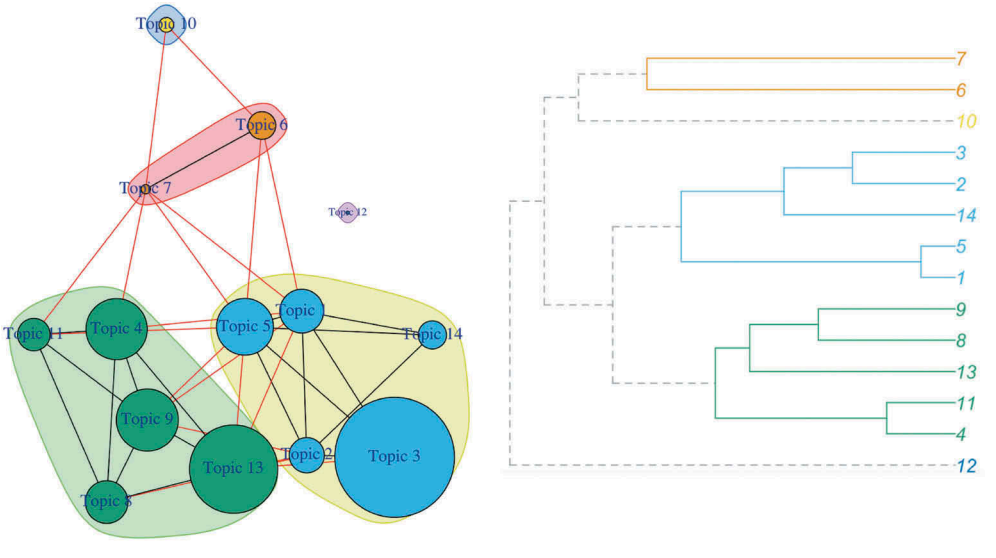


**Figure 7.** Community structure, 2009–2014. The communities and their topics are Relationship with Companies (6, 7); Bidding, Companies, Competition and Market (10); Protection of property and the environment (1, 2, 3, 5, 14). Claims and Property Regularization (4, 8, 9, 11, 13); and Formalism, Individual, Judicial Process (12).

ways over time. The topic interconnections were measured by the number of common keywords shared by the topics.

### Understanding legal complexities by community structure of topics

#### Network analysis

To build the network of word distributions across topics (topic interaction), we used the subjects, found by the LDA model, as network nodes and the common keywords between these topics as proxies for network edges between the nodes. Topic modelling enables the detection of the most frequent subjects per period and thus the most recurrent sources of conflict. In this network representation, the radius of the node is proportional to the frequency of appearance of the corresponding topic at any given period. For example, in 1981–2000 (Figure 5), groundwater was not an issue. Instead, problems related to protection of property and the environment were the major causes of conflict. On the other hand, in 2001–2008 (Figure 6), topics 4 and 13, related to the regularization of property, stand out as important causes of conflict. Finally, in 2009–2014 (Figure 7), topic 3, related to groundwater claims (as in the previous analysis for the entire corpus), acquired a clear relevance, represented by a larger node in the network. This period was marked by serious water scarcity in different regions of the country, which in turn eventually produced an overexploitation of groundwater.

#### Community structure of topics

The goal of community detection is to find the 'natural lines' along which a network separates, that is, the division of the nodes of a network into groups, clusters or communities according to the pattern of edges in the network. The number and size of the groups into which the network is divided are not specified by the domain specialist; they are determined by the network itself and computed by a community detection algorithm (Newman, 2006).

The identified communities correspond to the more common types of actions and fundamentals of the legal claims. The community detection algorithm appears to have picked out the global structure of water-conflict-related topics and their interrelationships using the analysis of network data alone. Moreover, this community structure for the network of topics evolves over time, tending to diversify among several dominant topics per cluster or network group. Further, new emerging communities with fewer links between them appear, suggesting more specialization and structure (and thus, greater complexity) in the types of legal demands.

## Discussion and conclusions

Market-focused approaches tend to overlook that water is a crucial resource for local communities and in many cases also has cultural meaning and helps shape collective identity. As a result, and particularly under conditions of growing scarcity, a water allocation framework based on market transactions is likely to be a source of considerable conflict. In fact, as public opinion and users' and academics' views suggest (Bauer, 2015, Vergara, 2015), the current diagnosis is of an ongoing and growing number of water-related conflicts throughout the country, while the institutional capacity to

resolve them remains weak. In this work, we have provided results based on data and text mining that confirm this shared perception.

The 1981 Chilean Water Code granted a strong role for the judiciary as an arbitrator of water rights disputes and in securing compliance with water management decisions. We therefore used available judicial decisions to assess the real magnitude and dynamics of water-related conflicts. Our results show an explosion of legal claims, particularly after the 2005 Water Code reform. One interpretation of this fact is that the modification of the Water Code was not the immediate catalyst of the rapid growth in lawsuits. Rather, by easing access to justice, the modification of the code 'opened the door' to reveal a number of important conflicts that were hidden behind the apparent economic benefits of the free market for water and which also arose due to climate drivers and increasing competition for this scarce resource. At the same time, some strong indicators of poor regulation of the market and powerful incentives for conflict emerged from our data analysis. For example, the state agency in charge of regulating the market, the DGA, granted so many new water rights that the rights have exceeded real demand, prompting market speculation.

Furthermore, the DGA has lost up to 73% of the legal cases in which it has been involved. This is highly suggestive of a state organization with inadequate technical and/or financial capacities to competently assign water rights and defend the state's interests in court. Furthermore, our results confirmed the excessive centralization of the institutional framework in dealing with water-related conflicts. This may be a sign of the lack of capacity of local courts to resolve local conflicts, which risks overloading the judicial system in the long run. This situation is at odds with the inspiring idea behind water markets, which considers as a basic and desirable principle the minimizing of state intervention.

In this work we adopted a new approach, through network analysis and topic modelling, to understand common types of actions, legal resources and the fundamentals of claims in water rights lawsuits. Considering court judgments as proxies for underlying water-related conflicts and using topic modelling, we found the most frequent subjects that represented the most typical sources of conflicts through time. For example, in 1981–2000, problems related to protection of property and the environment were major causes of conflicts; in 2001–2008, regularization of property (i.e., water rights) stands out as an important cause of disputes; and in 2009–2014, overexploitation of groundwater due to surface water scarcity and poor DGA control of groundwater wells were the most frequent cause of conflicts.

Our results for the community structure of the network of topics showed an overall map of conflicts through these three periods. Some overall patterns were clearly revealed, such as the evolution to fewer but more frequent (dominant) topics per cluster/network group, new emerging network groups, and fewer but stronger links between community members, revealing an increase in complexity. These observations allow us to capture the evolution and global trends of underlying conflicts related to the use of water.

In 2018 a new reform of the Water Code was proposed by the current government. The reform intends to change the structure of perpetual water rights to a type of temporary and administrative use permit, and to give stronger powers to the DGA to limit the exercise of water rights for environmental reasons (Celume, 2015). We believe that this is a great opportunity for a more critical appreciation of the complexities involved in water resources regulation and processes of development. One way to

accomplish this goal is by bridging science and public policy through data sciences and by taking the necessary actions and reforms to solve the real issues that generate conflicts.

## Note

1. Readers of the print version can view the figures in colour in the online article at https://doi.org/10.1080/02508060.2019.1599774

## Acknowledgements

We acknowledge the Commission for Scientific & Technological Research (CONICYT) for supporting our research funded by the Fifth National Competition for Research Centers in Priority Areas CONICYT/FONDAP/15130015; and CONICYT/FONDECYT/ Postdoctorado-3150558

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Fifth National Competition for Research Centers in Priority Areas of National Commission for Scientific & Technological Research CONICYT/FONDAP/15130015; and CONICYT/FONDECYT/ Postdoctorado-3150558.

## References

Aitken, D., Rivera, D., Godoy-Faúndez, A., & Holzapfel, E. (2016). Water scarcity and the impact of the mining and agricultural sectors in Chile. *Sustainability*, *8*(2), 128.

Bauer, C. J. (2005). In the image of the market: The Chilean model of water resources management. *International Journal of Water*, *3*(2), 146–165.

Bauer, C. J. (2015). Water conflicts and entrenched governance problems in Chile's market model. *Water Alternatives*, *8*(2), 147–172.

Berthard, S., Ghosh, S., Martin, J. H., & Summer, T. (2009). Topic model methods for automatically identifying out-of-scope resources. In *JCDL '09: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, 6, 19–28. New York, NY: ACM.

Blei, D., & Lafferty, J. (2009). Topic models. In A. Srivastava & M. Sahami (Eds.), *Text mining: Classification, clustering, and applications chapman & hall/CRC data mining and knowledge discovery series* (pp. 71–89). Boca Raton, FL.

Boisier, J. P., Rondanelli, R., Garreaud, R. D., & Muñoz, F. (2016). Anthropogenic and natural contributions to the Southeast Pacific precipitation decline and recent megadrought in central Chile. *Geophysical Research Letters*, *43*(1), 413–421.

Budds, J. (2004). Power, nature and neoliberalism: The political ecology of water in Chile. *Singapore Journal of Tropical Geography*, *25*(3), 322–342.

Budds, J. (2009). Contested H2O: Science, policy and politics in water resources management in Chile. *Geoforum*, *40*(3), 418–430.

Casella, G., & George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, *46*(3), 167.

Celume, T. (2015). Pilares sobre los que se sustenta la reforma al Código de Aguas chileno y desafíos pendientes. *Actas de Derecho de Aguas*, *5*, 39–50.

DGA. (2017a). *Derechos de aprovechamiento de aguas registrados en DGA*. Retrieved from http://www.dga.cl/productosyservicios/derechos_historicos/Paginas/default.aspx

DGA. (2017b). *Registro Público de Organizaciones de Usuarios*. Retrieved from http://www.dga.cl/administracionrecursoshidricos/OU/Paginas/default.aspx

Donoso, G. (2011). *WP6 IBE EX-POST case studies*. The Chilean water allocation mechanism, established in its Water Code of 1981. Deliverable No. D6.1 – IBE Review Reports. Lessons learned. Prepared under contract from the European Commission Grant Agreement No. 265213 FP7 Environment. Deliverable no.: D6.1 – IBE Review Reports Lessons learned. Retrieved from http://www.feem-project.net/epiwater/docs/d32-d6-1/CS30_Chile.pdf

Firdhous, M. F. M. (2010). Automating legal research through data mining. *International Journal of Advanced Computer Science and Applications - IJACSA*, *1*(6), 9–16.

Fukuyama, F. (2013). What is governance? *Governance*, *26*(3), 347–3686.

Garreaud, R. D., Alvarez-Garreton, C., Barichivich, J., Boisier, J. P., Christie, D., Galleguillos, M., & Zambrano-Bigiarini, M. (2017). The 2010–2015 megadrought in central Chile: Impacts on regional hydroclimate and vegetation. *Hydrology and Earth System Sciences*, *21*(12), 6307–6327.

Grüen, B., & Hornik, K. (2011). Topic models: An R package for fitting topic models. *Journal of Statistical Software*, *40*(13), 1–30.

Hearne, R., & Donoso, G. (2005). Water institutional reforms in Chile. *Water Policy*, *7*(1), 53–69.

Hearne, R., & Donoso, G. (2014). Water markets in Chile: Are they meeting needs?. In W. Easter & Q. Qiuqiong (Eds.), *Water markets for the 21st century: What have we learned?* (Vol. 11, pp. 103–126). Springer.

Jurafsky, D., & Martin, J. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2nd ed.). Engle-wood Cliffs, NJ: Prentice Hall.

Merkl, D., & Schweighofer, E. (1997). En route to data. Mining in legal text corpora: Clustering, Neural. Computation, and international treaties. Proceedings of the 8º International Workshopond Database and Expert Systems Aplications. Toulouse, 465–470.

Microjuris.com, (2017) *Access to database*. Retrieved from http://cl.microjuris.com/home.jsp

Montecinos, A., & Aceituno, P. (2003). Seasonality of the ENSO-related rainfall variability in central Chile and associated circulation anomalies. *Journal of Climate*, *16*(2), 281–296.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*, 8577–8582.

OECD. (2015). *Water resources allocation: Sharing risks and opportunities, OECD studies on water*. Paris: Author. doi:10.1787/9789264229631-en

Quintana, J. M., & Aceituno, P. (2012). *Changes in the rainfall regime along the extratropical west coast of South America (Chile): 30-43º S. Atmósfera, 25*(1), 1-22. Retrieved 10 April 2019, from http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-62362012000100001&lng=es&tlng=en

Retamal, R., Melo, O., & Arumi, J. L., & Parra, Ó. (2012). Sustainable water governance: From a sectoral management system to an integrated one? In D. Rivera (Ed.). *Chile: Environmental, political and social issues* (pp 33–68). Hauppage, NY: Nova Science

Rivera, D., Godoy-Faúndez, A., Lillo, M., Alvez, A., Delgado, V., Gonzalo-Martín, C., & García-Pedrero, Á. (2016). Legal disputes as a proxy for regional conflicts over water rights in Chile. *Journal of Hydrology*, *535*, 36–45.

Saleth, R. M., & Dinar, A. (2005). Water institutional reforms: Theory and practice. *Water Policy*, *7*, 1–19.

Scott, W. R. (1995). *Institutions and organizations*. Thousand Oaks, CA: SAGE.

Steyvers, M., & Griths, T. (2007). Probabilistic topic models. *Handbook of Latent Semantic Analysis*, *427*(7), 424–440.

Sunil, R. (2013). Knowledge discovery from legal documents dataset using text mining techniques. *International Journal of Computer Applications*, *66*(23), 32–34.

Sunil, R. (2014). Exploratory analysis of legal documents using unsupervised text mining techniques. *International Journal of Engineering Research & Technology*, *3*(2), 2264–2267.

Valdés-Pineda, R., Pizarro, R., García-Chevesich, P., Valdés, J. B., Olivares, C., Vera, M., … Helwig, B. (2014). Water governance in Chile: Availability, management and climate change. *Journal of Hydrology*, *519*, 2538–2567.

Vergara, A. (2002). Las aguas como bien público (no estatal) y lo privado en el derecho chileno: Evolución legislativa y su proyecto de reforma. *Revista de Derecho Administrativo Económico*, *IV*(1), 63–79.

Vergara, A. (2015). *Crisis institucional del agua. Descripción del modelo, crítica a la burocracia y propuesta de tribunales especiales* (2ª ed.). Santiago: Thomson Reuters, Ediciones UC.

Waltl, B., Matthes, F., Waltl, T., & Grass, T. (2016). *LEXIA - A data science environment for semantic analysis of German legal texts, IRIS: Internationales rechtsinformatik symposium*. Salzburg, Austria.

World Bank. (2011). *Diagnóstico de la gestión de los recursos hídricos*. Departamento de Medio Ambiente y Desarrollo Sostenible Region para América Latina y el Caribe, CHILE. Retrieved from http://www.dga.cl/eventos/Diagnostico%20gestion%20de%20recursos%20hidricos%20en%20Chile_Banco%20Mundial.pdf

World Bank. (2013). *Estudio para el mejoramiento del marco institucional para la gestión del agua*. Departamento de Medio Ambiente y Desarrollo Sostenible Region para América Latina y el Caribe, CHILE. Retrieved from http://reformacodigodeaguas.carey.cl/wp-content/uploads/2014/09/Informe-Banco-Mundial-Estudio-para-el-mejoramiento-del-marco-institucional.pdf

Wyner, A., Mochales-Palau, R., Moens, M. F., & Milward, D. (2010). Approaches to text mining arguments from legal cases. In E. Francesconi, S. Montemagni, W. Peters, & D. Tiscornia (Eds.), *Semantic processing of legal texts. Lecture notes in computer science* (Vol. 6036, pp. 60–79). Berlin, Heidelberg: Springer.